Journal of Biotechnology 154 (2011) 248–254

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

# Integrating molecular dynamics and co-evolutionary analysis for reliable target prediction and deregulation of the allosteric inhibition of aspartokinase for amino acid production

Zhen Chen, Sugima Rappert, Jibin Sun [1], An-Ping Zeng *

*Institute of Bioprocess and Biosystems Engineering, Hamburg University of Technology, Denickestrasse 15, D-21073 Hamburg, Germany*

ARTICLE INFO

*Article history:*
Received 14 February 2011
Received in revised form 2 May 2011
Accepted 9 May 2011
Available online 14 May 2011

*Keywords:*
Aspartokinase
Allosteric regulation
Molecular dynamics simulation
Statistical coupling analysis

ABSTRACT

Deregulation of allosteric inhibition of enzymes is a challenge for strain engineering and has been achieved so far primarily by random mutation and trial-and-error. In this work, we used aspartokinase, an important allosteric enzyme for industrial amino acids production, to demonstrate a predictive approach that combines protein dynamics and evolution for a rational reengineering of enzyme allostery. Molecular dynamic simulation of aspartokinase III (AK3) from *Escherichia coli* and statistical coupling analysis of protein sequences of the aspartokinase family allowed to identify a cluster of residues which are correlated during protein motion and coupled during the evolution. This cluster of residues forms an interconnected network mediating the allosteric regulation, including most of the previously reported positions mutated in feedback insensitive AK3 mutants. Beyond these mutation positions, we have successfully constructed another twelve targeted mutations of AK3 desensitized toward lysine inhibition. Six threonine-insensitive mutants of aspartokinase I–homoserine dehydrogenase I (AK1–HD1) were also created based on the predictions. The proposed approach can be widely applied for the deregulation of other allosteric enzymes.



## 1. Introduction

Allosteric regulation is one of the fundamental mechanisms that control almost all cellular metabolisms and gene regulation (Tsai et al., 2009). Deregulation of allsoteric inhibition has been a challenge in designing and optimizing metabolic pathways for the production of target metabolites such as amino acids. So far, this is achieved almost exclusively by multiple rounds of random mutation and selection. Despite the successful application of these approaches for the development of amino acid producers, they have several disadvantages. For example, undesirable mutations would be introduced which may cause growth retardation and by-product formation. Furthermore, well selectable phenotypes such as resistance to analogs of inhibitors are prerequisite for these processes. Thus, these approaches cannot be used for some allosteric enzymes which lack corresponding selectable phenotypes for the mutants. A rational approach that could be used to guide targeted reengineering of allosteric enzymes without screening or selection process is highly desired.

Recent advances in structural biology together with computational analysis are opening a new avenue toward understanding and rational reengineering of allosteric enzymes (Chen et al., 2010). Two approaches, molecular dynamic simulation (MD) and statistical coupling analysis (SCA) are especially useful for such purpose (Estabrook et al., 2005). Allosteric regulation is a dynamic process and thus MD can provide valuable information for correlated or anti-correlated motions among different structural elements relating dynamics to allostery (Smock and Gierasch, 2009). On the other hand, SCA can reveal correlated mutations of protein family and help to identify coupled residues contributing to the allosteric communication (Lockless and Ranganathan, 1999; Suel et al., 2003). Estabrook et al. (2005) demonstrated the usefulness of the combined approach of SCA and MD for identification of amino acid pairs essential for catalysis. In this work, we further show that such an integrated approach is efficient to define a cluster of residues that are essential for allosteric regulation and can be used for rational deregulation of allosteric inhibition.

