Structural Elucidation of Bi-Specificity of A Domains as a Basis to Activate Non-Natural Amino Acids

Heidi Kaljunen,[a] Stephan H. H. Schiefelbein,[b] Daniela Stummer,[b, c] Sandra Kozak,[a] Rob Meijers,[a] Guntram Christiansen,*[d] Andrea Rentmeister*[b, c]

Dedication ((optional))

Abstract: Many biologically active peptide secondary metabolites of bacteria are produced by modular enzyme complexes, the non-ribosomal peptide synthetases. Substrate selection occurs through an adenylation (A) domain, which activates the cognate amino acid with high fidelity. The recently discovered A domain of an Anabaenopeptin synthetase from *Planktothrix agardhii* (ApnA A_1) is capable of activating two chemically distinct amino acids (Arg and Tyr). Here we present crystal structures of the A domain that reveal how the enzyme can adapt both substrates. Analysis of the binding pocket led to the identification of three residues that are critical for substrate recognition. Systematic mutagenesis of these residues created A domains that became mono-specific, or changed substrate specificity to tryptophan. The non-natural amino acid substrate 4-azidophenylalanine is also efficiently activated by a mutant A domain, enabling the production of diversified non-ribosomal peptides for bioorthogonal labeling.

Cyanobacteria are a well-known source of bioactive secondary metabolites, notably peptides produced by non-ribosomal peptide synthetases (NRPS). NRPS are multimodular enzymes composed of repeating units of modules each subdivided into distinct catalytic domains that sequentially process the substrates to the final peptide.[1] This modularity allowed creation of recombinant peptide synthetases and production of novel anti-microbial peptides.[2]

The crystal structure of PheA, a phenylalanine adenylating domain of Gramicidin synthetase 1, led to the identification of a specificity-conferring code defining 10 residues in the active site involved in substrate specificity and catalysis.[3] The code provides a paradigm for substrate specificity for many A domain sequences prior to biochemical characterization[4] and can be used to alter specificity of A domains by changing single residues within the code.[5] However, substrate promiscuity of adenylation (A) domains can lead to families of structurally related peptides as reported for certain cyanobacteria.[2b]

We have recently reported on the first A domain of the Anabaenopeptin synthetase from *Planktothrix agardhii* strain PCC7821 (ApnA A_1), which is bi-specific for the two chemically distinct amino acids arginine (Arg) and tyrosine (Tyr) (Figure 1).[6] To understand substrate selection, we have now determined its crystal structure and verified ligand binding by soaking the apo crystals in solutions containing several ligand compositions (Suppl. Table 2).

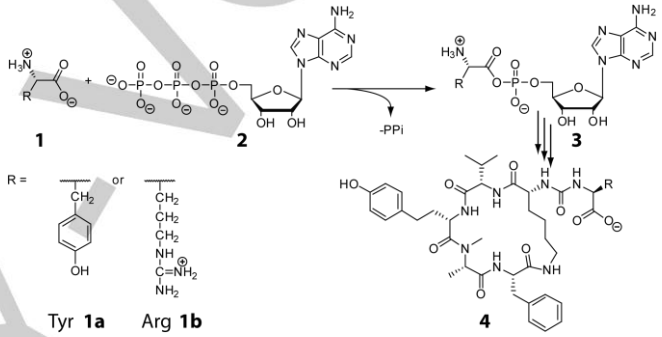


Figure 1 Amino acid selection and activation in ApnA A_1 from *Planktothrix agardhii*. Amino acid (**1**) activation with ATP (**2**) in NRPS catalyzed by A domains leads to aminoacyl-adenylate (**3**) which is condensated by additional domains and modules to anabaenopeptin (**4**). The domain is bispecific for Tyr (**1a**) and Arg (**1b**) leading to two different anabaenopeptins (**4**).

