CODEX DNA

BioXp™ variant libraries

DNA variant libraries are a powerful tool for manipulating protein structure for optimization studies in discovery biology, protein engineering, and several other disciplines. Codex DNA offers a broad menu of DNA variant library types, including scanning and combinatorial libraries.

Scanning libraries allow users to systematically query a target sequence by mutating adjacent codons one at a time. Codons can either be mutated to alanine in alanine-scanning libraries, or substituted with all possible amino acids at each desired position in site saturation libraries. Combinatorial libraries enable users to evaluate the combined effect of two or more simultaneous amino acid substitutions.

A
...AGTC**NNN**ATGGATCACAGTCA...
...AGTCACT**NNN**GATCACAGTCA...
...AGTCACTATG**NNN**CACAGTCA...
...AGTCACTATGGAT**NNN**AGTCA...
...AGTCACTATGGATCAC**NNN**CA...

B
...AGTCAC**NNK**ATGGA**H**CACAGTCAG...
...AGTCAC**NNN**ATGGA**R**CACAGTCAG...
...AGTCACT**SK**ATGGA**Y**CACAGTCAG...
...AGTCACTA**Y**ATGGA**B**CACAGTCAG...
...AGTCACTA**K**ATGGA**S**CACAGTCAG...

Figure 1. BioXp™ variant library types: A. Scanning libraries. Useful for quickly identifying amino acid sites that are sensitive to the changes made. Substitutions can include site saturation (NNN or NNK), or alanine-encoding codons.
B. Combinatorial libraries. Helpful for identification of the optimal combination of changes for the desired effect. Multiple, non-contiguous codon substitutions can be made with degenerate bases or sets of codons.

DNA variant libraries from Codex DNA can be built hands-free, in the user's lab, using the benchtop BioXp™ system, or ordered as a service. The BioXp™ system is the world's first and only commercially available push-button automated platform for on-demand DNA assembly and amplification. It enables labs to automate the synthesis of genes, clones, variant libraries and even genomes.

Due to the unique way that the BioXp™ system assembles dsDNA libraries, researchers can begin the downstream screening process more quickly saving significant time during the design-build-test cycle.

To demonstrate the quality of DNA variant libraries built on the BioXp™ system, we have showcased blinded data from two library projects built on the BioXp™ system by *Company A*. *Company A* performed independent sequence analysis for both projects.

