

1 Evolution Inspired Engineering of Megasyntetases

2

3 Kenan A. J. Bozhüyük^{1,2,3,&,*}, Leonard Präve^{1,3,&}, Carsten Kegler^{1,&}, Sebastian
4 Kaiser^{1,4}, Yan-Ni Shi³, Wolfgang Kutenlochner⁵, Leonie Schenk^{1,3}, T. M. Mohiuddin,³
5 Michael Groll⁵, Georg K. A. Hochberg^{4,6,7}, Helge B. Bode^{1,3,6,7,8,*}

6

7 ¹Max-Planck-Institute for Terrestrial Microbiology, Department of Natural Products in
8 Organismic Interactions, 35043 Marburg, Germany

9 ²Myria Biosciences AG, Basel, Switzerland

10 ³Molecular Biotechnology, Department of Biosciences, Goethe-University Frankfurt,
11 60438 Frankfurt, Germany

12 ⁴Evolutionary Biochemistry Group, Max-Planck-Institute for Terrestrial Microbiology,
13 35043 Marburg, Germany

14 ⁵Chair of Biochemistry, Center for Protein Assemblies, Technical University of Munich,
15 Ernst-Otto-Fischer-Straße 8, 85748 Garching, Germany

16 ⁶Center for Synthetic Microbiology (SYNMIKRO), Phillips University Marburg, 35043
17 Marburg, Germany

18 ⁷Department of Chemistry, Phillips University Marburg, 35043 Marburg, Germany

19 ⁸Senckenberg Gesellschaft für Naturforschung, 60325 Frankfurt, Germany

20 [&]These authors contributed equally to this work

21 ^{*}Correspondence authors: Kenan.Bozhueyuek@mpi-marburg.mpg.de,

22 helge.bode@mpi-marburg.mpg.de

23

24 Abstract

25 Many clinically used drugs are derived from or inspired by bacterial natural products
26 that often are biosynthesised via non-ribosomal peptide synthetases (NRPS), giant
27 megasyntetases that activate and join individual amino acids in an assembly line
28 fashion. Since NRPS are not limited to the incorporation of the 20 proteinogenic amino
29 acids, their efficient manipulation would allow the biotechnological generation of
30 complex peptides including linear, cyclic and further modified natural product
31 analogues, e.g. to optimise natural product leads. Here we describe a detailed
32 phylogenetic analysis of several bacterial NRPS that led to the identification of a new
33 recombination breakpoint within the thiolation (T) domain that is important for natural
34 NRPS evolution. From this, an evolution-inspired eXchange Unit between T domains
35 (XUT) approach was developed which allows the assembly of NRPS fragments over
36 a broad range of GC contents, protein similarities, and extender unit specificities, as
37 demonstrated for the specific production of a proteasome inhibitor designed and
38 assembled from five different NRPS fragments.

39 Introduction

40 Natural products (NPs) have been extensively studied for their therapeutic potential
41 given their remarkable chemical and structural diversity in nature. Not only are they
42 considered a rich reservoir of pharmacologically active lead compounds with
43 therapeutic potential, but with ~48% of all new medicines approved between 1981 and
44 2019 originating in nature, NPs play an important role in the drug discovery and
45 development process¹. In recent decades, the collective efforts of the scientific
46 community have led to tremendous progress in the identification of novel NPs to
47 evaluate their pharmacological properties and mode of action, that could not be easily
48 transferred into the development of new clinical drugs². One of many reasons why the
49 pharmaceutical industry stepped back from NP-based drug discovery.

50
51 Genetic engineering of natural products holds the potential for faster and more cost-
52 effective discovery of (tailor-made) biological drugs than conventional methods³. Many
53 bioactive bacterial NPs are derived from biosynthetic gene clusters (BGCs), genomic
54 bacterial islands encoding non-ribosomal peptide synthetases (NRPS)⁴. NRPS are
55 genetically encoded molecular assembly lines with many moving parts and reaction
56 centres that all work together, to produce a broad variety of valuable non-ribosomal
57 peptides (NRP) or even clinical drugs – such as penicillins⁵⁻⁷, bleomycin⁸, and
58 ciclosporin⁹. Given these outstanding biological activities that benefit global public
59 health, NRPS assembly lines would be an ideal target for synthetic biology, e.g. to
60 improve pharmacological properties of natural product leads for (pre-) clinical
61 development.

62
63 NRPS assembly lines consist of sequentially repeating modules of enzymatic domains,
64 each of which catalyses the incorporation and chemical modification of a specific
65 extender unit into the growing chain before the extended chain is passed on to the next
66 module⁴. Hundreds of different extender units, typically derived from amino acids, have
67 been described so far^{10,11}. Selection and activation of an extender unit within an NRPS
68 is catalysed by an adenylation (A) domain. The activated substrate is then covalently
69 attached to the post-translationally attached prosthetic thiol (phosphopantetheine)
70 group of a small thiolation (T) domain. Condensation (C) domains then link the
71 covalently bound substrates to the growing NRP-chain in a co-linear fashion. In
72 addition to these "core" domains that define the functional unit of an assembly line

73 module, tailoring domains may be present to modify NRP chain, that are
74 heterocyclization (Cy), epimerisation (E), *N*-methylation (MT), oxidation (Ox) or
75 reduction (R) domains. Finally, the full-length NRP is released from the enzymatic
76 machinery by hydrolysis or macrocyclization catalysed by a thioesterase (TE) domain.

77

78 The very logic of this assembly line mechanism inspired numerous rational efforts to
79 engineer megasynthases to produce natural product analogues or even artificial NP-
80 like compounds¹². Although early engineering attempts yielded biosynthesis clusters
81 that were either greatly impaired in their activity or non-functional, recent technical
82 advances and the growing body of structural data accelerated the development of
83 innovative synthetic biology strategies to engineer megasynthetases. Examples are
84 the identification of interchangeable catalytically functional domain units^{13,14}, CRISPR-
85 Cas9 based gene editing to engineer complex antibiotic assembly lines¹⁵, yeast cell
86 surface display assay to engineer the specificity of individual A domains,¹⁶ and splitting
87 megasynthases into individually expressible subunits to reduce their complexity and
88 size (up to several MDa) either *via* adding zinc-finger tags¹⁷ (DNA-templated NRPS)
89 or SYNZIPs¹⁸⁻²⁰ (heterospecific coiled-coil peptides). Furthermore, with the continuous
90 increase in publicly available genomic data and the extensive efforts of the community
91 to develop processing tools for BGC and NP identification²¹⁻²³, there is a new trend
92 towards assembly-line engineering using evolution-driven strategies. A number of
93 insightful studies have led to the conclusion that understanding the mechanisms by
94 which nature has evolved these often huge multifunctional enzyme machines will
95 further improve our ability to redesign assembly line proteins to achieve even greater
96 structural diversity while maintaining good production titres and could help us to
97 expand our therapeutic arsenal²⁴⁻²⁶. However, the evolutionary mechanisms to achieve
98 the exchange of individual extender units in NRP scaffolds are still poorly understood.

99

100 The genetic and architectural modularity of NRPS, but also of the biochemically distinct
101 yet mechanistically analogous polyketide synthases (PKS), is central for current
102 evolution models of these BGCs. Historically, the functional unit able to perform one
103 round of chain elongation of a PKS (KS-AT-(DH-KR-ER)-T) and NRPS (C-A-T) is
104 called a module. It is yet unclear whether this architectural and genetic unit also
105 corresponds to an evolutionary unit that has been preserved in megasynthases^{24,26}.
106 Phylogenetic and computational analyses of the PKS family have led to a proposed

107 redefinition of module boundaries from the "historical" KS-AT-(DH-KR-ER)-ACP to the
108 "alternative" AT-(DH-KR-ER)-ACP-KS^{27,28}, and highlighted the presence of genetic
109 repeats²⁹ (GRINS = genetic repeats of intense nucleotide skews) in a large number of
110 PKS. The latter play a putative role in accelerating diversification of closely related
111 BGCs by promoting gene conversion. For NRPS, studies on the underlying
112 evolutionary processes have only just begun.

