

## METHODS

# Avoiding and controlling double transformation artifacts

Moshe Goldsmith<sup>1</sup>, Csaba Kiss<sup>2</sup>, Andrew R.M. Bradbury<sup>2</sup> and Dan S. Tawfik<sup>1,3</sup>

<sup>1</sup>Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel and <sup>2</sup>B division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

<sup>3</sup>To whom correspondence should be addressed. E-mail: tawfik@weizmann.ac.il

**This article describes a set of standard control experiments for the authentication of new protein variants isolated through library selection and site-directed mutagenesis. These controls are specifically designed to rule out artifacts derived from ‘double transformants’—i.e. cells transformed with, or infected by, two different plasmids simultaneously. These seem to have been the source of past artifacts and, as demonstrated here, are far more common than generally recognized. By following standard protocols for cloning, plasmid isolation, subcloning, in combination with functional assays, the presence of such artifacts can be ruled out. This protocol needs to be applied for any new variant isolated from heterogeneous gene repertoires, and in particular for variants isolated by selection for either enzymatic activity, or binding.**

**Keywords:** artifacts/directed evolution/double transformations/library selection/plasmid DNA

*Comments, corrections, or additions to this paper are welcome on the peds blog, PEDS methods (www.oxfordjournals.org/PEDSmethods).*

Double transformations, in which a single cell (typically an *Escherichia coli* bacterium) is transformed simultaneously with two different plasmids, are deemed by most researchers to be very low frequency events, in particular when both plasmids carry the same antibiotic resistance gene, and/or origin of replication, or when antibiotic selection is applied for one of the plasmids only. Nevertheless, double transformants were reported to have yielded artifactual phenotypes, in particular with regard to novel variants isolated by screening heterogeneous gene libraries, or from libraries that were subjected to a functional selection (Altamirano *et al.*, 2002; Zeytun *et al.*, 2004). Consider, for example, a selected library clone carrying a mixture of two plasmids: One plasmid encodes a library variant conferring no function, and the other, a natural ‘wild type’ protein with the target function. The latter would lead to the selection of this clone by virtue of exhibiting the desired trait (e.g. ligand binding, enzymatic activity, fluorescence, etc.); whereas the former may provide the signal associated with a library hit (e.g. PCR amplification with the relevant primers and a sequence corresponding to a library variant).

The isolation of such a clone is obviously dependent on having a plasmid bearing a gene encoding a natural protein

with the desired function. However, ‘wild type’ proteins are often used as a positive control, or for setting up the selection system. In fact, laboratories that are routinely engaged in screening large repertoires and directed evolution often encounter ‘wild type’ contaminations that take over their libraries. Tracing a wild type cross-contamination may not be straight forward, in particular when one is not aware of the possibility of double transformants. Moreover, a common measure against such cross-contaminations is the use of a ‘wild type’ protein encoded by a plasmid that differs from the one used for library construction, so that the former would not be PCR amplified and carried over to the first round of selection, nor from one round to another. In such cases, the cryptic presence of the wild type plasmid does not interfere with, and might not even be detected by, colony PCR amplification or sequencing of the library variant present in the same clone.

It is generally assumed that, double transformants are extremely rare, thus making the above scenario highly unlikely, if not impossible. However, we have come to the realization that double transformants can be frequent, in particular when transformations are performed with high amounts of plasmid DNA. The latter is the routine in library selections, where one aims at the highest numbers of transformants possible. Conceivable mechanisms leading to double transformations can be either: the presence of a subpopulation of competent cells which is readily transformed by several plasmids at a time (‘hypercompetent’ cells), cellular penetration of several plasmids during plasmid electroporation due to a high local concentration of plasmid DNA, or the transformation of plasmid concatamers (Viret *et al.* 1991). Similarly, infection of bacteria by filamentous phage(mid) particles can lead to bacteria hosting multiple different plasmids. In fact the ability to introduce multiple phagemids into bacteria has been used as a method to increase diversity (Sblattero and Bradbury, 2000).

