Supporting Information

Precise Probing of Residue Roles by NRPS Code Swapping: Mutation, Enzymatic Characterization, Modeling, and Substrate Promiscuity of Aryl Acid Adenylation Domains

Fumihiro Ishikawa[1,*], Maya Nohara[1], Shinya Nakamura[2], Isao Nakanishi[2], and Genzoh Tanabe[1,*]

1. Laboratory of Pharmaceutical Organic Chemistry, Faculty of Pharmacy, Kindai University, 3-4-1 Kowakae, Higashi-Osaka, Osaka 577-8502, Japan

2. Laboratory of Computational Drug Design and Discovery, Faculty of Pharmacy, Kindai University, 3-4-1 Kowakae, Higashi-Osaka, Osaka 577-8502, Japan

[*]Correspondence and request for materials should be directed via email to Fumihiro Ishikawa (ishikawa@phar.kindai.ac.jp) or Genzoh Tanabe (g-tanabe@phar.kindai.ac.jp)

SUPPLEMENTARY DISCUSSION

Characterization of Enzymatic Activity of EntE Variants toward DHB Substrate. The EntE variant Asn340Cys exhibited catalytic properties distinct from those of the other single variants Tyr236Phe, Val339Ile, and Val339Leu; in the Asn340Cys variant, DHB recognition was unaffected, but the k_{cat} value for DHB activation was perturbed and was ~3-fold lower than that of wild-type EntE (Figure 2D and Tables 1). Asn340 is positioned at the edge of the DHB-binding site and points away from the DHB substrate, and this potentially controls the available space in the ATP-binding pocket. Kinetic analysis of the quadruple-mutant Tyr236Phe/Ser240Cys/Val339Leu/Asn340Cys revealed that the enzyme-catalyzed reaction was not saturated at the tested concentrations of DHB substrate (Figure 2I). The replacement of Val339 in the active site of the EntE variant Tyr236Phe/Ser240Cys/Asn340Cys with a Leu toward the NRPS codes of Sal-activating A-domain YbtE apparently disrupted the conformation of the DHB-binding site.

Characterization of Enzymatic Activity of EntE Variants toward Sal Substrate. The activity of the variant Asn340Cys measured with Sal (k_{cat}/K_m = 879 min^{-1} mM^{-1}) were higher than that of wild-type EntE with Sal (Table 1). These single mutations led to the k_{cat}/K_m values being 2-fold higher than that of wild-type EntE toward the non-cognate Sal substrate. The EntE variant Asn340Cys showed enzymatic activity toward Sal that was comparable to that of a native Sal-activating A-domain, MbtA.[1] By contrast, the k_{cat}/K_m value of the EntE variant Asn340Cys measured with the cognate DHB substrate was 365 mM^{-1} min^{-1}, respectively, which was 3.6-fold lower than that of wild-type EntE with DHB substrate (Table 1). The manipulation of this residue toward the NRPS codes of Sal-activating A-domains yielded enzyme-specificity switches toward Sal substrate of 8-fold (Asn340Cys) (Table 1). The k_{cat}/K_m value of Tyr236Phe/Ser240Cys/Val339Leu/Asn340Cys variant measured with Sal substrate was 11 mM^{-1} min^{-1} (Table 1). This result indicates that the active site of EntE could not accommodate the NRPS code of YbtE for accepting Sal substrate.

The NRPS codes of aryl acids. The resulting single variants activate Sal (k_{cat}/K_m = 879−1616 mM^{-1} min^{-1}) as efficiently as wild-type EntE activates DHB (k_{cat}/K_m = 1311 mM^{-1} min^{-1}) (Table 1). Moreover, the multiple variants (harboring double, triple, and quadruple mutations) displayed negligible reduction in k_{cat}/K_m values with Sal substrate and a 52–108-fold decrease in k_{cat}/K_m values with DHB, corresponding to a 22–58-fold switch in substrate specificity toward Sal (Table 1). These results demonstrated the specificity-conferring functions of positions 236, 240, and 339 in DHB- and Sal-activating A-domains. Furthermore, Asn340 plays no significant role in the recognition of aryl acid substrates. Amino acid residues at positions 235, 277, 306, 308, and 331 are invariant throughout aryl acid A-domains. On the basis of enzymatic kinetics studies, we performed a comparison of wild-type EntE and the EntE variant Tyr236Phe/Ser240Cys/Val339Ile339/Asn340Cys active site volumes. We constructed close-up

