Co-evolutionary analysis enables rational deregulation of allosteric enzyme inhibition in Corynebacterium glutamicum for lysine production

Zhen Chen, Weiqian Meyer, Sugima Rappert, Jibin Sun§ and An-Ping Zeng*

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

Running title: Deregulation of allosteric enzyme

*Corresponding author: Zeng, A.-P. E-mail: aze@tu-harburg.de
Tel: +49-40-42878-4183. Fax: +49-40-42878-2909

§Present address:
Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjing, 300308, P. R. China
Abstract

Product feedback inhibition of allosteric enzymes is an essential issue for developing highly efficient microbial strains for bioproduction. Here we use aspartokinase from *Corynebacterium glutamicum* (CgAK), a key enzyme controlling the biosynthesis of industrially important aspartate family amino acids, as a model to demonstrate a fast and efficient approach to deregulate allostery. In the last fifty years many researchers and companies have made considerable efforts to deregulate this enzyme from allosteric inhibition by lysine and threonine. However, only a limited number of positive mutants have been identified so far, almost exclusively by random mutation and selection. In this study, we used statistical coupling analysis of protein sequences, a method based on co-evolutionary analysis, to systematically clarify the interaction network within the regulatory domain of CgAK that is essential for the allosteric inhibition. A cluster of interconnected residues linking different inhibitors’ binding sites as well as other regions of protein have been identified, including most of the previously reported positions of successful mutations. Beyond these mutation positions, we have created other 14 mutants that can partially or completely desensitize CgAK from allosteric inhibition as shown by enzyme activity assay. Introducing only one of the inhibition-insensitive CgAK mutations (here Q298G) into wild-type *C. glutamicum* strain by homologous recombination resulted in accumulation of 58 g/L L-lysine within 30h of fed-batch fermentation in a bioreactor.
**Introduction**

Construction of a minimally mutated strain with high production rate and final titer represents a major effort and future direction of strain engineering (14). For such a purpose, only the beneficial mutations are ought to be integrated into the genome of a production strain. The unnecessary or detrimental mutations, which are often introduced during the processes of random mutation and selection, should be avoided in order to minimize their side-effects on cellular constitutions such as cell growth and substrate consumption. A fast identification and construction of beneficial mutations is thus a key issue for successful construction of a minimally mutated strain. For many industrial processes such as amino acids production, deregulation of allosteric inhibition is the initial and highly challenging step. A rational approach that can fast and efficiently identify and introduce targeted mutations to desensitize allosteric inhibition of enzymes is highly desirable to this end. In the present study, we describe a strategy employing co-evolutionary analysis of protein sequences to circum the problem of allosteric inhibition of enzymes and to demonstrate its application to create a minimally mutated lysine-producting strain. The allosteric enzyme aspartokinase is selected as a model enzyme for proof of concept.

Aspartokinase is the first and committed enzyme involved in the biosynthesis of several industrially important amino acids such as lysine, threonine and methionine (2). Aspartokinase catalyzes the phosphorylation of β-carboxyl group of aspartate using ATP and is feedback regulated by end products. It has different isozymes and can be regulated in different ways. For example, *Escherichia coli* has three aspartokinases,
two of them are conjugated with homoserine dehydrogenases. Aspartokinase I-homoserine dehydrogenase I (AKI-HDI) is allosterically inhibited by threonine and its synthesis is repressed by threonine plus leucine, while aspartokinase II-homoserine dehydrogenase II (AKII-HDII) is only repressed by methionine (8). The third aspartokinase isozyme in *E. coli*, aspartokinase III (AKIII), is a single functional enzyme which is inhibited and repressed by lysine (16, 17, 19). In *Corynebacterium glutamicum*, there is only one single aspartokinase (CgAK) which is concertedly inhibited by lysine and threonine (29).

CgAK has a α₂β₂-type heterotetrameric structure, containing equimolar α and β subunits which are encoded by in-frame overlapping genes (15). The α subunit consists of two domains, the N-terminal catalytic domain and C-terminal regulatory domain. The β subunit is identical to the regulatory domain of α subunit. The β subunit or the regulatory domain of the α subunit consists of two motifs called as ACT domains, which share a common fold of βαβαβ (5, 10). ACT domain appears in many allosteric enzymes, especially enzymes in the amino acids and purine synthesis pathways, serving as a small molecule binding side for allosteric regulation of the proteins (5). In CgAK, lysine and threonine are also bound at the interface between different ACT domains from α subunit and β subunit as indicated in Figure 1A (30).

