Biotechnol Lett
https://doi.org/10.1007/s10529-019-02645-x

ORIGINAL RESEARCH PAPER

Pinpointing the L-phenylalanine binding sites of TyrR using biosensors and computer-aided simulation

Danyang Bai · Dongqin Ding · Jinlong Li · Lina Cong · Dawei Zhang

Received: 6 July 2018 / Accepted: 9 January 2019

Abstract
Objectives To determine the binding sites for L-phenylalanine in TyrR protein via a rational mutation analysis combining biosensors and computer-aided simulation.
Results TyrR protein of *Escherichia coli* is the chief transcriptional regulator of several genes essential for the biosynthesis and transport of aromatic amino acids. The identification of ligand-binding sites is often the starting point for protein function annotation and structure-based protein design. Here we combined computer-aided prediction methods and biosensors to identify the ligand-binding sites for L-Phe in TyrR protein.
Conclusions Residues at positions 160, 173 and 184 of TyrR protein are important for transcriptional activation of target genes *tyrP* induced by L-Phe, which indicates that they are the bona fide L-Phe binding sites of TyrR protein.

Keywords TyrR protein · L-Phe binding site · Biosensor · Computer-aided simulation

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10529-019-02645-x) contains supplementary material, which is available to authorized users.

D. Bai · L. Cong (✉)
School of Biological Engineering, Dalian Polytechnic University, Dalian 116034, People's Republic of China
e-mail: linacong@163.com

D. Bai · D. Ding · J. Li · D. Zhang
Tianjin Institutes of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, People's Republic of China

D. Ding · D. Zhang
Key Laboratories of Systems Microbial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, People's Republic of China

D. Ding · D. Zhang (✉)
University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China
e-mail: zhang_dw@tib.cas.cn

Introduction

Many proteins have special pockets on their surface that act as binding sites for small-molecule ligands, with which proteins interact to perform their biological function. The identification of specific ligand-binding sites on proteins is often the starting point for protein function annotation and makes it possible to rationally design antagonists or to introduce targeted mutations to improve the protein function (Koyanagi et al. 2008). Furthermore, even if the main binding site has been identified and validated, it is often important to find additional potential binding sites where different biological effects are possible with appropriate targeting. However, due to technical difficulties and high cost of analysis, the structural details of protein–

Published online: 24 January 2019

ligand interactions are unknown for 40% of proteins in the PDB with experimentally solved 3D structures (Yang et al. 2013).

The TyrR protein of *E. coli* K-12 is a transcriptional regulator that plays a key role in the control of a group of at least eight transcription units (such as *aroF*, *aroG*, *aroL*, *aroP*, *tyrB*, *tyrR*, *mtr* and *tyrP*), which are involved in the biosynthesis and transport of the three aromatic amino acids (Deng et al. 2015). Moreover, TyrR protein functions as an activator of its own transcription in the presence of L-Phe (Liu-Smith and Meyskens 2016).

TyrR is a polypeptide of 513 amino acid residues that is a homodimer in solution (Yang et al. 2004). By introducing site-specific amino acid substitutions, Yang and coworkers identified a number of amino acid residues at positions 2, 3, 5, 7, 9, 10 and 16, which play an important role in activation (Yang et al. 1996). As previous functional mapping of TyrR was carried out by site-directed mutagenesis, one question remained to be answered: are there any other amino acid residues in the *N*-terminal domain that are essential for activation?

In view of the crucial regulatory role of TyrR protein (Deng et al. 2015; Koyanagi et al. 2009), we sought to answer this question by performing rational mutagenesis on the N-domain of TyrR and combining biosensors with computer simulations to precisely pinpoint the L-Phe binding sites of TyrR.

Materials and methods

Strains, plasmids and culture media

The relevant *E. coli* strains and plasmids used in this study are listed in Supplementary Table 1. The *tyrR* gene was deleted using a previously described CRISPR/Cas9 approach (Zhao et al. 2016) to yield the *tyrR* mutant strain B01. The p15AS plasmid was employed for the construction of the L-Phe biosensors.

M9 minimal medium was used in the growth experiments. Glucose was added to a final concentration of 5 g/l. Media were supplemented with chloramphenicol (30 μg/ml).

