

# Redesigning Enzymes Based on Adaptive Evolution for Optimal Function in Synthetic Metabolic Pathways

Yasuo Yoshikuni,<sup>1,4</sup> Jeffrey A. Dietrich,<sup>1,4</sup> Farnaz F. Nowroozi,<sup>1,4</sup> Patricia C. Babbitt,<sup>1,5,6</sup> and Jay D. Keasling<sup>1,2,3,4,\*</sup><sup>1</sup>UCSF/UCB Joint Graduate Group in Bioengineering<sup>2</sup>Department of Chemical Engineering<sup>3</sup>California Institute for Quantitative Biomedical Research (QB3)

University of California at Berkeley, Berkeley, CA 94720, USA

<sup>4</sup>Synthetic Biology Department, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94710, USA<sup>5</sup>California Institute for Quantitative Biomedical Research (QB3)<sup>6</sup>Department of Pharmaceutical Chemistry and Biopharmaceutical Sciences

University of California at San Francisco, San Francisco, CA 94134, USA

\*Correspondence: [keasling@berkeley.edu](mailto:keasling@berkeley.edu)

DOI 10.1016/j.chembiol.2008.05.006

## SUMMARY

Nature has balanced most metabolic pathways such that no one enzyme in the pathway controls the flux through that pathway. However, unnatural or nonnative, constructed metabolic pathways may have limited product flux due to unfavorable in vivo properties of one or more enzymes in the pathway. One such example is the mevalonate-based isoprenoid biosynthetic pathway that we previously reconstructed in *Escherichia coli*. We have used a probable mechanism of adaptive evolution to engineer the in vivo properties of two enzymes (3-hydroxy-3-methylglutaryl-CoA reductase [tHMGR] and many terpene synthases) in this pathway and thereby eliminate or minimize the bottleneck created by these inefficient or nonfunctional enzymes. Here, we demonstrate how we significantly improved the productivity (by ~1000 fold) of this reconstructed biosynthetic pathway using this strategy. We anticipate that this strategy will find broad applicability in the functional construction (or reconstruction) of biological pathways in heterologous hosts.

## INTRODUCTION

It is well known that many enzymes are able to catalyze highly specific chemical reactions with surprising accuracy and efficiency (Kraut et al., 2003). In metabolic pathways, these enzymes, each catalyzing different reactions in a series, often cooperate to minimize the unnecessary accumulation of metabolic intermediates while maximizing the production of the final product or optimizing the control of its production. Although natural metabolic pathways tend to be well balanced (with no one enzyme controlling flux), constructed (heterologous) metabolic pathways may not be so balanced, particularly when the enzymes for that metabolic pathway come from organisms that are very different from the host. If the function of any single enzyme were suboptimal, the enzyme would cause a significant bottleneck in the pathway, lower the yield of the final product,

or limit the growth of the host organism. The flux through any particular step is determined by the expressability, solubility, stability, substrate specificity, reaction selectivity, and activity (collectively referred to as in vivo properties) of the enzyme catalyzing that reaction as well as the concentration of substrates, cofactors, inhibitors, and products. Although one might be able to use traditional metabolic engineering techniques (e.g., expression modulation, promoter engineering, and codon optimization) to overexpress a rate-limiting enzyme in a metabolic pathway to improve flux through that pathway, a better strategy would be to improve the in vivo properties of that rate-limiting enzyme so that it need not be overexpressed and the amino acids and energy needed to overexpress the enzyme would not be expended and could be used for other cellular processes. In cases where the in vivo properties cannot be improved by traditional methods, improving in vivo properties may be the only method to improve pathway flux. Therefore, the in vivo enzyme properties are important determinants of the efficiency of metabolic pathways and should be a major concern in designing and constructing synthetic pathways.

We previously constructed a synthetic mevalonate pathway in *E. coli* (Figure S1, see the Supplemental Data available with this article online) and coupled it with a terpene synthase for production of amorpha-4,11-diene (a precursor to the antimalarial drug artemisinin) (Martin et al., 2003). Since other terpenoids have also found use as drugs, flavors, fragrances, and nutraceuticals and potential use as advanced biofuels, this microbial system has become an attractive alternative to extraction from the natural host for production of these chemically complex terpenoids. However, two reaction steps—catalyzed by an N-terminally truncated 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) reductase (tHMGR) (Donald et al., 1997) and many terpene synthases we have examined (Martin et al., 2001; Reiling et al., 2004)—have been identified as the primary bottlenecks in this synthetic pathway as the result of their unfavorable in vivo properties. That these two enzymes limited product formation was evident by the data showing that expressing amorpha-4,11-diene synthase, an enzyme with excellent in vivo properties, significantly reduced the accumulated farnesyl diphosphate (FPP) and allowed production of large titers of amorpha-4,11-diene (over 200 mg/l) and that vastly overexpressing the gene encoding tHMGR improved terpene production (Pitera et al., 2007). Although

amorphadiene synthase had favorable in vivo properties, most terpene synthases have much lower activity in vivo (Martin et al., 2001; Reiling et al., 2004). Not only do terpene synthases and tHMGR have poor in vivo properties, the substrates for these enzymes (HMG-CoA and FPP, respectively) appeared to be toxic and inhibit the growth of *E. coli* when they accumulated inside the cell (Martin et al., 2003; Pitera et al., 2007). However, the products of these enzymes (mevalonate and sesquiterpenes, respectively) do not interfere with host growth and are readily secreted into the medium.

The genes for the terpene biosynthetic pathway were encoded in three separate constructs. The *MevT* operon contains the genes encoding the enzymes in the top part of the pathway (acetyl-CoA to mevalonate: *atoB* encoding acetoacetyl-CoA thiolase, *ERG13* encoding HMG-CoA synthase [HMGS], and the 3' portion of *HMG1* [without the 5' portion of the gene encoding the membrane-binding N terminus of the enzyme] encoding tHMGR [Figure S1A]). The MBIS operon contains the genes encoding the bottom part of the pathway (mevalonate to FPP: *ERG12* encoding mevalonate kinase [MK], *ERG8* encoding phosphomevalonate kinase [PMK], *MVD1* encoding mevalonate diphosphate decarboxylase [MVD], *idi* encoding isopentenyl diphosphate isomerase, and *ispA* encoding prenyl diphosphate synthase [Figure S1B]). The third construct consists of the terpene synthase, which catalyzes the conversion of the prenyl diphosphates (geranyl-, farnesyl-, and geranylgeranyl diphosphates) to various terpenes with different regio- and stereochemistry (Figure S1C).

An example of a terpene synthase with poor in vivo properties is  $\gamma$ -humulene synthase (HUM). HUM is a sesquiterpene synthase from the gymnosperm *Abies grandis* that is known to produce at least 52 different sesquiterpene olefins from FPP through a wide variety of carbocation reaction mechanisms (Steele et al., 1998). On the basis of the theory of divergent molecular evolution, we previously demonstrated the redesign of many specific mutant variants that produce a single or a few terpenes (Martin et al., 2003; Yoshikuni et al., 2006a; Yoshikuni and Keasling, 2007). Because of its versatility and plasticity in catalysis, HUM might find use in industrial terpenoid production and is an excellent model to study the biochemistry of terpene synthases. However, the overexpression of HUM with this system resulted in the formation of less than 0.02 mg/l of terpenes even after expression optimization at 37°C. Detailed analyses have revealed that a large fraction of HUM misfolds and accumulates in the insoluble fraction. As a result, only a small fraction of this enzyme is functional in vivo. Although various genetic tools have been tried to improve expression of this enzyme, none of these strategies has worked well. Because of the potential importance of this enzyme and the difficulties in increasing its activity in vivo, we have chosen this enzyme and its specific mutant variants as models for redesign.

As has been discussed, since in vivo properties are strongly dependent on innate characteristics of enzymes, neither traditional nor more recently developed metabolic engineering strategies (Pitera et al., 2007) are likely to improve poorly functioning enzymes. Instead, protein engineering strategies that actually improve the function of these enzymes might be very useful or even become the sole solution to these problems. Protein engineering has been widely adopted as an effective design strategy

