Speeding up Directed Evolution: Combining the Advantages of Solid-Phase Combinatorial Gene Synthesis with Statistically Guided Reduction of Screening Effort

Sabrina Hoebenreich[#,†,‡], Felipe E. Zilly[†,#], Carlos G. Acevedo-Rocha[†,‡] Matías Zilly[§], and Manfred T. Reetz[*,†,‡]

† Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany
‡ Fachbereich Chemie, Philipps-Universität Marburg, Hans-Meerwein-Straße, 35032 Marburg, Germany
§ Fakultät für Physik, Universität Duisburg-Essen, Lotharstraße 1, 47048 Duisburg, Germany
[#]authors contributed equally

CONTENT:

PROCEDURE TO ESTIMATE PARENTAL CONTENT FROM A SEQUENCING CHROMATOGRAM

Intensities of the chromophores attached to the chain-terminators can be compared at each base position (personal communication with GATC, Germany). Peak high or trace values of the four different chromophores representing adenine (A), cytosine (C), guanidine (G) and thymine (T) are extracted e.g. using BioEdit (export trace values) or read out with Vector NTI (simple move the mouse over the peaks) or "Chromas Lite" for windows users. In the case of Mac users, it is possible to use the program "4 peaks". Values are presented in pie diagrams using excel (Figure S1) and compared to the expected values, which are determined by the chosen randomization scheme, i.e., NDC, NNK etc.

The process to qualitatively estimate parental contamination is explained by taking the examples of positions randomized using NDC (Figure S1). Most straight forward is the content of parental contamination assessed for such cases with distinct yes-no answers. For example when only one base is expected, like the C in the third position of NDC, but more are observed. More difficult and less precise in our experience are the estimations based on deviations from the expected ratio, since some bias is expected due to the different size in colonies resulting in non-equal contributions of single sequences to the plasmid pool. The qualitative assessment of the Sloning libraries concerning degeneracy quality was therefore less straight forward and the amount of parental sequences could not be derived from the obtained data.

In the case of A-PCR residues R47, T49 and Y51 showed around 85% cytosine (C) content in position 3 of the target codon, whereas we expected to see 100%. In addition, the first position shows a bias towards thymine (T). The second base position of residue T49 exhibit 10% C. This is not encoded in the NDC design but comes from the parental codon (ACG). Taking all these observations together parental contamination of the PCR-A sample is estimated to be 10-15%. In the case of B-PCR, we expected to see 50% of G and T, but we observed 20-25% A and a little bit of C in the third position of the codon, therefore the yield was 75-80% for this library.

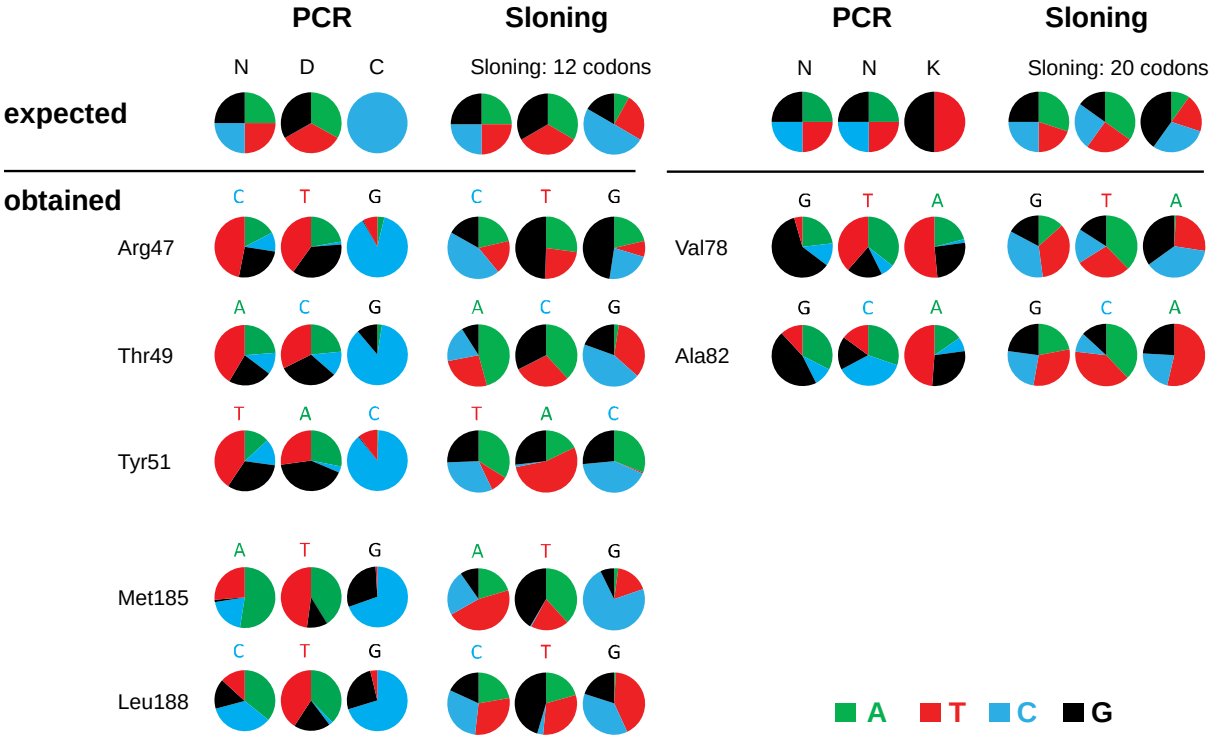

Figure S1. Quick Quality Control (QQC) of combinatorial libraries. Transformed *E. coli* cells were streaked out onto LB agar, followed by incubation at 37°C over night. The next day, all cells were harvested and the pool of plasmid DNA was extracted and sequenced in a single sequencing run to analyze diversity at the DNA level and to estimate

