Supplementary Information for: “Dynamic thiolation-thioesterase di-domain structure of a non-ribosomal peptide synthetase.”

Dominique P. Frueh¹, Haribabu Arthanari¹, Alexander Koglin¹, David A. Vosburg²†, Andrew E. Bennett¹, Christopher T. Walsh¹, Gerhard Wagner¹.

¹Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

Content:

- Table 1| NMR structure statistics
- Supplementary Figure 1| Domain organisation and amino-acid sequence.
- Supplementary Figure 2| Effect of the S48A mutation on the T/TE interaction.
- Supplementary Figure 3| Ensemble of structures.
  ♦ p6: Further discussion on the structure of the T-TE didomain of EntF.
- Supplementary Figure 4| comparison of NRPS T-domains.
- Supplementary Figure 5| comparison of selected T-domains.
- Supplementary Figure 6| comparison of NRPS TE-domains.
- Supplementary Figure 7| comparison of TE domains.
  ♦ p10: Discussion on dynamics in the T-TE didomain of EntF.
- Supplementary Figure 8| Distribution of \(^{1}\text{H}-^{15}\text{N}\) HSQC cross-peak signal intensities.
- Supplementary Figure 9| Solvent exchange.
- Supplementary Figure 10| Two conformers in the T-TE fragment.
  ♦ p16: Titrations of EntF T-TE with interacting proteins and protein domains.
- Supplementary Figure 11| Secondary shifts in the interaction with Sfp.
- Supplementary Figure 12| Interactions with phosphopantetheinyl transferases.
- Supplementary Figure 13| Interactions with condensation domains.
Table 1 NMR structure statistics

<table>
<thead>
<tr>
<th>Violations (mean and s.d.)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance constraints (Å)</td>
<td>0.24 +/- 0.12</td>
</tr>
<tr>
<td>Dihedral angle constraints (°)</td>
<td>3.12 +/- 0.57</td>
</tr>
<tr>
<td>Max. dihedral angle violation (°)</td>
<td>3.93</td>
</tr>
<tr>
<td>Max. distance constraint violation (Å)</td>
<td>0.47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deviations from idealized geometry</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond lengths (Å)</td>
<td>0.095</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.17</td>
</tr>
<tr>
<td>Impropers (°)</td>
<td>1.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average pairwise r.m.s.d. (Å) [residues 10–337]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy</td>
<td>2.74 +/- 0.36</td>
</tr>
<tr>
<td>Backbone</td>
<td>2.23 +/- 0.38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average r.m.s.d to mean [Cα alignment for residues in regular secondary structures] (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>residues 10–337</td>
</tr>
<tr>
<td>T domain</td>
</tr>
<tr>
<td>TE domain</td>
</tr>
</tbody>
</table>
Supplementary Figure 1] Domain organisation and amino-acid sequence. Top, The domain boundaries are indicated above the cartoon, whereas the numbering used in the paper is shown below the T-TE di-domain. Amino-acids in red were introduced for the structural studies described here. Bottom, Amino-acid sequence of full-length EntF. The domains are coloured according to the figure on top.
Supplementary Figure 2 | Effect of the S48A mutation on the T/TE interaction. 

a, Difference in chemical shifts between the apo T-TE and the S48A mutant T-TE. The residues that are close in space to residue 48 are the only residues to display significant differences due to the replacement of the electronegative hydroxyl group by a proton. These include the T-domain S48A neighboring residues (49 and 50) as well as the TE domain residues that are in contact with the T domain in the vicinity of S48A (residues 240-244). 

b, Differences of TE chemical shifts between the apo T-TE di-domain and the excised TE domain. 

c, Differences of TE chemical shifts between the S48A T-TE mutant and the excised TE domain. 

b and c indicate the effect of the T domain (wild type or mutant) on the TE domain. The preservation of the pattern indicates that the T/TE interaction is not significantly affected by the mutation.
Supplementary Figure 3 | Ensemble of structures. The 15 models with lowest energy and fewest constraint violations (out of a set of 200 calculated structures) are shown. a, Alignment on the Cα atoms in residues that belong to regular secondary structures in the T domain (rmsd = 0.6 Å). b, Same for the TE domain (rmsd = 1.0 Å). c, Same on the whole T-TE fragment (rmsd = 1.2 Å). The average pairwise r.m.s.d. for residues 10–337 was 2.74 Å for heavy atoms and 2.23 Å for backbone atoms.
Further discussion on the structure of the T-TE didomain of EntF.