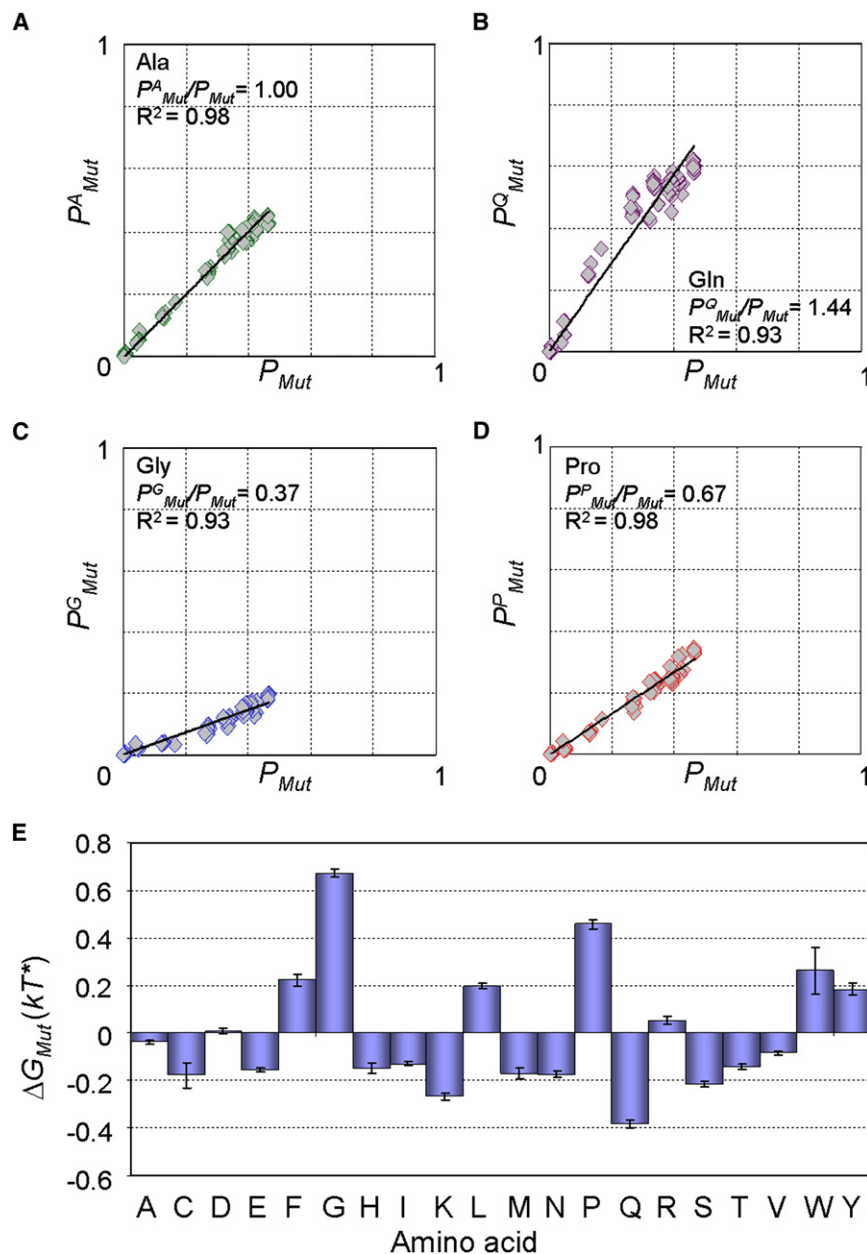
to improve many different enzyme properties. However, the application of conventional protein engineering strategies, such as directed evolution (Roodveldt et al., 2005) and rational and computational design (Daujotyte et al., 2003; Nasreen et al., 2006), are extremely difficult or nearly impossible for such enzymes as HUM and other terpene synthases because of the lack of both a high-throughput screen and enzyme structures with atomic level accuracy. Although several screening methodologies using fused reporter proteins have been previously described, the reporter activity is not necessarily correlated with the activity of the target proteins and, hence, a large number of mutants still need to be screened (Yoshikuni et al., 2006b).

In addition, since it is thought that each terpene synthase has evolved after a series of speciation events, the primary sequences of terpene synthases of diverse function are more related within an organism or closely related organisms than are terpene synthases of similar function but in very different organisms. Primary sequences for terpene synthases are limited and biased to the enzymes derived from model species. Thus, it is also extremely difficult to extract useful information from a conventional analysis of their evolutionary relationship (Amin et al., 2004; Lehmann et al., 2002; Steipe et al., 1994). Furthermore, because the function of these enzymes is highly plastic, any single mutation introduced into a terpene synthase could alter its product selectivity. Therefore, a methodology that does not affect product selectivity would be ideal. For these reasons, we have considered how evolution might essentially optimize and maintain the favorable in vivo properties of enzymes and thereby efficiencies of metabolic pathways. Here, we present a simple and fast redesign strategy for enzymes based on plausible mechanisms of molecular evolution and demonstrate its use on two enzymes (tHMGR, HUM, and its specific mutant variants) whose in vivo activities have been difficult to improve using traditional metabolic engineering techniques.

## RESULTS AND DISCUSSION

### Immutability of Amino Acids

In nature, pathways are thought to be constructed and gradually optimized through adaptive evolution of each enzyme in the pathway to render the host organisms adaptable to changing environments (Schmidt et al., 2003). In molecular evolution, the fixation probability of mutations is primarily determined by their fitness effects if a population size is small, which can be deleterious (opposed by purifying selection and likely discarded from a population), neutral or nearly neutral (genetic drift), or advantageous (generated through positive selection and fixed in a population) (Pal et al., 2006). It has been proposed that many enzymes are horizontally transferred from one species to another and are subsequently adapted to the new intracellular environment (Pal et al., 2005). Therefore, when mutations to a particular amino acid at a given position are kept to a minimum in this process (less mutable), then these amino acids are probably more essential for maintaining the in vivo properties of that enzyme. If such amino acids could be identified, then it would be possible to improve their function in the host organisms by properly realigning these amino acids on the basis of evolutionary relationships of that enzyme.



**Figure 1. Evolutionary Study of the Relative Stability for Each Amino Acid**

An immutability of each amino acid (likelihood to gain and loss by substitutions, insertions, and deletions) was calculated by comparing over 200 *E. coli* proteins involved in central metabolism and each of their homologs (30,000 total), and it was converted to free energy difference ( $\Delta G_{Mut}^X$ ;  $kT^*$  denotes arbitrary unit). A probability of mutation to each amino acid  $X (P_{Mut}^X)$  was plotted against that for all amino acids ( $P_{Mut}$ ) using the glutamate synthase large subunit protein family as an example; plots for alanine ( $P_{Mut}^A/P_{Mut} \cong 1$ ) (A), glutamine ( $P_{Mut}^G/P_{Mut} > 1$ ) (B), glycine ( $P_{Mut}^G/P_{Mut} < 1$ ) (C), and proline ( $P_{Mut}^P/P_{Mut} < 1$ ) (D) are shown. The average of the relative stability for each amino acid to mutations obtained from analyses of 209 different protein families (or superfamilies) is shown (E: mean  $\pm$  2 SE). The results clearly indicate that Gly and Pro were significantly more immutable over the course of evolution.

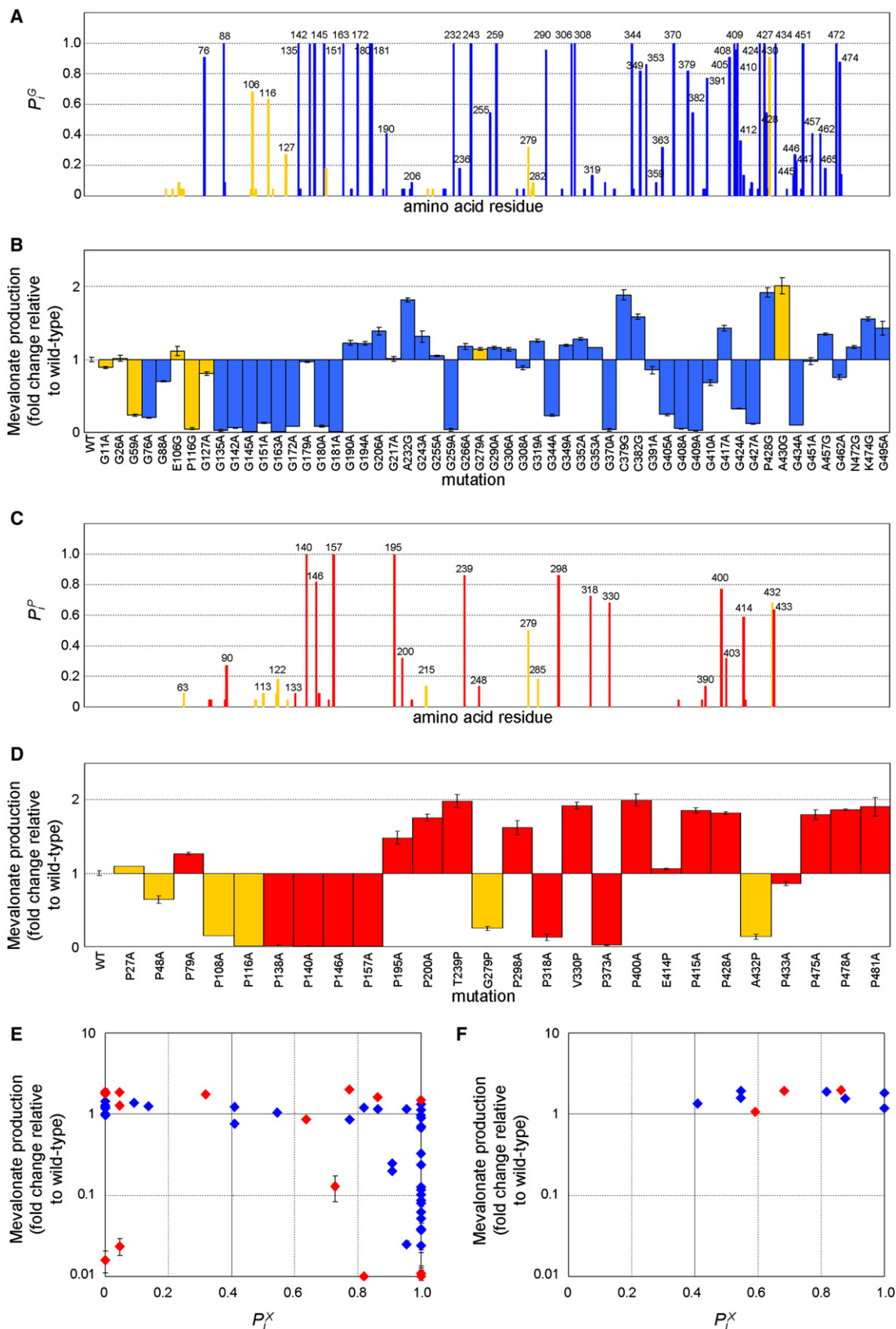
enzymes. The average  $P_{Mut}^X/P_{Mut}$  for each amino acid  $X$  was determined and then used to calculate the change in free energy due to the amino acid change ( $\Delta G_{Mut}^X$ ) (Figure 1E). The analysis clearly indicates that Gly and Pro are significantly less mutable, compared with other amino acids in *E. coli* central metabolic enzymes. The results imply that Gly and Pro are more essential and have been acquired or maintained or both during the adaptation of these enzymes toward function in *E. coli*.