the amount of parental construct originating from either PCR and/or insufficient *DpnI* digestion. For each library, at least 3000 colonies were obtained, a representative sample size. Parental codons are given above the pie charts.

MATHEMATICAL DERIVATION OF THE CASTER 2.0 WILD TYPE/PARENTAL BACKGROUND CALCULATION

First, we consider the case that no parental sequence is present in the libraries. Let the library be of size m, i.e., there are m variants, each with equal frequency contained in the library. Now if, randomly, a set of n colonies (transformants) are generated from the library wherein each colony comprises one variant, then the average number x of variants comprised in the colonies is

$$x(n) = m\left[1 - \left(1 - \tfrac{1}{m}\right)^n\right] \qquad \textbf{(eqn. 1)}$$

We do the proof by induction. Obviously, if we pick a single random colony from the library, then this colony comprises exactly 1 variant, i.e.

$$x(1) = 1 = m\left[1 - \left(1 - \frac{1}{m}\right)^1\right]$$

Thus the base case is true. Now we have to prove the inductive step, that, given the average number $x(n)$ of variants in a set of n colonies follows the above formula, also the average number $x(n+1)$ of variants in a set of $n+1$ colonies follows the same formula. Therefore we analyze $x(n+1)$. If the average number of variants in a set of n colonies is $x(n)$, then the probability, that the $n+1^{\text{st}}$ colony comprises one of the variants which are already contained in the other n colonies, is $\frac{x(n)}{m}$. On the other hand, the probability that in the $n+1^{\text{st}}$ colony we find a variant which is not contained in the other n colonies, is $1 - \frac{x(n)}{m}$. Thus, the average number variants contained in the set of $n+1$ colonies is

$$x(n+1) = \frac{x(n)}{m}x(n) + \left(1 - \frac{x(n)}{m}\right)(x(n)+1)$$

$$= \left[1 - \left(1 - \frac{1}{m}\right)^n\right] m\left[1 - \left(1 - \frac{1}{m}\right)^n\right] + \left(\frac{m\left(1 - \left[1 - \left(1 - \frac{1}{m}\right)^n\right]\right)}{m}\right)\left(m\left[1 - \left(1 - \frac{1}{m}\right)^n\right] + 1\right)$$

$$= m\left[1 - \left(1 - \frac{1}{m}\right)^n\right] + 1 - \left[1 - \left(1 - \frac{1}{m}\right)^n\right]$$

$$= m\left[1 - \left(1 - \frac{1}{m}\right)^{n+1}\right]$$

This completes the proof of the first formula. So, if we desire our colonies to contain a fraction $f = \frac{x}{m}$ of all library variants, then we have to choose at least

$$n = \frac{\log(1-f)}{\log\left(1 - \frac{1}{m}\right)}$$

colonies. This is the formula published in CASTER 1.0.[1]

As the second step we consider the case that the library contains certain fraction of undesired parental sequence as a result of the process of library generation. Let the probability that a colony chosen at random contains the parental sequence or "wild type" (WT) be w. For simplicity, we assume that the sequence of w is not part of the desired m variants of the library. Again, we analyze the average number $y(n)$ of variants in a set of n colonies. This number is the sum of the probabilities that out of the n colonies r comprise the wild types and $n-r$ comprise variants times the average number $x(n-r)$ of different variants in the $n-r$ colonies, *i.e.*:

$$y(n) = \sum_{r=0}^{n} \binom{n}{r} w^r (1-w)^{n-r}\, x(n-r),$$

where, for completeness $x(0) = 0$. Inserting **eqn. 1** and observing the binomial rule $(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r}$ we find

$$y(n) = m - m\left(1 - \frac{1-w}{m}\right)^n.$$

Again, if we desire our colonies to contain a fraction $f = \frac{y}{m}$ of all library variants, then we have to choose at least

$$n = \frac{\log(1-f)}{\log\left(1 - \frac{1-w}{m}\right)}$$

colonies. This is the formula contained in the new version of CASTER 2.0. We wish to emphasize that this estimate holds for the case that each of the variants occurs equally frequent in the library. If this is not the case (see examples in the main part of the text), this estimate is too optimistic.

LIBRARY SETUP AND OVERALL SIZE

When aiming for 90-95% sequence space coverage, but considering the yield of degeneracy from the QQC in cases of the PCR libraries (see main text), and assuming (at the onset of the project) a 100% yield for the Sloning libraries, the amount of picked colonies was as follows:

Library **A-PCR** (3xNDC) consisted of 54 plates, where wells A1 contains *E. coli* BL21-Gold(DE3) containing pETM11-P450-BM3 WT and B1 the same strain but harboring pETM11-P450-F87A. This setup results in a total number of 94x54 = **5076** screened transformants. The yield of the randomization was estimated to be 85% by the QQC.

