Acta Crystallographica Section D

Volume 70 (2014)

Supporting information for article:

Structure of the epimerization domain of tyrocidine synthetase A

Stefan A. Samel, Paul Czodrowski and Lars-Oliver Essen

Structure of the epimerization domain of tyrocidine synthetase A

Stefan A. Samel[a], Paul Czodrowski[b] and Lars-Oliver Essen[a]*

[a]Biomedical Research Centre/FB15, Philipps Universität, Hans-Meerwein-Strasse 4, Marburg, 35032, Germany

[b]Computational Chemistry, Merck KGaA, Frankfurter Strasse 250, Darmstadt, 64293, Germany

Correspondence email: essenl@staff.uni-marburg.de

Supporting Information

S1. **Modelling and pK calculations of the PCP-TycA-E bidomain**

To derive the approximate orientation and location of the cofactor 4'-phosphopantetheine within the active site channel of TycA-E, the structure of trichothecene 3-O-acetyltransferase TRI101 in complex with coenzyme A and the inhibitor T-2 (pdb code: 2zba) (Garvey *et al.*, 2008) was superposed onto the structure of TycA-E using the SSM algorithm as implemented in COOT (Emsley & Cowtan, 2004)

The homology model of the TycA-PCP domain was generated using the program Modeller 9v7 (Eswar *et al.*, 2008) as well as the sequence of TycA (GenBank entry AF004835) and the A/H-conformation of the TycC5-PCP domain (residues R1 to K82 of pdb code 2jgp, T5115-K5195 of SwissProt entry TYCC_BREPA, pair-wise sequence identity 32%). This yielded a model of the TycA PCP domain ranging from Q523 to R603. The experimental structure of the epimerization domain includes the position of its N-terminus up to the subsequent residue K604. The PCP domain could be thus directly linked to this residue giving already one fixed point as a constraint for obtaining valid PCP-E arrangements. In order to form a model for a catalytically competent state of the PCP-E bidomain, the PCP-domain was rotated around this point till the functional serine, S563, fulfilled the second constraint by being within bonding distance of the placed cofactor's phosphate group. Given the previously derived location of the 4'-Ppan cofactor only one initial orientation of the PCP-domain along the E-domain's surface was found to be sterically plausible. The MEAD 2.2.9 program suite (Bashford & Gerwert, 1992, Bashford, 1997) was used to solve the Poisson-Boltzmann equations and to derive the protonation states of the individual PCP- and E-domains. The PEOE_PB charges, which were optimized for pK value estimations, were used as partial charges in MEAD calculations (Czodrowski *et al.*, 2006). The robustness of obtained pK_a values on parameters and used calculation method was firstly checked by increasing the dielectric constant to 20 instead of 6 for the protein

portion, and secondly by using the H++-server with standard parameters (Gordon *et al.*, 2005). With calculated pK_a values of 9.91 and 8.93 (ε=6), respectively, H743 and H880 were found in the protonated state (ε=20: 7.80, 8.37) whereas E882 (ε=6: 0.87, ε=20: 2.89) is deprotonated. With calculated pK_a values of 8.8, 8.0 and <0.0 H++ predicted the same protonation states for H743, H880 and E882, respectively.

S2. Setup of molecular dynamics studies for the PCP-TycA-E bidomain

Molecular dynamics simulations were performed with the AMBER11 suite (Case *et al.*, 2010) using the ff99sb force field and a partial charge model for the Phe-4'-Ppan thioester that was derived by the program antechamber using the AM1-BCC method. Thus obtained partial charges of the thioester moiety (S: -0.20, C: 0.57, O: -0.53) are comparable to that derived by higher-level quantum-chemical calculations on alanyl-ethyl-thioester by R.E.D. tools using the RESP-A1 charge model (S: -0.36, C: 0.50, O: -0.47 (Dupradeau *et al.*, 2010)) or methyl-thioacetate (S: -0.34, C: 0.72, O: -0.56 (Deerfield II & Pedersen, 1995)). The PCP-E bidomain model was placed into a box of 90.5 x 98.0 x 95.6 Å size that was filled with 19026 TIP3 water molecules, 93 sodium and 78 chloride ions to ensure electroneutrality. The MD simulations were performed after minimization and 50000 steps preequilibration with a periodic NPT ensemble (T=300 K, p=1 atm, SHAKE constraints applied on bonds with hydrogens, non-bonded cutoff of 10 Å, particle-mesh Ewald method) for a total of 10^6 steps of 2 fs each. The trajectories were analysed using VMD (Humphrey *et al.*, 1996). After initial adjustments during the first 0.1 ns of the simulation the structure of the PCP-E bidomain remained essentially unchanged.

