Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation

Jin Hwan Park*†, Kwang Ho Lee*†‡, Tae Yong Kim*†, and Sang Yup Lee*†§¶

*Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 program), BioProcess Engineering Research Center, †Center for Systems and Synthetic Biotechnology, Institute for the Biocentury, and §Department of BioSystems and Bioinformatics Research Center, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea; and ‡R & D Center for Bioproducts, CJ Corporation, Seoul 157-724, Korea

Communicated by Lonnie O. Ingram, University of Florida, Gainesville, FL, March 20, 2007 (received for review January 3, 2007)

The L-valine production strain of *Escherichia coli* was constructed by rational metabolic engineering and stepwise improvement based on transcriptome analysis and gene knockout simulation of the *in silico* genome-scale metabolic network. Feedback inhibition of acetohydroxy acid synthase isoenzyme III by L-valine was removed by site-directed mutagenesis, and the native promoter containing the transcriptional attenuator leader regions of the *ilvGMEDA* and *ilvBN* operon was replaced with the *tac* promoter. The *ilvA*, *leuA*, and *panB* genes were deleted to make more precursors available for L-valine biosynthesis. This engineered Val strain harboring a plasmid overexpressing the *ilvBN* genes produced 1.31 g/liter L-valine. Comparative transcriptome profiling was performed during batch fermentation of the engineered and control strains. Among the down-regulated genes, the *lrp* and *ygaZH* genes, which encode a global regulator Lrp and L-valine exporter, respectively, were overexpressed. Amplification of the *lrp*, *ygaZH*, and *lrp-ygaZH* genes led to the enhanced production of L-valine by 21.6%, 47.1%, and 113%, respectively. Further improvement was achieved by using *in silico* gene knockout simulation, which identified the *aceF*, *mdh*, and *pfkA* genes as knockout targets. The VAMF strain (Val Δ*aceF* Δ*mdh* Δ*pfkA*) overexpressing the *ilvBN*, *ilvCED*, *ygaZH*, and *lrp* genes was able to produce 7.55 g/liter L-valine from 20 g/liter glucose in batch culture, resulting in a high yield of 0.378 g of L-valine per gram of glucose. These results suggest that an industrially competitive strain can be efficiently developed by metabolic engineering based on combined rational modification, transcriptome profiling, and systems-level *in silico* analysis.

systems biology | global regulator | exporter | *in silico* prediction

Most amino acid-producing bacterial strains have been constructed by random mutagenesis. A significant disadvantage of this approach is the possibility that the random distribution of mutations in regions not directly related to amino acid biosynthesis can cause unwanted changes in physiology and growth retardation. Rational metabolic engineering by specific targeted modifications can overcome this disadvantage. The recent development of omics technology, combined with computational analysis, now provides a new avenue for strain improvement (1–4) by providing new information extracted from a large number of data, which is termed "systems biotechnology" (5).

L-valine, an essential hydrophobic and branched-chain amino acid, is used as a component of cosmetics and pharmaceuticals as well as animal feed additives. L-valine has been produced by employing bacteria belonging to the genera *Brevibacterium*, *Corynebacterium*, and *Serratia*, which have been improved by random mutation and selection (6, 7). A recent report describes the production of L-valine by rationally constructed *Corynebacterium glutamicum*, in which a feedback inhibition-resistant small subunit of acetohydroxy acid synthase (AHAS; encoded by *ilvN*) was generated by site-directed mutagenesis; the *panB* (encoding pantothenate synthase) and *ilvA* (encoding threonine dehydratase) genes were deleted; and the *ilvBNC* operon, encoding acetohydroxy acid synthase (*ilvBN*), and acetohydroxy acid isomeroreductase (*ilvC*) were overexpressed (8).

Escherichia coli has been widely used for amino acid production. For example, randomly mutated and/or metabolically engineered *E. coli* strains have been used for the production of L-threonine (2), L-phenylalanine (9), and L-alanine (10). However, there has been no report on the production of L-valine by *E. coli* constructed by rational metabolic engineering. This situation may be because *E. coli* possesses more complex regulatory mechanisms for L-valine biosynthesis than other microorganisms, which makes it difficult to redirect the flux toward enhanced L-valine production. *E. coli* has three AHAS isoenzymes, key enzymes in the biosynthesis of branched-chain amino acids. Three *E. coli* AHAS isoenzymes differ in their biochemical properties and regulatory mechanisms (11). The isoenzymes AHAS I, II, and III are encoded by the *ilvBN*, *ilvGM*, and *ilvIH* genes, respectively. Expression of the *ilvGMEDA* operon is controlled by transcriptional attenuation mediated by all three amino acids, L-leucine, L-valine, and L-isoleucine, whereas the expression of the *ilvBN* operon is controlled by attenuation mediated only by L-valine and L-leucine. The sites responsible for feedback inhibition are located in the small subunits of AHAS I and III encoded by *ilvN* (12) and *ilvH* (13, 14), respectively. The isoenzyme AHAS II encoded by *ilvGM* is insensitive to L-valine (11). It has been known that in the *E. coli* K-12 strain, AHAS II isoenzyme is not expressed because of frameshift mutation in *ilvG* (15).

The global regulator leucine responsive protein (Lrp) is involved in the regulation of L-valine biosynthesis; it activates the expression of the *ilvIH* genes encoding the AHAS III isoenzyme (16, 17) and represses the expression of the *ilvGMEDA* operon encoding the AHAS II isoenzyme (18). Lrp is also known to repress the expression of the *livJ* and *livK* genes, encoding the branched-chain amino acid transporter (19). These regulatory roles of Lrp correlated well with the previous results from the global gene expression profiling in *E. coli* K12, which revealed that in an *lrp*-deleted strain, the expression levels of *ilvG* and *ilvM* genes were up-regulated by 2.17- and 3.07-fold, respectively, and that of the *livJ* gene by 2.32-fold (20). However, its role in L-valine production in *E. coli* has not been elucidated.