Library **B-PCR** (2xNNK) consisted of 32 plates, where well A1 and B1 harbors BM3 WT and mutant F87A, respectively. The total number of transformants screened was 94x32 = **3008**. The yield of the randomization was estimated to be around 80%, as judged by the QQC (Figure S1).

Library **C-PCR** (2xNDC) consisted of 8 plates, where well A1 and B1 harbors BM3 WT and mutant F87A, respectively. The total number of transformants screened is 94x8 = **752**. The yield of the randomization was estimated to be about 75% by the QQC (Figure S1).

Library **A-Sloning** (3x12 codons: 3x12 aa) consisted of 42 plates, where well A1 is empty, A2 contains *E. coli* BL21-Gold(DE3) transformed with pETM11, A3 has the same strain containing BM3-WT and A4 has the same strain but containing as control mutant BM3-F87A. The total number of transformants screened is 92x42 = **3864**.

Library **B-Sloning** (2x20 codons: 2x20 aa) consisted of 13 plates, where well A1 and B1 harbors BM3 WT and mutant F87A, respectively. The total number of transformants screened was 94x13 = **1222**.

Library **C-Sloning** (2x12 codons: 2x12 aa) consisted of 7 plates, where no controls were present. The total number of transformants screened was 96x7 = **672**.

In total, **14,594** transformants and 342 controls were screened by HPLC for this project. Some of the hits found in the PCR libraries are reported elsewhere.[2]

RESULTS OF THE RANDOM SAMPLE SEQUENCING

Table S1. Codon identity and occurrence frequency observed in the sequencing dataset of the random samples from both PCR and Sloning libraries A (R47/T49TY51) and C (M185/L188) as well as the sequencing results of Slonings internal quality control. PCR libraries were designed with NDC degeneracy and Sloning with 12 codon to 12 aa degeneracy at each position. Non-designed codons were observed. The different codon usage between Sloning and NDC degeneracy is highlighted in green.

.

			PCR Random Sample					Sloning Random Sample					Sloning Quality Control				
			R47	**T49**	**Y51**	**M185**	**L188**	**R47**	**T49**	**Y51**	**M185**	**L188**	**R47**	**T49**	**Y51**	**M185**	**L188**
	aa	**codon**	CGT	ACG	TAC	ATG	CTG	CGT	ACG	TAC	ATG	CTG	CGT	ACG	TAC	ATG	CTG
Allowed	Asn	AAC	4	5	7	9	20	8	9	11	5	5	3	3	5	6	4
Codons	Ser	AGC	4	0	2	5	5	5	7	12	5	5	3	3	9	2	5
	Ile	ATC	5	12	3	4	2	4	9	7	5	7	1	6	7	6	5
	His	CAC/CAT	1	2	3	11	8	5	10	4	5	10	2	5	1	8	6
	Arg	CGC/CGT	1	0	2	0	1	12	7	9	5	3	7	4	7	5	0
	Leu	CTC	0	3	7	3	2	6	12	3	9	13	3	7	3	3	4
	Asp	GAC	4	9	6	0	5	4	8	2	5	4	3	7	3	3	3
	Gly	GGC/GGA	4	3	8	1	0	5	7	7	15	2	4	2	4	4	5
	Val	GTC/GTG	9	4	8	4	1	9	2	4	5	3	6	1	3	1	1
	Tyr	TAC	10	10	9[a]	10	2	6	3	4	6	7	5	6	5	4	2
	Cys	TGC/TGT	6	5	8	3	2	11	5	5	7	11	12	7	4	5	10
	Phe	TTC	20	13	6	3	2	7	4	15	7	9	4	2	2	7	9
Sum			**68**	**66**	**69**	**53**	**50**	**82**	**83**	**83**	**79**	**79**	**53**	**53**	**53**	**54**	**54**
Non -	Cys	TGT	1														
Designed	Ser	TCC	1														
Codons	Ser	TCC		1													
	Thr	ACC		2													
	Leu	CTG		1													
	Leu	CTG			1		1[b]										
	Met	ATG				1[b]											
	Lys	AAG					1										
	Arg	CGG					1	1									
	His	CAT					1										
Parental	Arg	CGT	11														
Constructs	Thr	ACG		11													
	Tyr	TAC			11[a]												
	Met	ATG				10											
	Leu	CTG					10										
Sum (total)			**81**	**81**	**81**	**64**	**64**	**83**	**83**	**83**	**79**	**79**	**53**	**53**	**53**	**54**	**54**

[a] A total of 11 parental constructs were found and 9 mutants retaining the parental amino acid in position Y51. [b] Mutant L188R(CGG)/F87A and M185N/L185L(CTG)/F87A, which are not part of the library design and hence the targeted sequence space, were observed.