A deregulation of CgAK from allosteric inhibitions was considered as the most important step toward lysine production by *C. glutamicum* (8). Mutants with CgAK insensitive to lysine inhibition have been obtained by random mutation and selection with the phenotype of resistance to S-(2-aminoethyl)-L-cysteine (AEC), an L-lysine
structural analog (20, 23). This traditional method, however, is time and labor-consuming and not efficient enough. Only a limited number of positive mutations within less than 10 amino acid positions have been identified during the past 40 years and these points are subjects of more than 20 patents (8). The underlying mechanism upon such mutations is nevertheless unknown. These positions, however, may only represent a small percentage of the large sequence space which may influence the allosteric control of the enzyme. A fast and systematic method is desirable to explore the sequence space and to better understand the mechanism. To this end, we attempted to use the method of statistical coupling analysis (SCA) (9, 18, 26), a powerful co-evolutionary analysis approach, to clarify the allosteric interaction network within the regulatory domain of CgAK. This allosteric interaction network was then used as the basis for designing efficient targeted mutations.

SCA is a sequence-based statistical method for analyzing conserved cooperative actions between amino acids (8, 26). Based on a multiple-sequence alignment of protein families, SCA can identify a small subset of residues that are coevolved during protein evolution. These coevolved or correlated residues are considered to be especially important for the structure or function of proteins (1, 3). Prediction of these correlated residues can greatly reduce the sequence space of proteins and facilitate the identification of potential allosteric interaction network that mediates protein function. SCA has been successfully used to clarify the interconnected allosteric networks of several protein families, such as PDZ domains (18), G-proteins, and serine proteases.
However, this method has not been used for guiding targeted mutagenesis of industrial important allosteric proteins such as aspartokinases.

In this study, we used the SCA method to define a cluster of evolutionarily correlated residues within the regulatory domain of CgAK which is responsible for the allosteric inhibition. This small cluster of residues includes most of the previously proved positions for efficient mutations and enabled us to fast design another 14 deregulated CgAK mutants. Mapping of these residues onto the recently solved crystal structure of CgAK (30) helped us to understand the potential effects of these point mutations upon the regulation of this enzyme. Finally, we integrated one such point mutation (Q298G) into the wildtype *C. glutamicum* ATCC 13032 which resulted in the remarkably high lysine production (58 g/L) in fed-batch fermentation based on only one point mutation. The mutant showed similar growth behavior as the wildtype strain and can be used as a basis for further strain development.

**Materials and Methods**

**Multiple sequence alignment**

Sequences of the aspartokinase family proteins were collected from the UniRef90 database in UniProt Knowledgebase (http://www.uniprot.org/). The sequences are aligned with MUSCLE (7) and ClustalX (27) followed by structure-guided manual adjustment (6). The sequences of the regulatory domains of aspartokinase family were truncated from this dataset according to the sequence of β subunit of CgAK. Any
sequence sharing >90% similarity to another sequence was then removed in order to get the diverse distribution of samples. The sequence positions with gap frequency higher than 20% was deleted. The final alignment consisted of 500 sequences and 160 amino acid positions.

**Statistical coupling analysis**

The detailed procedure for the SCA calculation has been previously described (11, 26). The METLAB script can be downloaded from the publication of Halabi et al. (11). In summary, SCA gives a matrix that calculates the weighted sequence correlation between any two positions, \( C_{ij} = \phi_i \phi_j (f_{ij} - f_i f_j) \), where \( f_{ij} \) denotes the observed frequency of amino acid \( a \) at position \( i \) and \( \phi_i \) is the positional conservation-based weight. The positional conservation is defined as \( D_i = \ln\left(\frac{f_i(1 - q^a)}{(1 - f_i)q^a}\right) \), where \( q^a \) is the background frequency of this amino acid. \( \phi_i = \partial D_i / \partial f_i \).