views of the major residues in the active site of EntE and the EntE variant Tyr236Phe/Ser240Cys/Val339Ile/Asn340Cys that are involved in the substrate discrimination (Figure 7). Our modeling analysis indicated that the specificity-conferring residues at positions 236, 240, and 339 are likely to collectively control the substrate recognition toward DHB and Sal substrates, making the cavity slightly small near the C3 carbon of BA and engaging in appropriate interactions to accommodate Sal substrate (Figure 7). The analysis presented provides a deeper understanding of aryl acid adenylation domains. Furthermore, it provides a rational framework for reprogramming of aryl acid A-domains by site-directed mutagenesis and directed evolution to produce novel natural products.

DHB	RESP		Sal	RESP
O1	-0.759		O1	-0.742
O2	-0.759		O2	-0.742
C0	0.793		C0	0.716
C1	-0.057		C1	-0.016
C2	-0.227		C2	-0.217
C3	-0.264		C3	-0.182
C4	-0.273		C4	-0.196
C5	0.313		C5	-0.212
C6	0.134		C6	0.277
H2	0.172		H2	0.158
H3	0.149		H3	0.106
H4	0.166		H4	0.128
O5	-0.639		H5	0.123
H5	0.434			
O6	-0.596		O6	-0.614
H6	0.412		H6	0.412
Ring	-0.373		Ring	-0.546

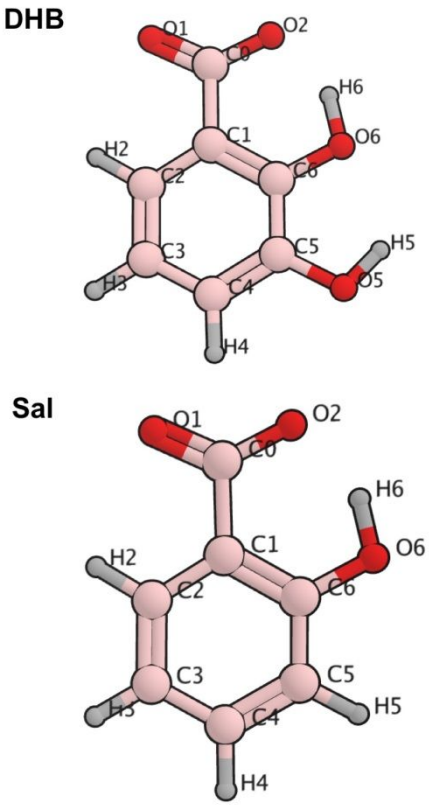

Figure S1. Comparison of RESP atomic partial charges. The total charge of the carbon atoms of the benzene ring is described in the "Ring" line.