Table S2. Codon identity and occurrence frequency observed in the sequencing dataset of the random samples from both PCR and Sloning libraries B (V78/A82), as well as Slonings internal quality control. PCR libraries were designed with NNK degeneracy and Sloning with 20 codon: 20 aa degeneracy at each position. Non-designed codons were observed. The different codon usage between Sloning and NNK degeneracy is highlighted in green.

		Codon Usage		PCR Random Sample		Sloning Random Sample		Sloning Quality Control	
		PCR	**Sloning**	**V78** GTA	**A82** GCA	**V78** GTA	**A82** GCA	**V78** GTA	**A82** GCA
Allowed Codons	Ala	GCG	GCG	1	0	1	2	2	3
		GCT		3	0				
	Arg	CGT	CGT	1	0	10	8	3	4
		AGG		5	1				
		CGG		0	0				
	Asn	AAT	AAT	5	10	7	4	1	5
	Asp	GAT	GAC	4	9	6	8	4	3
	Cys	TGT	TGC	4	1	4	1	0	2
	Gln	CAG	CAA	0	0	8	2	6	3
	Glu	GAG	GAG	0	3	2	3	3	1
	Gly	GGG	GGC	1	5	3	2	1	1
		GGT		4	1				
	His	CAT	CAC	1	1	10	3	7	4
	Ile	ATT	ATC	8	3	4	6	1	3
	Leu	CTG	CTG	1	0	2	9	4	1
		CTT		1	0				
		TTG		1	0				
	Lys	AAG	AAA	0	5	7	5	2	2
	Met	ATG	ATG	1	3	3	7	1	6
	Phe	TTT	TTC	4	4	0	5	0	1
	Pro	CCG	CCG	0	0	5	6	5	3
		CCT		1	0				
	Ser	TCT	AGC	0	0	1	6	1	2
		TCG		1	1				
		AGT		6	5				
	Thr	ACG	ACC	4	1	6	2	4	2
		ACT		2	0				
	Trp	TGG	TGG	1	3	3	1	3	1
	Tyr	TAT	TAT	3	4	5	3	2	3
	Val	GTT	GTT	1	1	3	6	1	1
		GTG		4	1				
	Stop	TAG	not included	0	2	--	--	--	--
Sum				**68**	**64**	**90**	**89**	**51**	**51**
Non - Designed Codons	Ala	GCA			4[a]				
	Gln		CAG			1			
	Arg		CGG				1		
	Phe		TTT				1		
Parental Constructs	Val	GTA		18					
	Ala	GCA			18				
Sum (total)				**86**	**86**	**91**	**91**	**51**	**51**

[a] Four unexpected mutants, outside of the target sequence space, were found. These retained the parental codon GCA (Ala) in position 82, which is not encoded in NNK, but exhibit 4 NNK randomized codons at position V78(GTA), i.e., CAT/His, TTT/Phe, GTT/Val or GTG/Val. The two later encode the parental amino acid Val. These findings suggest a hybridization bias, as mentioned in the main text and discussed elsewhere.[3, 4] Therefore, a total amount of 20 parental sequences is present in the B-PCR random sample.

CHI SQUARE TEST RESULTS

Table S3. Sequencing results translated into amino acids and analyzed by Chi Square Test.

	site A						site C			
PCR	**R47**		**T49**		**Y51**		**M185**		**L188**	
expected relative frequency	expected absolute frequency	observed frequency	expected absolute frequency	observed frequency	expected absolute frequency	observed frequency	expected absolute frequency	observed frequency	expected absolute frequency	observed frequency
1/12	5.667	6	5.500	5	5.750	8	4.417	3	4.167	2
1/12	5.667	4	5.500	9	5.750	6	4.417	0	4.167	5
1/12	5.667	20	5.500	13	5.750	6	4.417	3	4.167	2
1/12	5.667	4	5.500	3	5.750	8	4.417	1	4.167	0
1/12	5.667	1	5.500	2	5.750	3	4.417	11	4.167	8
1/12	5.667	5	5.500	12	5.750	3	4.417	4	4.167	2
1/12	5.667	0	5.500	3	5.750	7	4.417	3	4.167	2
1/12	5.667	4	5.500	5	5.750	7	4.417	9	4.167	20
1/12	5.667	1	5.500	0	5.750	2	4.417	0	4.167	1
1/12	5.667	4	5.500	0	5.750	2	4.417	5	4.167	5
1/12	5.667	9	5.500	4	5.750	8	4.417	4	4.167	1
1/12	5.667	10	5.500	10	5.750	9	4.417	10	4.167	2
chi-square probability		3.411E-08		3.843E-05		0.323		2.859E-04		2.702E-12

	site A						site C			
Sloning	**R47**		**T49**		**Y51**		**M185**		**L188**	
1/12	6.833	11	6.917	5	6.917	5	6.583	7	6.583	11
1/12	6.833	4	6.917	8	6.917	2	6.583	5	6.583	4
1/12	6.833	7	6.917	4	6.917	15	6.583	7	6.583	9
1/12	6.833	5	6.917	7	6.917	7	6.583	15	6.583	2
1/12	6.833	5	6.917	10	6.917	4	6.583	5	6.583	10
1/12	6.833	4	6.917	9	6.917	7	6.583	5	6.583	7
1/12	6.833	6	6.917	12	6.917	3	6.583	9	6.583	13
1/12	6.833	8	6.917	9	6.917	11	6.583	5	6.583	5
1/12	6.833	12	6.917	7	6.917	9	6.583	5	6.583	3
1/12	6.833	5	6.917	7	6.917	12	6.583	5	6.583	5
1/12	6.833	9	6.917	2	6.917	4	6.583	5	6.583	3
1/12	6.833	6	6.917	3	6.917	4	6.583	6	6.583	7
		0.413		0.232		0.006		0.211		0.036