Author contributions: J.H.P. and S.Y.L. designed research; J.H.P., K.H.L., and T.Y.K. performed research; K.H.L. and T.Y.K. contributed new reagents/analytic tools; J.H.P., T.Y.K., and S.Y.L. analyzed data; and J.H.P. and S.Y.L. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

¶To whom correspondence should be addressed. E-mail: leesy@kaist.ac.kr.

This article contains supporting information online at www.pnas.org/cgi/content/full/0702609104/DC1.

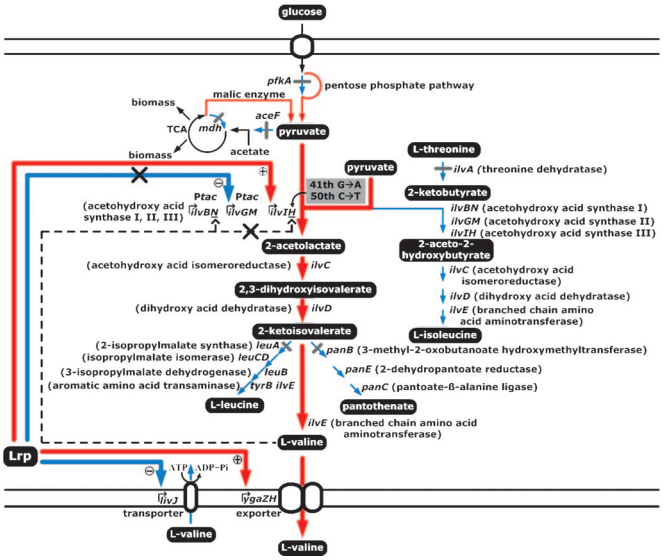

Fig. 1. The biosynthetic pathways of branched-chain amino acids in *E. coli*, regulations involved, and the strategies for constructing the L-valine-producing base strain. The shaded boxes represent mutations introduced into the genome, and the gray bars indicate the genes that were knocked out. Thin red arrows indicate the increased flux by knocking out the genes suggested by *in silico* simulation. Thick red arrows indicate the increased flux or activity by directly overexpressing the corresponding genes. Thick blue arrows indicate repression of gene expression by Lrp. Thin blue arrows indicate the decreased flux by knocking out the corresponding genes. Dotted lines indicate feedback inhibition. The black X indicates that the inhibition or repression is removed. The plus (+) and minus (−) symbols indicate activation and repression of gene expression, respectively.

In this study, rational metabolic engineering based on known regulatory and metabolic information and new knowledge generated by transcriptome analysis, combined with *in silico* simulation, was carried out to develop an *E. coli* strain capable of overproducing L-valine.

Results

Metabolic Engineering of *E. coli* for L-Valine Production. The *E. coli* W3110 strain was rationally engineered to produce L-valine by the following targeted genetic modifications. The biosynthetic pathways of branched-chain amino acids in *E. coli*, the regulations involved, and the strategies for constructing L-valine production strain are shown in Fig. 1. First, feedback inhibition and transcriptional attenuation were removed by site-directed mutagenesis. The sequences involved in these regulatory mechanisms were identified from previously published data or patents (13, 14, 21–23). Feedback inhibition in the small subunit of AHAS III encoded by the *ilvH* gene was removed by displacing the 41st base G with A and the 50th base C with T. Transcriptional attenuation sites were removed by displacing the attenuator leader region with the *tac* promoter by homologous recombination. The successfully mutated clone was identified by *sacB*-positive selection (24). Although the frameshift mutation in *ilvG* causes defects in the expression of *ilvG*, the transcriptional attenuator needs to be replaced with the *tac* promoter for constitutive expression of the downstream *ilvED* genes. The WL3110 strain, a *lacI*-deleted W3110 strain, was constructed to allow constitutive expression of genes under the *tac* promoter without adding an inducer.

Second, L-valine production was enhanced by increasing the carbon flux toward L-valine formation. The availability of pyruvate and 2-ketoisovalerate is important because it can influence the flux toward L-valine formation. Pyruvate is used as a common substrate for biosynthesis of L-valine, L-leucine, and L-isoleucine. To eliminate the formation of 2-ketobutyrate, the intermediate precursor of L-isoleucine biosynthesis competing with L-valine biosynthesis, the *ilvA* gene (encoding L-threonine dehydratase), was knocked out. The *panB* (encoding 3-methyl-2-oxobutanoate hydroxymethyltransferase) and *leuA* (encoding 2-isopropylmalate synthase) genes were also knocked out to increase 2-ketoisovalerate availability toward L-valine synthesis. This base strain was named the Val strain, of which genotype was described in supporting information (SI) Table 2.

After removing the feedback inhibition and attenuation followed by chromosomal inactivation of competing pathways, we amplified the AHAS I isoenzymes, which catalyze the first reaction of L-valine biosynthetic pathway. It is known that the AHAS I, encoded by the *ilvBN* genes, has a much higher affinity for pyruvate than for 2-ketobutyrate, compared with the isoenzymes AHAS II and AHAS III (25). Hence, the *ilvBN* operon was selected for amplification. The operon was cloned into pKK223–3 to make pKKilvBN. The Val strain was transformed with pKKilvBN and cultured in a bioreactor. The time profiles of cell growth and L-valine production during the batch fermentation of Val (pKKilvBN) are shown in SI Fig. 4. The growth rate of the recombinant Val strain was slightly lower than that of a control strain, WL3110 harboring pKK223–3 (SI Fig. 4). The Val (pKKilvBN) strain successfully produced 1.31 g/liter of L-valine. After this initial success in metabolic engineering of *E. coli* for L-valine production, further stepwise improvements were attempted by using this Val strain.