Construction of expression vectors

Recombinant DNA manipulation procedures, including PCR and Gibson assembly (Bordat et al. 2015), were performed according to standard protocols. For the construction of pSen, the *tyrR* gene was amplified by PCR using the primers tyrR-F and tyrR-R with the genomic DNA of MG1655 as the template. The pSen vector backbone was amplified from plasmid p15AS, using the primers p15AS-ver2-F and p15AS-ver2-R. Then, the whole DNA sequence was generated based on these DNA templates by Gibson assembly, and was directly used to transform *E. coli* DH5α. For the construction of pSenphe, the promoter and the first 15 nucleotides of *tyrP* were amplified by PCR using the primers PtyrP-F and PtyrP-R with the genomic DNA of MG1655 as the template. The *gfp* gene was amplified from plasmid pMA5 using the primers tyrP-gfp-F and tyrP-gfp-R. DNA fusion of PtyrP-gfp was achieved using the overlap extension PCR method, with the primers PtyrP-F and tyrP-gfp-R. The pSenphe vector backbone was amplified from the pSen plasmid using the primers p15AS-ver1-F and p15AS-ver1-R. After that, the complete DNA was generated based on these DNA templates by Gibson assembly, and directly used to transform *E. coli* DH5α. The resulting plasmid pSenphe was identified and validated via colony PCR and sequencing (GENEWIZ, China).

The vector backbone of pSenphe-12 was amplified from pSenphe, using the primers tyrP-12-F and tyrP-12-R, and was digested for 1 h at 37 °C using *Dpn*I enzyme (Thermo-Fisher Scientific, USA). The resulting DNA was directly used to transform into *E. coli* DH5α and the target plasmid pSenphe-12 was identified and validated via colony PCR and sequencing (GENEWIZ, China). The plasmids pSenphe-1201, pSenphe-1202, pSenphe-1203, pSenphe-1204, pSenphe-1205, pSenphe-1206, pSenphe-1207, pSenphe-1208, pSenphe-1209, pSenphe-1210, pSenphe-1211, pSenphe-1212, pSenphe-1213, pSenphe-1214 and pSenphe-1215 were constructed on the basis of pShenphe-12 using the same approach. To examine the effects of these mutations on TyrR activation, each of the pSenphe derivatives containing the mutant *tyrR* genes was transferred into strain B01.

For the construction of ptrc99a-aroF, the promoter of the *aroF* gene was amplified by PCR using the primers ParoF-F and ParoF-R, with the genomic DNA

of MG1655 as the template. The ptrc99a-aroF vector backbone was amplified from plasmid ptrc99a using the primers ptrc99a-ver-F and ptrc99a-ver-R. The complete DNA was generated based on these DNA templates by Gibson assembly, and was directly used to transform *E. coli* DH5α. The primers used in the study are listed in Supplementary Table 2.

Measurement of GFP fluorescence

A single colony was transferred into 5 ml of lysogeny broth (LB; 1% (w/v) peptone, 0.5% (w/v) yeast extract, and 0.5% (w/v) NaCl) and cultured for 12 h at 37 °C and 200 rpm. Subsequently, the cells were concentrated by centrifugation and washed three times to remove the LB broth, and finally, 5 mL of minimal medium with 5 g/l of glucose were inoculated with the LB pre-culture at a ratio of 1:50. The cultures in minimal medium were grown at 37 °C and 220 rpm. When the culture OD_{600} reached ~ 0.6, IPTG was added to a final concentration of 0.1 mM to induce expression of the biosensor. Various amounts of Phe–Phe were then added into the individual culture samples to control the intracellular L-Phe concentration. Furthermore, after a 5 h incubation, the fluorescence (483 nm excitation, 525 nm emission) of 200 µl of the diluted cultures was monitored via multimode microplate reader.

LacZ expression assay

The β-galactosidase reporter activity was assayed as described by Miller (Miller 1972).

Modelling and molecular dynamics simulation

The 3D structural model of TyrR was based on X-ray crystal structure (PDB ID: 2JHE). Energy minimization of the constructed model was done using AMBER16. The ff14SB.redq force field was applied for molecular dynamics simulation of final model. The complete simulation methodology used in this work is available in the supporting information. The MD trajectories were further analyzed, to identify the binding mode of L-Phe with TyrR.