After solving the EntF T-TE di-domain it is interesting to compare some features of related proteins. However, a detailed review of the structures of carrier proteins and thioesterases is beyond the scope of this publication. There are currently two structures available for peptidyl carrier T domains\(^1, 2\) and two for TE domains\(^3, 4\). The single-turn helix \(\alpha1^T\) previously found in the midst of loop \(L1_T\) of the EntF T-domain (see main text) is not observed in the excised TycC3 T domain nor in the TycC5 T domain of the TycC5-6 T-C di-domain (this T-domain essentially corresponds to the A/H state of TycC3 and is not shown in Supplementary Figure 4). However, both TycC structures feature a turn in the loop, at the same location. A comparison with the structure of the T domain of TycC3 in both A and A/H states is discussed below (Discussion on dynamics in the T-TE didomain of EntF).

Supplementary Figure 4| comparison of NRPS T-domains. a, structure of the EntF T-domain as observed in the T-TE di-domain. The mutated active site S48A is displayed in blue. b, A/H state of the TycC3 T-domain and c, A state of the same molecule. The S48A mutation was expected to select for the A state; in contrast, we observe that the T-domain conformation in the in EntF T-TE di-domain closely resembles the A/H state.
We note that such a single-turn helix has been observed for a number of free
standing ACP-type T domains, as shown in Supplementary Fig. 5. Indeed, although carrier
proteins exhibit a remarkably well-conserved fold, a right-handed helical bundle, their
structures differ substantially in this region. In particular, the loop may or may not contain
this single-turn helix. Our results indicate that this helix is involved in the T/TE interaction
for NRPS systems and may play a key role in domain recognition in general. Future protein
interaction studies, including mutation analysis, are however needed to determine how
general or specific the role of this helix is.

Supplementary Figure 5 | comparison of selected T-domains. a, Structure of the spinach fatty
acid synthase ACP-type T domain. b, Structure of the EntF PCP-type T domain as observed in the
T-TE di-domain. c, Structure of the oxytetracycline polyketide synthase ACP-type T domain.

Comparison between the EntF TE domain (in the T-TE fragment) and the Surfactin
TE (Srf TE), or the Fengamycin TE (Fen TE) shows that the fold of the core of the domains
is well conserved (Supplementary Figure 6). The major differences occur in the lid region,
also referred to as the “webbed fingers” in EntF TE. The tip of the fingers, which interact
with the T-domain, differs significantly between EntF TE and Srf TE. A flexible loop links
α4_{TE} and α5_{TE} in EntF, whereas this region features a turn in Srf TE (circled in red).

Although a significant portion of the lid region is invisible in Fen TE, the presence of a nick in α5_{TE} (the label follows the one used in the main text) is again indicative of structural diversification.

Supplementary Figure 6 | comparison of NRPS TE-domains. a, Structure of the EntF TE-domain as observed in the T-TE di-domain. b, Structure of Srf TE in the closed state. c, Structure of the Srf TE in the open state. d, Structure of Fen TE.

In addition to these major differences, others occur in the α/β sandwich cores. Loop L4_{TE} is substantially longer in EntF, in which it features a region with helical character (circled in red in Supplementary Figure 6). This may help in shaping the bottom of the canyon that will accommodate the (loaded) 4’-PP arm. A proline induces a nick in helix α6_{TE} of EntF and the helical turn in L12_{TE} and the beginning of α1_{TE} both are ill-defined.
This may again be the result of dynamic events perturbing those regions. Some or all of these features may reflect the fact that the EntF TE is part of an iterative enzyme in which different intermediates are stored covalently on S180, whereas the Srf and Fen TE domains have the task to cyclise and release the product presented by the upstream T domain. Finally, the region between $\beta_{6\text{TE}}$ and $\beta_{7\text{TE}}$ differs in all three proteins. Lack of complete assignments for this region, which is likely due to mobility, prevents a thorough comparison. However, the beginning of this stretch of residues is clearly different between EntF and the other two TEs with available structures.