#### Redistribution of Gly and Pro Residues at Predicted Positions in Proteins

As mentioned above, through prior work in our laboratory, we determined that tHMGR and HUM limited flux through the terpene biosynthetic pathway (Pitera et al., 2007; Yoshikuni et al., 2006a). Al-

though several groups have previously considered the immutability of amino acids, the results have not been consistent and appear to change with the types of proteins being analyzed (Endo et al., 2002; Graur, 1985). To reconcile these findings, we analyzed over 30,000 homologous sequences to over 200 different *E. coli* enzymes involved in central metabolism across multiple species (see Experimental Procedures for details). Because of the essential roles of central metabolic enzymes in host viability, the in vivo properties of these enzymes are expected to be optimal. The probabilities of mutations to each amino acid  $X$  between two sequences ( $P_{Mut}^X$ ) and that for all amino acids ( $P_{Mut}$ ) were determined (see Experimental Procedures for more information); interestingly,  $P_{Mut}^X$  and  $P_{Mut}$  were linearly correlated in many cases (Figures 1A–1D). Thus, we are able to determine which amino acid contributes most to the diversification of

though both enzymes were expressed at a high level, neither enzyme had sufficient in vivo activity. Thus, we focused our work on these two enzymes. We selectively redistributed Gly and Pro both in tHMGR and HUM on the basis of their evolutionary relationships to other HMGRs and terpene synthases, respectively. First, multiple sequence alignments (MSAs) for both tHMGR and HUM were constructed. The MSA for tHMGR was constructed using other HMGR sequences derived from Archaea, because HMGR from Archaea is produced in soluble form. The MSA for HUM was constructed using mono- and diterpene synthases derived from gymnosperms, because only few sesquiterpene synthases derived from gymnosperms have been discovered. Next, the probabilities of conservation for Gly ( $P_i^G$ ) and Pro ( $P_i^P$ ) at each  $i$ th residue of tHMGR and HUM were calculated (represented in blue and red bars, respectively, in Figures 2A and 2C for



tHMGR and in Figures 3A and 3C for HUM). Finally, substitutions involving Gly and Pro were introduced to both tHMGR and HUM according to the predicted profiles.

For the case that the product of particular enzyme does not impact cell physiology and the reaction step catalyzed by this enzyme is the primary bottleneck in a biosynthetic pathway, it can be assumed that improvements to the *in vivo* properties of the enzyme will be directly reflected in its product formation and that mutational effects can be easily evaluated. Therefore, the fitness effects of these mutations on tHMGR and HUM were monitored by the level of *in vivo* mevalonate and sesquiterpene production, respectively (represented by the blue and red bars, respectively, in Figures 2B and 2D for tHMGR and Figures 3B and 3D for HUM).

In both cases, the predictions were greater than 80% accurate (Gly- and Pro-related mutations are represented by the blue and red dots, respectively in Figures 3E and 3F, and Figures 4E and 4F); the exceptions were the residues that did not align well in MSAs (represented by the yellow bars in Figures 3A–3D and Figures 4A–4D). In the case of HUM, the effects of mutations to residues predominantly conserved in either mono- or diterpene synthases (represented by the green and purple bars in Figures 3A–3D, respectively) were difficult to evaluate, because changes to these residues substantially impacted enzyme activity. As such, the quality of the MSA would significantly affect the accuracy of the predictions. Although the MSA for HUM was constructed primarily from mono- and diterpene synthases, relatively high correlation between product formation and probability of conservation was observed (represented by the blue and red dots in Figures 3E and 3F for HUM). The correlation for tHMGR was not as good as that for HUM (represented by the blue and red dots in Figures 2E and 3F). Because the pathway containing the native tHMGR was already capable of producing mevalonate at almost 10% of its theoretical maximum, mevalonate production could be improved only a few fold. Thus, improvement to the *in vivo* properties of tHMGR did not have as significant impact on mevalonate production as improvements on HUM had on sesquiterpene production. It is also interesting to note that the method was good at predicting the effects of cumulative neighbor mutations. For example, introduction of each of the tHMGR mutants G353A and G349A resulted in increased mevalonate production even though the mutations were predicted to reduce mevalonate production. However, when these mutations were introduced into the G352A mutant, these mutations decreased mevalonate production (10%–40% reduction in mevalonate productivity, compared with that from G352A alone). Thus, the methodology accurately predicted the effect of cumulative neighbor mutations.

Finally, to determine whether another amino acid might be a better change at these positions, saturation mutagenesis was performed on G148, G227, G327, and G361 in HUM. In all cases, substitution of the native amino acid with Ala improved sesquiterpene production, and no further mutations were considered.

### Accumulation of Single Mutations Improved the *In Vivo* Enzyme Properties and Production of Various Terpenes

Mutations that improved the *in vivo* properties of tHMGR and HUM were subsequently recombined, and the effects of mutations on their *in vivo* properties were cumulative; we have constructed the tHMGR-G9 mutant containing the changes P200A/G206A/T239P/G319A/G352A/G417A/P428G/K474G/G495A and HUM-G6 mutant containing the changes K126P/R142G/G148A/G227A/G327A/G361A. When integrated into the host alone, tHMGR-G9 increased mevalonate production three fold, and HUM-G6 improved sesquiterpene production 80 fold (Figure 4B and Figure 5B, respectively). The improvements in product titer were due to increases in flux through the pathway as well as improvements in growth, as *E. coli* harboring tHMGR-G9 did not suffer the growth inhibition of accumulated HMG-CoA that the strain harboring wild-type tHMGR did (Figure 4A and Figure 5A). When the tHMGR-G9 and HUM-G6 mutants were integrated into the same host, host growth improved by three to four fold, and sesquiterpene production improved by nearly 1000 fold (Figures 5A and 5B).

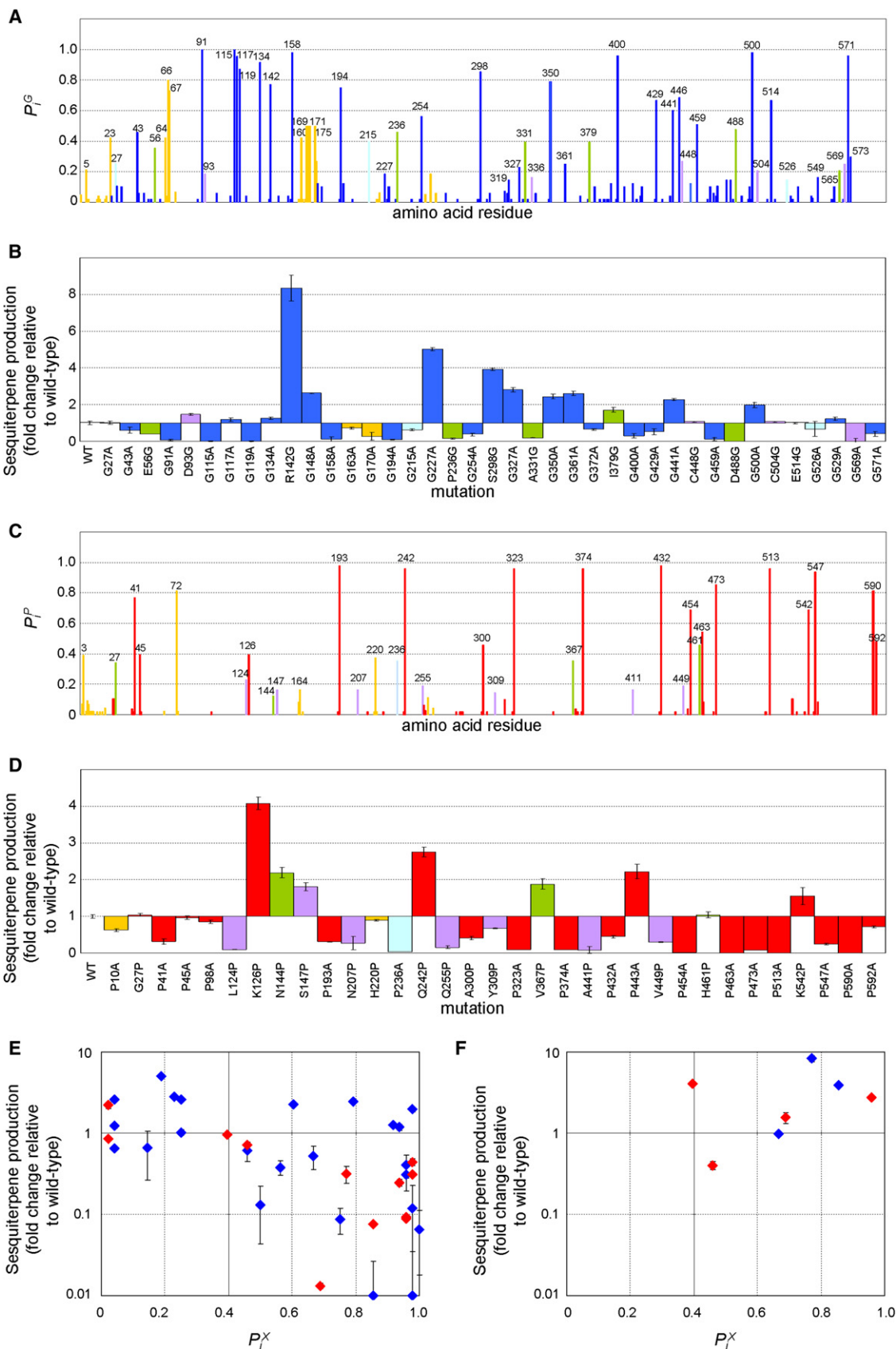
The same mutations that were introduced into HUM-G6 were also introduced into six previously designed, mutant variants of HUM that produced a few or a single product rather than the many products generated by HUM (Yoshikuni et al., 2006a); the effects of mutations on their *in vivo* properties were similar to that of HUM. The terpene productivities of the host harboring these improved enzymes were also dramatically improved (8–34 mg/l from ~0.02 mg/l; Figure 5C). Given that the conservation analysis is supported for over 200 different enzyme families (or superfamilies), we predict that similar mutations could improve the *in vivo* properties of many other enzymes.