113
114 In a recent *in silico* study, the evolution of bacterial NRPS across various phyla was
115 analysed²⁴. The authors not only showed that intragenomic recombination along with
116 speciation and horizontal gene transfer together with recombination are important
117 factors in NRPS evolution, but also enabled the authors to introduce a unifying model
118 for the evolution of the present-day variety of NRPs. Within the framework of this
119 model, it was suggested that single recombination events at multiple breakpoints within
120 the A domains of NRPS, referred to as subdomain swapping, are not only a widespread
121 phenomenon, as reported previously^{25,30-32}, but are also a major contributing factor to
122 the diversification and functionalisation of NRP families. Further key findings are that
123 stereochemical changes from the L- to the D-configuration in the final NRP seem to be
124 achieved by the combined exchange of T-C di- for T-E-C tri-domains; and that there is
125 a trend to keep intact both the native C-A linker region, physically connecting both
126 domains, and the A-T domain interface. However, the practical evidence of these
127 findings for successful NRPS engineering on a broad basis have not been shown yet.
128 Up to date, there are only a very limited number of examples where evolutionary
129 insights have been successfully used to engineer megasynthases – but only within a
130 very narrow range of genetic and chemical changes introduced into the underlying
131 BGCs and produced NPs, respectively^{15,32,33}.

132
133 Herein, we particularly focused on deciphering the evolutionary history of NRPS to
134 identify an evolution-inspired moiety that is best suited to enable NRPS engineering in
135 a unified and more efficient manner. In order to approach the problem from different
136 angles, a broad dataset of NRPS sequences from different phyla was analysed *in silico*
137 to identify recombination events, a fusion point screening was performed to identify
138 ideal engineering sites, the identified sites were broadly evaluated by reprogramming
139 NRPS enzymes, and finally this knowledge was used to design a pharmaceutically
140 active peptide *de novo*.

141

142 **Results**

143 **Deciphering the Evolutionary History of NRPS**

144 Homologous recombination is a pervasive biological process that affects sequences in
145 all living organisms and undoubtedly is the main driver for megasynthase
146 diversification³⁴. Sequences having undergone recombination, such as NRPS, will
147 display two different histories: one history for one part of their sequence, affected by
148 the recombination event, and one history for the other part. Consequently, the
149 evolutionary history of an alignment of homologous NRPS sequences cannot be
150 properly depicted by classical phylogenetic methods because only one bifurcating tree
151 is reconstructed. Therefore, we applied a previously established maximum likelihood
152 method that was explicitly designed to detect multiple phylogenetic histories caused
153 by recombination events. It uses a phylogenetic Hidden Markov Model (HMM) to
154 search for a specified number of independent evolutionary histories that together best
155 explain the alignment³⁴. The algorithm returns the site likelihoods for each tree for
156 every single position in an alignment, which can then be used to detect recombination
157 breakpoints. In our analysis, we searched for two different histories, expecting one to
158 broadly fit known A domain trees³⁵, and the other to broadly fit known C domain
159 phylogenies^{36,37}. To identify which sites, belong to which history, we then subtracted
160 the site-wise log likelihoods of the second tree from those of the first tree (Figs. 1a and
161 Supplementary Fig. 1) – positive values indicate sites that are better described by the
162 first phylogenetic history, sites with negative values are better described by the second
163 phylogenetic history.

164

165 We applied this method to a dataset comprising of >200 aligned amino acid sequences
166 of NRPS A-T-C tri-domains from *Photorhabdus* and *Xenorhabdus* species, as well as
167 representative NRPS from firmicutes, actinomycetes, cyanobacteria and other
168 proteobacteria (Supplementary Dataset 1). This analysis revealed two major insights:
169 First, sites in the A domain mostly preferred the first history, and sites in the C domain
170 strongly prefer the second history (Supplementary Fig. 2). This confirms our method
171 can detect that A and C domains have different evolutionary histories. And second, the
172 breakpoint between these two histories appears to lie somewhere inside the T domain
173 (Fig. 1a), though were exactly was not clear from this analysis: The difference in site
174 likelihoods between histories becomes significantly more negative (indicating a

175 preference for the second history) roughly in the middle of the T-domain, from around
176 zero (not preferring either tree) to values well below -50 log units (strongly preferring
177 the second tree).

178

179 To gain a better understanding of this potential recombination breakpoint, we repeated
180 our phylogenetic HMM analysis with just the T domain together with the A-T-linker,
181 again searching for two histories (Fig. 1b – d, and Supplementary Fig. 1;
182 Supplementary Dataset 2). We did this, because the first half of the T domain preferred
183 neither tree in our first analysis (Figs. 1c and Supplementary Fig. 3), potentially
184 because it doesn't exactly share the A or C domain's history. In this analysis, we see
185 a sharp boundary between the two trees within the conserved FFxxGGxS motif in the
186 T domain (Figs. 1c and Supplementary Fig. 4). Interestingly, the second history has a
187 topology similar to the C domain tree (Figs. 1d and Supplementary Fig. 5). It also
188 contains a clear split that separates T domains according the condensation reaction
189 catalysed by the downstream C domains (Fig. 1d). The first tree, however, is not similar
190 to either the C or A domain trees (Fig. 1c). Taken together, these observations suggest
191 that the T domain may be a frequent recombination site, with a particularly important
192 boundary in the conserved FFxxGGxS motif in the T domain.

193

194 To further confirm these *in silico* predictions and to avoid the result being a
195 computational artefact, we have analysed in detail examples of homologous NRPS
196 such as the PAX^{38,39}, endopyrrole A⁴⁰, rhizomide A⁴¹, and syringopeptin SP-25a⁴²
197 producing synthetases to obtain evidence of recombination events within the T
198 domains. In brief, this detailed analysis indeed supported the notion that recombination
199 events within T domains frequently occur either to introduce a stereochemistry change
200 (T-C vs T-C/E), and/or to exchange T-TE domains, and/or to increase/decrease the
201 size of the BGC and the respective NP scaffold. A detailed description of this analysis
202 can be found in the supporting information (Supplementary Figs. 6 – 9).

203

204 In summary, the results gained from the phylogenetic HMM (Fig. 1 and Supplementary
205 Figs. 1 – 5) and the detailed analysis of various BGCs results point towards a yet
206 undescribed recombination breakpoint.

207

208 **Fusion Point Screening**

209 The conserved core motif (FFxxGGxS) of the ~100 amino acid T domains is located at
210 the *N*-terminus (loop1) of the second helix ($\alpha 2$) holding the invariant serine residue that
211 becomes post-translationally modified by a phosphopantetheinyl (Ppant) transferase⁴³⁻
212 ⁴⁵. Although the T domain is the only NRPS domain without an autonomous catalytic
213 activity, the attachment of Ppant is a functional prerequisite, not only to covalently bind
214 activated extender units and the growing peptide chain, but also to pervade the active
215 sites of A and C domains. In addition, it is known from structural data that the first part
216 of the T-domain (T_{p1}), which is *N*-terminal to the core motif, mainly interacts with the A
217 domain via $\alpha 1$ and loop1 and the second half (T_{p2}), which is *C*-terminal to the core
218 motif, interacts with the C domain via $\alpha 4$ ⁴⁶. However, as computational recombination
219 analysis (Fig. 1) naturally does not come up with one specific splicing position but with
220 a sequence region that is likely to promote homologous recombination, initially a fusion
221 point screening was performed (Fig. 2) to verify fusion sites resulting in the best peptide
222 production.

223
224 As a starting point, we chose the GameXPeptide⁴⁷ (GxpS) and the xenoamicin⁴⁸
225 producing synthetases (XabABC) from *P. luminescens* TT01 and *X. stockiae* (Fig. 2a),
226 respectively, to produce seven recombinant NRPS (Fig. 2b, NRPS-1 to -7), each with
227 a different fusion site (I to VII, Fig. 2c), with the *in silico* predicted breakpoint
228 represented by fusion site IV. Briefly, this screening led to the identification of three
229 functional fusion sites (I, III, and IV) in NRPS-1, -3 and -4, that all produce the expected
230 lipopeptides **1-3**, differing only in the acyl starter originating from the fatty acid pool of
231 *E. coli*, with titres between 12 and 27 mgL⁻¹ (Fig. 2b, Supplementary Figs. 10 – 17, and
232 Supplementary Table S6). Of note, throughout the present work, all NRPS were
233 heterologously produced in *E. coli* DH10B::mtaA⁴⁹. The resulting peptides
234 (Supplementary Table S5) and yields were confirmed by HPLC-MS/MS and
235 comparison of retention times with synthetic standards (see Supplementary
236 Information).