**Preparation of expression vector for CgAK and site-directed mutagenesis**

For the overexpression and enzyme purification of CgAK, the *lysC* gene was amplified by PCR from genomic DNA of *C. glutamicum* ATCC 13032 using primers *lysC*_ndeI_F (5'-GAGCCTCATATGGCCCTGGTCGTACAGAAATA-3') and *lysC*_xhoI_R (5'-ATTAATCTCGAGGCGTCCGGTGCCTGCTCATATAA-3'). The DNA fragment was double-digested with NdeI and XhoI, and ligated with the NdeI-XhoI digested plasmid pET-22b (+) (Novagen). The expression vector was designated as pElysC. 14 point mutations of CgAK derived from SCA calculation were created using Quick-change site-directed mutagenesis kits (Stratagene) according to standard...
protocol, named as pElysC(I272E), pElysC(D274A), pElysC(E278V), pElysC(F283R),
pElysC(Q298G), pElysC(N299L), pElysC(T336L), pElysC(M365A), pElysC(N372A),
pElysC(N374A), pElysC(I375P), pElysC(E382A), pElysC(R384L), and
pElysC(S386A), respectively. All of the CgAKs, including wildtype and mutants, have
(His)$_6$-tags at C-terminus.

**Protein expression and purification of CgAK and its mutants**

CgAK and its mutants were overexpressed in *E. coli* BL21-CodonPlus (DE3)-RIL cells
(Stratagene). The recombinant cells bearing the expression vectors were grown in LB
media supplemented with 100 ug/ml ampicilin at 37 ℃. When the absorbance of the
culture at 600nm reached 0.6, gene expression was induced by adding 0.1mM isopropyl
β-D-thiogalactopyranoside (IPTG) and the culture was continued for an additional
12-14h at 20℃. The cells were harvested and washed with buffer A (20 mM Tris–HCl,
pH 7.5), and suspended in buffer B (20 mM Tris–HCl (pH 7.5), 150 mM NaCl, 0.5 M
ammonium sulfate). Suspended cells were disrupted by sonication (10x30s at 200w
with 60s cooling interval) and centrifuged at 20,000 g for 1h. The supernatant was
purified by Ni$^{2+}$-NTA column (GE Healthcare Bio-Sciences, Piscataway, NJ) for
obtaining samples for enzyme activity assay.

**Enzyme assay**

The enzyme activity of aspartokinase was detected by the method of Black and Wright
(2). The quantity of aspartate hydroxamate formed in the presence of hydroxylamine
was measured at 540nm. The standard assay reaction in 1ml contained 200 mM
Tris-HCl (pH 7.5), 10 mM MgSO₄·6H₂O, 10 mM Aspartate, 10 mM ATP, 160mM NH₂OH·HCl (neutralized with KOH), and appropriate amount of enzyme. After incubation at 30℃ for 30 min, the reaction was stopped by mixing with 1ml 5% (w/v) FeCl₃ solution and the absorbance at 540 nm was monitored. The kinetic analyses were carried out under the same conditions except for the substrate concentrations. For the characterization of the inhibition profiles of CgAKs, different concentrations of lysine and threonine were added in the standard reaction systems. The activities were expressed as relative activities normalized by the specific activities of CgAKs without inhibitions. One unit of enzyme activity is defined as the formation of one micromole of aspartate hydroxamate per minute under assay conditions. All of the measurements were repeated three times.

**Generation of lysine-producing strain *C.glutamicum* LC298 by homologous recombination**

The feedback-insensitive CgAK with mutation of Q298G was introduced into the strain *C.glutamicum* ATCC 13032 via homologous recombination. The fragment bearing a point mutation of Q298G in *lysC* gene was amplified from the previously described plasmid pElysC(Q298G) using two primers:

5´-GAGCCTGAATTGCCTCGGTGTCGTACAGAAATA-3´ and 5´-ATTATGGATCCGCGTCCGGTGCCTGCATAAA-3´. PCR product was cloned into the EcoRI/BamHI site of suicide plasmid pK18mobsacB (8) and the resulted plasmid was designated as pKlysC(Q298G). This plasmid was electroporated into *C.glutamicum* ATCC 13032 and the recombinants with single crossover homologous
recombination were selected for kanamycin resistance. The kanamycin-resistant transformant was grown without kanamycin for 1 day and the new recombinants with the second homologous recombination was selected via sucrose tolerance marker. Sequence analysis was performed for the validation of the second recombination. The strain with correct mutation was designated as *C. glutamicum* LC298.