Sequences were ordered through the Codex DNA portal. *Company A* received the reagents, loaded the BioXp™ system (loading the instrument takes approximately five minutes), and then initiated the library build by pressing *Start* on the instrument touch screen. After an instrument run of approximately eight hours, *Company A* collected the BioXp™ libraries, analyzing the product from each well directly using next-generation sequencing (NGS).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
AAG	0.06%	007%	0.08%	0.12%	0.13%	0.08%	0.12%	0.08%	0.13%	0.08%	0.11%	0.06%	0.09%	0.07%	0.12%	0.10%	0.11%	0.07%	0.11%	0.08%
CAG	0.07%	0.09%	0.11%	0.15%	0.25%	0.14%	0.18%	0.13%	0.11%	0.09%	0.12%	0.12%	0.13%	0.10%	0.18%	0.15%	0.14%	0.07%	0.12%	0.08%
GAG	0.13%	0.17%	0.20%	0.29%	0.21%	0.11%	0.15%	0.10%	0.13%	0.08%	0.10%	0.06%	0.06%	0.05%	0.15%	0.16%	0.18%	0.10%	0.19%	0.14%
TAG	0.06%	0.10%	0.12%	0.05%	0.14%	0.06%	0.12%	0.07%	0.11%	0.06%	0.10%	0.07%	0.06%	0.04%	0.14%	0.13%	0.13%	0.08%	0.12%	0.08%
ACG	0.06%	0.08%	0.10%	0.16%	0.27%	0.15%	0.20%	0.19%	0.27%	0.11%	0.15%	0.13%	0.14%	0.10%	0.16%	0.13%	0.12%	0.08%	0.12%	0.07%
CCG	0.06%	0.05%	0.10%	0.19%	0.54%	0.29%	0.35%	0.32%	0.42%	0.08%	0.35%	0.26%	0.23%	0.13%	0.23%	0.24%	0.18%	0.11%	0.15%	0.07%
GCG	0.13%	0.18%	0.24%	0.37%	0.36%	0.15%	0.21%	0.18%	0.25%	0.10%	0.14%	0.12%	0.10%	0.07%	0.17%	0.20%	0.18%	0.10%	0.22%	0.14%
TCG	0.07%	0.11%	0.13%	0.06%	0.24%	0.10%	0.17%	0.15%	0.21%	0.08%	0.17%	0.12%	0.09%	0.07%	0.16%	0.16%	0.13%	0.08%	0.14%	0.09%
AGG	0.12%	0.15%	0.20%	0.27%	0.20%	0.12%	0.17%	0.12%	0.17%	0.10%	0.07%	0.05%	0.06%	0.06%	0.14%	0.14%	0.16%	0.08%	0.21%	0.17%
CGG	0.15%	0.17%	0.25%	0.30%	0.30%	0.16%	0.22%	0.15%	0.23%	0.09%	0.12%	0.09%	0.12%	0.08%	0.18%	0.18%	0.17%	0.12%	0.22%	0.16%
GGG	0.33%	0.40%	0.52%	0.68%	0.43%	0.21%	0.29%	0.21%	0.26%	0.13%	0.05%	0.05%	0.05%	0.05%	0.24%	0.29%	0.33%	0.20%	0.40%	0.36%
TGG	0.14%	0.23%	0.27%	0.12%	0.21%	0.10%	0.16%	0.12%	0.18%	0.08%	0.09%	0.05%	0.05%	0.04%	0.17%	0.19%	0.20%	0.13%	0.23%	0.19%
ATG	0.08%	0.09%	0.11%	0.13%	0.12%	0.06%	0.10%	0.08%	0.17%	0.08%	0.08%	0.05%	0.08%	0.05%	0.12%	0.10%	0.10%	0.07%	0.14%	0.09%
CTG	0.07%	0.09%	0.11%	0.13%	0.20%	0.10%	0.15%	0.14%	0.24%	0.09%	0.15%	0.10%	0.12%	0.07%	0.17%	0.13%	0.13%	0.09%	0.15%	0.08%
GTG	0.17%	0.22%	0.26%	0.34%	0.19%	0.08%	0.14%	0.11%	0.19%	0.08%	0.07%	0.05%	0.04%	0.04%	0.17%	0.19%	0.19%	0.12%	0.25%	0.18%
TTG	0.08%	0.12%	0.13%	0.05%	0.11%	0.05%	0.12%	0.08%	0.15%	0.06%	0.09%	0.06%	0.05%	0.04%	0.14%	0.13%	0.12%	0.09%	0.16%	0.10%
AAT	0.03%	0.04%	0.03%	0.05%	0.12%	0.08%	0.12%	0.08%	0.12%	0.08%	0.13%	0.12%	0.10%	0.07%	0.11%	0.08%	0.08%	0.04%	0.06%	0.04%
CAT	0.03%	0.05%	0.05%	0.12%	0.25%	0.14%	0.21%	0.12%	0.18%	0.11%	0.24%	0.13%	0.15%	0.10%	0.17%	0.13%	0.13%	0.05%	0.07%	0.04%
GAT	0.07%	0.08%	0.12%	0.14%	0.14%	0.07%	0.12%	0.07%	0.10%	0.07%	0.12%	0.06%	0.07%	0.05%	0.13%	0.12%	0.12%	0.06%	0.10%	0.08%
TAT	0.03%	0.05%	0.05%	0.03%	0.12%	0.06%	0.12%	0.06%	0.10%	0.07%	0.13%	0.07%	0.06%	0.05%	0.13%	0.09%	0.10%	0.05%	0.08%	0.04%
ACT	0.03%	0.04%	0.04%	0.08%	0.25%	0.13%	0.21%	0.18%	0.26%	0.15%	0.19%	0.13%	0.13%	0.10%	0.14%	0.12%	0.10%	0.05%	0.07%	0.03%
CCT	0.03%	0.04%	0.04%	0.11%	0.55%	0.29%	0.39%	0.32%	0.44%	0.22%	0.38%	0.27%	0.24%	0.16%	0.23%	0.22%	0.18%	0.07%	0.10%	0.04%
GCT	0.07%	0.09%	0.10%	0.20%	0.29%	0.12%	0.19%	0.16%	0.20%	0.12%	0.17%	0.11%	0.12%	0.12%	0.13%	0.13%	0.14%	0.06%	0.11%	0.07%
TCT	0.03%	0.05%	0.04%	0.03%	0.21%	0.09%	0.16%	0.14%	0.20%	0.11%	0.20%	0.11%	0.09%	0.06%	0.14%	0.12%	0.11%	0.05%	0.12%	0.04%
AGT	0.07%	0.08%	0.09%	0.12%	0.13%	0.07%	0.13%	0.09%	0.13%	0.08%	0.09%	0.05%	0.06%	0.05%	0.11%	0.09%	0.09%	0.04%	0.10%	0.07%
CGT	0.08%	0.09%	0.11%	0.16%	0.23%	0.12%	0.21%	0.14%	0.20%	0.10%	0.16%	0.08%	0.09%	0.08%	0.17%	0.14%	0.08%	0.07%	0.11%	0.08%
GGT	0.16%	0.21%	0.23%	0.30%	0.23%	0.11%	0.18%	0.11%	0.15%	0.07%	0.03%	0.04%	0.04%	0.04%	0.12%	0.18%	0.19%	0.12%	0.19%	0.12%
TGT	0.08%	0.12%	0.11%	0.06%	0.15%	0.07%	0.13%	0.08%	0.14%	0.07%	0.10%	0.05%	0.05%	0.04%	0.13%	0.12%	0.13%	0.07%	0.11%	0.09%
ATT	0.04%	0.04%	0.04%	0.05%	0.09%	0.05%	0.10%	0.07%	0.14%	0.08%	0.09%	0.06%	0.06%	0.04%	0.11%	0.12%	0.07%	0.05%	0.06%	0.04%
CTT	0.03%	0.04%	0.05%	0.06%	0.17%	0.08%	0.15%	0.12%	0.23%	0.11%	0.18%	0.10%	0.10%	0.08%	0.15%	0.12%	0.11%	0.06%	0.09%	0.04%
GTT	0.09%	0.11%	0.11%	0.15%	0.13%	0.05%	0.11%	0.07%	0.13%	0.07%	0.12%	0.05%	0.04%	0.04%	0.14%	0.12%	0.13%	0.07%	0.12%	0.08%
TTT	0.04%	0.05%	0.05%	0.03%	0.09%	0.04%	0.12%	0.06%	0.13%	0.06%	0.10%	0.06%	0.04%	0.03%	0.12%	0.09%	0.09%	0.06%	0.08%	0.05%