Aspartokinase was chosen in this work as a model enzyme. It catalyzes the phosphorylation of aspartate and controls the biosynthesis of several industrially important amino acids such as lysine, threonine and methionine (Yoshida et al., 2007). In *Escherichia coli*, there exist three aspartokinase isozymes. Two of them, aspartokinase I–homoserine dehydrogenase I (AK1–HD1) encoded by *thrA gene* and aspartokinase III (AK3) encoded by *lysC* gene are allosteric

* Corresponding author. Tel.: +49 40 42878 4183; fax: +49 40 42878 2909.
*E-mail address:* aze@tu-harburg.de (A.-P. Zeng).
[1] Present address: Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjing, 300308, PR China.

0168-1656/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2011.05.005

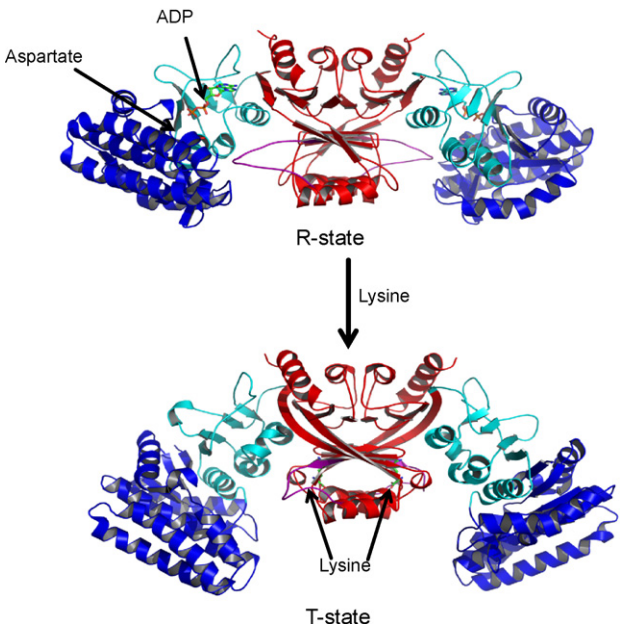


**Fig. 1.** Conformational transition of aspartokinase III (AK3) by allosteric regulation of lysine. The regulatory domain is shown in red. The N-lobe and C-lobe of the catalytic domain are denoted by blue and cyan, respectively. Residues 352–362 that showed the largest conformational change are colored in purple. The substrates (ADP, aspartate) and the inhibitor (lysine) are represented by CPK models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

enzymes and especially important for lysine and threonine production. AKI–HDI is allosterically inhibited by threonine and its synthesis is repressed by threonine plus leucine (Bearer and Neet, 1978) while AK3 is inhibited and repressed by lysine (Kotaka et al., 2006). In the last fifty years, considerable efforts have been made to deregulate these two enzymes from allosteric inhibition by random mutation and selection of mutants resistant to lysine analogue S-(2-aminoethyl)-L-cysteine (AEC) (Kikuchi et al., 1999; Miyata et al., 2001), or threonine analogue α-amino-β-hydroxyvaleric acid (AHV) (Lee et al., 2003). However, only a limit number of positive mutations have been identified so far, especially for AK1–HD1.

The crystal structures of AK3 complex with substrates (R-state) or lysine (T-state) have been solved (Kotaka et al., 2006). Lysine binding induces a large conformational change of AK3 (Fig. 1). AK3 consists of an N-terminal catalysis domain and a C-terminal regulatory domain. The regulatory domain possesses two motifs called as ACT domains (Chipman and Shaanan, 2001) which are responsible for the lysine binding. Mapping of the reported mutations into the three-dimensional structure of AK3 enables us to evaluate their roles for allosteric regulation. Interestingly, these mutated residues are located not only within lysine binding sites but also in other regions of the protein (Table 1 and Fig. 4A). This motivates us to carry out a more systematic analysis of the whole structure of AK3 to identify residues which may form an interacting network responsible for the allosteric regulaltion. Specific residues of this interacting network should be evaluated as potential targets for deregulation of the allostery.