Transcriptome Analysis of the L-Valine Production Strain. To understand the altered global gene expression levels in the recombinant Val strain producing L-valine, transcriptome profiling was performed. Transcriptome profiles of Val (pKKilvBN) and WL3110 (pKK223–3) were compared by using the samples taken at the OD_{600} of 6.55 (the former) and 6.77 (the latter). They were both at the exponential growth phase, and the specific growth rates were 0.14 h^{-1} for the former and 0.19 h^{-1} for the latter. At the time of sampling, the Val (pKKilvBN) strain produced 0.5 g/liter of L-valine. Genes that were found to be significantly up- or down-regulated are listed in SI Tables 3–5. The expression ratios of genes involved in central metabolic pathways and L-valine biosynthesis are shown in Fig. 2. The expression levels of genes involved in glycolysis and pentose phosphate pathway were not significantly changed, whereas those of most of the genes in the tricarboxylic cycle decreased. The increased expression of AHAS isoenzymes, which are directing pyruvate toward L-valine biosynthesis, may cause depletion of acetyl-CoA and thereby reduce the expression of tricarboxylic cycle genes.

As expected, the *ilvBN* genes, which were amplified both chromosomally (attenuation removal) and extrachromosomally (plasmid-born amplification), were significantly up-regulated by 54.46- (*ilvB*) and 32.50-fold (*ilvN*), respectively. The *ilvGM* genes, in which the native transcriptional attenuator leader region was replaced by the *tac* promoter, were also up-regulated by 7.41- (*ilvG*) and 2.21-fold (*ilvM*), respectively. Similarly, the *ilvCDE* genes that are involved in L-valine biosynthetic pathway were up-regulated by 3.74-, 4.28-, and 1.32-fold, respectively. However, the *leuABCD* genes involved in L-leucine biosynthesis were significantly down-regulated to 0.17-, 0.14-, 0.31-, and 0.17-fold, respectively. These results revealed that the targeted genetic modifications were successfully reflected at the transcript level. Other notable changes found from transcriptome profiling are described in *SI Text*.

Enhanced Production of L-Valine by Coamplification of the *ilvCED* Genes. Transcriptome profiling revealed that the expression levels of the *ilvCED* genes encoding acetohydroxy acid isomeroreductase (*ilvC*), branched-chain amino acid aminotransferase (*ilvE*), and dihydroxy acid dehydratase (*ilvD*) involved in the L-valine biosynthetic pathway (Fig. 1) were increased in the recombinant Val strain (Fig. 2). However, the extents of the increase were much less than those of the *ilvBN* genes (Fig. 2). This result is easily understandable

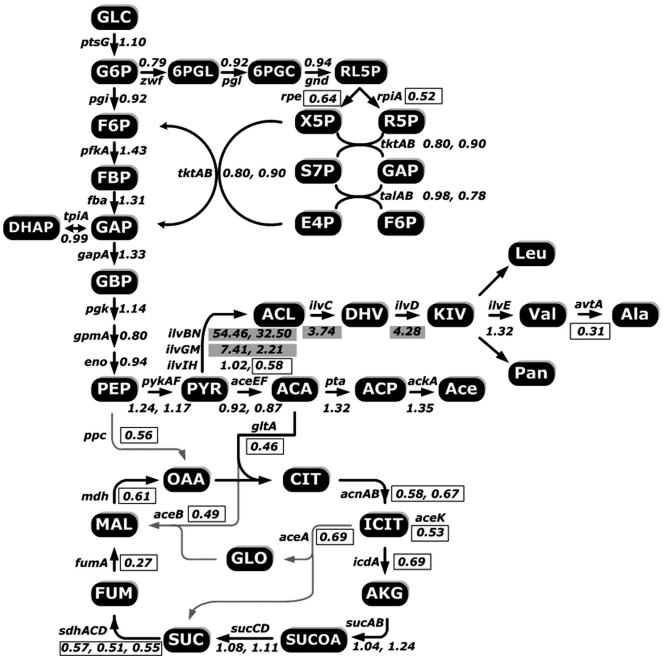

Fig. 2. Results of comparative transcriptome profiling. The numbers are the ratios of the expression levels in Val (pKKilvBN) vs. WL3110 (pKK223–3). The shaded and boxed numbers indicate significantly up- and down-regulated genes, respectively, in L-valine-producing strain Val (pKKilvBN). GLC, glucose; G6P, glucose-6-phosphate; 6PGL, gluconolactone-6-phosphate; 6PGC, 6-phospho gluconate; RL5P, ribulose-5-phosphate; X5P, xylurose-5-phosphate; R5P, ribose-5-phosphate; S7P, sedoheptulose-7-phosphate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; FBP, fructose-1, 6-bisphosphate; GAP, glyceraldehyde-3-phosphate; DHAP, dihydroxyacetonephosphate; GBP, 1,3-bisphosphoglycerate; PEP, phosphoenolpyruvate; PYR, pyruvate; ACA, acetyl-CoA; ACP, acetyl-phosphate; Ace, acetate; ACL, 2-acetolactate; DHV, 2,3-dihydroxyisovalerate; KIV, 2-ketoisovalerate; Val, L-valine; Leu, L-leucine; Pan, pantothenate; Ala, L-alanine; CIT, citrate; ICIT, isocitrate; AKG, α-ketoglutarate; SUCOA, succinyl-CoA; SUC, succinate; FUM, fumarate; MAL, malate; OAA, oxaloacetate.

because the *ilvBN* genes were amplified by plasmid-based overexpression, whereas the *ilvCED* genes were not; only the attenuation control was removed for the *ilvED* genes. Thus, we decided to coamplify the *ilvCED* genes to possibly increase L-valine production. The *ilvC* and *livED* genes were cloned sequentially following the *ilvBN* operon in pKKilvBN to make pKKilvBNCED. The Val strain was transformed with pKKilvBNCED and cultured in NM2 medium for 48 h. The final OD_{600} and L-valine concentration at the end of flask cultivation were 23.8 and 3.43 g/liter, respectively. Thus, the coamplification of the *ilvCED* genes was found to be beneficial for increasing L-valine production.