The structures reveal that the fold of ApnA A_1 is highly similar to the fold of the adenylate-forming enzyme family featuring two distinct subdomains, a large N-terminal domain (residues 1-408) and a small C-terminal domain (residues 413-547) linked via a flexible four-residue loop (Figure 2a).[3a, 7] Structural alignment with PheA revealed high similarity in the overall topology with a root mean square deviation (r.m.s.d.) of 0.777 Å for C_α-atoms of 283 residues. In the apo structure of ApnA A_1, the C-terminal domain is not visible in the electron density. In presence of AMP-PNP, a non-hydrolyzable ATP-analogue, the C-terminal domain becomes ordered and residues 409–445 and 451–492 can be built into the electron density. In comparison to PheA, the C-terminal domain is rotated about 140° towards the N-terminus and the enzyme appears to be in the T-state.[3a, 7]

When an apo-crystal of ApnA A_1 is soaked with 4 mM $MgCl_2$, 2 mM ATP and a saturated solution of L-tyrosine, a crystal structure is obtained with an adenylated L-tyrosine (Tyr-AMP) in the active site (Figure 2b). The ATP molecule is hydrolyzed, and the phosphate groups can be observed leaving the substrate binding pocket through a cavity lined by residues Phe199 and Phe419 (Suppl. Figure 1). The Tyr-AMP group occupies the same position as in the AMP-PNP binary complex. Using a similar soaking protocol, replacing L-tyrosine with 50 mM L-Arg resulted in a crystal structure of ApnA A_1 in complex with adenylated L-Arg (Arg-AMP) which diffracted to 2.0 Å (Figure 2c). Superposition of the two structures containing the

[a] H. Kaljunen, S. Kozak, Dr. R. Meijers
European Molecular Biology Laboratory Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany),
[b] S. H. H. Schiefelbein, Dr. D. Stummer, Prof. Dr. A. Rentmeister, University of Muenster, Department of Chemistry, Institute of Biochemistry, Wilhelm-Klemm-Straße 2, D-48149 Muenster, Germany
E-mail: a.rentmeister@uni-muenster.de
[c] Dr. D. Stummer, Prof. Dr. A. Rentmeister, Cells-in-Motion Cluster of Excellence (EXC 1003 – CiM), University of Münster, Germany
[d] Dr. G. Christiansen, University of Innsbruck, Research Institute for Limnology, Mondseestr. 9, A-5310 Mondsee, Austria
Email: guntram.christiansen@uibk.ac.at

Supporting information for this article is given via a link at the end of the document.((Please delete this text if not appropriate))

adenylated amino acids showed a near identical alignment of the protein backbone (r.m.s.d. 0.129 Å for 408 C_α-atoms), and the bound amino-acid-AMP ligands, suggesting that both amino acids are bound and activated in the same pocket of the protein. To achieve substrate complementarity, the Arg side chain adopts a conformation that mimics the shape of the Tyr ring (Figure 3a and Suppl. Figure 2a).

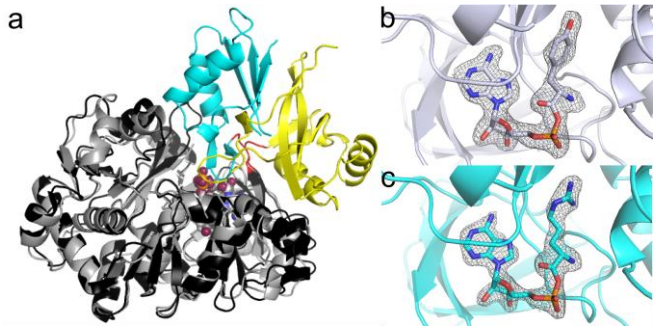


Figure 2. Crystal structure of ApnA A_1 of *P. agardhii* PCC7821 in complex with an adenylated amino acid substrate. (a) Superimposition of ApnA A_1 and PheA (PDB code 1AMU).[3a] The N- and C-terminal domains are coloured in black and yellow for ApnA A_1, and in light grey and light blue for PheA. The linker region is marked in red. The active site with the bound ligand AMP-Tyr (spheres, coloring according to atom type with carbons in grey and oxygens in firebrick red) is located in a cleft between the domains. (b)-(c) A difference electron density for the adenylate reaction intermediates was calculated by omitting the ligands from the model. The map for the tyrosyl adenylate intermediate (b) contoured at 0.12 e*Å^{-3} (σ = 2.0) and for the arginyl adenylate (c) at 0.14 e*Å^{-3} (σ = 3.0). The adenylated end products are represented as light grey (Tyr) and light blue (Arg) stick models.