**Fed-batch cultivation and analytical method**

The medium for the preculture and fermentation of *C. glutamicum* have been described elsewhere (20). The main fermentation was carried out in a 2L fermentor at 34℃. The feeding solution contained 50%(w/v) glucose, 4.5%(w/v) NH₄Cl, and 0.5mg/L D-biotin. The feeding started after the initial sugar was consumed and the feeding rate was controlled to keep the glucose concentration lower than 10g/L. The agitation speed was changed during the fermentation to keep sufficient oxygen supply. The pH was maintained at 7.3 by adding NH₄OH. The fed-batch fermentations were repeated three times for both strains (*C. glutamicum* ATCC 13032 and LC298).

Cell concentration was determined as optical density at 660nm and dry biomass measurement. The glucose concentration was determined by HPLC. L-lysine was quantified as described by Hsieh et al.(13).

**Results**

**Identification of highly correlated residues within the regulatory domain of CgAK**
The regulatory domain of CgAK was considered to be primarily responsible for the allosteric inhibition (15, 29). All of the lysine-insensitive mutations were identified within the regulatory domain of CgAK (20, 23, 29). Thus, we performed SCA to the regulatory domain of aspartokinase family in order to identify a small subset of residues which would be potential targets for point mutations. After deleting positions with high gap frequency, the final samples for SCA calculation contained 500 sequences and 160 amino acid positions corresponding to residues 250-409 of CgAK. The evolutionary correlation between these 160 positions was shown in form of a SCA matrix in Figure 1D. Residue pairs that are evolutionarily correlated would have high correlation values (red color) while uncorrelated residue pairs show low correlation value (blue color).

The 25 highly correlated positions are summarized in Table 1. Strikingly, most of the previously reported mutation points of feedback-resistance CgAKs are within these identified positions. Based on this cluster of residues, we have created other 14 point mutations (Table 1). All of the constructed CgAK mutants are partially or completely insensitive to the concerted inhibition of lysine and threonine (Figure 2 and Figure 3), indicating that these highly correlated residues are essential for the allosteric regulation of CgAK.

Based on the recently published crystal structure of CgAK (30), we are able to map these 25 highly correlated residues into the three-dimensional structure (Figure 1E). Interestingly, although they are sparsely distributed within the sequence of CgAK, this small cluster of residues forms an interconnected network mainly mediated by Van der Waals interaction or hydrogen bonds. This interconnected network not only covers the
most important residues of the inhibitors’ binding sites but also provides a route linking the two threonine binding sites and one lysine binding site, indicating their central roles for mediating the concerted inhibition of CgAK. For a better understanding of their roles in allosteric regulation, we describe the mutational effect of these residues according to their locations in more detail in the following sessions.

**Mutation of highly correlated residues within inhibitors’ binding sites**

The purified wildtype CgAK contains equal molar of alpha and beta subunit (Figure 1B) and is strongly inhibited by the concerted action of threonine and lysine (Figure 1C). Threonine or lysine alone only shows slight inhibition on CgAK. Mutation of points within threonine (or lysine) binding sites to reduce the binding affinity of threonine or lysine would be expected to destroy their concerted action upon CgAK and thus reduce the concerted inhibition.

CgAK has two threonine binding sites per $\alpha\beta$ dimer which are located at the interface between the regulatory domain of $\alpha$ subunit and $\beta$ subunit (30) (Figure 1A). The two threonine binding sites are equal from sequence level if we consider that the sequence of $\beta$ subunit is identical to the regulatory domain of $\alpha$ subunit. The highly correlated residues within threonine binding sites are shown in Figure 2A. Two mutations, G277A and A279V have been previously reported for the deregulation of CgAK by concerted inhibition and the *C. glutamicum* mutant with the latter point mutation showed AEC-resistance (29). Both G277 and A279 participate in the hydrogen-bond network with threonine through water molecules. Mutations of other highly correlated residues
within threonine binding sites in our experiment, either by destroying the hydrogen bonds or introducing steric hindrance, also reduce the concerted inhibition of CgAK by lysine and threonine (Figure 2C). Mutations of I272E, Q298G, N372A and I375P not only affect the threonine binding but also reduce the lysine inhibition. The concerted inhibitions of these mutants are also significantly removed.