	site B					site B			
PCR	**V78**		**A82**		**Sloning**	**V78**		**A82**	
expected rel. frequency	expected abs. frequency	observed frequency	expected abs. frequency	observed frequency	expected rel. frequency	expected abs. frequency	observed frequency	expected abs. frequency	observed frequency
2/32	4.250	4.000	4.000	0.000	0.050	4.500	7.000	4.450	5.000
1/32	2.125	4.000	2.000	1.000	0.050	4.500	7.000	4.450	4.000
1/32	2.125	4.000	2.000	9.000	0.050	4.500	6.000	4.450	2.000
1/32	2.125	0.000	2.000	3.000	0.050	4.500	1.000	4.450	6.000
1/32	2.125	4.000	2.000	4.000	0.050	4.500	4.000	4.450	6.000
2/32	4.250	5.000	4.000	6.000	0.050	4.500	3.000	4.450	7.000
1/32	2.125	1.000	2.000	1.000	0.050	4.500	8.000	4.450	2.000
1/32	2.125	8.000	2.000	3.000	0.050	4.500	10.000	4.450	3.000
1/32	2.125	0.000	2.000	5.000	0.050	4.500	5.000	4.450	6.000
3/32	6.375	3.000	6.000	0.000	0.050	4.500	10.000	4.450	8.000
1/32	2.125	1.000	2.000	3.000	0.050	4.500	2.000	4.450	9.000
1/32	2.125	5.000	2.000	10.000	0.050	4.500	6.000	4.450	8.000
2/32	4.250	1.000	4.000	0.000	0.050	4.500	2.000	4.450	3.000
1/32	2.125	0.000	2.000	0.000	0.050	4.500	1.000	4.450	2.000
3/32	6.375	6.000	6.000	1.000	0.050	4.500	3.000	4.450	2.000
3/32	6.375	7.000	6.000	6.000	0.050	4.500	3.000	4.450	6.000
2/32	4.250	6.000	4.000	1.000	0.050	4.500	5.000	4.450	3.000
2/32	4.250	5.000	4.000	2.000	0.050	4.500	4.000	4.450	1.000
1/32	2.125	1.000	2.000	3.000	0.050	4.500	3.000	4.450	1.000
1/32	2.125	3.000	2.000	4.000	0.050	4.500	0.000	4.450	5.000
1/32	2.125	0.000	2.000	2.000					
chi-square probability		0.004		2.793E-11			0.014		0.123

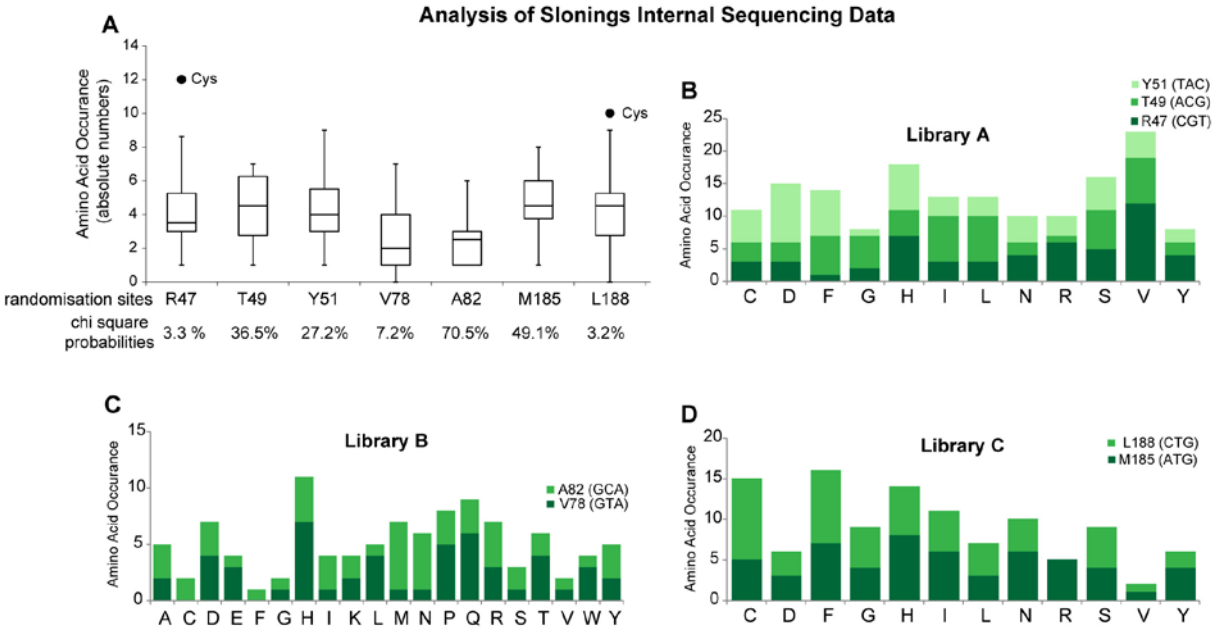

Figure S2. Analysis of Slonings internal quality control data. Libraries were sequenced prior to delivery by random sampling 50 transformants for each library. A) Box-plot and probabilities of the chi-square test; B,C,D) Amino acid diversity in the Sloning internal quality control. It should be noted that the amount of ~50 samples is a borderline for a statistically solid conclusion.