Residues Glu204 and Ser243 appear to play a key role in coordinating the correct orientation of the substrates. The interaction pattern of these residues differs depending on the nature and adenylation state of the amino acid substrate (Suppl. Figure 2b-d). In the pre-reactive state, Glu204 forms a single salt bridge with the L-Arg substrate. The side chain of Glu204 is kept in place by a nearby residue (Tyr183) that forms a hydrogen bond between the OE1 atom of Glu204 and the OH group of Tyr183. Upon adenylation, the side chain of Glu204 is adapting its conformation to accommodate the different functional groups of Tyr and Arg, resulting in a single hydrogen bond to the hydroxyl group of the Tyr phenol ring as opposed to a bidentate interaction with the guanidinium group of Arg. Importantly, after adenylation there is a water molecule that forms a bridge between the substrates and Ser243. Additionally, we observed a non-polar interaction between Ala307 and the phenyl ring of Tyr or the aliphatic part of the Arg side chain (Figure 3a). As a part of the specificity-conferrring code, position 307 (331 in PheA) possesses high variability between different A domains and is important for the substrate specificity of the enzyme.[3b] These interactions determine the substrate bi-specificity of ApnA A_1 from PCC7821: The end groups of the substrates interact with the side chains of Glu204 and Ser243 to accommodate the opposite charges and to stretch the substrate into a specific conformation. The residue Ala307 further restricts the conformational space between the end groups of the substrate and the catalytic site.

In the genus of *Planktothrix* two different groups of homologs of ApnA A_1 exist with one being monospecific on Arg. Comparison of the specificity conferring codes showed non-conservative substitutions at two positions (243S/D and 307A/E) which may thus be responsible for the different specificities (Suppl. Figure 9).[6] We observed that substitution S243D alone turned ApnA A_1 into an enzyme monospecific for Arg (Figure 3c). The variant had high overall activity and was more active on Arg than the wild type (WT). Variant A307E however was not active. Our data suggests that the 243S/D substitution may be responsible for the different Tyr/Arg specificities of the two naturally occurring groups of homologs.

Until now, no naturally occurring Tyr-monospecific ApnA A_1 could be identified. We anticipated that strengthening the hydrophobic interaction of Ala307 with the aromatic ring of the Tyr ligand should benefit a preference for Tyr over Arg. Fifteen out of 19 substitutions at position Ala307 led to inactive variants (Figure 3b and Suppl. Figure 3), confirming that Ala307 is a key factor for enzymatic function. However, substitutions with larger aliphatic side chains (Val, Leu) resulted in a substrate preference for Tyr (Figure 3c and Suppl. Figure 3).

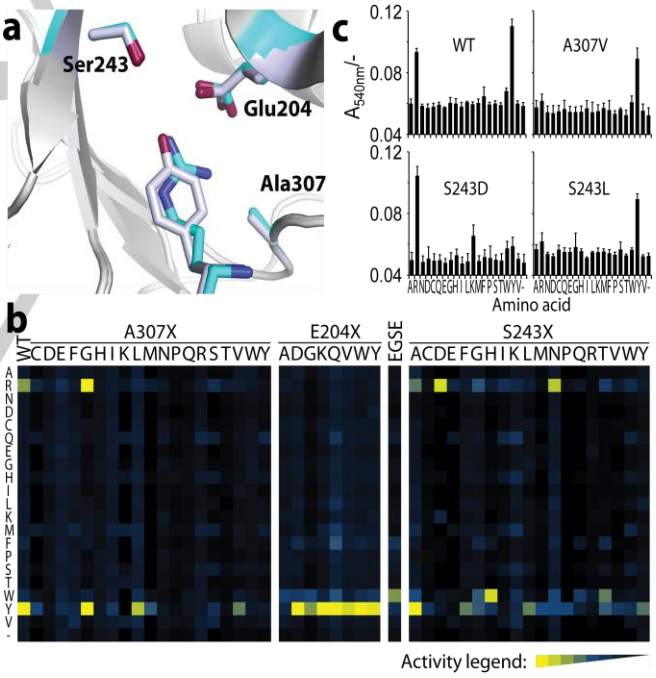


Figure 3. Identification and substitution of specificity related residues. (a) Alignment of crystal structures with Arg and Tyr showing critical residues in the substrate binding pocket in sticks. (b) Substrate profile of tested variants. Colors indicate activities according to the legend. EGSE: E204G/S243E (c) Activity tests of selected variants with different substrate profiles compared to WT. (b) and (c) show average values from two independent measurements in duplicate.