Results and disscussion

Construction and characterization of the L-Phe biosensor

The use of biosensors responsive to intracellular chemicals has opened the doors for solving many current challenging tasks, such as screening the best producing strains from a large library with wide genetic diversity. Such sensory-regulatory devices, mostly based on transcription factors, have been successfully used to detect the presence of many different metabolites. In contrast to the typical applications, we used the characteristics of the TyrR regulator, which is responsible for the activation of *tyrP* gene in the presence of L-Phe, to construct a pSenphe biosensor. The schematic diagram of pSenphe is presented in Fig. 1a. The pSenphe plasmid is composed of a green fluorescence protein (GFP) controlled by the *tyrP* promoter, and the CDS of the TyrR protein which activates the expression of *tyrP* in the presence of L-Phe. Using pSenphe, we can identify the L-Phe binding sites of TyrR by measuring the expression of the reporter gene from constructs with different mutant TyrR variants.

To test the performance of pSenphe, we investigated the correlation between the responsive fluorescence signal and the intracellular L-Phe concentration to confirm the functionality of this biosensor. The pSenphe sensor plasmid was introduced into strain *E. coli* MG1655 as well as the L-Phe producing strains TB1 and TB2. The L-Phe production of each strain was measured after 48 h of fermentation and the intracellular L-Phe concentrations were measured using HPLC (Fig. 1b). The GFP fluorescence values of the three strains were detected by fluorescence spectrometry at the end of the fermentation (Fig. 1b). As expected, there was an increase in specific GFP fluorescence between the control strain MG1655 and the producer strains TB1 and TB2 (Fig. 1b). However, the total fluorescence values of TB1 and TB2 were quite low, and there was no obvious difference between the fluorescence values of the two strains.

Improving the sensitivity of the biosensors

The number and location of DNA-binding sites for the TyrR protein vary for each gene. Nevertheless, they have been extensively characterized and all conform

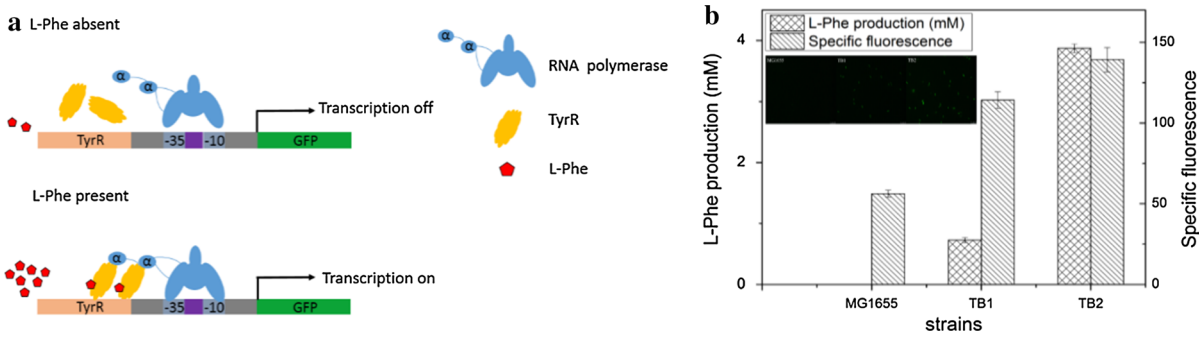

Fig. 1 Characterization and validation of the L-Phe biosensor in different L-Phe-producing strain. **a** Schematic diagram of the functional principle of the biosensor plasmid pSenphe. **b** The cytosolic L-Phe concentrations and fluorescence values of the three L-Phe producer strains MG1655, TB1 and TB2 and fluorescence microscopy images of cells carrying the L-Phe biosensor. The error bars show the standard deviations from three independent experiments

to the prototypical palindrome TGTAAN6TTTACA (Yang et al. 2004), referred to as the TyrR box. In the case of *tyrP*, the ability of TyrR protein to interact with RNA polymerase and activate transcription appears to be maximized when the upstream box is moved 3 or 12 to 14 residues upstream (Andrews et al. 1991). To improve the sensitivity of the pSenphe reporter, 12 residues (TCAGATGATCTG) were inserted into the two TyrR boxes, yielding the sensor plasmid pSenphe-12. To test the performance of pSenphe-12, we induced GFP expression by adding the Phe–Phe dipeptide to the culture medium in order to control the intracellular L-Phe concentration. The effect of Phe–Phe was tested in *E. coli* MG1655 containing pSenphe-12. The intracellular L-Phe concentration showed a linear correlation with the extracellular dipeptide concentration. Following 5 h after the addition of the dipeptide addition, GFP fluorescence was measured using fluorescence spectrometry.