Although CgAK has two potential lysine binding units per ™ß dimer, only one lysine molecule has been found in the crystal structure which locates at the interface between the ™ subunit and ℟ subunit (30) (Figure 1A). The highly correlated residues within lysine binding sites are shown as Figure 2B. Two mutations, T361A (29) and S381F (23) have been previously reported for the deregulation of CgAK by concerted inhibition and the C. glutamicum mutant with the latter point mutation showed AEC-resistance. The results are understandable as both of these two point mutations would destroy the corresponding hydrogen bonds and the binding of lysine would be inhibited. We constructed another mutation of highly correlated residues within this region, E382A. Strangely, this point mutation doesn’t change the inhibition profile of CgAK by lysine but change its inhibitions by threonine alone and lysine plus threonine (Figure 2C). Threonine partially inhibited the mutant while the concerted inhibition was reduced. The reason is unclear. The effects of mutations on I291, I293 and D294 have not been examined in this study. However, mutations of the corresponding points in aspartokinase III from E. coli release lysine inhibition (Chen et al., submitted).

**Mutation of highly correlated residues in other regions of protein**
Some highly correlated residues located in other regions of the regulatory domain are not directly involved in the inhibitors’ binding (Figure 3A). These residues, however, may play important roles in signal transduction between inhibitors binding sites and catalytic domains (24, 30). For example, the CgAK with mutation of S301F binding both lysine and threonine has been crystallized, indicating that the inhibitors’ binding was not destroyed upon such mutation (30). Thus, the potential reason for its down-regulation of CgAK is that the mutation destroys the allosteric interaction network and thus affects the allosteric signal transduction. Several AEC-resistance mutants have been reported to have mutation points belong to this group of residues, including S301F (29), T308A (29), T311I (20). We also created several other point mutations belonging to this group as shown in Figure 3B. Mutation of T336L doesn’t change the inhibition profiles of CgAK by lysine and threonine alone but reduces the concerted inhibition, revealing that it may destroy the signal transduction pathway of threonine as threonine only work with lysine in concerted behavior. Mutation of F283R completely removes the inhibitions of CgAK by lysine and threonine, indicating that this residue may be crucial for the signal transduction of both lysine and threonine. Interestingly, mutations of N299L, M365A, R384L and S386A transfer threonine into an allosteric activator and reduce the concerted inhibition. This phenomenon was also illustrated for the mutation of G277A (29). The reason is unclear now. All of these aspartokinase mutants having this property, however, showed relatively low activities compared to that of wildtype aspartokinase. One potential reason could be that these mutations destroy the interaction between the α subunit and β subunit which is
essential for enzyme activity. Threonine has been reported to be able to stabilize the protein complex and thus may recover part of the enzyme activity (29). A more clear understanding of this phenomenon needs the aid of protein dynamics which is under investigation in our group.

**Introduction of point mutation Q298G into \textit{lysC} gene for lysine over-production**

One of the feedback-insensitive mutations, Q298G, was introduced into the \textit{lysC} gene of \textit{C. glutamicum} ATCC 13032 through a two-step homologous recombination. The resulting strain is designated as \textit{C. glutamicum} LC298. The production of lysine by strain LC298 was investigated in fed-batch fermentation using the wildtype strain as a control. All of the experiments have been repeated three times. The results are depicted in Figure 4.

The growth profile of strain LC298 is similar to that of wildtype strain although the final growth level was somewhat lower. Strain LC298 started exponential growth at 4h and reached stationary phase after 16h. Lysine was mainly accumulated during the fed-batch stage (8-30h) and reached an average of 58 g/L within 30h. After 30h, the cell entered the decline phase and lysine production almost stopped. In contrast, there was no extracellular lysine accumulated in \textit{C. glutamicum} ATCC 13032. Thus, the lysine production in strain LC298 should be due to the deregulation of allosteric inhibition of CgAK which enabled the shift of metabolic flux from TCA cycle and biomass to lysine synthesis. The lysine production rate of strain LC298 is 1.9g/L/h for the first 30h which is comparable to the classically derived industrial strain B-6 (20) although the yield is
relatively low (0.17g lysine/g glucose and 1.38g lysine/g biomass). Several factors including precursor and NADPH availability, byproduct formation would be further engineered to increase lysine production and yield.