At amino acid position 243, the enzyme shows a remarkable high mutational robustness with 12 out of 19 amino acid substitutions resulting in active variants (Figure 3b and Suppl.

Figure 4). Substitutions with larger non-polar side chains like Leu or Met abolished activity on Arg, whereas smaller non-polar residues like Gly and Ala retained bi-specificity. This suggests that a larger aliphatic side chain at position 243 can still favorably interact with the aromatic ring of the tyrosine substrate but hampers the binding mode of arginine, presumably by displacing the buffering water molecule involved in substrate binding. However, interactions with the aromatic ring resulted in variants that retained enzymatic activity on Tyr. S243Y and S243F are also monospecific for Tyr most likely due to stacking interactions.

The importance of Glu204 on Arg binding becomes clear when the substrate profile of Glu204 variants is investigated. Out of the 8 variants tested, 7 had lost activity on Arg but were still active on Tyr. Only E204A was inactive on all substrates tested (Figure 3b, Suppl. Figure 5). These data suggest that position 204 is critical for bi-specificity, because Glu204 seems to be required to tolerate Arg. However, Glu204 is not essential for tyrosine activation. Interestingly, naturally occurring homologues of ApnA A_1 that are monospecific for tyrosine are not known and Glu204 is fully conserved in nature (Suppl. Fig. 8 and 9).

We noticed that variant E204G as well as several other variants (e.g., S243H, S243A, S243T) also had an increased activity on tryptophan (Trp). We hypothesized that combining the larger substrate binding pocket of E204G with S243E to maintain the net charge might allow the activation of bigger amino acids as substrates. Indeed, the double variant E204G/S243E had a strong substrate preference for Trp – the bulkiest of the 20 canonical amino acids – with an activity in the range of WT for its preferred substrates (Suppl. Figure 5). This observation was confirmed by kinetic measurements, revealing a 3,000-fold switch in substrate specificity to Trp and a higher catalytic efficiency of E204G/S243E for Trp activation (k_{cat}/K_m = 21 mM^{-1} min^{-1}) than WT for Tyr (k_{cat}/K_m = 7 mM^{-1} min^{-1}) (Table 1, Suppl. Table 1).

Since NRPSs are capable of using many amino acids beside canonical amino acids, we attempted to introduce activity on the non-natural amino acid 4-azidophenylalanine (Az, Figure 4a). Since Az is compatible with bioorthogonal click chemistry[8] it could in the long run allow production of novel anabaenopeptins **4** (Figure 1a) and labeling in the organism. The activity of variant E204G/S243E on Az was lower than for Trp but significantly higher than for Tyr and Arg, reaching the activity range of WT for Arg and a catalytic efficiency of 1.7 mM^{-1} min^{-1} (compared to 4 mM^{-1} min^{-1} for WT for Arg) (Figure 4b, Table 1).

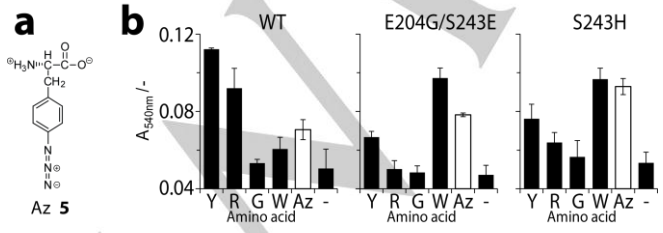


Figure 4. Activation of a non-natural amino acid by ApnA A_1 variants. (a) 4-azidophenylalanine Az (**5**). (b) Activity of WT and selected variants on **5** in comparison to Tyr, Arg, Gly, and Trp. Shown are average values from two independent experiments in duplicate.

Testing promising variants on Az revealed that S243A, S243H, E204G, and S243T also had considerable activity on this non-natural amino acid (Figure 4 and Suppl. Figure 6). S243H showed substrate preference for both Az and Trp with activities and k_{cat}/K_m values in the range of the WT enzyme for its preferred substrates. Our kinetic data indicate that the single substitution S243H switched the specificity 100-fold towards Az and the resulting variant activates this non-natural amino acid as efficiently WT activates tyrosine (k_{cat}/K_m = 7 mM^{-1} min^{-1} for WT on Tyr and S243H on Az). Thus, both E204G/S243E as well as S243H should allow incorporation of Trp and Az into new anabaenopeptins.