LIBRARY SCREENING RESULTS

Table S4. Numbers of transformants (data entries) qualifying as hits after threshold criteria were applied to the original screening data. Selectivity cut-off was ≥ 85%, with an area of at least 10,000 for the OHT product peak. In addition, a cut-off threshold for ≥35% total testosterone conversion (%HPLC) for 2β-OHT selective mutants and ≥40% total testosterone conversion for 15β-OHT selective mutants was applied. Since mutants were selected from the 95% datasets for evaluation, minor deviation to the numbers in Table 4-6 occur.

Library coverage	Selec-tivity	Site A		Site B	
		PCR	Sloning	PCR	Sloning
~95% (1 out of 1)	2β	148	17	11[a]	4[a]
	15β	0	0	12	17
78% (1 out of 2)	2β	63	15	4[a]	3[a]
	15β	0	0	6	11
63% (1 out of 3)	2β	30	12	3[a]	2[a]
	15β	0	0	5	9
53% (1 out of 4)	2β	24	11	3[a]	2[a]
	15β	0	0	5	6
45% (1 out of 5)	2β	24	11	3[a]	2[a]
	15β	0	0	2	5

	Site C (Activity)[b]	
	PCR	Sloning
~95% (1 out of 1)	10	76
78% (1 out of 2)	0	16
63% (1 out of 3)	0	12
53% (1 out of 4)	0	11
45% (1 out of 5)	0	7

[a] selectivity criterion was lowered from ≥85% to ≥80% 2β-OHT and to ≥30% total testosterone conversion. [b] total testosterone conversion ≥50%, as determined by HPLC.

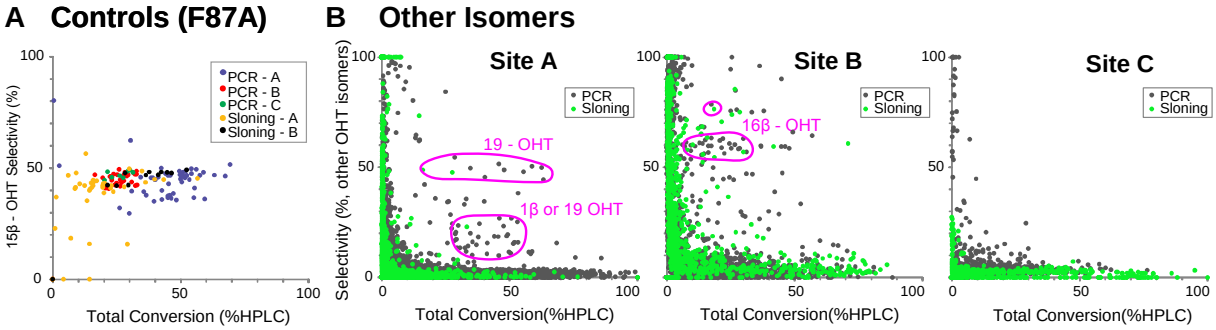

Figure S3. Analysis of controls and plots of the sum of additionally observed regio- and stereoselectivities. A) Plot of all positive controls (mutant F87A) present on each library plate. Library C-SLO did not contain control F87A. B) The sum of additionally observed regio- and stereoselectivities plotted against total testosterone conversion reveals that only very few mutants show other selectivities than 2β or 15β (as seen in Figure 4). Libraries A contain mutants producing 1β-OHT or 19-OHT in mixture with 2β-OHT, whereas in libraries B 16β-OHT was observed in mixture with 15β-OHT.

VISUALIZATION OF ONE OF THE REDUCED AND THE NON-REDUCED DATASETS

Figure S4 exemplarily shows, how the data scatters in the reduced datasets corresponding to 63% library coverage (at least 1 out of 3, 95% probability). Overall, these smaller datasets (63% coverage) show the same shape as the 95% coverage datasets. Therefore the reduction uniformly thins out the screening data, resulting in smaller libraries which offer the same probability of finding hits as the bigger libraries. Figure S5 provides a zoom into the region of the 63% datasets which harbor the desired variants with improved properties. The figure shows where some of the encountered improved variants occurred within the original library screening data. As seen in Figure S5, hits can be identified from the smaller library sets. Therefore we see library coverages around 50% as promising compromise between a medium-to-high library coverage (>50% of targeted mutants are tested) yet low screening effort (absolute numbers). Of course any lower coverage can be sufficient as well, as previous studies have shown[5-7]. But, not always is a good hit encountered early on in screening, nor is the screening process stopped when in a created amount of 96-deep well plates a hit is encountered early, since screening is a random event and the question wether there is something better in is nagging on the researcher.