Discussion

Deregulation of allosteric inhibitions of enzymes is a challenge in strain development (4). Random mutation and selection have been traditionally used for such a purpose and contributed greatly to the improvement of fermentation processes like amino acids production (21). However, this approach has serious disadvantages. For example, it is time-consuming and can introduce unnecessary mutations at the same time. More importantly, a well selectable phenotype such as resistance to analog of inhibitor is required for this process. In many cases, however, there are no reliable selectable phenotypes for the allosteric enzymes such as phosphoenolpyruvate carboxylase, an important anaplerotic enzyme inhibited by aspartate and malate. Thus, a more rational approach that can design targeted mutation is highly desired. We demonstrated in this study that SCA can be an effective way for solving the problem of allosteric regulation of industrial important enzymes.

SCA has been proved to be an efficient approach to clarify the co-evolutionary pattern of residues within protein family (26). SCA uses two levels of statistical analysis of protein sequences to identify important targets: position-specific conservation (first level) and position-correlated conservation (second level). The first level of statistics is traditionally used in the multiple sequence alignment to identify conserved residues at
certain position. The conservation of residues, however, doesn’t give direct information
to their function. SCA is thus more focusing on the second level of statistics to identify
the highly correlated residues in which the mutation of one residue would often result in
the corresponding change of the other during the evolution. These residues co-evolve
and often form interacted network in order to keep certain protein function or regulation.
As indicated in this study, a small cluster of residues within the regulatory domain show
a high correlation during evolution of aspartokinases. These residues form an
interconnected network linking different inhibitor binding sites and play important
roles in mediating the concerted inhibition of CgAK by lysine and threonine. Mutation
of any residues within this network affects the inhibition profiles of CgAK. The effects
of mutations in threonine and lysine binding sites are reasonably comprehensible as
they will affect the binding of lysine or threonine. These point mutations can be
evaluated if the crystal structure is available and the inhibitors binding sites can be
analyzed. However, mutations of residues not directly involved in the binding of
inhibitors, such as F283R and S301F, have also been found to be able to reduce the
concerted inhibition of CgAK. The crystal structure of CgAK mutant S301F showed
that this point mutation didn’t affect the inhibitors’ binding. These point mutations can
not be predicted and the phenomena can not be explained by a simple analysis of the
static crystal structures. Allosteric regulation is a dynamic process and these residues
may play important roles in signal transduction during this process (28). The prediction
of these residues illustrates the usefulness of SCA. It should be mentioned that the
crystal structure of CgAK was not available during the time we predicted and
experimental verified all designed mutations. The crystal structure was combined
during the preparation of this manuscript as it facilitated the explanation of the
phenomena. But it is not necessary for SCA prediction as SCA is a sequences-based
analysis. We also expect that the identified positions would be efficient for other
members of aspartokinase family.

After the identification of targeted position, the choice of amino acid substitution is also
important for the successful deregulation of enzyme inhibition. In case that there is no
crystal structure available, this could be made based on the analysis of amino acids
distribution in the specific position during the evolution (or in multiple sequences
alignment). The basic principle is to destroy the allosteric interaction network. One
simple strategy is to change the native amino acid to other amino acids which are rarely
or never used in this position. More specifically, changing the amino acid group to the
ones having different chemical properties would be expected to have more obvious
effect on the deregulating the allosteric inhibition. In the other hand, the selection of
amino acid substitution could also affect the specific activity of the enzyme. In our
cases, since all of the targeted residues are located in the regulatory domain of
aspartokinase and not directly involved in substrate binding, the mutations didn’t show
significant influence on the Km values for ATP and aspartate (Table 2). The reduction
of enzyme specific activities of some mutants could be attributed to the diminished
Vmax owing to the effect of protein folding, stability or other factors. The predicted
design of optimal substitution to largely reduce allosteric inhibition and minimize the
lost of enzyme activity would be the future direction.
The identification of these feedback insensitive CgAK mutants enabled us to create a new lysine producer based on only one point mutation in \textit{lysC} gene. This point mutation (Q298G), however, enabled \textit{C. glutamicum} LC298 to produce 58g/L lysine within 30h. The production rate is comparable to some of industrial lysine producers such as \textit{C. glutamicum} B-6 (20). A similar point mutation on \textit{lysC} gene, T311I, has been previously identified by comparative genomics (20). Introduction of this mutation into wildtype \textit{C. glutamicum} ATCC 13032 achieved 55 g/L lysine production in 30h.