Energy minimization of this model of a productively docked PCP-E bidomain, where the 4'-Ppan moiety positions bound D-phenylalanine between H743 and E882, followed by relaxation using a molecular dynamics run by AMBER11 over 2 ns gave only minor structural deviations of 1.71 Å for 6249 atoms. PCP-E domain interfaces were calculated using the Protein Interfaces, Surfaces and Assemblies Service at the European Bioinformatics Institute (http://www.ebi.ac.uk/msdsrv/prot_int/pistart.html) (Krissinel & Henrick, 2007).

S3. Analysis of domain and core motif interactions within the PCP-TycA-E bidomain

The PCP-E model implies few, but prominent interactions between residues of the epimerization domain and the substrate amino acid. During the simulation the thioester-bound D-phenylalanine adopts a U-shaped conformation, where the phenyl moiety packs against parts of the 4'-Ppan cofactor (Fig. 6). Furthermore, the Cα-proton is directed towards E882 (Fig. 6B) thus being consistent with a deprotonated state for this residue. The catalytic residue H743 and the residue, to whose side chain it

forms intimate vdW interactions, W901, are both strictly conserved and part of core motifs E2 and E5, respectively. As these residues are suitably positioned to form an oxyanion hole around the thioester's carbonyl group (Fig. 6B) they might facilitate the deprotonation in the first place via stabilization of the transiently formed enol(ate). Yet another conserved residue, I619 of core motif E1, packs against the side chain of W901. As apparent in Fig. S1A the residues from core motifs E4 (green) and E5 (yellow) form as constituents of the floor loop the channel's bottom and those of core motif E7 the active site channel's ceiling. The substrate-binding pocket, which was shown to exert only little substrate selectivity, consists of Q636, G748, N965, L967 and L990. The reported low substrate specificity of the TycA-E domain that is capable to accept also larger peptidyl-like substrates for epimerization (Stein *et al.*, 2006) is reflected by voluminous cavities remaining next to the D-Phe substrate.

The inter-domain interactions implied by the model of the PCP-E bidomain are mainly located in two regions covering an overall interface area of 837 Å^2. One of these is the extended surface constituted by the floor loop. Unlike the floor loop of C-domains the floor loop of the TycA-E domain protrudes latch-like from the portal region that leads to the active site by harbouring the E-domains core motifs E4 and E5 as elongating anterior and posterior parts. This floor loop (H884-T903) is ideally suited to dock to the PCP domain during the epimerization reaction. According to this model, the E-domain's floor loop forms van-der-Waals (vdW) contacts and H-bonds to the PCP domain's helices αIII (E581-Y588) and αIV (I591-T601) as well as to the helix-connecting loop αII-αIII (Fig. S1B) and accounts for 35 % of the E-domain's surface involved in the PCP-E interface. Prominent interactions are formed by helix αIII and include complementary salt bridges from E581/K579 to R752/E756 of the E-domain. The second area of PCP-E domain contacts (Fig. S1B, surface area: ~310 Å^2) includes the N-terminal part of the αII-helix (S563-Q577) and the preceding αI-αII loop. Here, the hydrogen bonding network around S563, the site to which the prosthetic 4'-Ppan group is attached to, includes an aspartic acid residue, D562, that is only conserved in PCP domains preceding epimerization domains (PCP^E). Linne *et al.* reported that epimerization domains require the PCP domains from epimerizing modules and identified a profound difference in the sequence of the conserved motifs around the 4'-Ppan-attachment site being "GGDSI" in case of PCP^E unlike "GGHSL" as found for PCP^C domains, i.e. PCP domains from non-epimerizing modules (Linne *et al.*, 2001). In the PCP-E domain interface D562 forms up to two salt bridges to the likewise conserved residues K926 and R930 from the α9-helix of the TycA-E domain. This region may hence be important to ensure proper positioning of the PCP domain for exact delivery of 4'-Ppan-bound substrate into the active site opening. Obviously, the replacement of a positively charged histidine for a negatively charged residue such as D562 must not ablate the interaction with the down-stream condensation domain of TycB.