Influence of Lrp on L-Valine Production. According to the transcriptome data, the expression of the *lrp* gene was down-regulated to 0.52-fold (SI Table 4). This value was also validated by real-time RT-PCR analysis, which showed its down-regulation to 0.45-fold. It has been known that Lrp activates the expression of the *ilvIH* operon, which encodes the AHAS III isoenzyme (16). Thus, down-regulation of *lrp* expression might have caused insufficient expression of the *ilvIH* operon. Indeed, the expression level of the *ilvH* gene decreased to 0.58-fold, whereas that of the *ilvI* gene stayed at the similar level. Furthermore, it has been known that exogenous L-leucine antagonizes the positive effect of Lrp on the expression of the *ilvIH* operon and certain Lrp-regulated genes (26, 27). Because L-leucine is added during the cultivation to complement the auxotrophy, this also might have happened in our case. In another study, it has been reported that when the intracellular concentration of Lrp is sufficiently high, all positively regulated genes become insensitive to L-leucine (28). Taken together, it was assumed that the L-leucine requirement in our system may cause insufficient activation of *ilvIH* expression by Lrp. For this reason, the *lrp* gene was selected for amplification to overcome the negative effect of L-leucine on *ilvIH* expression. However, we thought that overexpression of too many genes at the same time by using a high-copy-number plasmid may be detrimental to cell growth. Thus, the origin of replication of pKKilvBNCED was replaced with that of a medium-copy-number plasmid pBR322 to make pKBRilvBNCED (SI Fig. 5*A*).

Table 1. Effects of overexpressing or deleting the *lrp* and/or *ygaZH* genes on L-valine production

Strain	OD_{600}	L-valine (g/liter)*
Val (pKBRilvBNCED)	25.65 ± 0.35	3.73 ± 0.16
Val (pKBRilvBNCED, pTrc184)	23.97 ± 0.15	3.57 ± 0.06
Val (pKBRilvBNCED, pTrc184lrp)	23.42 ± 0.67	4.34 ± 0.03
Vlrp (pKBRilvBNCED)	22.05 ± 0.39	2.38 ± 0.03
Val (pKBRilvBNCED, pTrc184ygaZH)	24.59 ± 1.58	5.25 ± 0.14
VygaZH (pKBRilvBNCED)	24.15 ± 0.62	3.73 ± 0.09
Val (pKBRilvBNCED, pTrc184ygaZHlrp)	24.85 ± 0.07	7.61 ± 0.24

*The L-valine concentration was determined at 48 h of flask cultivation in NM2 medium containing 50 g/liter glucose.

To examine the ability to produce L-valine, the Val strain harboring pKBRilvBNCED was cultured in NM2 medium for 48 h. The final OD_{600} and L-valine concentrations were 25.65 and 3.73 g/liter, respectively. These values are slightly better than those obtained with the Val strain harboring pKKilvBNCED (see above). For the amplification of the *lrp* gene, plasmid pTrc184lrp, which is compatible with pKBRilvBNCED, was constructed (SI Fig. 4*B*). When the Val strain harboring pKBRilvBNCED and pTrc184lrp was cultured under the same condition, the final L-valine concentration obtained was 4.34 g/liter (Table 1), which is 21.6% higher than that obtained without *lrp* amplification. To see whether this increase is truly due to the availability of Lrp, the Vlrp strain, which is the *lrp*-deleted Val strain, was transformed with pKBRilvBNCED and cultured. The final L-valine concentration obtained was 2.38 g/liter, which is 36.2% lower than the control strain (Table 1). These results suggest that amplification of the *lrp* gene is beneficial for L-valine production in *E. coli*.

Enhanced Production of L-Valine by Overexpressing the *ygaZH* Genes. In *C. glutamicum*, which is the most widely used bacterium for amino acid production, the *brnFE* genes encode the exporter of branched-chain amino acids (29). The *E. coli* genes homologous to *brnFE* were suggested to be the *ygaZH* genes by homology search (29). The amino acid sequence of YgaZ showed 27% identity to that of BrnF (SI Fig. 6). Thus, we examined whether the *ygaZH* genes, which form an operon, encode an L-valine exporter. The *ygaZH* genes were knocked out in W3110 to make the WygaZH strain (SI Table 2). Then the minimal inhibitory concentration (MIC) test was performed for the W3110 and WygaZH strains by using an L-valine analogue, DL-norvaline, which is not metabolized by the cell (SI Table 6). The *ygaZH*-deficient WygaZH strain did not grow in the presence of 9 mM DL-norvaline, whereas the wild-type W3110 strain survived in the presence of 15 mM DL-norvaline. Furthermore, the W3110 strain harboring pTrc184ygaZH, which overproduces YgaZH, showed normal growth even in the presence of 50 mM DL-norvaline (SI Table 6). This result is the first experimental proof that YgaZH functions as an L-valine exporter.

According to the transcriptome data, the expression levels of the *ygaZ* and *ygaH* genes decreased to 0.75- and 0.61-fold, respectively, in the recombinant Val strain (SI Table 4). We thus reasoned that the low-level expression of the *ygaZH* genes might have been limiting the L-valine production. Hence, the *ygaZH* genes were overexpressed by using pTrc184ygaZH plasmid in the Val (pK-