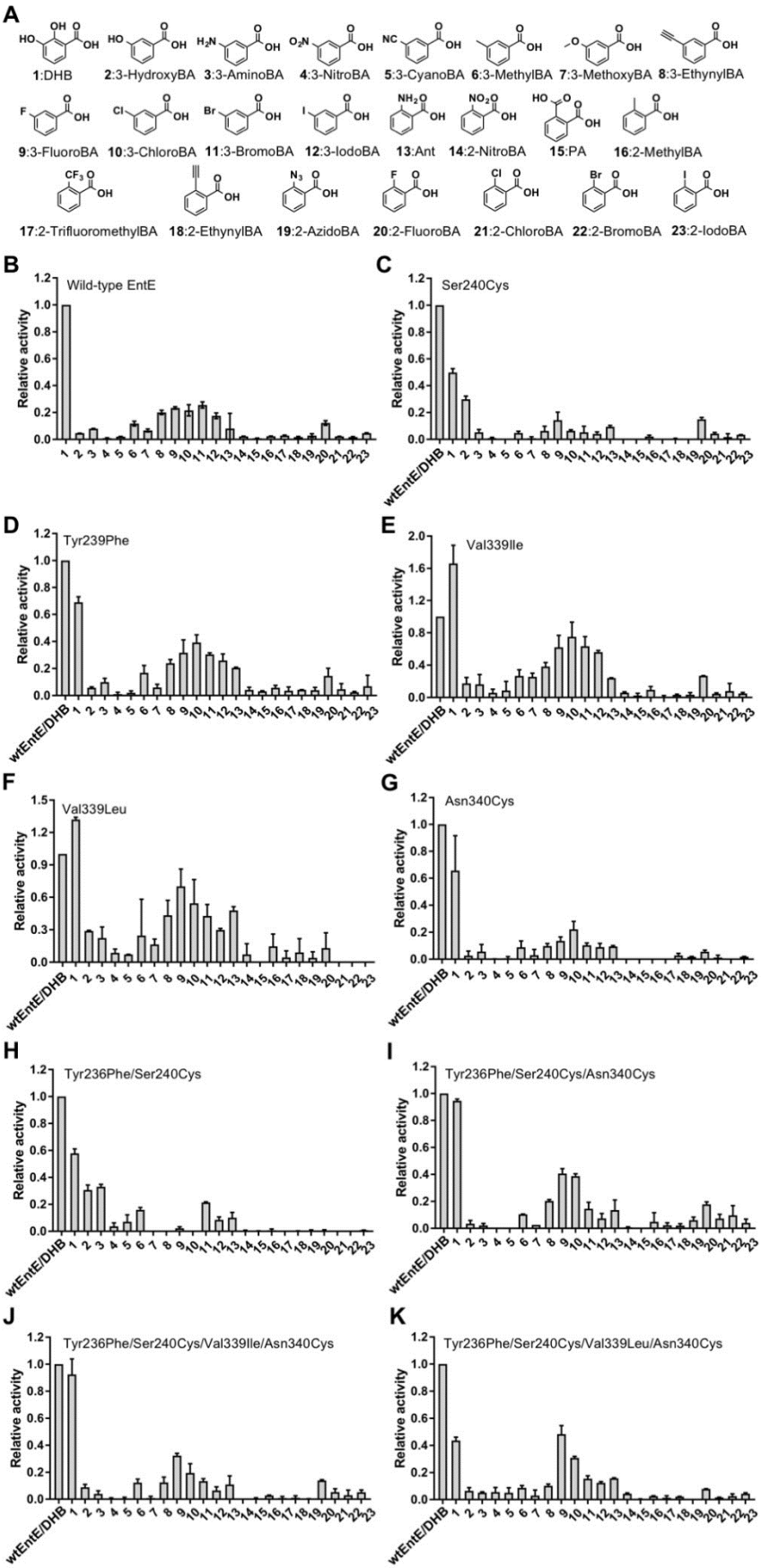
A
1:DHB
2:3-HydroxyBA
3:3-AminoBA
4:3-NitroBA
5:3-CyanoBA
6:3-MethylBA
7:3-MethoxyBA
8:3-EthynylBA
9:3-FluoroBA
10:3-ChloroBA
11:3-BromoBA
12:3-IodoBA
13:Ant
14:2-NitroBA
15:PA
16:2-MethylBA
17:2-TrifluoromethylBA
18:2-EthynylBA
19:2-AzidoBA
20:2-FluoroBA
21:2-ChloroBA
22:2-BromoBA
23:2-IodoBA
B
Wild-type EntE
C
Ser240Cys
D
Tyr239Phe
E
Val339Ile
F
Val339Leu
G
Asn340Cys
H
Tyr236Phe/Ser240Cys
I
Tyr236Phe/Ser240Cys/Asn340Cys
J
Tyr236Phe/Ser240Cys/Val339Ile/Asn340Cys
K
Tyr236Phe/Ser240Cys/Val339Leu/Asn340Cys
Relative activity
wtEntE/DHB

Figure S2. (A) Array of aryl acid substrates tested in this study. Substrate profiles of wild-type EntE (B) and the EntE variants Ser240Cys (C), Tyr236Phe (D), Val339Ile (E), Val339Leu (F), Asn340Cys (G), Tyr236Phe/Ser240Cys (H), Tyr236Phe/Ser240Cys/Asn340Cys (I), Tyr236Phe/Ser240Cys/Val339Ile/Asn340Cys (J), and Tyr236Phe/Ser240Cys/Val339Leu/Asn340Cys (K). Wild-type EntE and the EntE variants were used with 1 mM of the aryl acid substrates Control wells were treated identically, except that no aryl acid substrates were added to the reaction buffer. To estimate relative adenylation activities, we subtracted the 620-nm absorbance (A_{620}) values of reaction mixtures without the aryl acid substrate from the A_{620} values of reaction mixtures containing the aryl acid substrate. The adenylation activity was normalized to that of wild-type (wt) EntE toward DHB substrate.