Supplementary Figure 7| comparison of TE domains. a, Structure of the human FAS TE domain. b, Structure of the EntF TE-domain as observed in the T-TE di-domain. c, Structure of the erythromycin polyketide synthase TE domain, Ery TE. The differentiated “lid-region” is shown in green, with the two N-terminal helices of the Ery TE in light green.

There is currently no structural information on ACP-type T/TE interactions in related fatty acid synthases and polyketide synthases. The conserved fold of carrier proteins mentioned above together with a preserved core in type I TE domains may indicate that similar interactions take place in PKS and FAS systems. However, the lid-region of NRPS TE is drastically different in PKS and in FAS (Supplementary Fig. 7). For PKS, the helix that would correspond to $\alpha_{4\text{TE}}$ has a totally different orientation that generates a substrate channel, in contrast to NRPS TEs, and is linked to the subsequent helix via an extensive
loop. In addition, two helices attached at the N-terminus of the first beta strand pack with the lid region, thus mimicking a subdomain. In type I FAS TE, this region is even further differentiated and consists of a larger and purely alpha-helical sub-domain. Thus, for FAS thioesterases the substrates are expected to interact in the crevice at the interface of the two subdomains. All these differences indicate that further T/TE structural investigations are required for each system to allow for a complete understanding of termination processes in these multimodular assembly lines. Our results provide the first step in structural comparisons of T/TE domains interactions in these parented systems.

**Discussion on dynamics in the T-TE didomain of EntF.**

We mentioned in the main text that the EntF T domain was subject to extensive dynamics. Two conformational states have indeed been reported for the homologous excised TycC3 T domain\(^1\): an A/H state that is common to both apo (A) and holo (H) forms, and an A state that is only found in the apo form (see Supplementary Fig. 4). Helix \(\alpha_3T\) of the A/H state becomes a loop, \(L_3T\), in the A state, and the well-defined inter-helical interactions of the A/H state are relaxed in the A state (see Fig. 2 in the paper for its location). Although the S48A mutation in EntF was expected to lock the conformation into the A state as was reported for the TycC3 T domain\(^1\), the structure of EntF T in the T-TE di-domain is closer to the A/H form. All three helices of the T domain have orientations similar to those in the A/H state in the TycC3 T domain, and although helix \(\alpha_3T\) is not clearly formed, loop \(L_3T\) still displays helical characteristics. Thus, the TE domain likely stabilises the A/H state in which more hydrophobic residues protrude to form the dimer interface. In spite of this stabilisation, many regions of the T domain seem to experience slow internal motions, as indicated by the low intensities of their resonances. The location of these regions mostly coincides with the T/TE interaction surface (See Fig. 2 and
Supplementary Fig. 8). In this case, the observed reduction in intensity may be due to either local dynamics or to a fluctuation of the T/TE interaction or both (see below for additional support of a dynamic T/TE interface). Further support for intra-domain dynamics is provided by the distribution of residues undergoing fast NH exchange (Supplementary Fig. 9). While the core of the TE domain is not solvent accessible, both the T domain and the lid region of the TE domain are dynamic enough to undergo rapid exchange with solvent molecules.