Figure 2. Heat map showing occurrence of each of the 640 possible codons from the 20-position NNK scan.

Library 1: BioXp™ scanning library

Scanning NNK library

Scanning libraries are essential for protein function studies. *Company A* built a scanning library on the BioXp™ system. The library design included NNK codon substitutions at 20 amino acid positions. There are 32 possible codon combinations in an NNK substitution. Across 20 positions, there are a total of 640 possible codon combinations. Library sequences were analyzed with a heat map (figure 2), and by calculating codon distribution (figure 3).

As shown in figures 2 and 3, the BioXp™ scanning library has 100% coverage of all 640 possible codons — all possible trinucleotides are present at all 20 positions. As expected, a slight G/C bias is observed. Both the heat map and the codon distribution graph confirm that the BioXp™ system built a 20-position variant library with complete coverage and high diversity.

Library 1 summary

- NGS (528,627 reads) showed all possible codons observed at each of the 20 positions
- Each position is synthesized as an individual library, allowing for positional screening or pooling of positions for downstream analysis

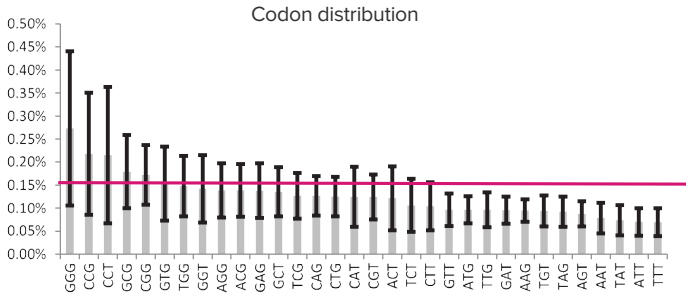

Figure 3. Codon distribution. The theoretical codon distribution is 0.156%. The observed codon distribution has a bias towards G/C.

