Journal of Biotechnology 157 (2012) 237–245

Contents lists available at SciVerse ScienceDirect

Journal of Biotechnology

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

Directed evolution to re-adapt a co-evolved network within an enzyme

John Strafford[a], Panwajee Payongsri[a], Edward G. Hibbert[a], Phattaraporn Morris[a], Sukhjeet S. Batth[a], David Steadman[b], Mark E.B. Smith[b], John M. Ward[c], Helen C. Hailes[b], Paul A. Dalby[a,*]

[a] Advanced Centre for Biochemical Engineering, Department of Biochemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK
[b] Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK
[c] Department of Structural and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK

ARTICLE INFO

Article history:
Received 24 August 2011
Received in revised form
19 November 2011
Accepted 22 November 2011
Available online 29 November 2011

Keywords:
Directed evolution
Coevolution
Transketolase
Enzyme engineering
Biocatalysis

ABSTRACT

We have previously used targeted active-site saturation mutagenesis to identify a number of transketolase single mutants that improved activity towards either glycolaldehyde (GA), or the non-natural substrate propionaldehyde (PA). Here, all attempts to recombine the singles into double mutants led to unexpected losses of specific activity towards both substrates. A typical trade-off occurred between soluble expression levels and specific activity for all single mutants, but many double mutants decreased both properties more severely suggesting a critical loss of protein stability or native folding. Statistical coupling analysis (SCA) of a large multiple sequence alignment revealed a network of nine co-evolved residues that affected all but one double mutant. Such networks maintain important functional properties such as activity, specificity, folding, stability, and solubility and may be rapidly disrupted by introducing one or more non-naturally occurring mutations. To identify variants of this network that would accept and improve upon our best D469 mutants for activity towards PA, we created a library of random single, double and triple mutants across seven of the co-evolved residues, combining our D469 variants with only naturally occurring mutations at the remaining sites. A triple mutant cluster at D469, E498 and R520 was found to behave synergistically for the specific activity towards PA. Protein expression was severely reduced by E498D and improved by R520Q, yet variants containing both mutations led to improved specific activity and enzyme expression, but with loss of solubility and the formation of inclusion bodies. D469S and R520Q combined synergistically to improve k_{cat} 20-fold for PA, more than for any previous transketolase mutant. R520Q also doubled the specific activity of the previously identified D469T to create our most active transketolase mutant to date. Our results show that recombining active-site mutants obtained by saturation mutagenesis can rapidly destabilise critical networks of co-evolved residues, whereas beneficial single mutants can be retained and improved upon by randomly recombining them with natural variants at other positions in the network.

1. Introduction

Targeted saturation mutagenesis of enzyme active-site residues can be used as an efficient alternative to error-prone PCR or DNA shuffling to engineer activity, specificity and enantioselectivity (Dalby, 2003; Park et al., 2005; Reetz et al., 2005; Paramesvaran et al., 2009; Bougioukou et al., 2009) as it targets the residues that most strongly influence catalysis and substrate binding. However, random point mutations selected only for activity often lead to a trade-off with protein stability (Beadle and Shoichet, 2002; Tokuriki et al., 2008) and their accumulation eventually leads to difficulties in introducing further mutations that improve function (Wang et al., 2002; Bloom et al., 2004, 2005).

The effects of random point mutations on protein stability or poor folding can be ameliorated by the addition of chaperones to allow the introduction of further activity enhancing mutations (Tokuriki and Tawfik, 2009). DNA shuffling can also be used to avoid the accumulation of destabilising mutations, as it tends to generate a greater proportion of functional variants with less trade off between activity and stability (Drummond et al., 2005). Alternatively, random mutagenesis to improve protein stability can be used to confer further evolvability for function (Bloom et al., 2006; Brown et al., 2010). For example, thermostability can be improved by consensus protein design approaches (Lehmann et al., 2002; Jackel et al., 2010), and also by the B-FIT method which targets saturation mutagenesis to residues with high crystallographic B-factors (Jochens et al., 2010; Reetz et al., 2010).

Abbreviations: GA, glycolaldehyde; HPA, hydroxypyruvate; PA, propionaldehyde; PP, pyrophosphate-binding domain; Pyr, pyrimidine-binding domain; SCA, statistical coupling analysis; TK, transketolase; TPP, thiamine pyrophosphate.

* Corresponding author. Tel.: +44 20 7679 2962.
E-mail address: p.dalby@ucl.ac.uk (P.A. Dalby).

0168-1656/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2011.11.017

Some residues form co-evolved networks of synergistically coupled residues that can be identified by statistical coupling (or correlated mutation) analyses (Lockless and Ranganathan, 1999; Kuipers et al., 2009). Residues in such networks can be much more critical to protein stability and folding than the majority (Socolich et al., 2005), or confer other functional roles (Gobel et al., 1994; Halabi et al., 2009), including allostery (Suel et al., 2003), mediation of conformational changes (Nascimento et al., 2008), catalysis (Estabrook et al., 2005) and substrate specificity (Kuipers et al., 2009).

Correlated mutation analyses of enzyme families can be used to guide directed evolution to residues that influence substrate specificity (Kuipers et al., 2009). Conversely, networks of structurally coupled residues that improve thermostability have even been created by iterative saturation mutagenesis (Reetz et al., 2009), and synergy has also been obtained by analysing the results of the first rounds of random mutagenesis and recombining those predicted to be mutually beneficial (Fox et al., 2007). However, inadvertently disrupting such networks may lead to significant loss of stability or function and therefore represent a potential barrier to recombining and accumulating beneficial mutations during directed evolution, as we have found here for the transketolase enzyme.