As shown in Fig. 2, the sensitivity of pSenphe-12 was better than that of its parent construct. The curve showing the dipeptide dose-fluorescence signal relationship had a sigmoid shape with a sensitive range from 0.2 to 5 mM extracellular Phe–Phe dipeptide. Accordingly, the fluorescence signal plateaued when the dipeptide concentration exceeded 5 mM even though more intracellular L-Phe accumulated at greater Phe–Phe concentrations.

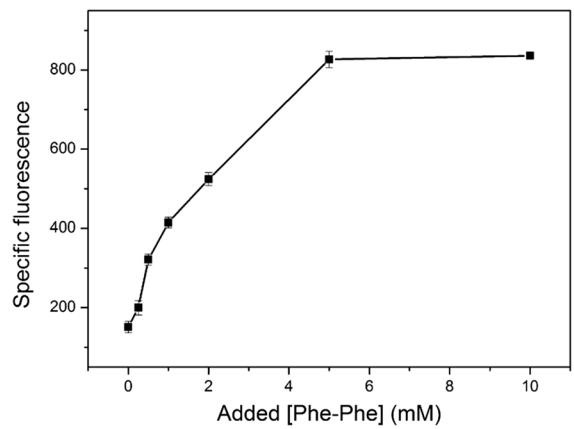

Fig. 2 Response of the biosensor towards external Phe–Phe dipeptide characterized by fluorescence spectrometry. The fluorescence signal plateaued when the dipeptide concentration exceeded 5 mM even though more intracellular L-Phe accumulated at greater extracellular Phe–Phe concentrations. The data are shown as the means ± SD; n = 3

Prediction of the L-Phe binding pockets of TyrR protein

The protein preparation wizard in Schrödinger software was used to process the protein structures. To avoid bias, all crystallographic water molecules were removed from the structures at the beginning, as were small molecules that were used for crystallization. The preparation was continued by assigning appropriate bond orders and formal charges to ligands and cofactors, as well as adding hydrogen atoms using the Schrödinger Maestro interface (Maestro 2018). SiteMap (SiteMap 2018) was employed to map possible binding pockets within the 3D structure of *E. coli* TyrR protein. After optimizing the structure, SiteMap identified two potential ligand binding sites (Fig. 3a). The docking algorithm Glide (Glide 2018) was used to perform the molecular docking experiments. The two predicted active sites encompassed

His173, Glu91, Glu160, Ser184 and Asn147 in the upper potential ligand binding site (Fig. 3b, c), as well as Arg15 and Gly12 in the lower potential ligand binding site (Fig. 3d).

Identification of binding sites that contribute to the binding of L-Phe to TyrR

To further identify the effects of the identified amino acid residues on TyrR activation, the coding sequence of each of the residues was mutated in the pSenphe-12 plasmid and introduced into the TyrR mutant strain B01. Whereas, the transcription unit *aroF* is repressed by TyrR in the presence of tyrosine. The strain with the aroF-lacZ fusion was used as a control to measure TyrR-mediated repression (Andrews et al. 1991). As shown in Fig. 4a, most changes had few effects on activation. These included substitutions of arginine at position 15 by either leucine or glutamine, substitutions of glutamate at position 91 by either proline or leucine, and substitutions of asparagine at position 147 by either leucine or glutamate. By contrast, substituting phenylalanine or leucine for glutamate at position 160 dropped the activation to about 65% of the WT value, while the substitution of leucine and valine for histidine at position 173 reduced the activation to about 73%. Changing serine at position 184 to valine or leucine reduced the activation to about 45%, suggesting that these residues play important roles. However, none of these mutations at residues 160 (B09-2), 173 (B11-2) and 184 (B13-2) significantly affects the ability of TyrR to repress the *aroF* promoter Fig. 4b, indicating that the lack of activation is not caused by a general instability of mutant proteins or the loss of DNA binding capacity. At the same time, we selected residues beyond the predicted pocket site for additional mutagenesis. Changing leucine at position 175 to glutamate had almost no effect on activation (Fig. 4a). By contrast, changing threonine at position 176 to glutamate (B16-2) led to a complete loss of the ability of TyrR protein to activate *tyrP* and repress *aroF*, indicating that this mutation may severely affect the overall conformation of the protein (Fig. 4a, b).