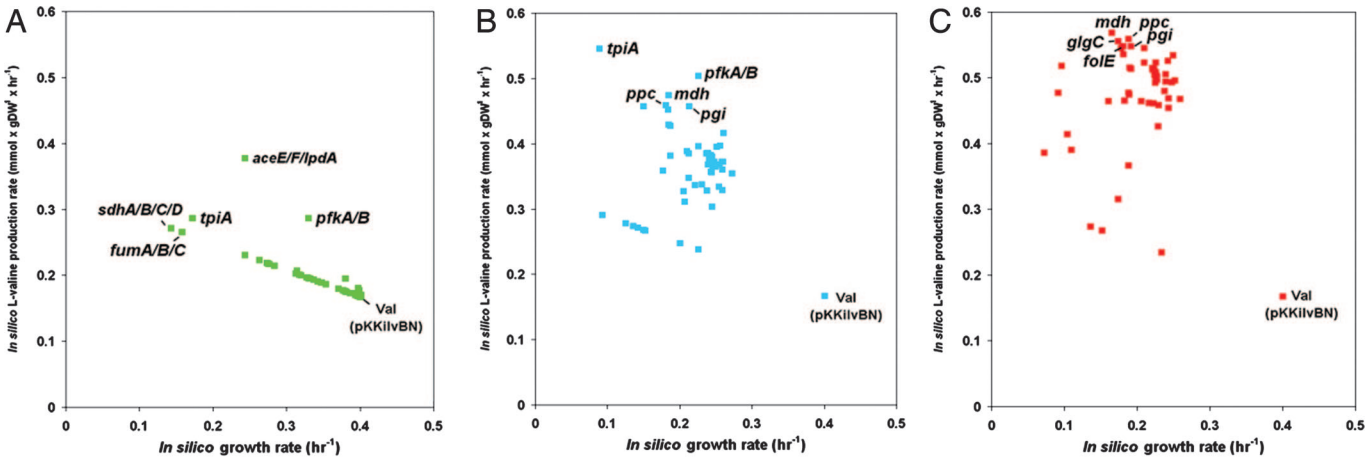

Fig. 3. Results of *in silico* gene knockout simulations by using the genome-scale metabolic model of *E. coli* MBEL979. The results of single (*A*), double (*B*), and triple (*C*) gene knockout simulations with respect to L-valine production and growth rates are shown. Only the five best candidates with respect to the L-valine production rate are shown for each stage of knockout simulation. Slashes indicate isoenzymes or subunits of the enzyme complex. The L-valine production and growth rates of the control Val strain harboring pKKilvBN are also indicated for comparison.

BRilvBNCED) strain. This recombinant Val strain produced 5.25 g/liter L-valine, which is 47.1% higher than that obtained without *ygaZH* amplification (Table 1). This result further suggests that the *ygaZH* genes encode an exporter of L-valine in *E. coli*. We then examined L-valine production in the VygaZH strain harboring pKBRilvBNCED to see whether L-valine production is hampered. Surprisingly, L-valine production was not affected in the VygaZH strain. This result suggests that other L-valine exporters exist and do not (or less efficiently) export DL-norvaline. Taken together, we can conclude that the YgaZH is one of the L-valine exporters in *E. coli*, which, on amplification, is beneficial for the enhanced production of L-valine.

Synergistic Effect of Lrp and YgaZH on L-Valine Production. Because the overexpression of the *lrp* and *ygaZH* genes resulted in the increased production of L-valine by 21.6% and 47.1%, respectively, we tried to overexpress both genes at the same time. The recombinant Val strain harboring pKBRilvBNCED and pTrc184ygaZHlrp produced 7.61 g/liter L-valine, which is 113% higher than that produced by the control strain. Thus, the effect of Lrp and YgaZH is more than additive, and the positive effect of Lrp seems to be further enhanced when accumulated L-valine is more rapidly exported by the overproduced YgaZH.

To understand the reason for this positive effect of coexpressing the *lrp* and *ygaZH* genes, the regulation of *ygaZH* expression by Lrp was investigated by real-time RT-PCR analysis by using the W3110 and Wlrp strains in response to 5 μM L-valine. In W3110, the expression of the *lrp*, *ygaZ*, and *ygaH* genes increased by 6.96-, 4.0-, and 3.2-fold, respectively, in response to 5 μM L-valine. However, the expression of the *ygaZ* and *ygaH* genes in the Wlrp strain decreased to 0.7- and 0.5-fold, respectively. These results suggest that the induction of the *ygaZH* expression by L-valine is likely to be mediated by Lrp, which functions as an activator of *ygaZH* expression.

Metabolic Engineering Based on *in Silico* Gene Knockout Studies. To further improve the Val strain, *in silico* gene knockout simulation studies were performed (SI Table 7). Fig. 3 presents the results of the *in silico* gene knockout simulation with respect to L-valine production and growth rates. It was straightforward to select the first knockout target as the *aceF* (or *aceE* or *lpdA*) gene (encoding pyruvate dehydrogenase); the L-valine production rate was increased by 2-fold to 0.377 mmol·gDW^{-1}·h^{-1} in this pyruvate dehydrogenase knockout mutant strain (Fig. 3*A*) by increasing the pyruvate pool. Consequently, we selected the *aceF* gene as the first knockout candidate because it affects the pyruvate dehydrogenase activity most significantly (10). After deleting the *aceF* gene, the second knockout candidate was identified by simulating the additional gene knockout effects on metabolic flux distribution by using the Val (Δ*aceF*) strain. Knocking out the *tpiA* gene was predicted to increase the L-valine production rate most significantly, but the growth rate was significantly reduced as well (Fig. 3*B*). Thus, the next best target, the *pfkA*/*B* gene, was selected as the second gene knockout candidate because it allows a considerable increase in the L-valine production rate with less reduction in the growth rate. The *pfkA*/*B* knockout seems to increase the availability of NADPH, an important cofactor for L-valine production, by pushing more carbon flux toward the pentose phosphate pathway. For the actual inactivation of phosphofructokinase, the *pfkA* gene was deleted because it encodes 6-phosphofructokinase-1, which possesses 90% of the activity (30). Using the Val (Δ*aceF* Δ*pfkA*) strain, the third gene knockout simulation was performed (Fig. 3*C*). There were several gene knockout candidates that were able to increase the L-valine production rate further. Among these, knocking out the *mdh* gene resulted in the highest increase of the L-valine production rate. Thus, the *mdh* gene was chosen as the third knockout target to inactivate malate dehydrogenase (31). By doing so, the pyruvate pool is expected to be indirectly increased by the enhanced flux of malate to pyruvate by a malic enzyme. Further knockout simulation was not attempted because the triple knockout mutant strain already showed significantly lower growth rate than the wild-type strain. These three predicted genes were knocked out by chromosomal gene inactivation by using the λ *red* operon (*SI Text*). Consequently, the VAMF strain, which is Val (Δ*aceF* Δ*mdh* Δ*pfkA*), was constructed. For the better growth of the VAMF strain lacking in pyruvate dehydrogenase, sodium acetate (3 g/liter) was supplemented.