### The Effects of Gly- and Pro-Related Mutations on Enzyme Activity

To investigate how Gly and Pro redistributions contribute to such a large improvement in sesquiterpene production, product selectivity, steady state kinetics, and *in vivo* sesquiterpene production from S-tagged versions of HUM, HUM-G3 (K126P/R142G/G227A), and HUM-G6 were measured. (The S-tag was removed for the steady state kinetics study). Although terpene synthases are known to be very plastic (Yoshikuni et al., 2006a; Yoshikuni

#### Figure 2. MSA-Predicted Gly and Pro Redistributions in tHMGR and Their Effect on Mevalonate Production

Distributions of Gly, represented in blue bars in (A), and Pro, represented in red bars in (C), to have effect on the *in vivo* properties of tHMGR were predicted on the basis of an MSA constructed using the primary sequences of HMGR derived from Archaea as a guide (sharing 30%–40% sequence identity). Residues that did not align in the MSA are represented by yellow bars in these figures. According to these profiles, Gly → Ala, Xaa → Gly, Pro → Ala, and Xaa → Pro substitutions were introduced into tHMGR, and functional consequences for these substitutions were monitored by *in vivo* mevalonate production. The effects caused by Gly → Ala and Xaa → Gly mutations are represented by the blue bars in (B), the effects caused by Pro → Ala and Xaa → Pro mutations are represented by the red bars in (D), and the effects caused by mutations to the residues that did not align in the MSA are represented by the yellow bars in (B) and (D) (mean ± SD of triplicate measurements). The correlation between the fitness effects caused by mutations and  $P_i^*$ , are plotted in (E) and (F). Gly → Xaa and Pro → Xaa mutations are represented by the blue and red dots, respectively, in (E), and the Xaa → Gly and Xaa → Pro mutations are represented by the blue and red dots, respectively, in (F). These results show that over 80% of the aligned mutations were accurately predicted from these profiles using  $P_i = 0.4$  as a threshold.  $P_i = 0.4$  was chosen for threshold because the MSA was constructed using homologous sequences with 30%–40% identity.



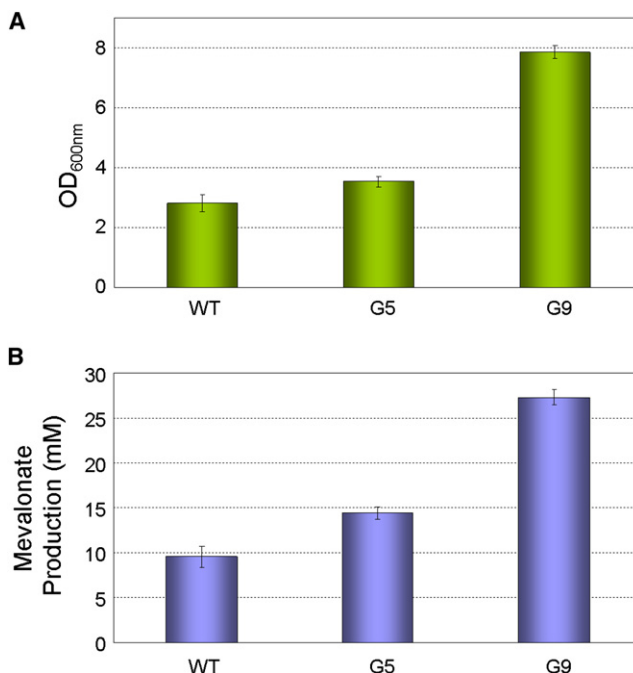
et al., 2006b), the product selectivity of HUM-G6 was comparable to that of wild-type HUM (Table 1). In addition, its activity ( $k_{\text{cat}}/K_m$ ) was similar to that of wild-type HUM (Table 1). In contrast, in vivo sesquiterpene production from S-tagged HUM-G6 increased nearly two fold over that of the non-S-tagged HUM-G6. The improvement in sesquiterpene production from HUM to HUM-G6 increased with temperature (3.3-fold at 20°C, 10-fold at 30°C, and 220-fold at 37°C; Figure 6A). Of all the temperatures tested, wild-type HUM showed the highest terpene production at 30°C, whereas HUM-G6 showed the highest terpene production at 37°C. Given that more enzyme was found in the soluble fraction at 37°C than at 30°C (Figure 6B), it appears that wild-type HUM does not fold properly at higher temperatures (many enzymes found even in the soluble fraction at 37°C were not functional) and that Gly and Pro redistributions increased the ability of HUM-G6 to functionally fold in the *E. coli* intracellular environment at higher temperature. In fact, quantification of in vivo enzyme levels in both the soluble fraction and crude lysate revealed that the mutations significantly improved the concentration of soluble enzymes (Figure 6C).

## SIGNIFICANCE

We successfully redesigned the in vivo properties of tHMGR, HUM, and its six specific mutant variants to improve in vivo sesquiterpene production without affecting the respective catalyzed reactions. Because of their unfavorable in vivo properties, these enzymes have been previously found to be significant bottlenecks in the terpene biosynthetic pathway. From our initial analysis of over 30,000 homologs to over 200 *E. coli* enzymes involved in central metabolism, Gly and Pro were initially suggested to be essential for their in vivo properties. These amino acids were then selectively redistributed in tHMGR and HUM on the basis of their evolutionary relationships. The results show that this protein engineering strategy is simple and fast, but yet powerful and effective to redesign in vivo enzyme properties without altering product distributions, a task that would have been extremely difficult or nearly impossible using conventional protein engineering strategies. Since proper distributions of these amino acids can be predicted largely from their evolutionary relationships and their redistribution dramatically impacted the flux through the heterologous, biosynthetic pathway, it is likely that proper distributions of these amino acids are essential for many enzymes expressed in *E. coli* and could have been achieved primarily as a result of adaptive evolution. Together with our previous studies (Yoshikuni et al., 2006a; Yoshikuni and Keasling, 2007), these results

### Figure 3. MSA-Predicted Gly and Pro Redistributions in HUM and Their Effect on Sesquiterpene Production

Distributions of Gly, represented by blue bars in (A), and Pro, represented by red bars in (C), to have effect on the in vivo properties of HUM were predicted on the basis of an MSA constructed using the primary sequences of mono-, sesqui-, and diterpene synthases derived from gymnosperms as a guide. Residues that did not align, aligned only in monoterpene synthases, and aligned only in diterpene synthases in MSA are represented by yellow, green, and purple bars, respectively. According to this profile, Gly → Ala, Xaa → Gly, Pro → Ala, and Xaa → Pro substitutions were introduced into HUM, and the fitness effects of these substitutions were monitored by in vivo sesquiterpene production. The effects caused by Gly → Ala and Xaa → Gly mutations are represented by the blue bars in (B), the effects caused by Pro → Ala and Xaa → Pro mutations are represented by red bars in (D), and mutations to residues that did not align in MSA are represented by yellow bars in (B) and (D) (mean ± SD of triplicate measurements). The correlation between the fitness effects by mutations and  $P_i^*$  is shown in (E) and (F). Gly → Xaa and Pro → Xaa mutations are represented by blue and red dots, respectively, in (E), and Xaa → Gly and Xaa → Pro mutations are represented by blue and red dots, respectively, in (F). These results indicate that over 80%–90% of the aligned mutations were accurately predicted from these profiles with  $P_i = 0.4$  as a threshold;  $P_i = 0.4$  was chosen for threshold because the MSA was constructed using homologous sequences with 30%–40% identity.