Comparing with that strain, our strain showed a little higher lysine production and but better cell growth indicating that the point mutation of Q298G on \textit{lysC} has lower effect on cell growth. Thus, this strain can be used as a good start-strain for introduction of more beneficial mutations.

Construction of minimally mutated strain is a direction for future industrial strain development (4, 14, 25). An attractive idea based on the comparative genomics has been proposed by Ike et al. (14). This approach, however, has several limitations. For example, it requires the availability of different industrial strains and the beneficial mutations should appear in the selected industrial strain. What is more, it is difficult to differentiate the beneficial or unbeneificial mutations without understanding their potential roles. The strategy herein provides an alternative way for achieving such purpose without the need of industrial strains. Our approach is based on more rational tools combining sequences and structure analysis to improve the defined enzymes in metabolic pathways. We have used the same approach to modify several other important enzymes including homoserine dehydrogenase and phosphoenolpyruvate.
carboxylase which also greatly increase lysine production (Chen et al., data not published). It can be concluded that co-evolutionary analysis is an efficient approach to accelerate the process of industrial strain development.

References


FIGURE LEGENDS

FIGURE 1. Structure and properties of aspartokinase from *C. glutamicum* (CgAK) and results of co-evolutionary analysis of aspartokinase family proteins. (A) Overall structure of CgAK. The catalytic domain and regulatory domain of the α subunit are denoted with green and cyan colors, respectively. The β subunit is shown as magentas. Lysine and threonine are presented by the CPK model. CgAK has a α₂β₂-type heterotetrameric structure. Only one αβ dimer from PDB 3AAW is indicated here. (B) SDS-PAGE of the purified CgAK. M, marker; P, purified protein. (C) The inhibition profiles of wildtype CgAK by lysine and threonine. The specific activity under standard assay conditions (Table 2) was taken as 100%. (D) SCA matrix of the regulatory domain of aspartokinase family proteins. The position numbering is indicated according to the sequence of CgAK. The color scale linearly maps the correlated value from 0 to 2. (E) Mapping of the 25 highly correlated positions into the structure of CgAK.

FIGURE 2. Distribution of the highly correlated residues within inhibitors’ binding sites and results of point mutations. (A) Distribution of the highly correlated residues within threonine binding sites. Residues from the β subunit are shown by asterisks and numbered as their corresponding positions of α subunit. (B) Distribution of the highly correlated residues within lysine binding sites. (C) The inhibition profiles of CgAKs by point mutations of the highly correlated residues within inhibitors’ binding sites. The specific activities under standard assay conditions (Table 2) were taken as 100%.
FIGURE 3. Distribution of the highly correlated residues within other regions of the CgAK protein and results of point mutations. (A) Distribution of the highly correlated residues within other regions of protein. (B) The inhibition profiles of aspartokinase with point mutations of highly correlated residues within other regions of protein. The specific activities under standard assay conditions (Table 2) were taken as 100% (Table 2).

FIGURE 4. Profiles of fed-batch fermentation (L-lysine production, cell growth and glucose consumption) by C. glutamicum wildtype strain ATCC 13032 and mutant strain LC298. Data represent mean values and standard deviation from three independent cultures.
Table 1. Locations of the 25 highly correlated positions and point mutations.