## 2. Materials and methods

### 2.1. Molecular dynamic simulation (MD) and cross-correlation analysis

The starting structure for the MD simulation of AK3 with lysine was based on the crystal structure of T-state AK3 (PDB code 2J0X). Aspartate and other ligands were removed from the structure and the missing residues were repaired using MODELLER 9v5 (http://salilab.org/modeller/). For the MD simulation of AK3 without lysine, lysine was also removed from the previous structure. Dynamic trajectories were computed using AMBER 10.0 with parm99SB force field (Duan et al., 2003). Protein was solvated in a box of TIP3P water molecules (Jorgensen, 1981) with the minimal distance of 1.5 nm from the protein to the box wall. $Na^+$ ions were added to neutralize the systems. 1500 steps of steepest-descent energy minimization and 2500 steps Newton–Raphson minimization were performed before the MD simulation. The systems were then heated to 300 K, followed by 500 ps equilibration and 10 ns MD simulations. The particle mesh Ewald method (Darden et al., 1993) was used to calculate the long-range electrostatics interactions. Non-bonded interactions were cutoff at 12.0 Å and updated every 25 steps. The SHAKE method (Ryckaert et al., 1977) was applied to constrain all covalent bonds involving H atoms. Each simulation was coupled to a 300 K thermal bath at 1.0 atm of pressure by applying the algorithm of Berendsen et al. (1984). The temperature and pressure coupling parameters were set as 0.2 and 0.05 ps, respectively. The integration step was set to 2 fs and the coordinates were saved every 0.1 ps, giving a total number of 100,000 structures for each trajectory.

Cross-correlation analysis of the trajectories from 2 ns to 10 ns was performed to evaluate the dynamical correlation between any two residues. The cross-correlation coefficient is defined as $C_{(ij)} = \left(\left\langle \Delta r_i \bullet \Delta r_j \right\rangle\right) / \left(\left\langle \Delta r_i^2 \right\rangle^{1/2} \left\langle \Delta r_j^2 \right\rangle^{1/2}\right)$, where $\Delta r_i$ and $\Delta r_j$ denotes the displacement vectors of residue $i$ and $j$ and the angle brackets denote ensemble average. The coordinate sets of 2 ns were used as the references. $C_{(ij)} = 1$ indicates that the motions of two residues are completely correlated (same phase) while $C_{(ij)} = -1$ indicates that the motions of two residues are completely anti-correlated (opposite phase). The extent of correlation was calculated using AMBER10.0.

### 2.2. Multiple sequence alignment (MSA) and statistical coupling analysis (SCA)

Sequences of the aspartokinase family proteins were collected from the UniRef90 database in UniProt Knowledgebase (http://www.uniprot.org/). Any sequence sharing >90% similarity to another sequence was removed in order to get a diverse distribution of samples. The sequences were aligned with MUSCLE (Edgar, 2004) followed by structure-guided manual adjustment (Doolittle, 1996). The sequence positions with gap frequency higher than 20% were deleted. The final alignment consisted of 340 sequences and 424 positions.

SCA measures the evolutional correlation between any two residue positions $C_{ij}^{ab} = \phi_i \phi_j (f_{ij}^{ab} - f_i^a f_j^b)$, where $f_i^a$ and $f_j^b$ denote the observed frequency of amino acid $a$ and $b$ at position $i$ and $j$. $f_{ij}^{ab}$ represents the joint frequency of having $a$ at position $i$ and $b$ at position $j$, and $\phi_i$ and $\phi_j$ are the positional conservation-based weights. The detailed procedure for the SCA calculation has been described elsewhere (Halabi et al., 2009; Suel et al., 2003). The SCA matrix was calculated using the METLAB script derived from the publication of Halabi et al. (2009).

### 2.3. The SCA·MD matrix

SCA·MD matrix is to measure the positional correlation based on both evolutional and dynamical contribution. It was created by multiplying the individual elements of the SCA matrix with the corresponding elements of the truncated cross-correlation matrix of MD trajectory of AK3 with lysine.