**Figure 4. Integration of Redesigned tHMGR into *E. coli* and the Resulting Mevalonate Production**

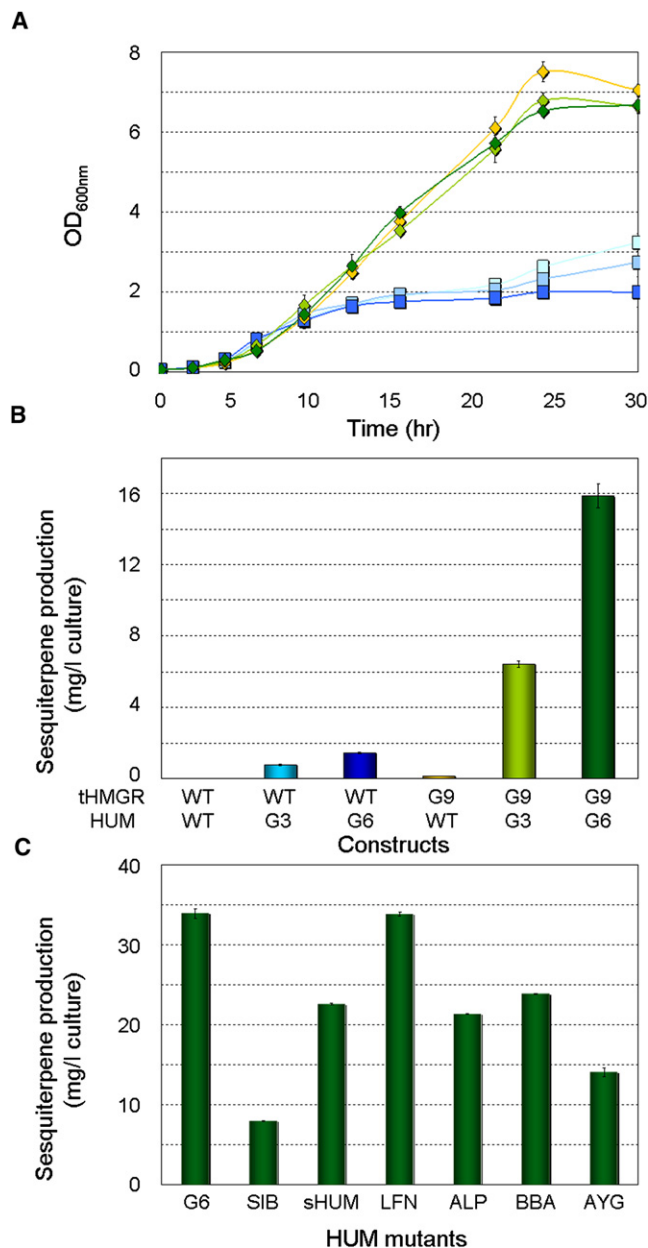
The growth (A) and mevalonate production (B) for strains harboring pBADMeVt containing tHMGR-WT (yeast wild-type HMG1p of its membrane binding domain truncated), tHMGR-G5 (G206A/G319A/G352A/G417A/G495A), and tHMGR-G9 (P200A/G206A/T239P/G319A/G352A/G417A/P428G/K474G/G495A) were measured. Both the growth and mevalonate production improved approximately 2.5–3-fold, and the increase in growth is proportional to the increase in mevalonate production. We previously demonstrated that accumulation of HMG-CoA inhibits cell growth. Thus, improving the in vivo properties of tHMGR reduced HMG-CoA accumulation and thus the toxicity of the pathway.

provide evidence that protein engineering strategies based on the theories of molecular evolution are very useful for design and construction of synthetic systems.

## EXPERIMENTAL PROCEDURES

### Reagents and Equipments

All enzymes and chemicals were purchased from New England Biolabs and Sigma-Aldrich, respectively, unless otherwise stated. An HP6890 gas chromatograph equipped with a 5973 mass selective detector (Hewlett Packard) or flame ionization detector, a CyclosilB capillary column (30 m × 250 μm i.d. × 0.25 μm thickness; Agilent Technologies) or DB5-MS capillary column (30 m × 250 μm i.d. × 0.25 μm thickness; Agilent Technologies), and a Combi PAL auto sample-injector (LEAP Technologies) were used for analysis. An



**Figure 5. Cointegration of Redesigned HUM and tHMGR into the Synthetic Pathway and the Resulting In Vivo Sesquiterpene Production**

*E. coli* DH1 harboring pBADMeVt (tHMGR-WT or tHMGR-G9 [P200A/G206A/T239P/G319A/G352A/G417A/P428G/K474G/ G495A]), pBBRMBIS, and pTrcHUM15 (containing HUM-WT, HUM-G3 [K126P/R142G/G227A], or HUM-G6 [K126P/R142G/G148A/G227A/G327A/G361A]) was used for in vivo sesquiterpene production. The growth curve (A) and sesquiterpene production at 24 hr after inoculation (B) are shown. HUM-WT, HUM-G3, and HUM-G6 cointegrated with tHMGR are shown in light blue, medium blue, and dark blue, respectively, and those with tHMGR-G9 are shown in yellow, light green, and dark green, respectively. The strain containing tHMGR-G9 grew 3-fold higher and produced 3-fold more mevalonate (Figure 4), resulting in a synergistic improvement in overall sesquiterpene production. The mutations in HUM-G6 were also applied to specific variants of HUM previously constructed in our laboratory (SIB, sibirene synthase; sHUM, specific  $\gamma$ -humulene synthase; LFN, longifolene synthase; ALP,  $\alpha$ -longipinene synthase; BBA,  $\beta$ -bisabolene

LS6500 multipurpose scintillation counter (Beckman Coulter) was used for enzyme kinetics.

#### Analysis of Amino Acid Composition Changes in Proteins across Multiple Species

To examine the relative importance of each of 20 different amino acids (X: Ala, Cys, Asp...), we calculated relative immutability (likelihood to gain and loss by substitutions, deletions, and insertions) for each amino acid from over 200 different protein families (or superfamilies), and it was fitted to free energy difference ( $\Delta G_{Mut}^X$ ). These protein families are all involved in central metabolism, including glycolysis, citric acid cycle, pentose phosphate pathway, oxidative phosphorylation, fatty acid metabolism, amino acid metabolism, and nucleic acid metabolism. In each protein family (or superfamily) (F), each homologous protein sequence (H) was searched using the basic local alignment search tool for proteins (BLASTP: <http://www.ncbi.nih.gov>). In pair-wise alignment between a particular *E. coli* protein sequence and its homologous protein sequence derived from a particular species, the probability of mutations ( $P_{Mut,H,F}^X$ ) for each amino acid (X) was calculated on the basis of the composition of all aligned pairs of amino acids described by the following formula:

$$P_{Mut,H,F}^X = \frac{N_{Mut,H,F}^X}{N^X}$$

where  $N_{Mut,H,F}^X$  and  $N^X$  denote the number of amino acid pairs only one and at least one of which is X in each pair-wise alignment, respectively. The pair-wise alignments used herein covered at least 80% of and shared over 50% identity to the corresponding *E. coli* protein sequences. For this calculation, we assumed that if proteins had evolved without any constraint, the  $P_{Mut,H,F}^X$  should be identical to that for all amino acids ( $P_{Mut,H,F}$ ) described as the following formula:

$$P_{Mut,H,F} = \frac{N_{Mut,H,F}}{N}$$

where  $N_{Mut,H,F}$  denotes the number of mutated amino acid pairs, and  $N$  denotes the number of all amino acid pairs in each pair-wise alignment. We, then, plotted  $P_{Mut,H,F}^X$  against  $P_{Mut,H,F}$ . On average, 164 ( $2 \times SE = 22$ ) plots (pair-wise sequence alignments) were made for each of over 200 protein families.  $P_{Mut,H,F}^X/P_{Mut,H,F}$  is defined as the slope for the linear regression of the data in the plot. The free energy difference of each amino acid X for the mutations in each protein family (or superfamily) F ( $\Delta G_{Mut,F}^X$ ) was then calculated according to Boltzmann statistics as follows:

$$\frac{P_{Mut,F}^X}{P_{Mut,F}} = \exp\left(\frac{-\Delta G_{Mut,F}^X}{kT^*}\right)$$

where  $kT^*$  denotes an arbitrary constant. In this analysis, we calculated  $\Delta G_{Mut,F}^X$  only when the  $R^2$  of the  $P_{Mut,F}^X/P_{Mut,F}$  plot was greater than 0.9 or 0.7 if the plots were distributed symmetrically about the trend line to minimize the bias introduced by duplicate or nearly duplicate plots.

#### Design Methodology to Improve In Vivo Properties of Enzymes Using MSA as a Guide

To predict where to distribute Gly, Pro, and Xaa (where Xaa denotes any amino acid residues other than Gly and Pro), we first created an MSA for both HUM and tHMGR using MUSCLE ([http://phylogenomics.berkeley.edu/cgi-bin/muscle/input\\_muscle.py](http://phylogenomics.berkeley.edu/cgi-bin/muscle/input_muscle.py)). The primary sequence of HUM from the gymnosperm *Abies grandis* was aligned with other mono-, sesqui-, and diterpene synthases derived from gymnosperms. Although many sesquiterpene synthases have been isolated from angiosperms, mono- and diterpene synthases from gymnosperms are more closely correlated to HUM at the primary sequence level. The primary sequence of tHMGR derived from yeast, which was solubilized for expression in *E. coli* by removal of the N terminus, was aligned with other orthologous sequences derived from archaeal species, as the archaeal HMGR is produced in a soluble form as opposed to the membrane-bound form found in most eukaryotes. The conservation probability

synthase; and AYG,  $\alpha$ -ylangene synthase) were cointegrated with tHMGR-G9 (C). The resulting specific terpene production at 48 hr after inoculation improved dramatically for each variant and is compared to total sesquiterpene production for G6. All data represent mean  $\pm$  SD of triplicate measurements.