Table 1. Measured values of K_m and k_{cat} for ApnA A_1 domains from *P. agardhii* PCC7821 in the amino-acid-mediated ATP/PP_i exchange reaction.[9] Shown are average values of multiple independent experiments.

Variant	Substr.	K_m [mM]	k_{cat} [min^{-1}]	k_{cat}/K_m [mM^{-1} min^{-1}]	Specificity[a]	Specificity switch[b]
WT	Tyr	0.044	0.3	7	1	-
WT	Arg	0.16	0.6	4	0.5	-
WT	Trp	1.6	0.3	0.18	0.03	-
WT	Az	1.1	0.9	0.8	0.1	-
EGSE[c]	Tyr	0.3	0.08	0.3	1	-
EGSE[c]	Arg	n.d.	n.d.	n.d.	-	-
EGSE[c]	Trp	0.07	1.5	21	80	3,000
EGSE[c]	Az	0.36	0.6	1.7	6	60
S243H	Tyr	0.32	0.2	0.6	1	-
S243H	Arg	6.2	0.4	0.07	0.1	0.2
S243H	Trp	0.39	2.2	6	9	300
S243H	Az	0.6	4.5	7	12	100

n.d.: Not determinable. [a] Specificity is relative specificity for indicated substrates with respect to tyrosine: $(k_{cat(substrate)}/K_{m(substrate)})/(k_{cat(Tyr)}/K_{m(Tyr)})$. [b] The specificity switch shows $specificity_{(variant)}/specificity_{(wt)}$. for the indicated amino acid. [c] E204G/S243E

In conclusion, our study sheds light on how the bi-specific A domain ApnA A_1 from *P. agardhii* strain PCC7821 governs the selection of the apparently contrasting substrates Arg and Tyr. Structural analysis revealed that the architecture of the active site forces Arg to adopt a Tyr-like conformation and thereby explains the bi-specificity. Strikingly, two of the three residues in the active site, identified by structural investigations to be critical for bi-specificity, are also subject for naturally occurring point mutations (Ser243 and Ala307).[6] Expanding the naturally occurring variation within these positions of the specificity conferring code revealed that additional substrate specificities for Tyr only and Trp could be obtained. Independent of the biosynthetic pathway all characterized cyanobacterial secondary metabolite operons contain an ATP binding cassette (ABC) transporter homolog. These transporter proteins most likely

shuttle their corresponding allocrite over a membrane, where it is unknown if it will be the thylakoid rather than the cytoplasmic membrane resulting in complete different localization. Engineering specific single and double variants accepting the non-natural amino acid Az should make it possible to reprogram ApnA A_1 by genetic engineering in its natural host with the potential to functionalize anabaenopeptins within the producing organism. Combined with selective bioorthogonal labeling it will help to identify potential transport routes within the cell. Insights into the localization of NRS peptides may help answering the question of the biological function of secondary metabolites in cyanobacteria.

Experimental Section

Experimental Details can be found in the Supplementary Information.

Acknowledgements

We would like to thank Julia Sandberg-Meinhardt, Lucas Lang (both from University of Hamburg) for excellent technical assistance. We would like to thank Stephane Boivin and the sample preparation and characterization facility (SPC) for assistance in crystallization and soaking experiments. We would like to thank Gleb Bourenkov, Thomas Schneider, and Michele Cianci for generous support at the P13 and P14 EMBL beamlines at the PETRA synchrotron situated at DESY in Hamburg. A.R. thanks the DFG for support with an Emmy Noether fellowship (RE 2796/2-1) and the Fonds der Chemischen Industrie for a “Sachkostenzuschuss”. This work was supported by funding from the European Community’s Seventh Framework Program under Contract 227764 (P-Cube) and partly supported by the Deutsche Forschungsgemeinschaft (DFG), Cells-in-Motion Cluster of Excellence (EXC 1003 – CiM), University of Münster, Germany. S. H. H. S. thanks the IRTG 2027 for a doctoral fellowship. G.C. thanks the Austrian Science Fund for financial support (P24070) and Rainer Kurmayer for general support.

Author contributions

H. K., S. H. H. S., D. S., G. C., and S. K. performed experiments. H. K., D. S., R. M., G. C., A. R. designed experiments. All authors discussed the results. H. K. and R. M. solved the crystal structures. H. K., S. H. H. S., G. C., and A. R. wrote the paper with help from D. S. and R. M.