Determination of TyrR activation

In order to verify the binding mode of L-Phe in the pocket encompassing the residues 173, 160 and 184, two modes of L-Phe interaction with residues 173, 160,

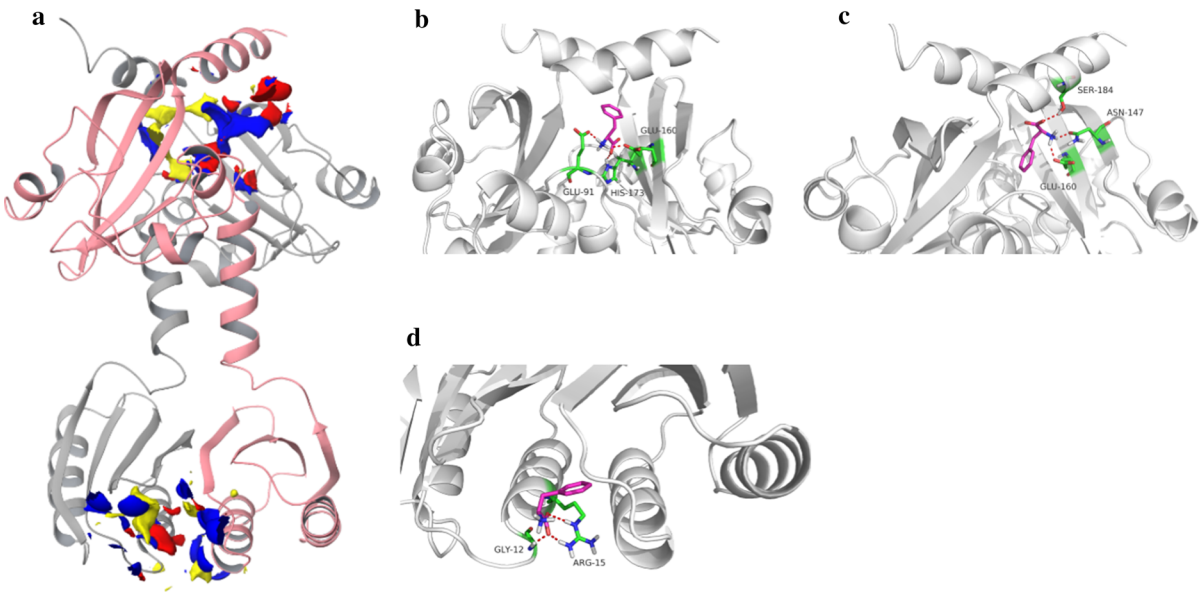

Fig. 3 Model structure of the TyrR protein. **a** Potential binding sites identified using the SiteMap program in Schrödinger software. Red represents the surface of the hydrogen bond acceptors (hbacceptor-accptr), blue represents the surface of the hydrogen bond donors (hbdonor-donor), and yellow represents hydrophobic areas (hydrophob-phil). **b**, **c** Two binding modes of L-Phe in the upper active site of TyrR. **d** Binding modes of L-Phe in the lower active site of TyrR. Carbon atoms are depicted in green, oxygen in red and nitrogen in blue. Dotted lines represent hydrogen-bonding interactions between L-Phe and TyrR