237
238 Taken together, the *in silico* observations (Fig. 1) along with the results from the *in vivo*
239 conducted fusion site screening (Fig. 2) led us to the hypothesis that both, T-C-A units
240 (fusion point I), T_{p1} -C-A- T_{p2} units (fusion points III & IV), and combinations thereof may
241 serve as ideal starting points to do rational evolution-inspired megasynthetase
242 engineering. However, after reviewing crystal structure data of A-T and T-C didomains

243 we decided to proceed with fusion site I and IV, because fusion sites III and IV are both
244 located directly adjacent (III) and within (IV) the conserved T domain motif,
245 respectively, and the two variable positions in between the conserved motif
246 (FFxxGGxS) are potentially contributing to a functional A-T interface⁴⁶.

247

248

249 **Evolution Inspired eXchange Units for NRPS Engineering**

250 To further verify the *in silico* identified (IV) and *in vivo* verified (I & IV) fusion sites on a
251 broad scale we targeted the NRPS FitAB (Fig. 3, NRPS-8) and FtrAB (Supplementary
252 Fig. 18, NRPS-17; Supplementary Dataset 3) producing the NRPs fitayylide and
253 faTTTVIR from *X. innexii* and *X. mauleonii*, respectively, as well as GxpS (Fig. 4). In
254 sum we created 16 recombinant FitAB derivatives (NRPS-9 to -18, Fig. 3), one ftrAB
255 derivative (NRPS-19 and -20, Fig. S18), and eight GxpS derivatives (Fig. 4) applying
256 fusion site I, IV, or both. The building blocks to engineer NRPS-8, NRPS-17 and GxpS
257 were selected to cover a broad range of bacterial genera (*Xenorhabdus*,
258 *Photorhabdus*, *Serratia*, *Myxococcus*, *Pseudomonas* and *Bacillus*) with GC contents
259 between 50 to 72 % GC to reveal if the identified fusion sites have the potential to
260 mimic horizontal gene transfer along with recombination on a rational scale suitable to
261 re-engineer NRPS.

262

263 Interestingly all recombinant NRPS (including a NRPS-PKS assembly line where the
264 PKS is responsible for the polyunsaturated starter acyl moiety [NRPS-19]), showed
265 catalytic activity producing a broad range of cyclic and linear peptides (**4-39**) at titres
266 ranging from 2.5 (NRPS-23a) to 136 mgL⁻¹ (NRPS-22a) and from 2.5 (NRPS-18b) to
267 98 mgL⁻¹ (NRPS-22b) for fusion site I and IV, respectively (Figs. 3, 4, Supplementary
268 Figs. 18 – 55, and Supplementary Tables 7 – 9). In addition, and as already indicated
269 from our initial fusion point screening (Fig. 2), no trend concerning a preferred fusion
270 site could be observed.

271

272 Noteworthy, both fusion sites of the evolution-inspired exchange units between T
273 domains (XUT) enabled us to create chimeric NRPS from completely unrelated BGCs
274 for the first time with respect to taxonomy and GC content. Other methods, such as A
275 subdomain swaps^{15,31} or the previously introduced eXchange unit concepts^{13,14},
276 enabled efficient reprogramming of NRPS only within a narrow range of related BGCs.

277 Nevertheless, a correlation between GC content of the introduced NRPS building
278 blocks and peptide production can be observed (Fig. 3 and 4). Whereas building blocks
279 of genera with a similar or slightly higher (50 to 65 %) GC content (i.e., NRPS-13 and
280 NRPS-14; Fig. 3) are generally well tolerated, building blocks originating from the high-
281 GC branch (~70 %, i.e. NRPS-17, -23, and -24; Fig. 3 and 4) are resulting in impaired
282 assembly lines when recombined with NRPS originating from *Xenorhabdus* and
283 *Photorhabdus*. The initial reduction of catalytic activity when building blocks of different
284 GC-content are recombined with each other might also occur naturally during
285 homologous recombination after a horizontal gene transfer event.

286

287 **Evolution Inspired eXchange Units allow targeted peptide production**

288 In order to validate the strength of these evolution-inspired exchange units (XUT), an
289 artificial biosynthetic assembly line producing a novel pharmacological active peptide
290 against a well characterised target was designed *de novo*. We chose the eukaryotic
291 proteasome as target which plays pivotal roles in protein homeostasis affecting cell
292 cycle, signal transduction and general cell physiology. Proteasomes are a family of *N*-
293 terminal nucleophilic hydrolases consisting of two sets of seven copies of α and β
294 subunits that assemble into a barrel-shaped complex (Fig. 5)⁵⁰. Peptides inhibiting the
295 proteasome, such as the clinically used bortezomib⁵¹, can lead to apoptosis, making
296 the human proteasome a target for anti-cancer chemotherapy. Similar to well-known
297 strategies from the pharmaceutical industry, we used the lipopeptide aldehyde
298 fellutamide B⁵² as inspiration from nature that is not only active against the eukaryotic
299 proteasome of humans and yeast, but is also the most potent inhibitor of the
300 *Mycobacterium tuberculosis* proteasome tested to date. Fellutamide B consist of a C8-
301 3OH acyl chain, L-Asn, L-Gln, and a L-Leu-aldehyde. The aldehyde moiety is
302 responsible for the reversible binding to the active site threonine (Thr1) of the
303 proteasome. From an NRPS engineering perspective, in particular the introduction of
304 reactive groups, denoted as warheads, is a major challenge. As an alternative to TE
305 domains, nature applies thioester reductase (R) domains⁵³⁻⁵⁵, not only to release the
306 synthesised peptide, but also to introduce the aldehyde function by catalysing an
307 NAD(P)H dependent two-electron reduction of the thioester.

308

309 For the final XUT proof-of-concept experiment we *in silico* designed an artificial three-
310 modular assembly line composed from NRPS building blocks derived from five

311 different origins (Fig. 5a): a C_{start} domain to introduce the acyl chain and A domains
312 with specificities (*N*- to *C*-terminus) for L-Gln (A1), L-Ala (A2), and L-Leu (A3). To
313 achieve the reduction of leucine into an aldehyde, the R-domain of the tilivalline
314 producing NRPS (XtvB)⁵⁶ from *X. indica* was used as termination domain using the
315 fusions sites I, III, IV, and VII. The resulting assembly lines NRPS-25 to -28 all showed
316 catalytic activity producing the desired lipopeptide aldehydes **40–42**, differing only in
317 the acyl group used as a starter, with titres between ~1 and ~22 mgL⁻¹ (Fig. 5b,
318 Supplementary Figs. 56 – 62, and Supplementary Table 10). Compared to the NRPS
319 generated with fusion sites I (NRPS-25) and VII (NRPS-28), the NRPS generated with
320 fusion sites III (NRPS-26) and IV (NRPS-27) produced about tenfold more peptides.
321 Whereas low titres of NRPS-28 are in good agreement with our initial fusion point
322 screening (Fig. 2), the peptide amount biosynthesised by NRPS-25 was unexpectedly
323 low. The impaired formation of a functional A-T domain-domain interface⁴⁶, in the case
324 of NRPS-25, and a functional T-R⁵³ domain-domain interface in the case of NRPS-28
325 could serve as an explanation for this result, as shown previously^{56,57}. Furthermore,
326 these results highlight the advantages of the evolution-inspired fusion sites III and IV
327 compared to fusion sites I and VII, which are located within the A-T and T-C linker
328 regions, respectively.

329

330 In order to test whether the new-to-nature lipopeptide aldehyde **41** is indeed able to
331 inhibit the yeast 20S proteasome core particle (yCP) by binding the active site Thr1,
332 the half maximum inhibitory concentration (IC₅₀) and co-crystallization of yCP together
333 with **41** (yCP:41 complex) was performed (Supplementary Fig. 63 and Supplementary
334 Table 11). Both experiments confirmed the expected activity of **41** against the yCP β5
335 subunit at 3.6 ± 0.8 μM and a binding mode to Thr1 equivalent to that of fellutamide B
336 (Fig. 5c-e). In summary, this proof-of-concept experiment not only revealed that
337 reactive groups efficiently can be introduced by applying the novel XUT approach but
338 also that tailormade bioactive peptides can be created *de novo* in a retro-biosynthetic
339 manner.