Library 2: BioXp™ combinatorial library

Ten-position variant library

Ten positions that were known to be associated with a previously identified functional region within a protein were randomized using IUPAC letters. When expressed, the variant proteins could be evaluated for a range of physical characteristics and enzymatic activity.

A	0%	100%	0%	41%	19%	0%	100%	23%	58%	22%	0%	55%	0%	0%	0%	100%	100%	0%	0%	0%	0%	42%	33%	100%	100%	0%	100%	18%
G	100%	0%	0%	0%	25%	0%	0%	23%	0%	25%	0%	0%	0%	100%	100%	0%	0%	0%	0%	0%	100%	58%	40%	0%	0%	100%	0%	31%
C	0%	0%	0%	59%	35%	0%	0%	36%	0%	34%	67%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	27%
T	0%	0%	100%	0%	21%	100%	0%	18%	42%	19%	33%	45%	100%	0%	0%	0%	0%	100%	100%	100%	0%	0%	27%	0%	0%	0%	0%	24%
	G	A	T	M	N	T	A	N	W	N	Y	W	T	G	G	A	A	T	T	T	G	R	O	A	A	G	A	N
	G	A	T	A/C	A/C/ T/G	T	A	A/C/ T/G	A/T	A/C/ T/G	C/T	A/T	T	G	G	A	A	T	T	T	G	A/G	A/G/T	A	A	G	A	A/C/ T/G

Figure 4. Variant library used IUPAC coding to vary ten positions in a 28 bp stretch of a gene.

Figure 4 shows the sequence distribution across the ten positions that were mutated in the library build. All non-variant sequences are shown in yellow; NGS demonstrated perfect sequence within this region for non-variant positions of this BioXp™ library. Variant positions are highlighted in green. At positions with a binary variation (A/C, C/T, A/T, A/G), 50% representation is expected for each base. Position 23 had a 3-nucleotide variation (A/G/T), with an expected distribution of 33% for each base. Remaining variant sites were designed to incorporate all bases with an expected distribution of 25% for each base. As shown in the image, a roughly equal proportion of each variant nucleotide is present at each variant position, indicating that this BioXp™ library demonstrates high diversity. The calculated diversity for this sequence is 24,576.

This combinatorial library contained a single stretch with five adjacent variant positions. Figure 5 shows the expected and observed nucleotide bases for these five positions. Consistent with typical library builds, the data show that there is a slight observed bias for incorporation of cytosine at positions 8, 10 and 11. Overall, the BioXp™ library exhibits the presence of all expected variation in approximately the expected ratio for all variant positions.

Library 2 summary

- Of 586,343 reads, 549,230 (94%) were unique
- The observed distribution matches the expected distribution

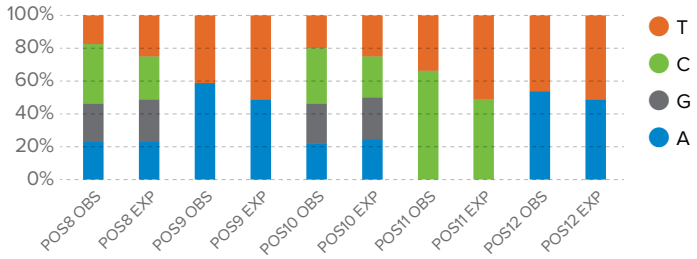

Figure 5. Comparison of observed vs expected bases at five positions.

Benefits of building libraries on the BioXp™ system

Library validation by *Company A* demonstrates that the BioXp™ system builds high-quality, high-diversity scanning and combinatorial libraries. In addition to validating the quality of BioXp™ libraries, *Company A* noted that building libraries on the BioXp™ system "radically expedited their workflow." Library builds using traditional service providers would take *Company A* two weeks to one month. Building libraries on the BioXp™ system instead greatly reduced that time, taking only five days. Other advantages of building variant libraries on the BioXp™ system include:

- Drastically accelerating the design-build-test cycle
- Low cost — Codex DNA libraries are competitively priced. To learn more, visit www.codexdna.com
- Each position of a BioXp™ scanning library is built as an individual library, allowing for greater flexibility in downstream analysis
- For scanning libraries, continuous and discontinuous scans are permitted

Codex DNA, Inc.
9535 Waples Street, Suite 100
San Diego, CA 92121 USA
hello@codexdna.com | 858.228.4115 | codexdna.com

Part number 45022 | Revision 1.1 | Effective 30SEP2020

CODEX DNA