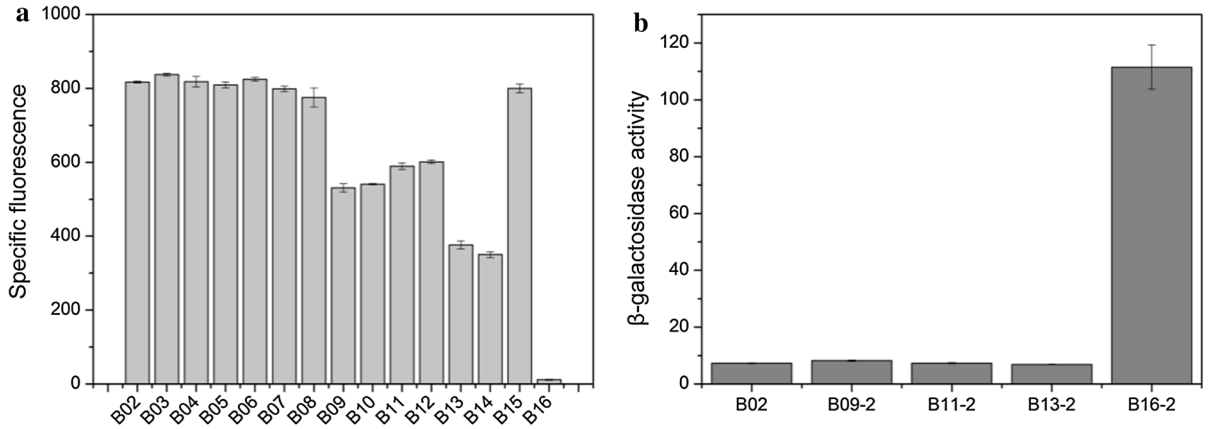

Fig. 4 Identification of the L-Phe binding sites of TyrR protein. **a** Response of biosensors containing WT and different mutated TyrR proteins measured by fluorescence spectrometry; B02(WT), B03(TyrR R15L), B04(TyrR R15Q), B05(TyrR E91P), B06(TyrR E91L), B07(TyrR N147L), B08(TyrR N147E), B09(TyrR E160F), B10(TyrR E160L), B11(TyrR H173L), B12(TyrR H173 V), B13(TyrR S184 V), B14(TyrR S184L), B15(TyrR L175E) and B16(TyrR T176E); **b** β-Galactosidase activity of strains B02, B09-2, B11-2, B13-2 and B16-2. Cells were grown in minimal medium with 5 mM L-Phe. β-Galactosidase activity was expressed in Miller units (Miller 1972). Error bars show the standard deviation of three independent experiments

91, and 184, 160, 147 were assessed using molecular dynamics (MD) simulations. Based on the initial structure shown in Fig. 3b and c, MD encompassing 50 ns at 37 °C were carried out. Finally, cluster analysis was carried out for each MD trajectory, and three representative conformations were selected (Fig. 5). Figure 5a, b, c are the three representative conformations of L-phe with TyrR [H173, E160, E91] and Fig. 5d, e, f are the three representative conformations of L-phe with TyrR [S184, E160, N147].

The frequency of individual conformations in each cluster was calculated in order to explore the possibility of emerging interactive conformations. A, Cluster 1: L-Phe interacts with carboxyl groups of 160-site side chains and amino groups of 91-site main chains to form hydrogen bonds (Frac. 49.9%). B, cluster 2: L-Phe interacts with carboxyl groups of 160-site side chains and amino groups of 173-site side chains to form hydrogen bonds (Frac. 45.0%). C, cluster 3: L-Phe interacts with carboxyl groups of 160-site side chains and amino groups of 173-site side chains to form hydrogen bonds (Frac. 5.1%). D, Cluster 1: L-Phe interacts with carboxyl groups of 160-position side chains to form hydrogen bonds, but does not interact directly with position 184 (Frac. 80.5%). E, Cluster 2: L-Phe interacts with carboxyl groups of 160-position side chains to form hydrogen bonds, but does not interact directly with position 184 (Frac. 13.5%). F, Cluster 3: L-Phe interacts with carboxyl groups of 160-position side chains to form hydrogen bonds, but not directly with position 184 (Frac. 6.0%).

The results of kinetics analysis showed that L-Phe interacts with the residues at site 173 and 160, but does not interact effectively with residue 184 at high frequency, indicating that L-Phe does not necessarily bind directly to the residue at site 184. Therefore, do mutations at position 184 affect the binding pose of L-Phe in the pocket? We took out the representative conformations of L-Phe in contact with sites 173 and 160, and verified the mutations S184L/V. The mutations S184L/V led to clashes with the benzene ring when L-Phe binds to 173 and 160, which may affect the normal binding of L-Phe in this pocket (Supplementary Fig. 1). We therefore speculated that residues 173, 160 and 184 are the binding sites of L-Phe, whereby 173 and 160 may provide hydrogen bonds for the binding of L-Phe, while the serine at 184 provides space for the benzene ring.