340

341 **Conclusion**

342 Despite all the technical advances and our knowledge of the fundamental biochemical
343 and structural properties of assembly line enzymes⁴, their engineering has remained a
344 major challenge⁵⁸. Nature, however, appears to have been successful at engineering

345 biosynthetic pathways through the process of BGC evolution using a broad range of
346 mechanisms. Previous studies either comprehensively analysed a diverse range of
347 NRPS families or focused on deciphering the evolution of one specific NRP family.
348 Both approaches have dramatically improved our understanding of the underlying
349 mechanisms of megasynthetase evolution. Pioneering studies for example proposed
350 the *N*-terminal expansion of modules in BGC evolution⁵⁹, highlighted the role of the A
351 domains in NRPS diversification^{24,30,32}, and introduced models explaining the
352 mechanisms resulting in present day NRPS families^{24,33,60}. However, most of these
353 studies have not succeeded in developing these findings into an overall rational
354 engineering approach. When the available datasets describing the evolution of NRPS
355 synthesising syringopeptin⁵⁹, jessipeptin, virginafactin, chicofactin, and syringafactin⁶⁰
356 were reanalysed, we clearly could identify the T domain as an additional recombination
357 hot spot.

358

359 Compared to these previous studies the major aim of this work was not the
360 identification of the exact mechanisms that led to present day NRPS families, but to
361 understand the major driving forces in NRPS evolution and how these insights can be
362 leveraged to improve rational engineering of assembly line enzymes. Based on our
363 findings we propose a yet undescribed recombination breakpoint within the conserved
364 core motif of T domains (fusion sites III and IV), resulting in the XUT T_{p1} -C-A- T_{p2} .
365 Interestingly, the XUT approach is completely in line with recent structural findings on
366 the catalytic cycle of NRPS^{46,61} and, with exceptions, mostly consistent with the
367 recently introduced unifying model for the evolution of the present-day variety of NRPs²⁴.
368 Although we are convinced that A subdomain exchanges are another important driver
369 for NRPS evolution, our data does not suggest such a recombination, probably
370 because the two data sets and the method of analysis are fundamentally different.
371 From an applied engineering perspective, XUT appears to be much more versatile
372 compared to A subdomain swaps^{15,31,32}, allowing the rational recombination of
373 completely unrelated NRPS building blocks over a broad range of GC contents (from
374 50% to 70 %), protein similarities (< 39 %), and extender unit specificities (Figs. 3 and
375 4).

376

377 To conclude, the XUT approach enables the mimicking of horizontal gene transfer
378 followed by a recombination event, opening up avenues for the expansion of structural

379 diversity that we can address through rational engineering – even beyond natural
380 diversity. This is clearly illustrated by the example of the artificial proteasome inhibitor
381 (Fig. 5) leading to the first time rational *de novo* design of a new-to-nature
382 pharmacologically active peptide.

383

384 **Acknowledgements**

385 This work was supported by an ERC Advanced Grant (835108), the LOEWE Research
386 center TBG funded by the State of Hesse, the Max-Planck Society (all to H.B.B.), and
387 the German Research Foundation (SFB1035, project number: 201302640, A02 to
388 W.K. and M.G.). The authors thank the staff of beamline X06SA at the Paul Scherrer
389 Institute, SLS, Villigen (Switzerland), for their assistance during data collection and are
390 grateful to all Bode lab members for continuous discussions about further
391 developments of NRPS engineering methods.

392

393 **Author contributions**

394 K.A.J.B., L.P., C.K., L.S. and T.M.M. planned and performed all NRPS engineering
395 experiments and peptide productions. Y.-N.S. isolated all peptides and elucidated their
396 structure. S.K., C.K. and G.K.A.H. performed all phylogenetic analyses. W.K. and M.G.
397 performed proteasome assays and crystallization. K.A.J.B. and H.B.B. conceived all
398 experiments and wrote the paper with input from all authors.

399

400 **Competing interests**

401 A patent describing the XUT approach was filed by the Goethe University Frankfurt.
402 K.A.J.B. and H.B.B. are cofounder and shareholder of Myria Biosciences AG, of which
403 K.A.J.B. is also CSO.

404

405 **References**

- 406 1 Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four
407 Decades from 01/1981 to 09/2019. *J Nat Prod* **83**, 770-803, doi:10.1021/acs.jnatprod.9b01285
408 (2020).
- 409 2 Huang, M., Lu, J. J. & Ding, J. Natural Products in Cancer Therapy: Past, Present and Future.
410 *Nat Prod Bioprospect* **11**, 5-13, doi:10.1007/s13659-020-00293-7 (2021).
- 411 3 David, F. *et al.* A Perspective on Synthetic Biology in Drug Discovery and Development-Current
412 Impact and Future Opportunities. *SLAS Discov* **26**, 581-603, doi:10.1177/24725552211000669
413 (2021).
- 414 4 Sussmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew*
415 *Chem Int Ed Engl* **56**, 3770-3821, doi:10.1002/anie.201609079 (2017).

- 416 5 Tahlan, K., Moore, M. A. & Jensen, S. E. delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine
417 synthetase (ACVS): discovery and perspectives. *J Ind Microbiol Biotechnol* **44**, 517-524,
418 doi:10.1007/s10295-016-1850-7 (2017).
- 419 6 Byford, M. F., Baldwin, J. E., Shiau, C. Y. & Schofield, C. J. The Mechanism of ACV Synthetase.
420 *Chem Rev* **97**, 2631-2650, doi:10.1021/cr960018l (1997).
- 421 7 Zhang, J. & Demain, A. L. ACV synthetase. *Crit Rev Biotechnol* **12**, 245-260,
422 doi:10.3109/07388559209069194 (1992).
- 423 8 Umezawa, H. Structure and action of bleomycin. *Prog Biochem Pharmacol* **11**, 18-27 (1976).
- 424 9 Fahr, A. Cyclosporin clinical pharmacokinetics. *Clin Pharmacokinet* **24**, 472-495,
425 doi:10.2165/00003088-199324060-00004 (1993).
- 426 10 Caboche, S., Leclere, V., Pupin, M., Kucherov, G. & Jacques, P. Diversity of monomers in
427 nonribosomal peptides: towards the prediction of origin and biological activity. *J Bacteriol* **192**,
428 5143-5150, doi:10.1128/JB.00315-10 (2010).
- 429 11 Walsh, C. T., O'Brien, R. V. & Khosla, C. Nonproteinogenic amino acid building blocks for
430 nonribosomal peptide and hybrid polyketide scaffolds. *Angew Chem Int Ed Engl* **52**, 7098-7124,
431 doi:10.1002/anie.201208344 (2013).
- 432 12 Bozhuyuk, K. A., Micklefield, J. & Wilkinson, B. Engineering enzymatic assembly lines to
433 produce new antibiotics. *Curr Opin Microbiol* **51**, 88-96, doi:10.1016/j.mib.2019.10.007 (2019).
- 434 13 Bozhuyuk, K. A. J. *et al.* Modification and de novo design of non-ribosomal peptide synthetases
435 using specific assembly points within condensation domains. *Nat Chem* **11**, 653-661,
436 doi:10.1038/s41557-019-0276-z (2019).
- 437 14 Bozhuyuk, K. A. J. *et al.* De novo design and engineering of non-ribosomal peptide synthetases.
438 *Nat Chem* **10**, 275-281, doi:10.1038/nchem.2890 (2018).
- 439 15 Thong, W. L. *et al.* Gene editing enables rapid engineering of complex antibiotic assembly lines.
440 *Nat Commun* **12**, 6872, doi:10.1038/s41467-021-27139-1 (2021).
- 441 16 Niquille, D. L. *et al.* Nonribosomal biosynthesis of backbone-modified peptides. *Nat Chem* **10**,
442 282-287, doi:10.1038/nchem.2891 (2018).
- 443 17 Huang, H. M., Stephan, P. & Kries, H. Engineering DNA-Templated Nonribosomal Peptide
444 Synthesis. *Cell Chem Biol* **28**, 221-227 e227, doi:10.1016/j.chembiol.2020.11.004 (2021).
- 445 18 Abbood, N., Duy Vo, T., Watzel, J., Bozhueyuek, K. A. J. & Bode, H. B. Type S Non-Ribosomal
446 Peptide Synthetases for the Rapid Generation of Tailormade Peptide Libraries. *Chemistry*,
447 e202103963, doi:10.1002/chem.202103963 (2022).
- 448 19 Bozhueyuek, K. A. J., Watzel, J., Abbood, N. & Bode, H. B. Synthetic Zippers as an Enabling Tool
449 for Engineering of Non-Ribosomal Peptide Synthetases*. *Angew Chem Int Ed Engl* **60**, 17531-
450 17538, doi:10.1002/anie.202102859 (2021).
- 451 20 Klaus, M., D'Souza, A. D., Nivina, A., Khosla, C. & Grininger, M. Engineering of Chimeric
452 Polyketide Synthases Using SYNZIP Docking Domains. *ACS Chem Biol* **14**, 426-433,
453 doi:10.1021/acscchembio.8b01060 (2019).
- 454 21 Kautsar, S. A., Blin, K., Shaw, S., Weber, T. & Medema, M. H. BiG-FAM: the biosynthetic gene
455 cluster families database. *Nucleic Acids Res* **49**, D490-D497, doi:10.1093/nar/gkaa812 (2021).
- 456 22 Blin, K. *et al.* antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic*
457 *Acids Res* **49**, W29-W35, doi:10.1093/nar/gkab335 (2021).
- 458 23 Blin, K., Shaw, S., Kautsar, S. A., Medema, M. H. & Weber, T. The antiSMASH database version
459 3: increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids*
460 *Res* **49**, D639-D643, doi:10.1093/nar/gkaa978 (2021).
- 461 24 Baunach, M., Chowdhury, S., Stallforth, P. & Dittmann, E. The Landscape of Recombination
462 Events That Create Nonribosomal Peptide Diversity. *Mol Biol Evol* **38**, 2116-2130,
463 doi:10.1093/molbev/msab015 (2021).
- 464 25 Booth, T. J., Bozhüyük, K. A. J., Liston, J. D., Lacey, E. & Wilkinson, B. Bifurcation drives the
465 evolution of assembly-line biosynthesis. *bioRxiv*, 2021.2006.2023.449585,
466 doi:10.1101/2021.06.23.449585 (2021).
- 467 26 Nivina, A., Yuet, K. P., Hsu, J. & Khosla, C. Evolution and Diversity of Assembly-Line Polyketide
468 Synthases. *Chem Rev* **119**, 12524-12547, doi:10.1021/acs.chemrev.9b00525 (2019).