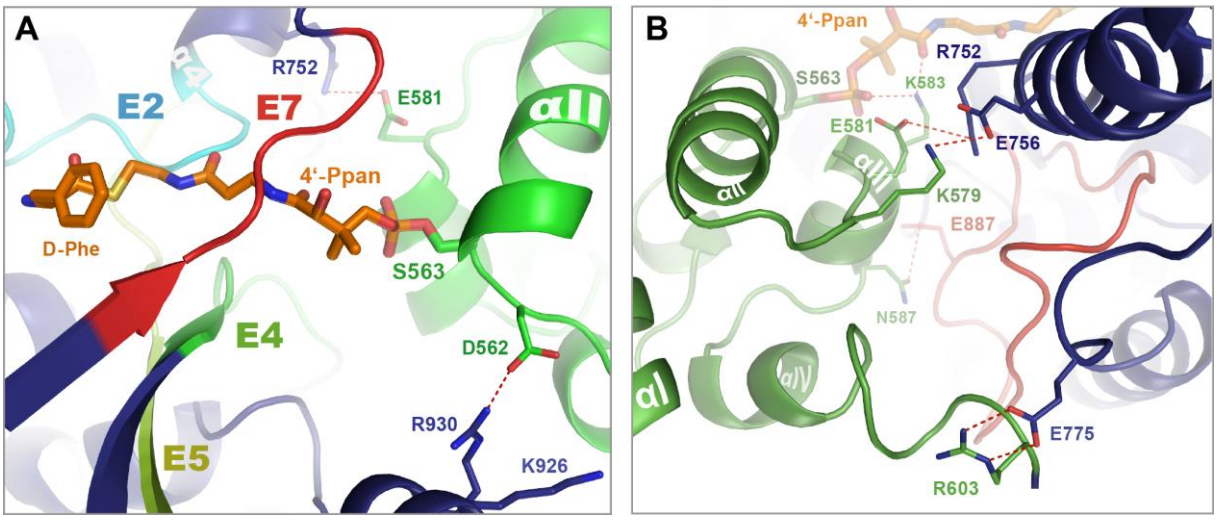

Figure S1 Interfaces of the of the productively docked TycA PCP-E bidomain model. *A.* Surroundings of the 4'-Ppan cofactor. The cofactor attached to the side chain of S563 enters the epimerization domain via an entrance formed by the core motifs E4 (floor loop) and E7 (bridge region). A cluster of H-bonds appears to be important for the proper relative domain positioning. It includes E581, D562 and S563 of the PCP domain and residues R752, K926 and R930 of the epimerization domain. The core motifs are coloured according to the colour code used in Fig. 3 (E2: light blue, E4 green, E5: yellow, E7: red). *B.* The PCP-E domain interface and the floor loop. The model suggests several interactions between residues of the PCP domain with that of the epimerization domain (blue) and the floor loop (red).

Supplementary References

Bashford, D. (1997). An object-oriented programming suite for electrostatic effects in biological molecules: An experience report on the MEAD project, Vol. 1343, Scientific Computing in Object-Oriented Parallel Environments, edited by Y. Ishikawa, R. R. Oldehoeft, J. V. W. Reynders & M. Tholburn, pp. 233-240.

Bashford, D. & Gerwert, K. (1992). *J. Mol. Biol.* **224**, 473-486.

Case, D. C., Darden, T. A., Cheatham III, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A. & Kollman, P. A. (2010). *AMBER 11*.

Czodrowski, P., Dramburg, I., Sotriffer, C. A. & Klebe, G. (2006). *Proteins* **65**, 424-437.

Deerfield II, D. W. & Pedersen, L. G. (1995). *J. Mol. Struct.* **358**, 99-106.

Dupradeau, F.-Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Grivel, N., Lelong, D., Rosanski, W. & Cieplak, P. (2010). *Phys. Chem. Chem. Phys.* **12**, 7821-7839.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126-2132.

Eswar, N., Eramian, D., Webb, B., Shen, M. Y. & Sali, A. (2008). *Methods Mol. Biol.* **426**, 145-159.

Garvey, G. S., McCormick, S. P. & Rayment, I. (2008). *J. Biol. Chem.* **283**, 1660-1669.

Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S. & Onufriev, A. (2005). *Nucleic Acids Res.* **33**, W368-W371.

Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Molec. Graph.* **14**, 33-38.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774-797.
Linne, U., Doekel, S. & Marahiel, M. A. (2001). *Biochemistry* **40**, 15824-15834.
Stein, D. B., Linne, U., Hahn, M. & Marahiel, M. A. (2006). *Chembiochem* **7**, 1807-1814.