**Table 1. Product Selectivity of HUM and Its Mutant Variants**

Name <sup>a</sup>	Mutations	Kinetic Parameters <sup>b</sup>			Product Distributions <sup>c,d</sup>					
		$k_{\text{cat}}$ ( $10^{-3} \text{ s}^{-1}$ )	$K_m$ ( $\mu\text{M}$ )	$k_{\text{cat}}/K_m$ ( $10^3 \text{ M}^{-1} \text{ s}^{-1}$ )	1	2	3	4	5	6
WT	None	12.00 ± 0.34	2.01 ± 0.17	5.96	8.3	7.2	14.9	26.1	34.0	9.5
G3	K126P, 142G, G227A	7.62 ± 0.21	4.66 ± 0.39	1.64	7.2	6.3	16.5	27.9	31.7	10.4
G6	K126P, 142G, G148A, G227A, G327A G361A	1.71 ± 0.17	0.69 ± 0.13	2.47	7.1	6.8	15.2	27.4	32.8	10.7
SIB	K126P, 142G, G148A, G227A, G327A G361A, F312Q, M339A, M447F	ND	ND	ND	0.2	2.8	2.2	<b>80.1</b>	13.8	0.9
HUM	K126P, 142G, G148A, G227A, G327A G361A, M339N, S484C, M565I	ND	ND	ND	5.6	11.7	5.9	0.7	<b>75.6</b>	0.6
LFN	K126P, 142G, G148A, G227A, G327A G361A, A317N <sup>e</sup> , A336S, S484C, I562V	ND	ND	ND	12.8	3.4	<b>62.1</b>	1.6	11.9	8.1
ALP	K126P, 142G, G148A, G227A, G327A, G361A, A336C, T445C, S484C, I562L, M565L	ND	ND	ND	<b>60.2</b>	4.6	13.7	0.4	14.6	6.5
BBA	K126P, 142G, G148A, G227A, G327A, G361A, A336V, M447H, I562T	ND	ND	ND	1.6	0.2	3.9	0.5	4.7	<b>89.1</b>
AYG	K126P, 142G, G148A, G227A, G327A, G361A, S484A, Y566F	ND	ND	ND	14.6	<b>27.5</b>	0.5	0.6	47.1	9.6

Bold type indicates products that enzymes were designed to produce.

<sup>a</sup> WT, wild type  $\gamma$ -humulene synthase; G3, third generation of mutant  $\gamma$ -humulene synthase; G6, sixth generation of mutant  $\gamma$ -humulene synthase; SIB, sibirene synthase; HUM, new  $\gamma$ -humulene synthase; LFN, longifolene synthase; ALP,  $\alpha$ -longipinene synthase; BBA,  $\beta$ -bisabolene synthase; AYG,  $\alpha$ -ylangene synthase.

<sup>b</sup> ND: parameters were not determined for a particular mutant.

<sup>c</sup> 1,  $\alpha$ -longipinene; 2,  $\alpha$ -ylangene; 3, longifolene; 4, sibirene; 5,  $\gamma$ -humulene; 6,  $\beta$ -bisabolene.

<sup>d</sup> All product distributions were represented for 1–6 as 100%; these are corresponding to more than 85%–95% and to 75% of total products in mutants and wild type (including G3 and G6), respectively.

<sup>e</sup> A317N occurred during recombination, and improved in vivo terpene production without a change in product distribution.

for Gly ( $P_i^G$ ) and Pro ( $P_i^P$ ) at column  $i$  in a given MSA was calculated on the basis of the composition of Gly and Pro at column  $i$  as follows:

$$P_i^X = \frac{N_i^X}{N_i}$$

where  $N_i^X$  and  $N_i$  denote the number of amino acid  $X$  (Gly or Pro) and the total number of aligned amino acids at position  $i$  in each column of the MSA, respectively. The fitness effects contributed by these mutations were predicted to be dependent on the value of  $P_i^X$ ; we used  $P_i^X = 0.4$  as a threshold because each sequence used to construct MSA shared approximately 35%–40% identity with HUM or tHMGR. We compared the value of  $P_i^X$  and the fitness effects resulting from single mutations; when  $P_i^X \geq 0.4$ , the mutation to amino acid  $X$  is predicted to show neutral, nearly neutral, or positive fitness effects and away from amino acid  $X$  to show neutral, nearly neutral, or negative fitness effects. When  $P_i^X \leq 0.4$ , the mutation to amino acid  $X$  is predicted to show neutral, nearly neutral, or negative fitness effects and away from amino acid  $X$  to show neutral, nearly neutral, or positive fitness effects. For example, if  $X$  is Gly,  $P_i^{\text{Gly}}$  is larger than 0.4, and the  $i$ th residue is Gly, the residue remains Gly; if the  $i$ th residue is any residue other than Gly, the residue is changed to Gly. If  $P_i^{\text{Gly}}$  is smaller than 0.4 and if the  $i$ th residue is Gly, the residue is changed to Ala; if the  $i$ th residue is any residue other than Gly, the residue remains unchanged.

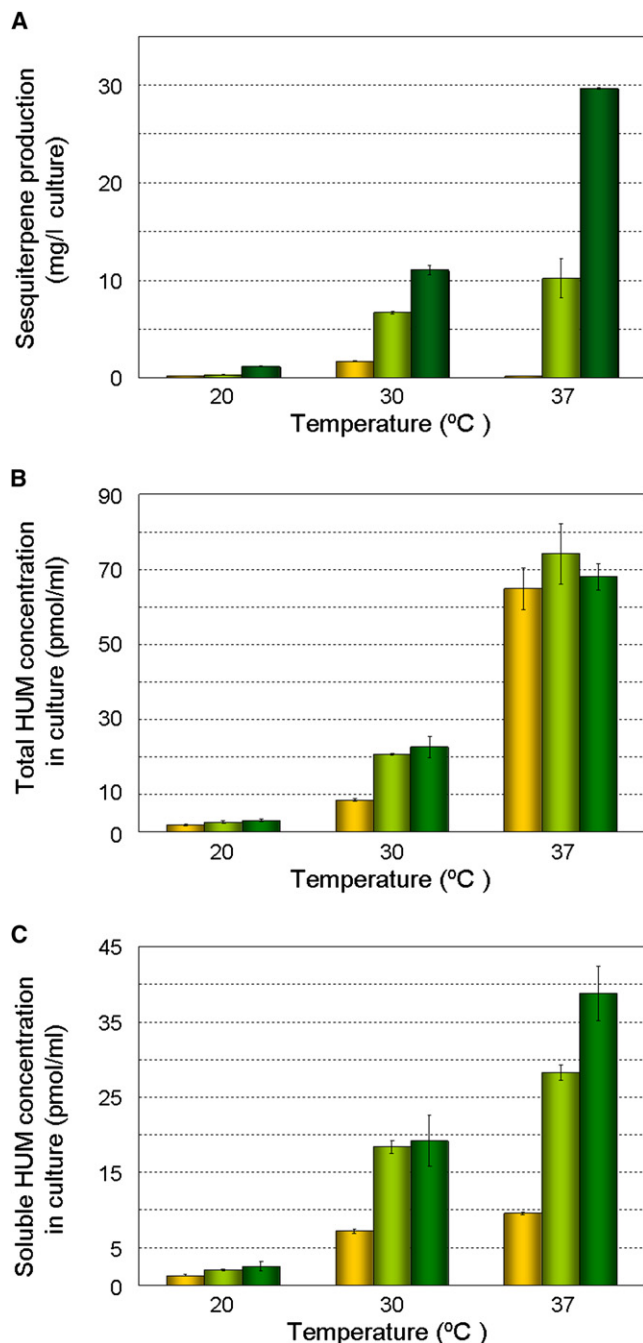
#### Strains and Plasmids

*E. coli* strains DH10B and DH1 were used for both mevalonate and sesquiterpene production, and BL21(DE3) was used for protein overexpression and purifica-

tion. Plasmids pBADMEVt (Martin et al., 2003) and their mutant variants were used for mevalonate production. Plasmid pBBRMBIS (Martin et al., 2003) was used for FPP production. Plasmids pTrcHUM (Yoshikuni et al., 2006a), pTrcHUM15, and their mutant variants were used for sesquiterpene production (Figure S1). Plasmid pTrcSHUM15 and its mutant variants were used for quantification of protein concentrations in vivo. Plasmids pETHUM (Yoshikuni et al., 2006a) and its mutant variants were used for protein overexpression and purification. All genes in pBADMEVt and pTrcHUM have been previously synthesized, and their codon usage has been optimized.