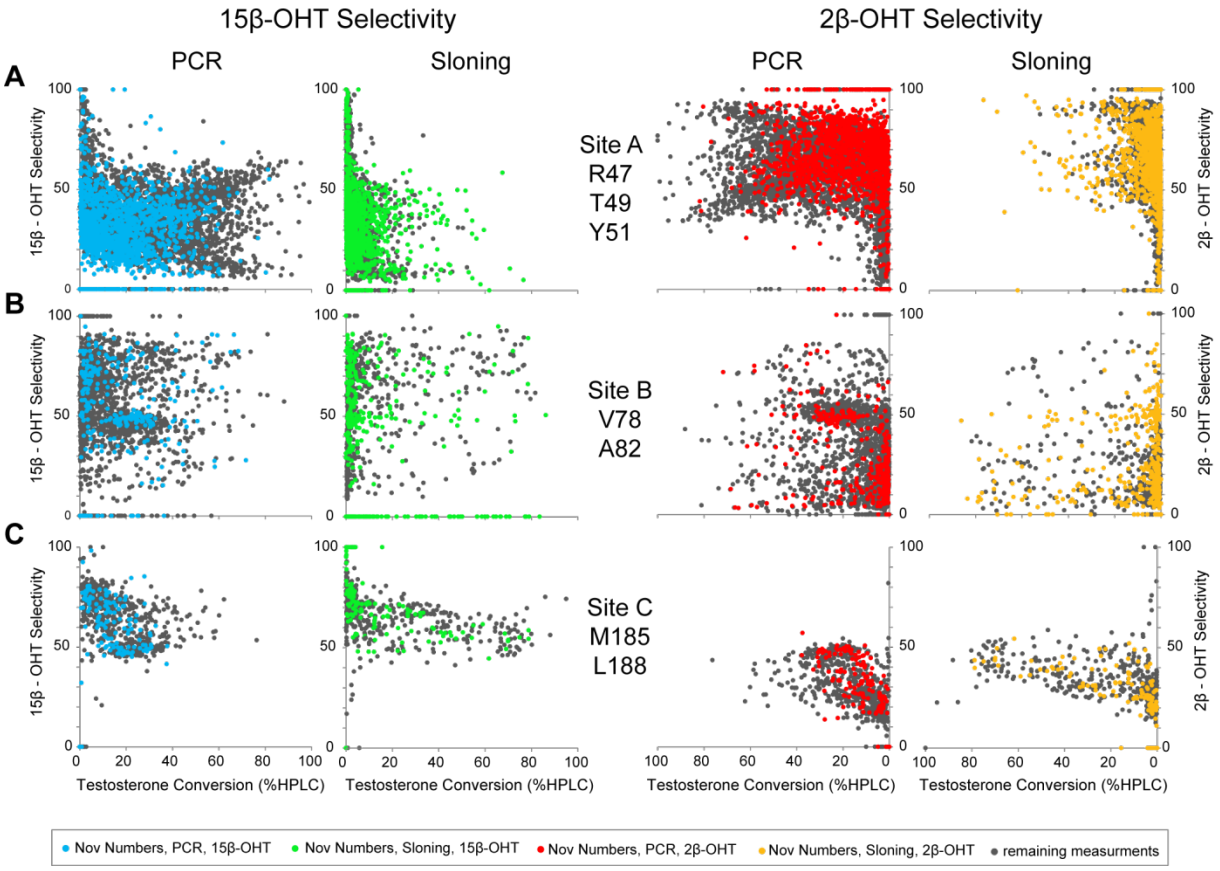

Figure S4. Library screening results with 63% coverage. Total testosterone conversion (%HPLC) of the six combinatorial libraries is shown as a function of either 15β-OHT or 2β-OHT regioselectivity. Colored entries show the data corresponding to 63% library coverage (at least 1 out of 3 best), while gray entries represent the *remaining* measurements from the 95% coverage dataset. TopLib predicts 1726, 399 and 143 transformants for libraries A, B and C respectively when aiming for 95% probability, 100% yield and no redundancy. Even though the A-PCR library pattern is stretched horizontally, the shape of the pattern is essentially the same. A longer reaction time (72h instead of 24h) resulted in an overall higher conversion level, causing the observed stretch. Despite this, selectivity values were generally not affected by the overall activity and therefore the scattering patterns are comparable.