To examine the performance of the VAMF strain, batch cultures of the following four recombinant strains were carried out in NM1 medium containing 20 g/liter glucose: Val (pKBRilvBNCED), VAMF (pKBRilvBNCED), Val (pKBRilvBNCED, pTrc184ygaZHlrp), and VAMF (pKBRilvBNCED, pTrc184ygaZHlrp). The final L-valine concentrations obtained with these four strains were 2.75, 4.00, 3.33, and 7.55 g/liter, respectively. The triple knockout mutation allowed a 45.5% increase in L-valine production over an engineered Val strain in the presence of plasmid pKBRilvBNCED. Furthermore, the L-valine concentration obtained with the VAMF (pKBRilvBNCED, pTrc184ygaZHlrp) was 2.27-fold higher than that

obtained with the corresponding recombinant Val strain. The L-valine yield achieved with VAMF (pKBRilvBNCED, pTrc184ygaZHlrp) was as high as 0.378 g of L-valine per gram of glucose. These results suggest that metabolic engineering based on the results of an *in silico* genome-scale prediction of knockout targets is an efficient approach toward strain improvement.

Discussion

Traditionally, amino acid-producing strains have been developed by random mutagenesis and selection. Although various large-scale analytical techniques such as transcriptome and proteome analysis are now available, it is difficult to apply these techniques on the randomly mutated industrial strains for further strain improvement because of unknown mutations in their genome. Thus, the aim of this study was to construct a genetically well defined engineered *E. coli* strain based on known metabolic and regulatory information, transcriptome analysis, and *in silico* genome-scale knockout simulation. Although *E. coli* has more complex regulatory mechanisms for L-valine biosynthesis than any other microorganisms, many known regulatory mechanisms could be engineered toward enhanced L-valine production. After the systematic metabolic engineering of *E. coli*, we were able to achieve a high yield of 0.378 g of L-valine per gram of glucose, which is even higher than that of industrial *C. glutamicum* or *E. coli* strains developed by random mutation and selection. More importantly, the engineered strain developed in this study can be further improved because all of the modifications are clearly defined, which is not easy for those randomly mutagenized strains.

We first constructed the L-valine-producing *E. coli* base strain by rational metabolic engineering by using the metabolic and regulatory information available in the literature. Then this base strain was improved stepwise by metabolic engineering based on new information obtained from transcriptome analysis and *in silico* gene knockout simulation. First, based on the published metabolic and regulatory information, we removed feedback inhibition and attenuation by chromosomal replacement and knocked out the genes responsible for the major competing pathways. It should be mentioned that removal of another feedback inhibition may be necessary to further increase L-valine production; the feedback inhibition in the small subunit of AHAS I encoded by *ilvN* was not removed in this study because the site responsible for inhibition has not been reported yet. This Val strain was used as a base strain for L-valine production. Then several genes encoding those enzymes directly involved in L-valine biosynthesis were amplified. After the manipulation of all these known and rationally selected genes, no further engineering targets could be easily identified. Thus, we performed comparative transcriptome profiling on the control and L-valine-producing strains to examine the differences in global transcript levels between the two strains.

Among the up-regulated genes was the *ilvC* gene encoding acetohydroxy acid isomeroreductase. It has been known that the expression of the *ilvC* gene is induced by acetolactate. This induction is mediated by a positive transcriptional activator encoded by the *ilvY* gene (32). In the production strain, acetolactate was accumulated by amplified AHAS I encoded by the *ilvBN* genes. Thus, the accumulated acetolactate seems to result in IlvY-mediated induction of the *ilvC* gene expression. Among the 35 genes that showed significant down-regulation, we could easily remove the *leuABD* genes from the possible gene amplification targets; their down-regulation is beneficial for increased L-valine production as we originally designed. We went through the list of 32 other down-regulated genes to select those that may positively affect L-valine production. It should be admitted that there is no general method to select the amplification target genes among the down-regulated genes. First, the *lrp* gene, which encodes a global regulator Lrp, was selected as a target; this was because Lrp positively regulates the expression of genes involved in L-valine biosynthesis, such as *ilvIH*. Because the Val strain is auxotrophic to L-leucine, we need to add L-leucine exogenously. However, L-leucine binds to Lrp and prevents it from activating L-valine biosynthesis genes. Thus, it was thought that a sufficient level of Lrp is required to activate the expression of L-valine biosynthesis genes in the presence of exogenous L-leucine. It should be noted that *E. coli* has three acetohydroxyacid synthase isoenzymes (AHAS I, II, III), which are encoded by the *ilvBN*, *ilvGM*, and *ilvIH* genes, respectively. Among them, AHAS II is not expressed in *E. coli* K-12 due to the frameshift mutation in the *ilvG* gene. Because the plasmid-born amplification of many genes can be a burden to cell growth, we decided to overexpress only one of them. Thus, we overexpressed the *ilvBN* genes encoding AHAS I, which has a higher affinity for pyruvate, the most important precursor for L-valine formation. Additionally, by using the Lrp, which up-regulates the expression of the *ilvIH* gene (encoding AHAS III), further amplification of AHAS was possible without any detrimental effect of plasmid-born overexpression. Furthermore, because the expression of the *ilvGM* operon is repressed by Lrp (18), the native promoter of the *ilvGM* operon, to which Lrp binds, was replaced with the *tac* promoter to abolish the repression. As a result, we used only the positive effect of Lrp on L-valine production by the overexpression of the *lrp* gene. As expected, the overexpression of the *lrp* gene enhanced L-valine production by 21.6% (Table 1). Opposite results were obtained in the *lrp* knockout mutant, which further proves that Lrp is necessary for efficient L-valine production in *E. coli*.