Since reduced expression of HUM slightly improved sesquiterpene production, an extra seven base pairs were introduced between the ribosome-binding-site (RBS) and the start codon at the *NcoI* site of pTrcHUM. The RBS region was amplified by polymerase chain reaction (PCR): 98°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec, repeated 30 times. The reaction mixture contained 1 × Phusion buffer, 2 mM dNTP, 0.5  $\mu\text{M}$  forward (5'-GCGCGTTGGTGC GATATC-3') and reverse (5'-CATGCCATGAGCCTTATTCTGTTCTCTGTGT GAAATTG-3') primers, 2.5 U Phusion DNA polymerase (Finezyme), and 50 ng pTrcHUM as a template in a total volume of 100  $\mu\text{l}$ . The amplified fragments were then digested with EcoRV/*NcoI* and inserted into the corresponding site of pTrcHUM to form pTrcHUM15.

pTrcSHUM15 was constructed on the basis of pTrcHUM15 backbone. The S-tag was fused to the N-terminal of HUM. The RBS region in pTrcHUM15 was amplified by PCR: 98°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec, repeated for 30 times. The reaction mixture contained 1 × Phusion buffer, 2 mM dNTP, 0.5  $\mu\text{M}$  forward (5'-GCGCGTTGGTGC GATATC-3') and reverse



**Figure 6. Investigation of the Temperature Effects for Gly and Pro Mutations**

S-tagged versions HUM-WT (yellow), HUM-G3 (light green), and HUM-G6 (dark green) were co-integrated with tHMGR-G9 into the synthetic pathway to examine the effect of temperature on accumulated Gly and Pro mutations. Sesquiterpene production by these strains (A), total HUM protein concentration in culture (B), and soluble HUM protein concentration in culture (C) at 24 hr after inoculation are shown. The  $OD_{600}$  reached approximately 2.5 at 20°C, 8.0 at 30°C, and 6.5 (G6)-8.0 (WT) at 37°C. Interestingly, sesquiterpene productivity from HUM-G6 improved almost 1.7-fold with an N-terminal S-tag. The higher the temperature, the more HUM was produced. At 37°C, the amount of HUM-G6 in the soluble fraction was significantly higher than that of HUM-WT. All data represent mean  $\pm$  SD of triplicate measurements.

(5'-GCAGCAGCGGTTTCTTTCATGGAGCTTATTCTGTTC-3') primers, 2.5 U Phusion DNA polymerase (Finezyme), and 50 ng pTrcHUM15 as a template in total volume of 100  $\mu$ l. The S-tag was amplified by PCR: 98°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec, repeated for 30 times. The reaction mixture contained 1 $\times$  Phusion buffer, 2 mM dNTP, 0.5  $\mu$ M forward (5'-GAAACA GAATAAGCTCCATGAAAGAAACCGCTGCTGC-3') and reverse (5'-CATGC CATGGAACCGCTGGC-3') primers, 2.5 U Phusion DNA polymerase (Finezyme), and 50 ng pET29 (Novagen) as a template in a total volume of 100  $\mu$ l. These two amplified fragments were spliced by overlap PCR: 98°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec, repeated for 30 times. The reaction mixture contained 1 $\times$  Phusion buffer, 2 mM dNTP, 0.5  $\mu$ M forward (5'-GCGCGTTGGTGGGATATC-3') and reverse (5'-CATGCCATG GAACCGCTGGC-3') primers, 2.5 U Phusion DNA polymerase (Finezyme), and the above-mentioned fragments as a template in a total volume of 100  $\mu$ l. The spliced fragment was then digested with EcoRV/NcoI and inserted into the corresponding site of pTrcHUM to form pTrcSHUM15.

#### GC-MS Analysis for In Vivo Mevalonate Production

To screen the single mutation library, a single colony harboring pBADMeVT (wild-type tHMGR or its mutant variants) was inoculated into Luria Bertani (LB) medium containing  $Cm^{50}$  and grown overnight at 37°C. An aliquot (50  $\mu$ l) of this seed culture was inoculated into fresh LB medium (5 ml) containing  $Cm^{50}$  and 13.3 mM (+)-L-arabinose and was grown for 24 hr at 37°C. An aliquot of culture (560  $\mu$ l) was mixed with 140  $\mu$ l of 0.5 M HCl to dehydrate the mevalonate to form mevalonolactone, and 700  $\mu$ l of ethyl acetate was then added to the sample. The mixture was vortexed for 5 min, and the ethyl acetate was analyzed by GC-MS using a GC oven temperature program of 90°C for 1 min, then ramping 30°C/min to 250°C for CyclosilB capillary column analysis. Mevalonolactone was identified from its mass spectrum and retention time by comparison to an authentic standard.

As for the final mevalonate production assay, a single colony harboring pBADMeVT (wild-type tHMGR or its mutant variants) was inoculated into LB medium containing  $Cm^{50}$  and grown overnight at 37°C. An aliquot (500  $\mu$ l) of this seed culture was inoculated into fresh modified m9 medium (50 ml, see above formulation) containing  $Cm^{50}$ . Two hours after the inoculation, (+)-L-arabinose was added to the final concentration of 13.3 mM. Mevalonate production was analyzed as mentioned above.

#### GC-FID and GC-MS Analysis for In Vivo Sesquiterpene Production

To screen the single mutation library, a single colony harboring pTrcHUM (wild-type HUM or its mutant variants) and pBBRMBIS was inoculated into LB medium containing 50  $\mu$ g/ml carbenicillin ( $Cb^{50}$ ) and 50  $\mu$ g/ml kanamycin ( $Km^{50}$ ) and grown overnight at 37°C. An aliquot (50  $\mu$ l) of this seed culture was inoculated into fresh LB medium (5 ml) containing 10 mM D/L-mevalonate,  $Cb^{50}$ , and  $Km^{50}$ , overlaid with 500  $\mu$ l dodecane, and was grown for 24 hr at 37°C. An aliquot of dodecane (50  $\mu$ l) was diluted into 200  $\mu$ l of ethyl acetate, and the mixture was analyzed by GC-MS or GC-FID using a GC oven temperature program of 80°C for 1 min, then ramping 30°C/min to 110°C, 5°C/min to 160°C, and 130°C/min to 250°C for CyclosilB capillary column analysis and of 80°C for 3 min, then ramping 5°C/min to 160°C, and 120°C/min to 300°C for DB-5MS capillary column analysis. Camphor was used as an internal standard. Sesquiterpenes were identified from their mass spectra and GC retention times by comparison to available authentic standards and spectra in libraries previously reported in the literature.

As for the final sesquiterpene production assay, a bacterial system containing three plasmids was used (Martin et al., 2003). A single colony harboring pTrcHUM15 or pTrcSHUM15 (wild-type HUM or its mutant variants), pBBRMBIS, and pBADMeVT (wild-type tHMGR [Donald et al., 1997] or its mutant variants) was inoculated into LB medium containing  $Cb^{50}$ ,  $Km^{50}$ , and chloramphenicol ( $Cm^{50}$ ) and was grown for overnight at 37°C. An aliquot of this seed culture was inoculated into 50 ml of fresh, modified, M9 medium (M9 salt, 75 mM MOPS, 3% glycerol, 5 g/l yeast extract, 2 mM  $MgSO_4$ , 1 mg/l thiamine, 10  $\mu$ M  $FeSO_4$ , 0.01 mM  $CaCl_2$ , and micronutrient [pH 7]) to the final  $OD_{600}$  of 0.05 containing  $Cb^{50}$ ,  $Km^{50}$ , and  $Cm^{50}$ , overlaid with 10 ml of dodecane. Two hours after the inoculation, isopropyl- $\beta$ -D-thiogalactopyranosid (IPTG) and (+)-L-arabinose were added to the final concentrations of 1 mM and 13.3 mM, respectively. Sesquiterpene production was analyzed as mentioned above.

**Site-Directed Mutagenesis of tHMGR by Overlap PCR**

Site-directed mutagenesis for tHMGR was performed using overlap PCR (Table S1 for the primer sequences used herein). DNA fragments encoding the N and C termini of the mutation were amplified by PCR: 98°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec, repeated for 30 times. The reaction mixture contained 1× Phusion buffer, 2 mM dNTP, 0.5 μM forward and reverse primers, 2.5 U Phusion DNA polymerase (Finezyme), and 50 ng pBADMeVt as a template for tHMGR in a total volume of 100 μl. The amplified DNA fragment was gel purified using a gel purification kit (QIAGEN) or treated with DpnI and purified using a PCR purification kit (QIAGEN). These two amplified DNA fragments were spliced via overlap PCR: 98°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec, repeated for 30 times. The reaction mixture contained 1× Phusion buffer, 2 mM dNTP, 0.5 μM forward and reverse primers, 2.5 U Phusion DNA polymerase (Finezyme), and 50 ng of the above-mentioned DNA fragments as a template in a total volume of 100 μl. The fully amplified HMGR fragment was digested with SpeI/HindIII and inserted into the corresponding site of pBADMeVt.

**Site-Directed Mutagenesis of HUM by Overlap PCR**

Site-directed mutagenesis of HUM was performed using overlap PCR (Table S2 for the primer sequences used herein). DNA fragments encoding the N and C terminus of the mutation were amplified by PCR: 98°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec, repeated 30 times. The reaction mixture contained 1× Phusion buffer, 2 mM dNTP, 0.5 μM forward and reverse primers, 2.5 U Phusion DNA polymerase (Finezyme), and 50 ng pTrcHUM in 100 μl as a template for γ-humulene synthase. Amplified DNA was gel purified using a gel purification kit (QIAGEN) or treated with DpnI and purified using a PCR purification kit (QIAGEN). These two amplified DNA fragments were spliced via overlap PCR: 98°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec, repeated for 30 times. The reaction mixture contained 1× Phusion buffer, 2 mM dNTP, 0.5 μM forward and reverse primers, 2.5 U Phusion DNA polymerase (Finezyme), and 50 ng of the above-mentioned DNA fragments as a template in a total volume of 100 μl. The fully amplified HUM fragment was digested with NcoI/XbaI and cloned into the corresponding site in pTrcHUM.