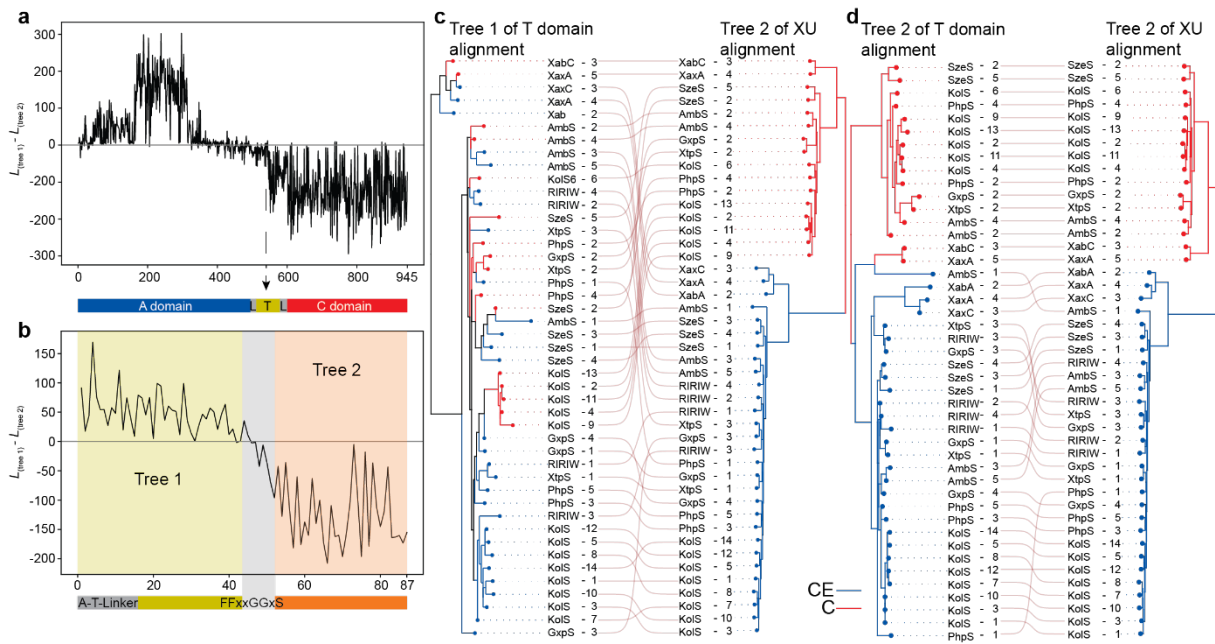
- 469 27 Vander Wood, D. A. & Keatinge-Clay, A. T. The modules of trans-acyltransferase assembly lines
470 redefined with a central acyl carrier protein. *Proteins* **86**, 664-675, doi:10.1002/prot.25493
471 (2018).
- 472 28 Keatinge-Clay, A. T. Polyketide Synthase Modules Redefined. *Angew Chem Int Ed Engl* **56**,
473 4658-4660, doi:10.1002/anie.201701281 (2017).
- 474 29 Nivina, A., Herrera Paredes, S., Fraser, H. B. & Khosla, C. GRINS: Genetic elements that recode
475 assembly-line polyketide synthases and accelerate their diversification. *Proc Natl Acad Sci U S*
476 *A* **118**, doi:10.1073/pnas.2100751118 (2021).
- 477 30 Calcott, M. J., Owen, J. G. & Ackerley, D. F. Efficient rational modification of non-ribosomal
478 peptides by adenylation domain substitution. *Nat Commun* **11**, 4554, doi:10.1038/s41467-
479 020-18365-0 (2020).
- 480 31 Kries, H., Niquille, D. L. & Hilvert, D. A subdomain swap strategy for reengineering
481 nonribosomal peptides. *Chem Biol* **22**, 640-648, doi:10.1016/j.chembiol.2015.04.015 (2015).
- 482 32 Crüsemann, M., Kohlhaas, C. & Piel, J. Evolution-guided engineering of nonribosomal peptide
483 synthetase adenylation domains. *Chem. Sci.*, 1041-1045, doi:10.1039/C2SC21722H (2013).
- 484 33 Meyer, S. *et al.* Biochemical Dissection of the Natural Diversification of Microcystin Provides
485 Lessons for Synthetic Biology of NRPS. *Cell Chem Biol* **23**, 462-471,
486 doi:10.1016/j.chembiol.2016.03.011 (2016).
- 487 34 Boussau, B., Gueguen, L. & Gouy, M. A mixture model and a hidden markov model to
488 simultaneously detect recombination breakpoints and reconstruct phylogenies. *Evol Bioinform*
489 *Online* **5**, 67-79, doi:10.4137/ebo.s2242 (2009).
- 490 35 Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation
491 domains in nonribosomal peptide synthetases. *Chem Biol* **6**, 493-505, doi:10.1016/S1074-
492 5521(99)80082-9 (1999).
- 493 36 Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. Phylogenetic analysis of
494 condensation domains in NRPS sheds light on their functional evolution. *BMC Evol Biol* **7**, 78,
495 doi:10.1186/1471-2148-7-78 (2007).
- 496 37 Wheadon, M. J. & Townsend, C. A. Evolutionary and functional analysis of an NRPS
497 condensation domain integrates beta-lactam, D-amino acid, and dehydroamino acid synthesis.
498 *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2026017118 (2021).
- 499 38 Fuchs, S. W., Proschak, A., Jaskolla, T. W., Karas, M. & Bode, H. B. Structure elucidation and
500 biosynthesis of lysine-rich cyclic peptides in *Xenorhabdus nematophila*. *Org Biomol Chem* **9**,
501 3130-3132, doi:10.1039/c1ob05097d (2011).
- 502 39 Vo, T. D., Spahn, C., Heilemann, M. & Bode, H. B. Microbial Cationic Peptides as a Natural
503 Defense Mechanism against Insect Antimicrobial Peptides. *ACS Chem Biol* **16**, 447-451,
504 doi:10.1021/acscchembio.0c00794 (2021).
- 505 40 Niehs, S. P. *et al.* Genome Mining Reveals Endopyrroles from a Nonribosomal Peptide
506 Assembly Line Triggered in Fungal-Bacterial Symbiosis. *ACS Chem Biol* **14**, 1811-1818,
507 doi:10.1021/acscchembio.9b00406 (2019).
- 508 41 Wang, X. *et al.* Discovery of recombinases enables genome mining of cryptic biosynthetic gene
509 clusters in Burkholderiales species. *Proc Natl Acad Sci U S A* **115**, E4255-E4263,
510 doi:10.1073/pnas.1720941115 (2018).
- 511 42 Isogai, A. *et al.* Structural analysis of new syringopeptins by tandem mass spectrometry. *Biosci*
512 *Biotechnol Biochem* **59**, 1374-1376, doi:10.1271/bbb.59.1374 (1995).
- 513 43 Sieber, S. A. & Marahiel, M. A. Molecular mechanisms underlying nonribosomal peptide
514 synthesis: approaches to new antibiotics. *Chem Rev* **105**, 715-738, doi:10.1021/cr0301191
515 (2005).
- 516 44 Zhou, Z., Lai, J. R. & Walsh, C. T. Directed evolution of aryl carrier proteins in the enterobactin
517 synthetase. *Proc Natl Acad Sci U S A* **104**, 11621-11626, doi:10.1073/pnas.0705122104 (2007).
- 518 45 Lai, J. R., Koglin, A. & Walsh, C. T. Carrier protein structure and recognition in polyketide and
519 nonribosomal peptide biosynthesis. *Biochemistry* **45**, 14869-14879, doi:10.1021/bi061979p
520 (2006).