Another amplification target gene, the *ygaZH* genes encoding a protein showing significant homology to the known *Corynebacterium* L-valine exporter, could be selected from the results of transcriptome profiling. Recently, YgaZH was suggested to be an L-valine exporter in the patent (33). We independently identified this gene by BLAST search and proved by knocking-out and amplification studies that it is truly involved in the export of L-valine. Interestingly, the level of L-valine production did not change in the *ygaZH*-deleted L-valine production strain (Table 1). This result suggests that YgaZH may not be the major L-valine exporter in *E. coli*. Nonetheless, the overexpression of the *ygaZH* genes allowed enhanced production of L-valine. Also, we were able to show that the expression of the *ygaZH* genes is activated by Lrp. The cooverexpression of the *lrp* and *ygaZH* genes was synergistic and led to a 2.13-fold increase in L-valine production up to a level of 7.61 g/liter from 50 g/liter glucose in flask culture. Recently, the L-leucine exporter in *E. coli* encoded by the *yeaS* gene has been published (34). It was shown that the MICs of L-valine in *E. coli* TG1, the *yeaS*-amplified TG1, and the *yeaS*-deleted TG1 strains were 0.002, 0.005, and 0.001 mg/ml, respectively. This result suggests that the *yeaS* gene product may also function as an L-valine exporter in *E. coli*. However, the small difference in MIC also suggests that the *yeaS* gene product may not be a major L-valine exporter in *E. coli* either. Identifying the major L-valine exporter in *E. coli* will be a good strategy for further improvement of L-valine production.

Further improvement of the L-valine-producing strain was achieved by using the results obtained from *in silico* gene knockout simulation. There has been no report on the successful application of *in silico* prediction strategy to improving amino acid-producing strains. By sequential *in silico* gene knockout studies, we were able to identify a triple mutant strain that allowed a higher L-valine production rate with a reasonable growth rate. In the VAMF strain, which is a base Val strain with additional knockouts in the *aceF*, *mdh*, and *pfkA* genes, the flux through pyruvate was increased by directly blocking the pyruvate and malate dehydrogenases. Deletion of the *pfkA* gene increases the flux through the pentose phosphate pathway and thus produces more NADPH, which is required for L-valine biosynthesis. The VAMF strain harboring pKBRilvBNCED and pTrc184ygaZHlrp showed a 126.7% increase in L-valine production, compared with the Val strain harboring the same plasmids. These results suggest that the beneficial effects of flux redistribution on L-valine production achieved in VAMF strain

could be further strengthened by implementing regulatory and export engineering. It should be noted that the *in silico* simulation currently used does not account for any regulatory mechanisms and exporting effects. Hence, it is important to combine the results of transcriptome profiling and *in silico* metabolic flux analysis, which are complementary to each other.

In summary, rational metabolic engineering based on known metabolic and regulatory information, and new information derived from transcriptome analysis and *in silico* simulation, allowed for the development of an *E. coli* strain capable of efficiently producing L-valine. An impressively high yield of 0.378 g of L-valine per gram of glucose could be achieved by using this metabolically engineered strain in batch culture. The approaches described in this article are a good example of systematically engineering strains for the enhanced production of amino acids, and they can also be applied in general for developing strains for the efficient production of other metabolites.

Materials and Methods

Bacterial Strains, Plasmids, and Chromosome Manipulation. The strains and plasmids constructed and used in this study are listed in SI Table 2. The primers used for gene cloning and deletion are listed in SI Table 8. Detailed procedures for the construction of plasmids are shown in *SI Text* and SI Figs. 5 and 7. All DNA manipulations were carried out by following standard protocols (35). Detailed procedures for site-directed mutagenesis and gene inactivation are described in *SI Text*. All the genes described in this paper are listed in SI Table 9.

Cultivation and Transcriptome Profiling. Batch fermentation of wild-type and engineered *E. coli* strains was performed in semidefined NM1 medium at 31°C (pH 6.0). Flask cultures were carried out in 250-ml baffled flasks containing NM2 medium at 31°C and 250 rpm with shaking incubator (SI-900; Jeio Tech, Daejeon, Korea). Culture media, culture conditions, and the procedure for determining a minimal inhibitory concentration of DL-norvaline are described in detail in *SI Text*. Transcriptome profiling was performed by using the TwinChip *E. coli*-6K oligo chip (Digital Genomics, Seoul, Korea) following the manufacturer's protocol. Detailed procedures can be found in *SI Text*. The mean signal intensity values of the duplicate spots were averaged and then normalized by the global normalization method. The genes showing P values <0.05 were considered to be significant.

***In Silico* Analysis.** The *in silico* gene-perturbation studies were performed by using the genome-scale metabolic model *E. coli* MBEL979 (SI Table 7 and ref. 36) consisting of 979 metabolic reactions and 814 metabolites (144 extracellular metabolites and 670 intermediates), which is a slightly modified network of *i*JR904 reported by Palsson *et al*. (37). Constraints-based flux analysis was performed on the rationally engineered Val strain, in which the *ilvA*, *leuA*, and *panB* genes were knocked out. To implement the gene deletion during the *in silico* metabolic flux analysis, we provided a constraint that sets the flux through a corresponding reaction to zero. Further gene knockout targets were identified by employing the minimization of metabolic adjustment (MOMA) algorithm by using the linear and quadratic programming methods by using a GAMS script (38). The base steady-state solution required for the quadratic programming for MOMA was calculated for each simulation by first solving the linear programming problem of maximizing the growth rate. For constraints-based flux analysis of the Val strain, the glucose-uptake and L-valine production rates of 4.37 and 0.17 mmol·gDW^{-1}·h^{-1}, which are the experimentally measured values at the exponential growth phase (≈10–11 h), were used as constraints. The uptake rates of L-isoleucine, L-leucine, and D-pantothenate, which were provided in the medium, were set to 5 mmol·gDW^{-1}·h^{-1}, which were large enough to ensure cell growth. The simulation of metabolic flux distribution in multiple gene knockout mutants was performed by a sequential approach (39). A gene was first identified whose deletion resulted in the maximum improvement in L-valine production, and double and triple mutants were subsequently sought by searching for the genes that upon knockout resulted in further improvement in L-valine production without too significant a reduction in the growth rate. To investigate the effects of sodium acetate addition on metabolic flux distribution, the acetic acid-uptake rate of 5 mmol·gDW^{-1}·h^{-1} was arbitrarily set, which is larger than the actual acetic acid-uptake rate.