**Quantification of In Vivo HUM Concentrations**

A single colony harboring pTrcSHUM15 (wild-type or its mutant variant), pBBRMBIS, and pBADMeVt was inoculated into LB medium containing Cb<sup>50</sup>, Km<sup>50</sup>, and Cm<sup>50</sup> and was grown overnight at 37°C. An aliquot of this seed culture was inoculated into fresh modified M9 medium (50 ml, see above formulation) containing Cb<sup>50</sup>, Km<sup>50</sup>, and Cm<sup>50</sup> to the final OD<sub>600</sub> of 0.05 and was grown at 37°C. Two hours after the inoculation, IPTG and (+)-L-arabinose was added to the final concentrations of 1 mM and 13.3 mM, respectively. The cultures were then grown at 20°C, 30°C, and 37°C. An aliquot of culture (1 ml) was taken and centrifuged at 14,000 × g. The resulting pellet was resuspended into Bugbuster containing recommended amount of Lysozyme (Novagen) to the final OD<sub>600</sub> of 20, and it was incubated for half an hour at room temperature. This lysis solution was centrifuged for 10 min at 14,000 × g. Total cell lysate (24 μl) (both soluble and insoluble fractions) and the soluble fraction only (24 μl) were mixed with 75 μl of 8 M guanidinium chloride and 1 μl of 4 M dithiothreitol. These solutions were incubated for another hour at room temperature. The concentration of HUM was determined by FRET Works S-tag assay kit following the recommended protocols (Novagen). In vivo sesquiterpene production from each culture was measured as described above.

**Protein Expression and Purification of HUM**

Wild-type HUM and its variants were cloned into pET29 and transformed into BL21 (DE3). Each transformant was inoculated into LB medium (5 ml) containing Km<sup>50</sup> and was grown overnight at 30°C. An aliquot (2 ml) of this seed culture was inoculated into fresh terrific broth (TB) medium containing Km<sup>50</sup> (500 ml), and the culture was grown at 30°C. When the culture reached OD<sub>600</sub> of 0.6–0.8, 0.1 mM of IPTG was added, and the culture was grown at 20°C for another 16 hr. Cells were harvested by centrifugation at 6,000 × g for 15 min. The pellet was suspended in 50 ml of BugBuster (Novagen) containing 20 U DNaseI and bacterial protease inhibitor cocktail II (Novagen) and was incubated for an hour at 4°C. The solution was centrifuged at 20,000 × g for 30 min and then was filtered through a 0.45-μm filter. S-tag™ Thrombin purification kit (Novagen) was used for the purification following the protocol

recommended by Novagen. All purifications were done at half scale. The eluted protein solution was dialyzed twice (PIERCE, MW 3,000 Da) against 1 l of buffer containing 10 mM Tes (pH 7.0), 10 mM MgCl<sub>2</sub>, 1 mM DTT, and 5% glycerol overnight. The protein concentration was measured using the Bradford method. We obtained approximately 3 ml of 25–500 μg/ml of protein solution with about 95% purity (confirmed by SDS-PAGE gel, data not shown).

**Enzyme Kinetics**

The kinetics studies of HUM and its variants were performed following a slightly modified protocol from that previously reported by Little and Croteau (2002). Kinetics for each enzyme was measured in a 40 μl reaction containing 0.15–0.4 μM enzyme, in buffer described in the previous section and overlaid with dodecane. The concentration of FPP varied from 0.229 to 58.6 μM with a fixed ratio of [<sup>3</sup>H]FPP. Seven to nine different concentrations of FPP were used for each enzyme (n = 3). The reaction mixture was incubated for 20 min at 31°C. To stop the reaction, 40 μl of a solution containing 4 M NaOH and 1 M EDTA was added and mixed. To extract sesquiterpene products, the reaction mixture was vortexed for 2 min, and 400 μl of dodecane was taken from the solution and mixed with 15 ml of scintillation fluid. Radioactivity was measured by scintillation counting. k<sub>cat</sub>, K<sub>m</sub> and k<sub>cat</sub>/K<sub>m</sub> were calculated using Enzyme Kinetics!Pro (ChemSW).

**SUPPLEMENTAL DATA**

Supplemental Data include two tables and one figure and can be found with this article online at <http://www.chembiol.com/cgi/content/full/15/6/607/DC1/>.

Received: November 9, 2007

Revised: April 29, 2008

Accepted: May 2, 2008

Published: June 20, 2008

**REFERENCES**

- Amin, N., Liu, A.D., Ramer, S., Aehle, W., Meijer, D., Metin, M., Wong, S., Gualfetti, P., and Schellenberger, V. (2004). Construction of stabilized proteins by combinatorial consensus mutagenesis. *Protein Eng. Des. Sel.* 17, 787–793.
- Daujotyte, D., Vilkaitis, G., Manelyte, L., Skalicky, J., Szyperski, T., and Klimasauskas, S. (2003). Solubility engineering of the HhaI methyltransferase. *Protein Eng.* 16, 295–301.
- Donald, K.A., Hampton, R.Y., and Fritz, I.B. (1997). Effects of overproduction of the catalytic domain of 3-hydroxy-3-methylglutaryl coenzyme A reductase on squalene synthesis in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 63, 3341–3344.
- Endo, T., Fedorov, A., de Souza, S.J., and Gilbert, W. (2002). Do introns favor or avoid regions of amino acid conservation? *Mol. Biol. Evol.* 19, 521–522.
- Graur, D. (1985). Amino acid composition and the evolutionary rates of protein-coding genes. *J. Mol. Evol.* 22, 53–62.
- Kraut, D.A., Carroll, K.S., and Herschlag, D. (2003). Challenges in enzyme mechanism and energetics. *Annu. Rev. Biochem.* 72, 517–571.
- Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S.F., Pasamontes, L., van Loon, A.P., and Wyss, M. (2002). The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.* 15, 403–411.
- Little, D.B., and Croteau, R.B. (2002). Alteration of product formation by directed mutagenesis and truncation of the multiple-product sesquiterpene synthases delta-selinene synthase and gamma-humulene synthase. *Arch. Biochem. Biophys.* 402, 120–135.
- Martin, V.J., Yoshikuni, Y., and Keasling, J.D. (2001). The in vivo synthesis of plant sesquiterpenes by *Escherichia coli*. *Biotechnol. Bioeng.* 75, 497–503.
- Martin, V.J., Pitera, D.J., Withers, S.T., Newman, J.D., and Keasling, J.D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* 21, 796–802.

- Nasreen, A., Vogt, M., Kim, H.J., Eichinger, A., and Skerra, A. (2006). Solubility engineering and crystallization of human apolipoprotein D. *Protein Sci.* *15*, 190–199.
- Pal, C., Papp, B., and Lercher, M.J. (2005). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat. Genet.* *37*, 1372–1375.
- Pal, C., Papp, B., and Lercher, M.J. (2006). An integrated view of protein evolution. *Nat. Rev. Genet.* *7*, 337–348.
- Pitera, D.J., Paddon, C.J., Newman, J.D., and Keasling, J.D. (2007). Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metab. Eng.* *9*, 193–207.
- Reiling, K.K., Yoshikuni, Y., Martin, V.J., Newman, J., Bohlmann, J., and Keasling, J.D. (2004). Mono and diterpene production in *Escherichia coli*. *Biotechnol. Bioeng.* *87*, 200–212.
- Roodveldt, C., Aharoni, A., and Tawfik, D.S. (2005). Directed evolution of proteins for heterologous expression and stability. *Curr. Opin. Struct. Biol.* *15*, 50–56.
- Schmidt, S., Sunyaev, S., Bork, P., and Dandekar, T. (2003). Metabolites: a helping hand for pathway evolution? *Trends Biochem. Sci.* *28*, 336–341.
- Steele, C.L., Crock, J., Bohlmann, J., and Croteau, R. (1998). Sesquiterpene synthases from grand fir (*Abies grandis*): comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of delta-selinene synthase and gamma-humulene synthase. *J. Biol. Chem.* *273*, 2078–2089.
- Steipe, B., Schiller, B., Pluckthun, A., and Steinbacher, S. (1994). Sequence statistics reliably predict stabilizing mutations in a protein domain. *J. Mol. Biol.* *240*, 188–192.
- Yoshikuni, Y., and Keasling, J.D. (2007). Pathway engineering by designed divergent evolution. *Curr. Opin. Chem. Biol.* *11*, 233–239.
- Yoshikuni, Y., Ferrin, T.E., and Keasling, J.D. (2006a). Designed divergent evolution of enzyme function. *Nature* *440*, 1078–1082.
- Yoshikuni, Y., Martin, V.J., Ferrin, T.E., and Keasling, J.D. (2006b). Engineering cotton (+)-delta-cadinene synthase to an altered function: germacrene D-4-ol synthase. *Chem. Biol.* *13*, 91–98.