- 521 46 Reimer, J. M. *et al.* Structures of a dimodular nonribosomal peptide synthetase reveal
522 conformational flexibility. *Science* **366**, doi:10.1126/science.aaw4388 (2019).
- 523 47 Nollmann, F. I. *et al.* Insect-specific production of new GameXPeptides in photorhabdus
524 luminescens TTO1, widespread natural products in entomopathogenic bacteria. *Chembiochem*
525 **16**, 205-208, doi:10.1002/cbic.201402603 (2015).
- 526 48 Zhou, Q. *et al.* Structure and biosynthesis of xenoamicins from entomopathogenic
527 *Xenorhabdus*. *Chemistry* **19**, 16772-16779, doi:10.1002/chem.201302481 (2013).
- 528 49 Schimming, O., Fleischhacker, F., Nollmann, F. I. & Bode, H. B. Yeast homologous
529 recombination cloning leading to the novel peptides ambactin and xenolindicin. *Chembiochem*
530 **15**, 1290-1294, doi:10.1002/cbic.201402065 (2014).
- 531 50 Baumeister, W., Walz, J., Zuhl, F. & Seemuller, E. The proteasome: paradigm of a self-
532 compartmentalizing protease. *Cell* **92**, 367-380, doi:10.1016/s0092-8674(00)80929-0 (1998).
- 533 51 Scott, K., Hayden, P. J., Will, A., Wheatley, K. & Coyne, I. Bortezomib for the treatment of
534 multiple myeloma. *Cochrane Database Syst Rev* **4**, CD010816,
535 doi:10.1002/14651858.CD010816.pub2 (2016).
- 536 52 Lin, G., Li, D., Chidawanyika, T., Nathan, C. & Li, H. Fellutamide B is a potent inhibitor of the
537 *Mycobacterium tuberculosis* proteasome. *Arch Biochem Biophys* **501**, 214-220,
538 doi:10.1016/j.abb.2010.06.009 (2010).
- 539 53 Deshpande, S., Altermann, E., Sarojini, V., Lott, J. S. & Lee, T. V. Structural characterization of
540 a PCP-R didomain from an archaeal nonribosomal peptide synthetase reveals novel
541 interdomain interactions. *J Biol Chem* **296**, 100432, doi:10.1016/j.jbc.2021.100432 (2021).
- 542 54 Kavanagh, K. L., Jornvall, H., Persson, B. & Oppermann, U. Medium- and short-chain
543 dehydrogenase/reductase gene and protein families : the SDR superfamily: functional and
544 structural diversity within a family of metabolic and regulatory enzymes. *Cell Mol Life Sci* **65**,
545 3895-3906, doi:10.1007/s00018-008-8588-y (2008).
- 546 55 Persson, B. & Kallberg, Y. Classification and nomenclature of the superfamily of short-chain
547 dehydrogenases/reductases (SDRs). *Chem Biol Interact* **202**, 111-115,
548 doi:10.1016/j.cbi.2012.11.009 (2013).
- 549 56 Tietze, A., Shi, Y. N., Kronenwerth, M. & Bode, H. B. Nonribosomal Peptides Produced by
550 Minimal and Engineered Synthetases with Terminal Reductase Domains. *Chembiochem* **21**,
551 2750-2754, doi:10.1002/cbic.202000176 (2020).
- 552 57 Calcott, M. J. & Ackerley, D. F. Portability of the thiolation domain in recombinant pyoverdine
553 non-ribosomal peptide synthetases. *BMC Microbiol* **15**, 162, doi:10.1186/s12866-015-0496-3
554 (2015).
- 555 58 Brown, A. S., Calcott, M. J., Owen, J. G. & Ackerley, D. F. Structural, functional and evolutionary
556 perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines.
557 *Nat Prod Rep* **35**, 1210-1228, doi:10.1039/c8np00036k (2018).
- 558 59 Medema, M. H., Cimermancic, P., Sali, A., Takano, E. & Fischbach, M. A. A systematic
559 computational analysis of biosynthetic gene cluster evolution: lessons for engineering
560 biosynthesis. *PLoS Comput Biol* **10**, e1004016, doi:10.1371/journal.pcbi.1004016 (2014).
- 561 60 Gotze, S. *et al.* Structure elucidation of the syringafactin lipopeptides provides insight in the
562 evolution of nonribosomal peptide synthetases. *Chem Sci* **10**, 10979-10990,
563 doi:10.1039/c9sc03633d (2019).
- 564 61 Reimer, J. M., Haque, A. S., Tarry, M. J. & Schmeing, T. M. Piecing together nonribosomal
565 peptide synthesis. *Curr Opin Struct Biol* **49**, 104-113, doi:10.1016/j.sbi.2018.01.011 (2018).
- 566 62 Drake, E. J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide
567 synthetases. *Nature* **529**, 235-238, doi:10.1038/nature16163 (2016).
- 568 63 Hines, J., Groll, M., Fahnestock, M. & Crews, C. M. Proteasome inhibition by fellutamide B
569 induces nerve growth factor synthesis. *Chem Biol* **15**, 501-512,
570 doi:10.1016/j.chembiol.2008.03.020 (2008).

571

572

573 **Figures**



574

575

576 **Fig. 1. Evolutionary analysis of ATC tri-domains and T domains of representative NRPS.** (a)

577 Likelihood difference plot of two phylogenetic trees of ATC tri-domains (also called XUs) that together

578 best describe the alignment using a phylogenetic hidden Markov model. Positive numbers indicate that

579 sites are better describe by tree 1, negative numbers indicate sites that are better described by tree two.

580 (b) Likelihood difference plot as in a, but for an alignment of T domain plus A-T linker. Partitions detected

581 by the hidden Markov model are indicated in different colours according to tree number. Recombination

582 breakpoint is annotated in grey and lies around two conserved glycine residues. (c, d) Comparison of