MATERIALS AND METHODS

Preparation of Overexpression Constructs. The gene entE was PCR-amplified from pKK223-3 containing entE, kindly provided by Prof. Michael D. Burkart, University of California, San Diego, USA; the gene was cloned into pET28b expression vector and used as the DNA template for site-directed mutagenesis. The mutant enzymes EntE (Tyr236Phe), EntE (Val339Ile), EntE (Val339Leu), and EntE (Asn340Cys) were constructed from pET28b-entE template PCR mutagenesis by using these primers (respectively): EntE (Y236F)_F (5′-CCGGCGGCTCATAACTTTGCCATGAGTTCGCCAGG-3′) and EntE (Y236F)_R (5′-CCTGGCGAACTCATGGCAAAGTTATGAGCCGCCGG-3′); EntE (V339I)_F (5′-GGCGGAAGGGCTGATTAACTACACCCGACTTGATG-3′) and EntE (V339I)_R (5′-CATCAAGTCGGGTGTAGTTAATCAGCCCTTCCGCC-3′); EntE (V339L)_F (5′-GGCGGAAGGGCTGTTAAACTACACCCGACTTGATG-3′) and EntE (V339L)_R (5′-CATCAAGTCGGGTGTAGTTTAACAGCCCTTCCGCC-3′); and EntE (N340C)_F (5′-GGCGGAAGGGCTGGTGTGCTACACCCGACTTGATG-3′) and EntE (N340C)_R (5′-CATCAAGTCGGGTGTAGCACACCAGCCCTTCCGCC-3′). Site-directed mutant EntE (Tyr236Phe/Ser240Cys) was constructed from pET28b-entE (Tyr236Phe/Ser240Cys) template PCR mutagenesis by using primers EntE (Y236F/S240C)_F (5′-CTTTGCCATGAGTTGCCCAGGATCGCTGGG-3′) and EntE (Y236F/S240C)_R (5′-CCCAGCGATCCTGGGCAACTCATGGCAAAG-3′). Site-directed mutant EntE (Tyr236Phe/Ser240Cys/Asn340Cys) was constructed from pET28b-entE (Tyr236Phe/Ser240Cys) template PCR mutagenesis by using primers EntE (Y236F/S240C/N340C)_F (5′-GGCGGAAGGGCTGGTGTGCTACACCCGACTTGATG-3′) and EntE (Y236F/S240C/N340C)_R (5′-CATCAAGTCGGGTGTAGCACACCAGCCCTTCCGCC-3′). Site-directed mutants EntE (Tyr236Phe/Ser240Cys/Val339Ile/Asn340Cys) and EntE (Tyr236Phe/Ser240Cys/Val339Leu/Asn340Cys) were constructed from pET28b-entE (Tyr236Phe/Ser240Cys/Asn340Cys) template PCR mutagenesis by using, respectively, primers EntE (Y236F/S240C/V339I/N340C)_F (5′-GGCGGAAGGGCTGATTTGCTACACCCGACTTGATG-3′) and EntE (Y236F/S240C/V339I/N340C)_R (5′-CATCAAGTCGGGTGTAGCAAATCAGCCCTTCCGCC-3′), and EntE (Y236F/S240C/V339L/N340C)_F (5′-GGCGGAAGGGCTGTTATGCTACACCCGACTTGATG-3′) and EntE (Y236F/S240C/V339L/N340C)_R (5′-CATCAAGTCGGGTGTAGCATAACAGCCCTTCCGCC-3′). Site-directed mutagenesis was verified by DNA sequencing.