To assess the frequency of double transformants, we transformed a plasmid mixture into electro-competent *E.coli* cells (Table I). The two plasmids were of similar size (3365 bp for pSKSL and 2868 bp for pUC19) and contained the same origin of replication. While the ampicillin (Amp) resistance plasmid (pUC19) simulated the library clones and was applied in large excess, the kanamycin (Kan) resistance plasmid (pSKSL) simulated the wild type cross-contamination. The mixture contained a 1:1000 mass ratio of pSKSL(Kan):pUC19(Amp) plasmids, which translates into a molecular ratio of 1:1250. We measured the number of surviving colonies on different antibiotic selection plates and thereby quantified the frequency of single and double transformants.

As the amount of transformed DNA increased, so did the frequency of double transformants (Fig. 1). Double transformants could still be selected when a 5 ng plasmid mix was transformed despite the low transformation efficiency ( $1E + 1*10^4$  cfu/ng). Notably, at high DNA levels ( $>50$  ng

**Table I.** Transformations of plasmid mixtures

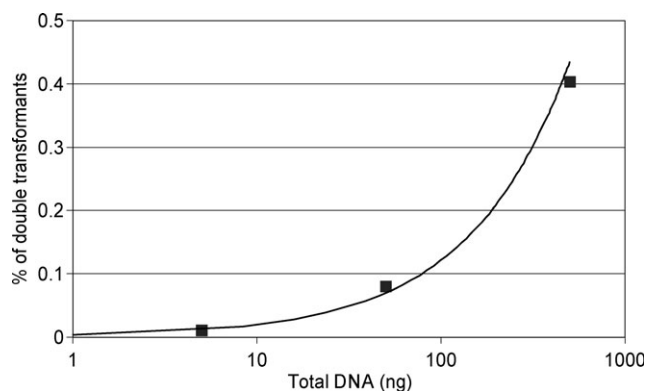
Transformed DNA		No. of cfu's		
ng	No. of plasmid molecules	Amp selection	Kan selection	Amp + Kan selection
500	$1.7 \times 10^{11}$	313 500	1284	1267
50	$1.7 \times 10^{10}$	65 350	114	53
5	$1.7 \times 10^9$	47 500	37	5
0.5	$1.7 \times 10^8$	8500	6	0
0.05	$1.7 \times 10^7$	1852	0	0

A mixture of pUC19 plasmid, conferring Amp resistance and pSKSL plasmid (Tomer and Livneh, 1999), conferring Kan resistance, at a ratio of 1250:1, respectively, was electroporated to electrocompetent JM109 *E. coli* cells ( $3.1 \times 10^7$  cells per transformation). The resulting transformation was incubated in 1 ml LB medium for 1 h at 37°C, and serially diluted upto 100 000-fold. The different dilutions were then plated on LB agar plates containing either Kan (50 µg/ml), Amp (100 µg/ml), or both. Plates were incubated for 14 h at 37°C and colonies counted.

per transformation), all the clones carrying the contaminating plasmid pSKSL(Kan) also carried the plasmid added in excess pUC19(Amp). Clearly, if the contaminating plasmid confers a selective advantage, it may take-over the transformed population during the selection even when present at extremely low numbers. It is also likely to appear as a 'double transformant' carrying the contaminating plasmid along with a library variant.

It should be noted that the experiment described above might in fact underestimate the frequency of double transformants. Using plasmids with different origins of replication, size and copy number, we have observed double transformants frequencies of up to 10% of the total number of transformants (data not shown). In addition, the results of Velappan *et al.* (2007) show that double or even triple transformant contamination can be remarkably persistent: over five overnight growths with high copy number origins.

Realizing that double transformants can be frequent (which might, in fact, come as no news to many) has concerned us, and promoted the introduction of a set of standard control experiments that is applied routinely following the isolation of new variants from library selections or screens. We would like to share this protocol with the community of protein engineers through this article, and have it reviewed, updated and endorsed by other groups, with the aim of formulating a



**Fig. 1.** The percent of double transformants observed with different amounts of transformed plasmid DNA.

set of protocols, that would become a standard practice in the field. We hope that this would be useful for those who are new to this area, and may minimize the waste of time, energy and spirit that accompanies the isolation of artifacts.