Finally, we further analyzed the interactions of residues of 15 site with L-Phe in the same way. However, the results indicated that L-Phe does not interact with R15 (Supplementary Fig. 2).

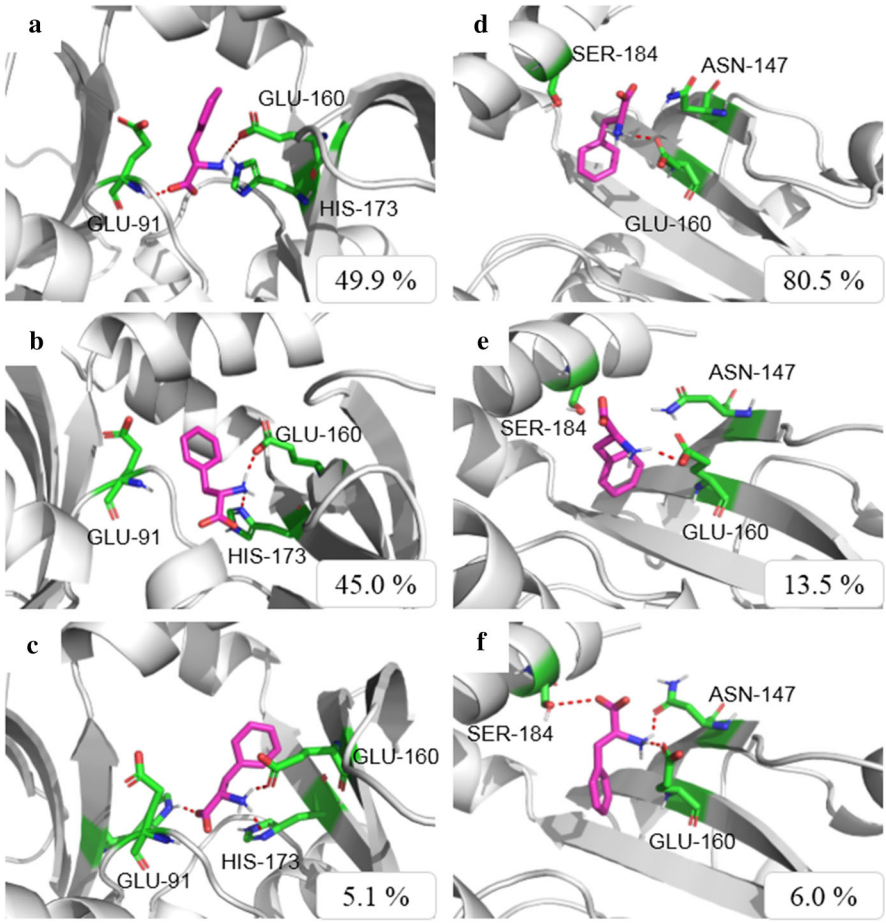

Fig. 5 Representative conformations of three clusters in MD simulations. **a** Cluster 1: L-Phe interacts with carboxyl groups of 160-site side chains and amino groups of 91-site main chains to form hydrogen bonds (Frac. 49.9%). **b** Cluster 2: L-Phe interacts with carboxyl groups of 160-site side chains and amino groups of 173-site side chains to form hydrogen bonds (Frac. 45.0%). **c** Cluster 3: L-Phe interacts with carboxyl groups of 160-site side chains and amino groups of 173-site side chains to form hydrogen bonds (Frac. 5.1%). **d** Cluster 1: L-Phe interacts with carboxyl groups of 160-position side chains to form hydrogen bonds, but does not interact directly with position 184 (Frac. 80.5%). **e** Cluster 2: L-Phe interacts with carboxyl groups of 160-position side chains to form hydrogen bonds, but does not interact directly with position 184 (Frac. 13.5%). **f** Cluster 3: L-Phe interacts with carboxyl groups of 160-position side chains to form hydrogen bonds, but not directly with position 184 (Frac. 6.0%)

Conclusions

An L-Phe biosensor was used to identify a number of new amino acid substitutions which render the TyrR protein defective in transcriptional activation. We concluded that the residues at positions 173, 160 and 184 are the binding sites of L-Phe, whereby 173 and 160 likely provide hydrogen bonds for the binding of L-Phe, while 184 is spatially close to and makes room for the benzene ring of L-Phe. By contrast, the replacement of threonine at position 176 with glutamate led to the complete loss of the activation- and repression functions of the TyrR protein, which suggests that its tertiary structure does not tolerate substitutions at this position.