Transketolase (TK) catalyses the enantioselective synthesis of α,α′-dihydroxy ketones from donors D-xylulose-5-phosphate or β-hydroxypyruvate (HPA), and acceptors D-ribose-5-phosphate, D-erythrose-4-phosphate or other aldehydes (Sprenger et al., 1995). HPA renders the reaction irreversible making it useful for industrial biocatalytic syntheses (Demuynck et al., 1991; Hobbs et al., 1993; Morris et al., 1996; Schenk et al., 1998; Turner, 2000; Shaeri et al., 2006; Ingram et al., 2007; Smith et al., 2010). Saturation mutagenesis was previously targeted independently to 10 active-site residues in contact with cofactor or substrate, and also to the 10 least conserved second shell residues (Hibbert et al., 2007), guided by TK structures (Littlechild et al., 1995; Nilsson et al., 1997), and an analysis of thiamine pyrophosphate (TPP)-dependent enzyme phylogeny (Costelloe et al., 2008). Various single mutants gave improved activity towards glycolaldehyde (GA) (Hibbert et al., 2007), or accepted propionaldehyde (PA) better with either enhanced or reversed enantioselectivity (Hibbert et al., 2008; Smith et al., 2008; Hailes et al., 2010). Several D469 mutations also accepted a range of linear aliphatic (Cazares et al., 2010) and aromatic (Galman et al., 2010) aldehydes. Chemical denaturation, dimer interface mutations and biocatalytic process conditions that destabilise *Escherichia coli* TK have also been well characterised (Brocklebank et al., 1999; Martinez-Torres et al., 2007; Aucamp et al., 2008).

Here, we show that the creation of double mutants from our previous activity-enhancing single mutants within the active site of *E. coli* transketolase, leads to loss of enzyme function. Activities and soluble expression levels of protein trade-off smoothly in the single mutants, whereas many of the double mutants disrupt both properties more extensively, suggesting that they fall below a critical threshold in stability or folding that then manifests as a loss of soluble protein expression and/or inclusion body formation, as observed previously (Calloni et al., 2005). Statistical coupling analysis (SCA) (Gobel et al., 1994; Lockless and Ranganathan, 1999; Suel et al., 2003) of a multiple sequence alignment for 382 related TPP-dependent enzymes revealed that 10 of the 11 active-site double mutants alter at least one of nine residues that form a single co-evolved network.

To test whether naturally occurring variants within this network could improve activity towards PA, or enhance our existing non-naturally occurring D469 mutants, we created a small library of single, double and triple mutants spanning seven of the co-evolved residues. Only naturally occurring mutations were allowed at each site, except D469 for which only variants previously found to improve activity towards PA were used. The use of only natural variants in so-called smart libraries has been shown recently to be an efficient way of improving enzyme activity and enantioselectivity (Jochens and Bornscheuer, 2010). A triple mutant cluster at D469, E498 and R520 was found to behave synergistically and a double mutant D469S/R520Q was obtained with significantly enhanced k_{cat}, useful for biocatalytic applications. The R520Q mutation was also found to double the specific activity of our previously identified mutant D469T to create our most active TK variant to date.

2. Materials and methods

All chemicals were obtained from Sigma and used as supplied, except β-hydroxypyruvate (HPA) which was prepared as the lithium salt by modification of a previously described procedure (Morris et al., 1996).

2.1. Statistical coupling analysis of TPP-dependent enzyme protein sequences

The alignment of 382 sequences of the homologous PP and Pyr domains from 17 TPP-dependent enzyme types including transketolase, was described previously (Costelloe et al., 2008). The statistical coupling analysis (SCA) was carried out as described previously using the SCA Matlab toolkit (Version 1.5) with default parameters (Suel et al., 2003). Perturbations were allowed that resulted in sub-alignments greater than 0.21 as a fraction of the total alignment size. Following clustering, the statistical-coupling matrix was iteratively focused on areas of high signal and re-clustered. Further details are given in Supplemental Experimental Procedures.

2.2. Double mutant and library construction

All defined mutants and libraries were introduced into the tktA gene in plasmid pQR791 (includes a His tag), using the Quikchange method as described previously (Hibbert et al., 2007), then transformed into XL10-gold cells (Stratagene, La Jolla, CA) for expression of the TK mutants. All defined mutants were confirmed by DNA sequencing. Defined double and triple mutants were constructed from existing single mutant plasmids as templates and then stepwise mutagenesis using primers with optimal mutagenic codons. Seven of the nine cluster residues identified by SCA above were subjected to partial random mutagenesis (Table 1). Pro493 and Pro486 were excluded. Based on the alignment of the 382 TPP-dependent enzyme sequences, the four most common naturally occurring residues at six of the seven sites, and the previously identified S, A, L and T variants of D469, were each introduced individually into the wild-type plasmid. All primer sequences are given in Table S1 (available online). The 467/495 double-mutant library was constructed using an equimolar mixture of the four G467 mutant plasmids as template and an equimolar mixture of the four D495 mutagenic primers. The converse reaction was also performed with the D495 plasmids and G467 mutagenic primers, and the products combined (1:1) prior to transformation. A similar strategy was used to create the triple mutant library of D469/E498/R520 and each of the double mutant libraries D469/E498, E498/R520 and D469/R520. Colonies from each library were picked and cultured in individual wells of 96-deep-square-well plates, then divided into 96-well reaction plates as described previously (Hibbert et al., 2008). The total library contained 28 confirmed single mutants in triplicate, 90 random colonies from each of the four double-mutant libraries, and 540 random triple-mutant colonies. Each plate contained three wild-type colonies and three blank wells for internal reference.