The protocols below are designed specifically for analysis of the 'final' clones, which are the end products of the screening or selection process. However, some of them (e.g. Colony PCR) can also be used for intermediate selection steps (e.g. clones to be used for further DNA shuffling or mutation), to avoid contamination of the next round with clones that have no advantageous, and possibly harmful mutations. We have noted that the presence of such clones in the DNA shuffling of fluorescent proteins leads to a failure to improve their selected properties, indicating that care in the selection of clones for further diversification is well rewarded in more rapid evolution.

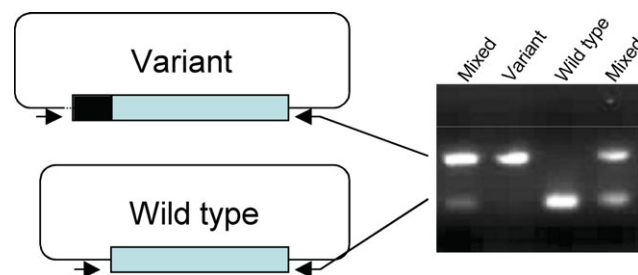
## The proposed standard control experiments

### Step 1: Clone Isolation

Following the identification of a 'positive' clone by functional screening or selection (be it ligand binding, formation of an enzymatic product, fluorescence, etc.), the positive clone is isolated and used for culture inoculation. The culture is then streaked on an agar plate, and a few colonies are removed and grown for further analysis. This step is important because artifacts of the kind discussed above can either arise from a single colony of a single type of cells containing two plasmids, or, from a colony containing two types of cells, each with a different plasmid. The latter is commonly observed when cells are plated at high density.

### Step 2: Colony PCR

The selected phenotype is then confirmed through a functional assay, and the presence of a library variant verified through 'colony PCR' on a few isolated colonies using primers that anneal specifically to the library construct. The size of the amplified DNA should be in accordance with a library insert. This step can be made more efficient if the primers used to amplify the selected library variant, amplify a fragment of different size than that of the wild type control. This can be engineered by inserting an additional coding (e.g. a peptide tag, such as myc or SV5), or non-coding sequence, in either the wild type control or in the library genes. As exemplified in Fig. 2, amplification of wild type sequences can then be clearly distinguished from



**Fig. 2.** A method to easily identify the contamination of evolved variant clones with wild type clones. By including an additional sequence (black box) between the primer annealing site and the coding sequence during creation of the variant library, a difference in length after PCR easily distinguishes wild type sequences from evolved sequences (right panel).

mutated sequences. Still, it is worth noting that a minor wild type contamination may be invisible at this stage due to amplification biases and therefore additional steps, such as described below, must be taken to ensure clonal purity.

### Step 3: Isolation of plasmid DNA

A single colony isolated through step 1 and verified by step 2 is used to inoculate a liquid culture from which plasmid DNA is subsequently isolated. Loading a sample of the plasmid DNA on an agarose gel can help assess its purity at this stage. Note that, the utilization of the original colony (from either Step 1 or 2) as is (i.e. by inoculation of new cultures) for sequencing, or overexpression and purification the protein, is a major source of artifacts and irreproducibility.

### Step 4: Sub cloning

The gene of interest is excised from the isolated plasmid, purified and re-ligated to a fresh vector. This step is important to eliminate the possibility that an unexpected mutation in the vector is responsible for the observed phenotype, or to resolve different library clones that had been linked together during concatemer formation in the cell (Viret *et al.*, 1991). The ligation product is then freshly transformed using very low DNA amounts (<1 ng; Table I), or a low-efficient transformation method, in order to reduce the likelihood of double transformants. The transformed bacteria are plated under antibiotic selection, but with no other selection pressure, and in particular by avoidance of the functional selection applied for the isolation of this clone in the first instance. This is particularly important where the expressed protein has a negative effect on bacterial growth, and can be selected against.

### Step 5: Reproducing the phenotype in unselected subclones

Several colonies derived from this fresh transformation (Step 4) are picked at random, and assayed for the selected function. They should all exhibit the target trait and also appear positive by 'colony PCR' (step 2). In case of mixed results, the above procedure (steps 1–5) can be repeated starting from one, or more, colonies that show the desired phenotype and appear positive by PCR. In any case, the resulting subclones must all meet the above criteria.