Supplementary Figure 8 | Distribution of $^1$H-$^{15}$N HSQC cross-peak signal intensities. Bottom: All residues. Top: Structured part excluding the N-terminal flexible segment. The locations of secondary structures are indicated on top, and elements discussed in the text are assigned. The average value ± standard deviation are indicated with solid and broken lines, respectively. Note that high intensities indicate fast motions (ps-ns, e.g. in N- and C-termini and exposed loops), whereas low intensities or missing signals indicate slow motions (μs-ms). The regions highlighted are those discussed in the main text: $\alpha'_{T}$, S48A and L50, L3$_{T}$, $\beta_{TE}$, the end of $\alpha_{TE}$ and L9$_{TE}$ are all involved in the T/TE interaction surface. Their low intensities indicate a modulation of their magnetic environment, which is suggestive of slow motions. This is in agreement with a transient contact between the two domains. In addition, L12$_{TE}$ and the residues of $\alpha_{4TE}$ that face it in the structure also have lower intensities. These two regions form the canyon discussed in the main text. Interestingly, the kink in helix $\alpha_{6TE}$ also indicates slow dynamics. This further supports the presence of several slightly different conformations in these regions.
**Supplementary Figure 9| Solvent exchange.** a, The location of rapidly exchanging amide protons (red) shows that the T domain and the finger region must be mobile. A double-headed arrow emphasizes that the flap formed by the $\alpha_{4TE}$ and $\alpha_{5TE}$ helices is relatively mobile and opens frequently. b, residual amide proton signals after 1 hr in D$_2$O, pH 6.7.
Dynamics seems to be a key feature of the S48A T-TE fragment. Further evidence for this derives from a comparison of NMR data of the T-TE di-domain and the TE domain alone. Signals belonging to a second form of the T-TE molecule in slow exchange with the main state were identified, although their intensities were too weak to allow for detailed structural studies or even complete assignments (Supplementary Fig. 10). Comparison with the spectrum of the excised TE domain showed that these signals correspond to a large extent to those in the single TE domain (Fig. 3b,c). This suggests the presence of an open conformer, T-TE$^0$, in which the TE and T domains interact little or not at all. Surprisingly, the residues that feature two states are not only located in the T/TE interface, but also in the ‘finger’ region (Fig. 3a,b), in agreement with a mobility of that region. When the PPTase Sfp is added to a solution of the S48A T-TE di-domain, all signals of the T-TE$^0$ conformer increase (they were absent from the reference spectrum, under measuring conditions), suggesting that conformational changes in the ‘finger’ regions are at least in part related to the open and closed form of the T-TE di-domain. Residual dipolar couplings$^9,10$ measured for the T-TE fragment (data not shown) indicated a complex dynamic behaviour of the T-TE di-domain, and consequently were not used for structure refinement or domain orientation. Thus, in addition to the internal mobility described above, the domain interaction in the T-TE fragment itself is dynamic. Unfortunately, the residues affected by these motions display low signal intensities (Supplementary Fig. 8), preventing detailed dynamic studies by NMR.

These dynamic events were observed in the inactive S48A mutant, which corresponds to the apo form of the molecule. In the holo protein, the presence of the phosphopantetheinyl arm is likely to influence both structural and dynamic properties of the molecule. For instance, additional interactions between the 4’-PP arm and the TE domain may stabilize the closed form of T-TE and select for a conformational state, in particular in
the canyon delimited by L12\textsubscript{TE} and \(\alpha_4\textsubscript{TE}\), in which the prosthetic arm is expected to reside. Similar conformational changes, or selection of conformational states, are likely to occur when the 4’-PP arm and S180 are loaded with the various intermediates. Indeed, although the structure of the S48A mutant places the two active sites at a distance that can be spanned by the prosthetic group, the position of \(\alpha_4\textsubscript{TE}\) in the closed form would not allow for the arm to navigate in and out of the canyon, as needed during enterobactin synthesis. We thus expect conformational changes in this region to occur in the major closed state of the molecule as well, as observed (Supplementary Figs. 8 and 9).