Keywords: adenylation domain • cyanobacteria • non-ribosomal peptide synthetase • crystal structure • substrate specificity

[1] M. Strieker, A. Tanovic and M. A. Marahiel, *Curr. Opin. Struct. Biol.* **2010**, *20*, 234-240.

[2] a) S. Doekel, M. F. C. L. Gal, J. Q. Gu, M. Chu, R. H. Baltz and P. Brian, *Microbiology* **2008**, *154*, 2872-2880. b) N. A. Magarvey, Z. Q. Beck, T. Golakoti, Y. S. Ding, U. Huber, T. K. Hemscheidt, D. Abelson, R. E. Moore and D. H. Sherman, *ACS Chem. Biol.* **2006**, *1*, 766-779; c) T. Stachelhaus and M. A. Marahiel, *J. Biol. Chem.* **1995**, *270*, 6163-6169.

[3] a) E. Conti, T. Stachelhaus, M. A. Marahiel and P. Brick, *EMBO J.* **1997**, *16*, 4174-4183. b) T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chem. Biol.* **1999**, *6*, 493-505.

[4] M. Rottig, M. H. Medema, K. Blin, T. Weber, C. Rausch and O. Kohlbacher, *Nucleic Acids Res.* **2011**, *39*, W362-W367.

[5] a) C. Y. Chen, I. Georgiev, A. C. Anderson and B. R. Donald, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3764-3769; b) K. Eppelmann, T. Stachelhaus and M. A. Marahiel, *Biochemistry* **2002**, *41*, 9718-9726; c) B. W. Stevens, R. H. Lilien, I. Georgiev, B. R. Donald and A. C. Anderson, *Biochemistry* **2006**, *45*, 15495-15504; d) J. Thirlway, R. Lewis, L. Nunns, M. Al Nakeeb, M. Styles, A. W. Struck, C. P. Smith and J. Micklefield, *Angew. Chem., Int. Ed.* **2012**, *51*, 7181-7184; e) B. Villiers and F. Hollfelder, *Chem. Biol.* **2011**, *18*, 1290-1299; f) H. Kries, R. Wachtel, A. Pabst, B. Wanner, D. Niquille and D. Hilvert, *Angew. Chem., Int. Ed.* **2014**, *53*, 10105-10108.

[6] G. Christiansen, B. Philmus, T. Hemscheidt and R. Kurmayer, *J. Bacteriol.* **2011**, *193*, 3822-3831.

[7] H. Yonus, P. Neumann, S. Zimmermann, J. J. May, M. A. Marahiel and M. T. Stubbs, *J. Biol. Chem.* **2008**, *283*, 32484-32491.

[8] a) M. L. Tsao, F. Tian and P. G. Schultz, *ChemBioChem* **2005**, *6*, 2147-2149; b) V. Boehrsch, R. Serwa, P. Majkut, E. Krause and C. P. R. Hackenberger, *Chem. Commun.* **2010**, *46*, 3176-3178; c) T. Mukai, M. Wakiyama, K. Sakamoto and S. Yokoyama, *Protein Sci.* **2010**, *19*, 440-448.

[9] W. Gevers, H. Kleinkauf and F. Lipmann, *Proc. Natl. Acad. Sci. U. S. A.* **1968**, *60*, 269-276.

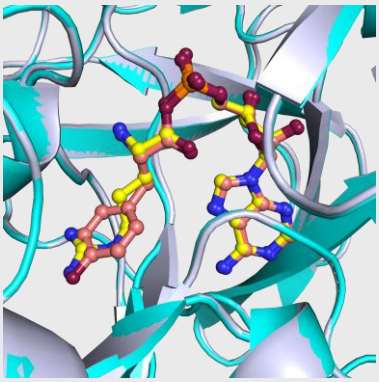
Entry for the Table of Contents (Please choose one layout)

Layout 1:

COMMUNICATION

Snuggling Substrates:

The structural basis for bi-specificity of an adenylation domain on the two contrasting substrates arginine and tyrosine was revealed by crystal structure analysis. Three residues in the active site were identified to be critical for substrate recognition and substituted to switch specificity.



*Author(s), Corresponding Author(s)**

Page No. – Page No.

Title

Layout 2:

COMMUNICATION

((Insert TOC Graphic here))

*Author(s), Corresponding Author(s)**

Page No. – Page No.

Title

Text for Table of Contents