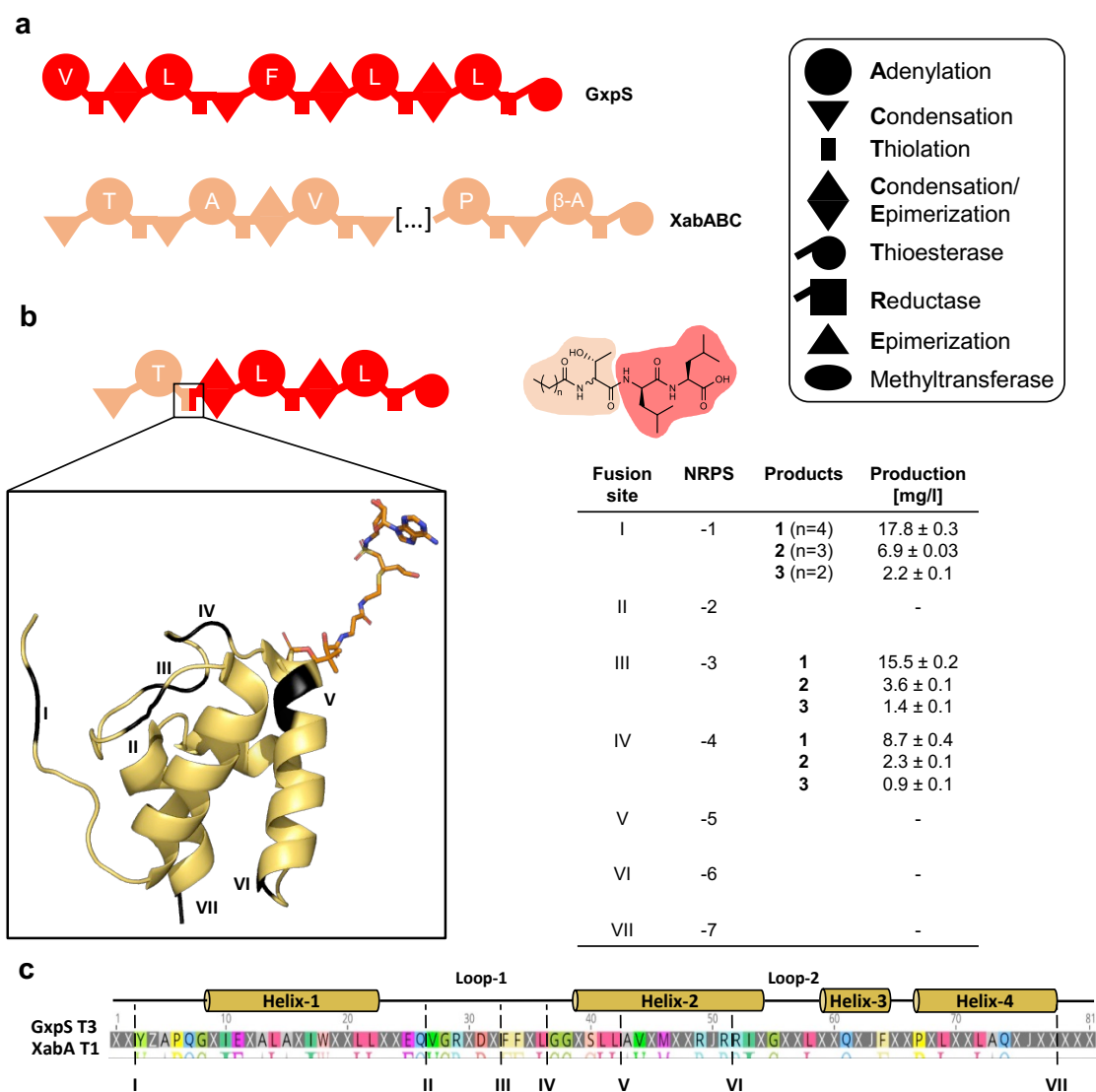
583 Tree 1 from the T domain alignment with Tree 2 from the XU alignment (left) and Tree 2 from the T

584 domain alignment with Tree 2 from the XU alignment (right). Names indicate abbreviation of NRPS and

585 numbers the XU within that NRPS. Lines connect the same NRPS and XUs between the two trees. Red

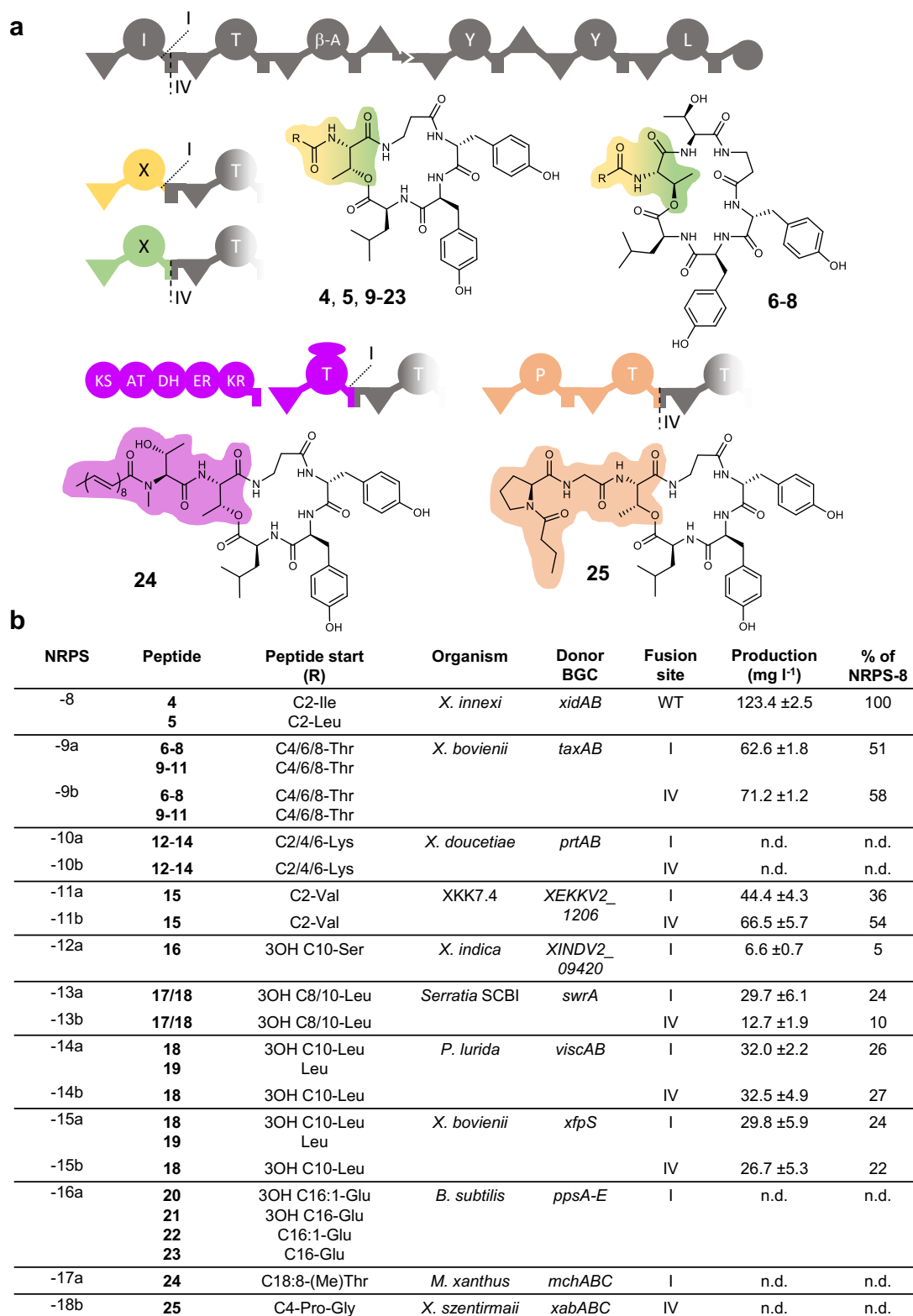
586 branches label XUs that contain ${}^L C_L$ domains, blue branches label XUs with C/E (dual C) domains.