<table>
<thead>
<tr>
<th>Point mutations</th>
<th>Locations of the mutated points</th>
<th>Correlated value*</th>
<th>Other Properties</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>I272E</td>
<td>Threonine binding sites</td>
<td>1.75</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>D274A</td>
<td>Threonine binding sites</td>
<td>1.51</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>G277A</td>
<td>Threonine binding sites</td>
<td>1.45</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>E278V</td>
<td>Threonine binding sites</td>
<td>1.49</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>A279V</td>
<td>Threonine binding sites</td>
<td>1.78</td>
<td>AEC- resistant</td>
<td>29</td>
</tr>
<tr>
<td>F283R</td>
<td>Other positions</td>
<td>1.47</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>I291**</td>
<td>Lysine binding sites</td>
<td>1.39</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>I293**</td>
<td>Lysine binding sites</td>
<td>1.88</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>D294**</td>
<td>Lysine/Threonine binding sites</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q298G</td>
<td>Threonine binding sites</td>
<td>1.88</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>N299L</td>
<td>Other positions</td>
<td>1.45</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>S301F</td>
<td>Other positions</td>
<td>1.55</td>
<td>AEC- resistant</td>
<td>29</td>
</tr>
<tr>
<td>T308A</td>
<td>Other positions</td>
<td>1.75</td>
<td>AEC- resistant</td>
<td>29</td>
</tr>
<tr>
<td>T311I</td>
<td>Other positions</td>
<td>1.86</td>
<td>AEC- resistant</td>
<td>20</td>
</tr>
<tr>
<td>T336L</td>
<td>Other positions</td>
<td>1.39</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>T361A</td>
<td>Lysine binding sites</td>
<td>1.59</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>F364A</td>
<td>Other positions</td>
<td>1.57</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>M365A</td>
<td>Other positions</td>
<td>1.67</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>N372A</td>
<td>Threonine binding sites</td>
<td>1.75</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>N374A</td>
<td>Threonine binding sites</td>
<td>1.55</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>I375P</td>
<td>Threonine binding sites</td>
<td>1.75</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>S381F</td>
<td>Lysine binding sites</td>
<td>1.76</td>
<td>AEC- resistant</td>
<td>23</td>
</tr>
<tr>
<td>E382A</td>
<td>Lysine binding sites</td>
<td>1.76</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>R384L</td>
<td>Other positions</td>
<td>1.74</td>
<td></td>
<td>This study</td>
</tr>
<tr>
<td>S386A</td>
<td>Other positions</td>
<td>1.88</td>
<td></td>
<td>This study</td>
</tr>
</tbody>
</table>

* The highest value was shown.

**The corresponding mutant has not been constructed or reported.
Table 2. Specific activities and kinetic parameters of aspartokinase mutants.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Specific activity* (mU/mg protein)</th>
<th>Km for ATP** (mM)</th>
<th>Km for Aspartate (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>wildtype</td>
<td>280±15</td>
<td>0.48±0.08</td>
<td>0.25±0.03</td>
</tr>
<tr>
<td>I272E</td>
<td>146±11</td>
<td>0.56±0.14</td>
<td>0.26±0.08</td>
</tr>
<tr>
<td>D274A</td>
<td>283±19</td>
<td>0.51±0.22</td>
<td>0.31±0.04</td>
</tr>
<tr>
<td>E278V</td>
<td>242±33</td>
<td>0.49±0.10</td>
<td>0.28±0.11</td>
</tr>
<tr>
<td>P283R</td>
<td>88±11</td>
<td>0.55±0.12</td>
<td>0.38±0.08</td>
</tr>
<tr>
<td>Q298G</td>
<td>294±15</td>
<td>0.46±0.17</td>
<td>0.25±0.06</td>
</tr>
<tr>
<td>N299L</td>
<td>118±16</td>
<td>0.51±0.04</td>
<td>0.28±0.08</td>
</tr>
<tr>
<td>T336L</td>
<td>205±12</td>
<td>0.50±0.09</td>
<td>0.27±0.05</td>
</tr>
<tr>
<td>M365A</td>
<td>131±13</td>
<td>0.49±0.25</td>
<td>0.31±0.07</td>
</tr>
<tr>
<td>N372A</td>
<td>174±12</td>
<td>0.48±0.08</td>
<td>0.27±0.12</td>
</tr>
<tr>
<td>N374A</td>
<td>199±23</td>
<td>0.52±0.12</td>
<td>0.38±0.11</td>
</tr>
<tr>
<td>I375P</td>
<td>125±15</td>
<td>0.60±0.18</td>
<td>0.32±0.13</td>
</tr>
<tr>
<td>E382A</td>
<td>257±18</td>
<td>0.50±0.13</td>
<td>0.24±0.05</td>
</tr>
<tr>
<td>R384L</td>
<td>101±14</td>
<td>0.53±0.11</td>
<td>0.29±0.11</td>
</tr>
<tr>
<td>S386A</td>
<td>97±8</td>
<td>0.63±0.08</td>
<td>0.31±0.15</td>
</tr>
</tbody>
</table>

* The standard assay condition containing 200 mM Tris-HCl (pH 7.5), 10 mM MgSO₄·6H₂O, 10 mM Aspartate, 10 mM ATP, 160mM NH₂OH·HCl.

** Kinetic analyses were carried out under the same conditions except for the substrate concentrations.