Acknowledgements This work was financially supported by National Key R&D Program of China (2018YFD0901001), Tianjin Science Fund for Distinguished Young Scholars (17JCJQJC45300), Nature Science Foundation of Tianjin City (CN) (16JCYBJC23500), Tianjin Science and Technology Project (15PTCYSY00020) and the Science and Technology Service Network (STS) Initiative of Chinese Academy of Sciences.

Supporting information Supplementary Table 1—Strains and plasmids used in this study.

Supplementary Table 2—Primers used in this study.

Supplementary Figure 1—The effect of 184 site of TyrR on the binding mode of L-phe. The representative conformation of L-phe with 173 and 160 sites is verified by mutation S184L/V. The mutations S184L and S184V would clash with the benzene ring of L-Phe when binds 173 and 160 sites.

Supplementary Figure 2—Representative conformations of three clusters in MD simulations for 15 site of TyrR protein.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Andrews AE, Dickson B, Lawley B, Cobbett C, Pittard AJ (1991) Importance of the position of TYR R boxes for repression and activation of the tyrP and aroF genes in *Escherichia coli*. J Bacteriol 173:5079–5085

Bordat A, Houvenaghel MC, German-Retana S (2015) Gibson assembly: an easy way to clone potyviral full-length infectious cDNA clones expressing an ectopic VPg. Virol J 12:89. https://doi.org/10.1186/s12985-015-0315-3

Deng Z et al (2015) TyrR, the regulator of aromatic amino acid metabolism, is required for mice infection of Yersinia pestis. Front Microbiol 6:110. https://doi.org/10.3389/fmicb.2015.00110

Glide (2018) Glide. Schrödinger, LLC, New York, p 2018

Koyanagi T, Katayama T, Suzuki H, Kumagai H (2008) Altered oligomerization properties of N316 mutants of *Escherichia coli* TyrR. J Bacteriol 190:8238–8243. https://doi.org/10.1128/jb.00889-08

Koyanagi T, Katayama T, Suzuki H, Onishi A, Yokozeki K, Kumagai H (2009) Hyperproduction of 3,4-dihydroxyphenyl-L-alanine (L-Dopa) using erwinia herbicola cells carrying a mutant transcriptional regulator TyrR. Biosci Biotech Bioch 73:1221–1223. https://doi.org/10.1271/bbb.90019

Liu-Smith F, Meyskens FL (2016) Molecular mechanisms of flavonoids in melanin synthesis and the potential for the prevention and treatment of melanoma. Mol Nutr Food Res 60:1264–1274. https://doi.org/10.1002/mnfr.201500822

Maestro (2018) Maestro. Schrödinger, LLC, New York, p 2018

Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

SiteMap (2018) SiteMap. Schrödinger, LLC, New York, p 2018

Yang J, Camakaris H, Pittard AJ (1996) Further genetic analysis of the activation function of the TyrR regulatory protein of *Escherichia coli*. J Bacteriol 178:1120–1125. https://doi.org/10.1128/jb.178.4.1120-1125.1996

Yang J, Hwang JS, Camakaris H, Irawaty W, Ishihama A, Pittard J (2004) Mode of action of the TyrR protein: repression and activation of the tyrP promoter of *Escherichia coli*. Mol Microbiol 52:243–256. https://doi.org/10.1111/j.1365-2958.2003.03965.x

Yang J, Roy A, Zhang Y (2013) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 29:2588–2595. https://doi.org/10.1093/bioinformatics/btt447

Zhao D et al (2016) Development of a fast and easy method for *Escherichia coli* genome editing with CRISPR/Cas9. Microb Cell Fact 15:205. https://doi.org/10.1186/s12934-016-0605-5