Protein Expression and Purification. Recombinant EntB (ArCP) was expressed and purified as described previously.[2] The in vitro phosphopantetheninylation of apo-ArCP was conducted as described.[2] The EntE variants were overproduced in E. coli BL21 (DE3) cells. Overnight cultures were used to inoculate 1 L of LB medium supplemented with 50 μg/mL kanamycin. Cultures were allowed to grow until their 600-nm absorbance (A600) was 0.45–0.80 at 37 °C, induced with IPTG added to a final concentration of 0.1 mM, and allowed to grow for a further 3 h at 37 °C. Cells were pelleted and resuspended in lysis buffer (20 mM Tris–HCl, pH 8.0, 0.5% Triton-X, and protease-inhibitor cocktail), and then lysed by sonication on ice by using an ultrasonic disruptor (UD201, Tomy Digital Biology Co., Ltd, Japan). The resulting cell lysates were centrifuged to remove insoluble debris. The supernatants were loaded onto Ni-NTA agarose columns (Qiagen) and eluted using a gradient of 20–500 mM imidazole. Eluted proteins were visualized by means of SDS-PAGE with Coomassie staining (Colloidal Coomassie Blue Stain) and quantitated using Bradford method.[3] Fractions containing the recombinant proteins were pooled and dialyzed against the assay buffer (20 mM Tris–HCl, pH 8.0, 1 mM $MgCl_2$, and 1 mM TCEP), after which 10% glycerol (v/v) was added and the proteins were stored at −80 °C.

Determination of Kinetic Parameters toward DHB and Sal Substrates. Kinetic parameters were determined using a coupled hydroxamate-MesG continuous spectrophotometric assay (Figure 2A).[1] Standard assay conditions were as follows. Reactions contained varying amounts of EntE proteins (0.5–1 μM) to maintain initial velocity conditions, 20 mM Tris (pH 8.0), 2.5 mM ATP, 1 mM $MgCl_2$, 1 mM TCEP, 150 mM hydroxylamine (pH 7.0), 0.1 U of purine nucleoside phosphorylase (Sigma–Aldrich, N8264), 0.04 U of inorganic pyrophosphatase (Sigma–Aldrich, I1643), 0.2 mM MesG (Berry & Associates), and varying concentrations of substrates. The reactions (100 μL) were run in 96-well half-area plates (Corning, 3881) and the cleavage of MesG was monitored by measuring the sample A_{355} on an EnVision Multilabel Reader (PerkinElmer). Working stocks of hydroxylamine were prepared fresh by combining 500 μL of 4 M hydroxylamine, 250 μL of water, and 250 μL of 7 M NaOH on ice. Steady-state kinetic parameters for the substrates with enzyme were determined using standard assay conditions as described above. The enzyme and substrate concentrations used were the following: EntE (Tyr236Phe), 1 μM with DHB (3.1–100 μM) and 500 nM with Sal (2.5–160 μM); EntE (Val339Ile), 750 nM with DHB (3.1–50 μM) and 500 nM with Sal (2.5–160 μM); EntE (Val339Leu), 750 nM with DHB (3.1–100 μM) and 500 nM with Sal (5–160 μM); EntE (Asn340Cys), 1 μM with DHB (3.1–100 μM) and 500 nM with Sal (5–160 μM); EntE (Tyr236Phe/Ser240Cys), 500 nM with DHB (125–3000 μM) and Sal (10‑320 μM); EntE (Tyr236Phe/Ser240Cys/Asn340Cys), 500 nM with DHB (125–3000 μM) and Sal (10–320 μM); EntE (Tyr236Phe/Ser240Cys/Val339Ile/Asn340Cys), 500 nM with DHB (125–2000 μM) and Sal (10–320 μM); and EntE (Tyr236Phe/Ser240Cys/Val339Leu/Asn340Cys), 1 μM with DHB (500–

4000 μM) and Sal (63–3000 μM). In all experiments, the total DMSO concentration was ≤2.0%. Initial velocities were fit to the Michaelis-Menten equation by using Prism 5 (GraphPad Software).