### Step 6: Controlling for double infections

An oft-used selection method involves the fusion of the library variants to a coat protein and their subsequent display on phage. In this case, multiple infection events can be avoided by ensuring that the multiplicity of infection after selection is less than one. This is not usually a problem in early selection rounds as much as in later rounds, in which significant amplification of phage titers has occurred. When carrying out phage titration, it is important to titrate the phage, and add to an excess of bacteria, not infect bacteria and titrate the infected bacteria.

### Step 7: Functional analysis of the selected protein

The above set of controls are aimed at establishing the link between genotype and phenotype of the selected clones. However, further verification should be provided through detailed analysis of the selected protein variant. At this level, general protocols would be largely useless since no two proteins are alike. The crux of this analysis should be to demonstrate unique features of the selected protein that differ from

its wild type equivalent. These experiments should be designed specifically to rule out the possibility that a minor contamination of the wild type protein is giving rise to the observed function.

The fact that most selected as well as designed proteins show low activity, often, orders of magnitudes lower than their natural counterparts, complicates such analyses. Consider an evolved/selected enzyme with a catalytic efficiency that is 1000-fold lower than its natural counterpart. Indications of protein purity (gels etc.) are useless as measures to eliminate the possibility of wild type contamination. So are the measurements of  $k_{cat}$  or  $K_d$  values—both are protein concentration dependent: a 0.1% contamination of wild type protein would simply give rise to a 1000-fold lower  $k_{cat}$  or  $K_d$  than the wild type. However, demonstrating a different value of  $K_M$  (or  $k_{off}$  for a binding protein), and especially, a different pattern of substrate selectivity, or ligand binding specificity, can be more indicative. Other properties may apply to individual proteins, e.g. interaction with inhibitors, or pH and temperature dependency of the new variant that differ from its natural counterpart.

### Step 8: Sequencing of final clones

It is essential to sequence all final clones, and note carefully any sequences that appear to be mixed (i.e. with two overlapping peaks, corresponding to two different base, at the same position), especially if these are found in the regions diversified. This is sometimes the only sign that a clone is not monoclonal, but a mixture. Sequencing the clones in both directions can also help identify previously unseen contaminations.

### Relevance to site-directed mutagenesis

Most, if not all, approaches to site-directed mutagenesis do not completely eliminate the wild type template used for the production of mutants. Under the circumstances described above, double transformants carrying the wild type template as well as a desired mutant could be obtained and contaminate subsequent studies. Nevertheless, we believe that the application of steps 1–5 is a recommended routine, in the case of mutants generated for the purpose of structure-function analyses, or protein engineering and design.

To conclude, we hope that the procedures described herein will be considered by many researchers in the field of protein engineering and may serve as standard practice to guide newcomers to the field. We seek responses, and potential additions and alterations to this procedure, and submission of protocols describing other standard practices in protein engineering.

### Acknowledgements

Financial support by the Weizmann-Pasteur Institute to DST, and Department of Energy GTL funds to AB, are gratefully acknowledged.

### References

- Altamirano, M.M., Blackburn, J.M., Aguayo, C. and Fersht, A.R. (2002) *Nature*, **417**, 468.
- Sblattero, D. and Bradbury, A. (2000) *Nat. Biotechnol.*, **18**, 75–80.
- Tomer, G. and Livneh, Z. (1999) *Biochemistry*, **38**, 5948–5958.
- Viret, J.F., Bravo, A. and Alonso, J.C. (1991) *Microbiol. Rev.*, **55**, 675–683.

**M.Goldsmith et al.**

Velappan,N., Sblattero,D., Chasteen,L., Pavlik,P. and Bradbury,A.R. (2007)  
*Protein Eng. Des. Sel.*, Mar 1; [Epub ahead of print doi:10.1093/peds/  
gzm005].

Zeytun,A., Jeromin,A., Scalettar,B.A., Waldo,G.S. and Bradbury,A.R.  
(2004) *Nat. Biotechnol.*, **22**, 601.

**Received May 13, 2007; revised May 13, 2007;  
accepted May 14, 2007**

**Edited by Arne Skerra**