Clearly, the modulation of the dynamics that was observed upon addition of the PPTase Sfp only relates to the apo form of the protein, since the holo form of T-TE need not (and does not\textsuperscript{1}) interact with the priming enzyme. On the other hand, it is hard to predict how the phosphopantetheiny1 arm will affect the dynamic interaction between the T and C domains. Our results indicate that, even in the absence of the 4’-PP arm, there is an interaction between the C and T domains, which modulates the T/TE interaction without disrupting it. In the full-length protein, and during enterobactin synthesis, dynamic events similar to those reported here must take place since the domains need to interact differently for different catalytic steps. In particular, the 4’-PP arm, which is present in the holo protein, need to navigate between the various domains during the biosynthesis and thus only interacts \textit{transiently} with each domain. The structure of the apo (mutant) T-TE presented here, together with its associated dynamics, may thus also represent the T/TE interaction in the holo protein, albeit when the prosthetic group is not interacting with the TE domain.
Supplementary Figure 10 | Two conformers in the T-TE fragment. Top, For many residues, signals indicative of a second form of the T-TE di-domain were observed in $^1$H-$^{15}$N HSQC spectra. These very weak signals could only be detected after long acquisition times (ca. 12h for the spectrum above, recorded at 900 MHz with a cryoprobe). For the signals belonging to the TE domain, a comparison with the spectrum of the excised TE domain suggests that these correspond to a conformer in which the two domains do not interact (see paper). Bottom, For the very intense signals of the N-terminal fragment of the protein (see Supplementary Figure 8), we could observe exchange-mediated cross-peaks in a NOESY spectrum with a mixing time of 200 ms (bottom right). This correlation proves that these small signals indeed represent two inter-converting conformations and not a chemically different form of the molecule. The nOe cross-peaks that occur within a conformer are indicated by rectangles, while the exchange cross-peaks that correlate the two conformers are indicated by ovals.
Titrations of EntF T-TE with interacting proteins and protein domains. The T-TE samples were dissolved in the same buffer as the titrant, and their resonances were reassigned. Spectra with various amount of titrant were recorded, starting from the highest concentration. We could therefore verify that no degradation induced by the titrant occurred during the titration, which would have led to spurious modifications of the spectra. The T-TE fragment was titrated with 1, 0.5 and 0.25 equivalents of Sfp, with 2 and 1 equivalents of each EntD and AcpS, with 4 and 2 equivalents of the TycC5 C-domain, and with 4, 3, 2 and 1 equivalents of the EntF C-domain. The reported chemical shift differences are calculated with $\Delta \omega^1_{(+)} = \left( (\Delta \omega_H)^2 + \left( \frac{1}{10} \Delta \omega_N \right)^2 \right)^{1/2}$, where $\Delta \omega_i$ is the chemical shift difference between the two species for nucleus $i$. 

$\text{doi: 10.1038/nature07162}$
Supplementary Figure 11| Secondary shifts in the interaction with Sfp. **Bottom,** Difference in chemical shifts of the S48A T-TE signals with and without Sfp, for the major conformer of S48A T-TE. **Top,** The colour code depicts the magnitude of the chemical shift difference from blue (lowest) to red (highest). The residues in green disappeared upon addition of Sfp. As mentioned in the text, in addition to an increase in the population of the second T-TE form (with no T/TE domain interaction), the signals of the major conformer are shifted. This indicates residues that are subject to a modification of their environment in the fast-exchange time-scale (μs). These can be due to either a weak interaction with Sfp, or to a modulation of the T-TE internal dynamics by the Sfp interaction. Interestingly many of these regions are subject to both modifications: the stabilisation of the “open” conformer and the secondary modulation. The secondary shifts also reveal that helix α1T is affected by Sfp, as has been previously reported on the excised TycC3 T domain1.
Supplementary Figure 12 | Interactions with phosphopantetheinyl transferases. To further investigate the relevance of the selection of the open state by PPTases, three representative transferases were added to a solution of T-TE: the non-specific Sfp, the enterobactin synthetase-dedicated EntD and the acyl carrier specific AcpS. a, S48A T-TE in absence of Sfp. b, with one equivalent of Sfp, the population of the open state increases (signals denoted with suffix B). c, S48A T-TE in absence of EntD. d, with two equivalent of EntD, a selection of the open state is again observed. e, S48A in absence of AcpS. f, Even with two equivalents of AcpS, no selection of the open state is observed. As expected, the interaction with the T-domain, and the accompanying selection of the open state, is only observed with peptidyl carrier protein phosphopantetheinyl transferases (Sfp and EntD) and not with the specific acyl carrier protein synthase AcpS, which selectively modifies acyl carrier proteins. We note that, even with an excess of PPTase, the population of the open state seems lower in presence of EntD than Sfp. This apparent delayed selection of the open state by EntD may be the result of the presence of surfactant, which was needed to solubilise the protein at high concentrations.
Supplementary Figure 13 | Interactions with condensation domains. The native interaction between the EntF C and T domains was compared to the heterologous interaction between the EntF T-domain and the TycC5 C-domain. A T/C recognition is only observed in the case of the native EntF interaction suggesting a T/C specific recognition. Reference spectra are in black, while the spectra of the S48A T-TE di-domain in presence of the respective C-domain are in red. a, S48A T-TE with four equivalents of EntF C-domain. b, S48A T-TE with 4 equivalents of TycC5 C-domain. In contrast to interactions with peptidyl PPTases, no selection of the open state was observed.
References