587



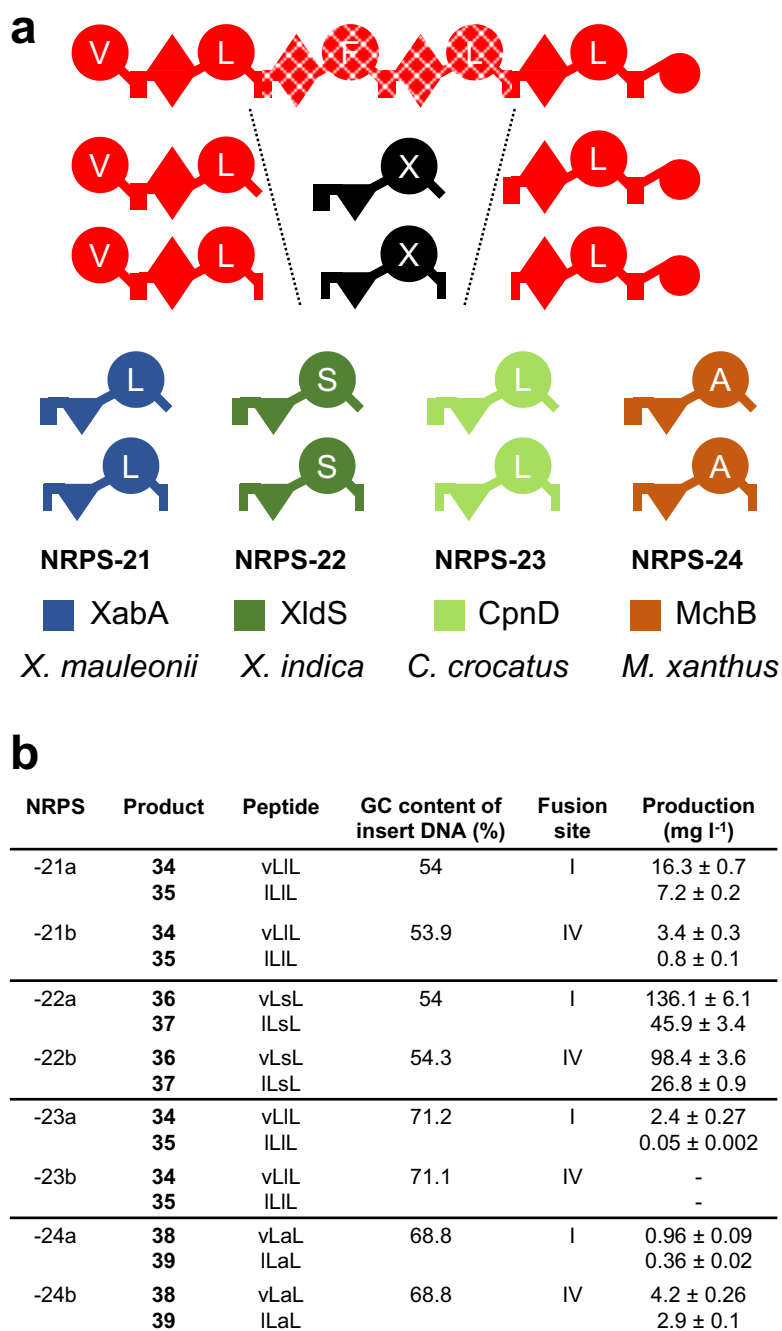
588
589
590
591
592
593
594
595
596
597

Fig. 2. Fusion point screening of a NRPS hybrid assembled within the T domain. (a) Schematic representation of precursor NRPS GxpS and XabABC, producing GameXPeptides and xenoamicins, respectively. All domain functions are explained in the box. (b) Schematic representation of the XabA-GxpS hybrid NRPS, produced lipopeptide and compounds titres. The colour code of the peptide structure follows the NRPS colour code. The different fusion sites within the T domain are highlighted in black at the respective positions in the crystal structure of the T domain EntF (PDB 4ZXJ)⁶². (c) Sequence alignment of GxpS T3 and XabA T1 with secondary structures of the T domain and fusion sites indicated.



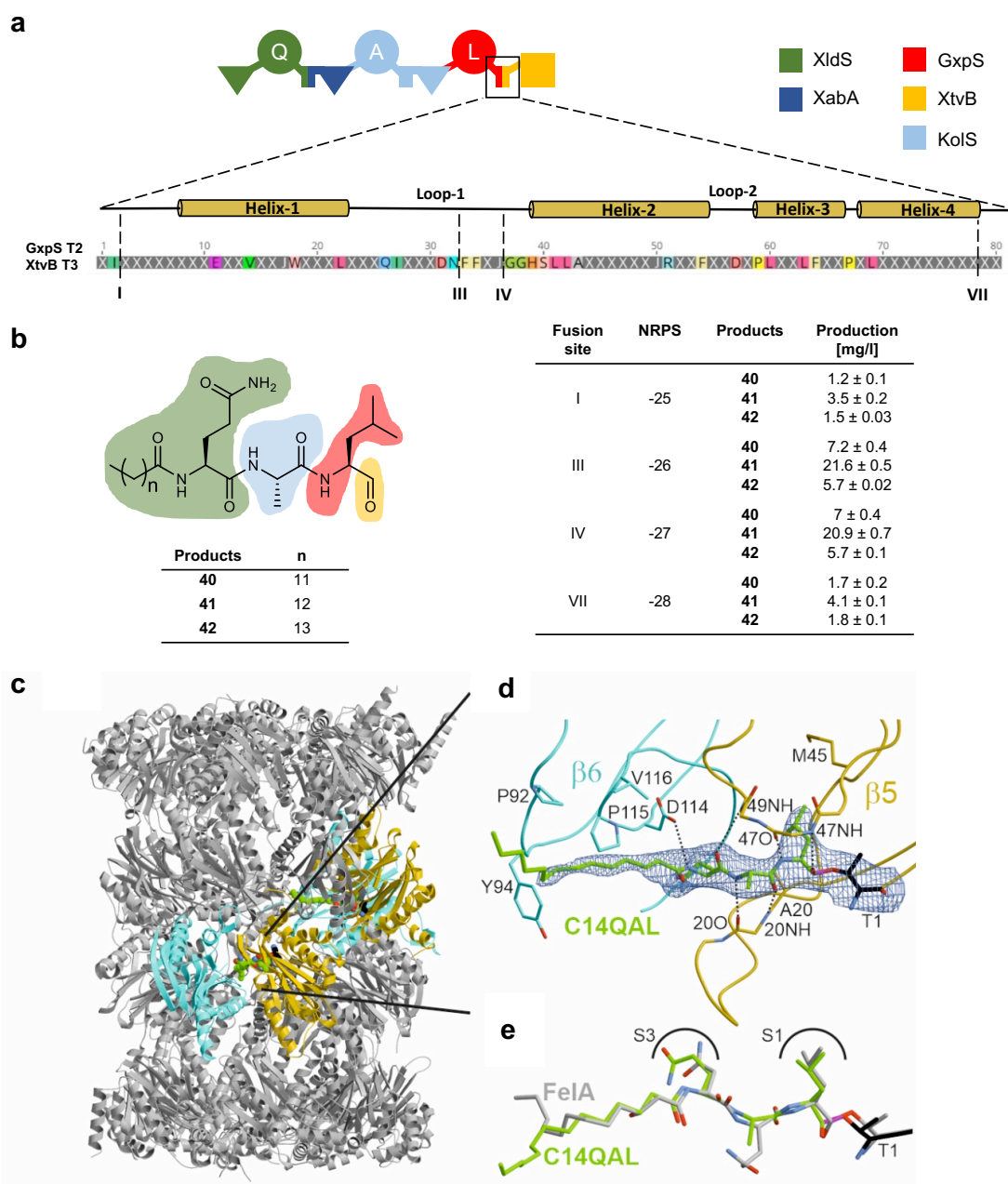
598
599
600
601
602
603
604
605
606
607

Figure 3. Evolution inspired exchange units replacing NRPS starting modules. (a) Schematic representation of the FitAB NRPS producing fitayyllide A (**4**) and B (**5**) and selected alternative starting modules from other NRPS with indicated fusion sites I and IV. Amino acid specificities are assigned for all A domains. KS (ketosynthase), AT (acyltransferase), DH (dehydratase), ER (enoylreductase). Selected structures of the produced peptides are shown so that in conjunction with the table in (b) all peptide structures can be deduced. Production data relative to the WT NRPS-8 and the absolute peptide yields are based on triplicate production cultures. The origin of the alternative starting module, their cognate gene cluster and the fusion point for each starter module is shown. Production was observed for all NRPS derivatives, but production titres were not determined for all of them (n.d.).



608
609
610
611
612
613
614
615

Fig. 4. Evolution inspired exchange units replacing internal modules. (a) Schematic representation of the precursor NRPS GxpS producing GameXPepptides. A T₂-T₄ fragment was exchanged with different XUTs from the xenoamicine, xenolindicine, crocapeptin and myxochromide producing NRPS XabA, XldS, CpnD and MchC, respectively. (b) Peptides, GC content of the inserted XUTs, corresponding fusion sites and production titres of the respective peptides.



616

617 **Fig. 5. XUT approach for the design of a proteasome inhibitor.** (a) Schematic representation of
 618 reassembled NRPS-25 to -28 composed of NRPS fragments from XldS, XabA, KolS (kolossin), GxpS
 619 and XtvB. The terminal T domain is shown as a sequence alignment of GxpS T2 and XtvB T3 indicating
 620 secondary structures and fusion sites. (b) Production titres corresponding to the fusion sites within the
 621 terminal T domain. The colour code in the peptides follows that of the NRPS fragments used. (c) Crystal
 622 structure of the yeast 20S proteasome in complex with **41** (spherical model, green carbon atoms) bound
 623 to the chymotrypsin-like active sites ($\beta 5$ subunits, gold, PDB ID 8BW1). (d) Illustration of the $2F_o - F_c$
 624 electron density map (blue mesh, contoured to 1σ) of **41** (depicted as C14-QAL) covalently linked
 625 through a hemiacetal bond (magenta) to Thr10^Y. Protein residues interacting with **41** are highlighted in
 626 black. Dots illustrate hydrogen bonds between **41** and protein residues. (e) Superposition of **41** (depicted
 627 as C14-QAL) (green) and fellutamide A (grey, PDB ID s3D29)⁶³ complex structures highlighting similar
 628 conformations at the chymotrypsin-like active site.