Analytical Procedure. Cell growth was monitored by measuring the absorbance at 600 nm (OD_{600}) by using an Ultrospec3000 spectrophotometer (Amersham Pharmacia Biotech, Uppsala, Sweden). Real-time RT-PCR was performed as described in *SI Text*, and the results were analyzed by using the $2^{-\Delta\Delta Ct}$ method (40). Glucose concentration was measured by using a glucose analyzer (model 2700 STAT; Yellow Springs Instrument, Yellow Springs, OH). The concentrations of amino acids were analyzed by an amino acid analyzer (Sykam S433; Sykam, Eresing, Germany).

We thank B. L. Wanner (Purdue University, West Lafayette, IN) for providing the plasmid pKD46. This work was supported by the Korean Systems Biology Project of the Ministry of Science and Technology (M10309020000-03B5002-00000) and the LG Chem Chair Professorship.

1. Kromer JO, Sorgenfrei O, Klopprogge K, Heinzle E, Wittmann C (2004) *J Bacteriol* 186:1769–1784.
2. Lee JH, Lee DE, Lee BU, Kim HS (2003) *J Bacteriol* 185:5442–5451.
3. Lange C, Rittmann D, Wendisch VF, Bott M, Sahm H (2003) *Appl Environ Microbiol* 69:2521–2532.
4. Huser AT, Chassagnole C, Lindley ND, Merkamm M, Guyonvarch A, Elisakova V, Patek M, Kalinowski J, Brune I, Puhler A, Tauch A (2005) *Appl Environ Microbiol* 71:3255–3268.
5. Lee SY, Lee DY, Kim TY (2005) *Trends Biotechnol* 23:349–358.
6. Nakayama K, Kitada S, Kinoshita S (1961) *J Gen Appl Microbiol Tokyo* 7:52–69.
7. Plachy J (1975) *Folia Microbiol* 20:346–350.
8. Elisakova V, Patek M, Holatko J, Nesvera J, Leyval D, Goergen J, Delaunay S (2005) *Appl Environ Microbiol* 71:207–213.
9. Gerigk M, Bujnicki R, Ganpo-Nkwenkwa E, Bongaerts J, Sprenger G, Takors R (2002) *Biotechnol Bioeng* 80:746–754.
10. Lee M, Smith GM, Eiteman MA, Altman E (2004) *Appl Microbiol Biotechnol* 65:56–60.
11. Umbarger HE (1996) in Escherichia Coli *and* Salmonella: *Cellular and Molecular Biology*, eds Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE (ASM Press, Washington, DC), 2nd Ed, pp 442–457.
12. Eoyang L, Silverman PM (1986) *J Bacteriol* 185:5442–5451.
13. Livshits VA (2004) US Patent 6,737,255; *Chem Abstr* 135:367699.
14. Vyazmensky M, Sella C, Barak Z, Chipman DM (1996) *Biochemistry* 35:10339–10346.
15. Lawther RP, Calhoun DH, Adams CW, Hauser CA, Gray J, Hatfield GW (1981) *Proc Natl Acad Sci USA* 78:922–925.
16. Platko JV, Willins DA, Calvo JM (1990) *J Bacteriol* 172:4563–4570.
17. Wang Q, Calvo JM (1993) *J Mol Biol* 229:306–318.
18. Rhee KY, Parekh BS, Hatfield GW (1996) *J Biol Chem* 271:26499–26507.
19. Haney SA, Plakto JV, Oxender DL, Calvo JM (1992) *J Bacteriol* 174:108–115.
20. Hung SP, Baldi P, Hatfield GW (2002) *J Biol Chem* 277:40309–40323.
21. Lawther RP, Hatfield GW (1980) *Proc Natl Acad Sci USA* 77:1862–1866.
22. Lawther RP, Wek RC, Lopes JM, Pereira R, Taillon BE, Hatfield GW (1987) *Nucleic Acids Res* 15:2137–2155.
23. Wek RC, Hauser CA, Hatfield GW (1985) *Nucleic Acids Res* 13:3995–4010.
24. Schweizer HP (1992) *Mol Microbiol* 6:1195–1204.
25. Barak Z, Chipman DM, Gollop N (1987) *J Bacteriol* 169:3750–3756.
26. Calvo JM, Matthews RG (1994) *Microbiol Rev* 58:466–490.
27. Marasco R, Varcamonti M, Cara L, Ricca E, DeFelice M, Sacco M (1994) *J Bacteriol* 176:5197–5201.
28. Ernsting BR, Denninger JW, Blumenthal RM, Matthews RG (1993) *J Bacteriol* 175:7160–7169.
29. Kennerknecht N, Sahm H, Yen MR, Patek M, Saier MH, Jr, Eggeling L (2002) *J Bacteriol* 184:3947–3956.
30. Daldal F (1983) *J Mol Biol* 168:285–305.
31. Cronan JE, LaPorte DC (1996) in Escherichia Coli *and* Salmonella: *Cellular and Molecular Biology*, eds Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE (ASM Press, Washington, DC), 2nd Ed, pp 206–216.
32. Wek RC, Hatfield GW (1988) *J Mol Biol* 203:643–663.
33. Tabolina EA, Rybak KV, Khourges EM, Voroshilova EB, Gusyatiner MM (2005) US Patent 2005/0239175; *Chem Abstr* 137:231479.
34. Kutukova EA, Livshits VA, Altman IP, Ptitsyn LR, Zyiatdinov MH, Tokmakova IL, Zakataeva NP (2005) *FEBS Lett* 579:4629–4634.
35. Sambrook J, Russell DW (2001) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), 3rd Ed.
36. Lee SY, Woo HM, Lee D-Y, Choi HS, Kim TY, Yun H (2005) *Biotechnol Bioproc Eng* 10:425–431.
37. Reed JL, Vo TD, Schilling CH, Palsson BO (2003) *Genome Biol* 4:R54.
38. Segre D, Vitkup D, Church GM (2002) *Proc Natl Acad Sci USA* 99:15112–15117.
39. Alper H, Jin YS, Moxley JF, Stephanopoulos G (2005) *Metab Eng* 7:155–164.
40. Livak K, Shmittgen TD (2001) *Methods* 25:402–408.