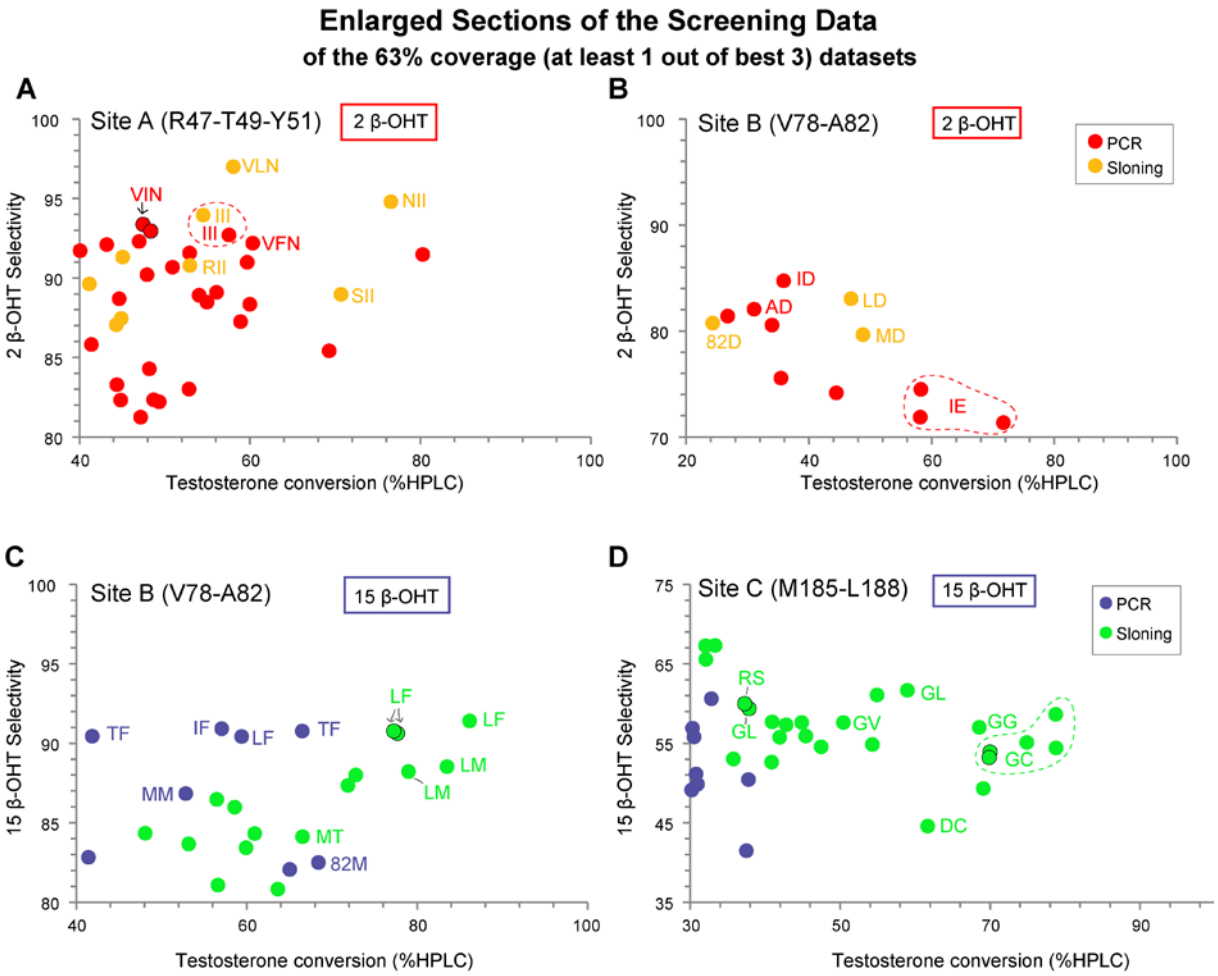

Figure S5. Enlarged section of Figure S4 "hit regions". For reasons of clarity, only the data entries of the 63% coverage datasets are presented. This figure shows the original library results, from where a researcher has to select transformants to evaluate by re-culturing the biocatalytic properties. When sequence identity was available, it is given in one letter amino acid code. Mutants of same identity are circled for clarity. **A)** Residues R47, T49 and Y51 were mutated in libraries A. Only the 2β-OHT region is shown, because libraries A did not contain data entries in the 15β-OHT region as seen in Figure S4. **B and C)** Residues V78 and A82 were mutated in BM3-F87A based libraries. Data entries were found in both the 2β-OHT and the 15β-OHT regions. **D)** Residues M185 and L188 were mutated in libraries C. No regio- and stereoselective data entries were found in the libraries. However, certain entries showed increased total testosterone conversion.

References:

(1) Reetz, M. T., and Carballeira, J. D. (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes, *Nat. Protoc. 2*, 891-903.

(2) Kille, S., Zilly, F. E., Acevedo, J. P., and Reetz, M. T. (2011) Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution, *Nature Chemistry 3*, 738-743.

(3) Horne, M. T., Fish, D. J., and Benight, A. S. (2006) Statistical thermodynamics and kinetics of DNA multiplex hybridization reactions, *Biophys. J. 91*, 4133-4153.

(4) SantaLucia, J., Jr., Allawi, H. T., and Seneviratne, P. A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability, *Biochemistry 35*, 3555-3562.

(5) Parra, L. P., Agudo, R., and Reetz, M. T. (2013) Directed evolution by using iterative saturation mutagenesis based on multiresidue sites, *ChemBioChem 14*, 2301-2309.

(6) Reetz, M. T., Kahakeaw, D., and Lohmer, R. (2008) Addressing the numbers problem in directed evolution, *ChemBioChem 9*, 1797-1804.

(7) Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D., and Vogel, A. (2005) Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test, *Angew. Chem. Int. Ed. Engl. 44*, 4192-4196.