Transfer of Sal to ArCP Domain Catalyzed by EntE variants. Reaction mixtures (50 μL) contained recombinant holo-ArCP (8 μM), EntE variant (1 μM), Sal (1 mM), 5 mM DTT, 10 mM $MgCl_2$, and ATP (2.5 mM) in 75 mM Tris (pH 7.5). In all experiments, the total DMSO concentration was maintained at 1.0%. After addition of all components, reactions were incubated for 30 min at 37 °C, precipitated with acetone, resolubilized in ddH_2O, and subject to matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) analysis.

Construction of Three-Dimensional Structural Models of EntE variants. The X-ray structures of EntE protein (Protein Data Bank (PDB) code 3RG240) and DhbE–DHB complex (PDB code 1MD930) were retrieved from PDB. The following model construction scheme was employed using Molecular Operating Environment.[4] The EntE–DHB model complex was constructed by superposition of the aforementioned two structures. Disordered structures were complemented using Structure Preparation module. Each EntE variant structure was mutated from EntE protein structure, and the stable sidechain structure of the mutated residue was searched using Rotamer Explorer module. All complex structures including Sal were manually modeled from the corresponding DHB complexes, after which geometry optimization was performed using AMBER10:EHT forcefield[4] to each model complex structure.

Estimation of Atomic Partial Charges of Compounds. Each structure of DHB and Sal was optimized in HF-6-31G* level using Gaussian09.[5] The ESP (Pop=MK) charge was calculated at the same level after structure optimization, and then the RESP charge was re-calculated using Antechamber.[6]

Malachite Green Phosphate Assay.[7]

Standard assay conditions: Reactions contained wild-type EntE (1 μM) or the EntE variants (1 μM), 20 mM Tris (pH 8.0), 0.2 mM ATP, 1 mM $MgCl_2$, 1 mM TCEP, 150 mM hydroxylamine (pH 7.0), 0.04 U of inorganic pyrophosphatase (Sigma–Aldrich, I1643), and aryl acid substrates (1 mM). The reactions (80 μL) were run in 96-well half-area plates (Corning, 3881). Hydroxylamine solution was prepared as described above. The reaction was initiated by adding ATP. After 30-min incubation, the reaction was quenched by adding 20 μL of the working reagent of malachite green phosphate assay kit (BioAssay Systems). After incubation for 30 min at room temperature, A_{620} was measured on an EnVision Multilabel Reader (PerkinElmer). The A_{620} value of the reaction mixture in the absence of substrate was subtracted from the A_{620} value of each reaction mixture in the presence of substrate to estimate adenylation activities.

Substrate profile of wild-type EntE and the EntE variants: Wild-type EntE and the EntE variants were used at 1 μM with 1 mM benzoic acid (BA) derivatives (3-hydroxyBA, 3-aminoBA, 3-nitroBA, 3-cyanoBA, 3-methylBA, 3-methoxyBA, 3-ethynylBA, 3-fluoroBA, 3-chloroBA, 3-

bromoBA, 3-iodoBA, Ant, 2-nitroBA, PA, 2-methylBA, 2-trifluoromethylBA, 2-ethynylBA, 2-azidoBA, 2-fluoroBA, 2-chloroBA, 2-bromoBA, and 2-iodoBA). In all experiments, the total DMSO concentration was mainteined at 1.0%.

References

1. Wilson, D. J. and Aldrich, C. C. (2010) A continuous kinetic assay for adenylation enzyme activity and inhibition. *Anal. Biochem. 404*, 56–63.
2. Ishikawa, F., Kasai, S., Kakeya, H., and Tanabe, G. (2017) Visualizing the adenylation activities and protein-protein interactions of aryl acid adenylating enzymes. *ChemBioChem 18*, 2199–2204.
3. Bradford, M. M. (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem. 72*, 248–254.
4. Molecular Operating Environment, ver. 2019.01, *Chemical Computing Group*, Montreal Canada.
5. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A. Jr., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Keith, T., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. Gaussian 09, Revision C.01. 2010, Gaussian, Inc., Wallingford CT.
6. Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model. 25*, 247–260.
7. McQuade, T. J., Shallop, A. D., Sheoran, A., DelProposto, J. E., Tsodikov, O. V., and Garneau-Tsodikova, S. (2009) A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. *Anal. Biochem. 386*, 244–250.