Comprehensive profiling of four base overhang ligation fidelity by T4 DNA ligase and application to DNA assembly

Vladimir Potapov[1], Jennifer L. Ong[1], Rebecca B. Kucera[2], Bradley W. Langhorst[2], Katharina Bilotti[1], John M. Pryor[1], Eric J. Cantor[2], Barry Canton[3], Thomas F. Knight[3], Thomas C. Evans, Jr.[1], Gregory J. S. Lohman[1,*]

[1] Research Department, New England Biolabs, Ipswich, MA, 01938, USA
[2] Applications and Product Development, New England Biolabs, Ipswich, MA, 01938, USA
[3] Ginkgo Bioworks, Boston, MA, 02210, USA

* To whom correspondence should be addressed. Tel:1-978-998-7916; Fax: 978-921-1350; Email: lohman@neb.com

Supporting Information

Supporting Note – Assembly results, 37°C 18 h

Supporting Tables

Supporting Table S1. Ten-fragment Golden Gate insert sequences.

Supporting Table S2. Ten-fragment Golden Gate assembly junction sequences

Supporting Table S3. Junction coordinates and sequences for high- and low-fidelity 12-fragment *lac* cassettes.

Supporting Table S4. Junction coordinates and sequences for high-fidelity 24-fragment *lac* cassette.

Supporting Table S5. Number of ligation events per experiment.

Supporting Table S6. Observations of blue/white colonies for 12- and 24-fragment *lac* cassette assemblies.

Supporting Table S7. Validation of high fidelity twelve fragment *lac* cassette assembly reactions by omission of each fragment. In the absence of any single one of the twelve insert fragments, no blue colonies are observed.

Supporting Table S8. Predicted high fidelity four-base overhang sets for use with Golden Gate assembly methods (based on 18 h at 37°C).

Supporting Table S9. Validation of high fidelity twenty-four fragment *lac* cassette assembly reactions by omission of each fragment. In the absence of any single one of the twelve insert fragments, no blue colonies are observed.

Supporting Table S10. Ligation fidelity substrate and ligation library sequences.

Supporting Figures

Supporting Figure S1. Ligation fidelity of T4 DNA Ligase for four-base overhangs (1 h at 25°C)

Supporting Figure S2. Ligation fidelity of T4 DNA Ligase for four-base overhangs (1 h at 37°C)

Supporting Figure S3. Ligation fidelity of T4 DNA Ligase for four-base overhangs (18 h at 37°C)

Supporting Figure S4. Frequency of specific base-pair mismatches by position, additional ligation conditions.

Supporting Figure S5. Ligation fidelity of T7 DNA Ligase for four-base overhangs (18 h at 25°C)

Supporting Figure S6. Ligation fidelity of T7 DNA Ligase for four-base overhangs (18 h at 37°C)

Supporting Figure S7. Predicted versus observed fragment linkages in a multi-fragment (Golden Gate) assembly for the DP and FP 10-fragment assemblies. Predicted frequencies of junctions are based on the fidelity library data generated for the four-base overhang substrate ligated with T4 DNA ligase at 25°C for 18 h. The experimental observations shown are for assembly of the 10-fragment HF and LF sets with Golden Gate Assembly mix, 37°C 5 min/16°C 5 min, 30 cycles.

Supporting Figure S8. Distribution of assembly sizes for the 10-fragment Golden Gate assemblies (18h at 37°C).

Supporting Figure S9. Predicted versus observed fragment linkages in a multi-fragment (Golden Gate) assembly for the HF, LF, DP and FP 10-fragment assemblies. Predicted frequencies of junctions are based on the fidelity library data generated for the four-base overhang substrate ligated with T4 DNA ligase at 37°C for 18 h. The experimental observations shown are for assembly of the 10-fragment HF, LF, DP, and FP sets with Golden Gate Assembly mix (18 h at 37°C).

Supporting Figure S10. Predicted mismatch ligation potential between all palindromic overhangs. Palindromic overhangs generated by Type IIP restriction enzymes are used in traditional restriction enzyme cloning methods. The fidelity profile (18 h at 25°C) was used to predict any likely mismatch ligation between these overhangs. No significant cross talk is predicted, but the TATA and TTAA overhangs are expected to be low efficiency in ligation compared to all other palindromes.

Supporting Data

Supporting File 1. Ligation fidelity of T4 DNA Ligase for four-base overhangs (1 h at 25°C)
Supporting File 2. Ligation fidelity of T4 DNA Ligase for four-base overhangs (1 h at 37°C)
Supporting File 3. Ligation fidelity of T4 DNA Ligase for four-base overhangs (18 h at 25°C)
Supporting File 4. Ligation fidelity of T4 DNA Ligase for four-base overhangs (18 h at 37°C)
Supporting File 5. Ligation fidelity of T7 DNA Ligase for four-base overhangs (18 h at 25°C)

Supporting File 6. Ligation fidelity of T7 DNA Ligase for four-base overhangs (18 h at 37°C)

Supporting File 7. HF 10-fragment assembly data (Cycled)

Supporting File 8. LF 10-fragment assembly data (Cycled)

Supporting File 9. DP 10-fragment assembly data (Cycled)

Supporting File 10. FP 10-fragment assembly data (Cycled)

Supporting File 11. HF 10-fragment assembly data (37°C incubation)

Supporting File 12. LF 10-fragment assembly data (37°C incubation)

Supporting File 13. DP 10-fragment assembly data (37°C incubation)

Supporting File 14. FP 10-fragment assembly data (37°C incubation)

Supporting Note – Assembly results, 37°C 18 h

Given the higher ligation fidelity observed at 37°C, assembly reactions with prolonged incubation (18 h) at 37°C in lieu of cycling were also tested. The results (Supporting Figure S10) are largely consistent with the results of cycling with a few key exceptions. Firstly, the incidence of ligation errors in the LF set was much lower with the higher ligation temperature, consistent with the observation that fidelity was much improved at 37°C in the multiplexed fidelity profiles. While the HF set assembled under these conditions again showed >99.9% of all observed assemblies formed by correct Watson-Crick pairings, the LF set had only 31.2% of all assemblies containing at least one mispair; the specific mismatches observed were the same, just present in lower prevalence. Overall, the cycled assembly conditions were in good agreement with predictions made by the fidelity library ligated at 37°C. This result indicates that if a suitable orthogonal set of overhangs is chosen, either method will result in high fidelity assembly, but if sets prone to mismatch ligation events are chosen, a ligation temperature of 16°C will result in significantly more failed assemblies. Under the high temperature incubation conditions, while the FP still showed a large increase in truncations at the predicted low-efficiency junction 7, the increase in truncations at the 100% GC junction 6 was much less noticeable.

The higher fidelity observed with 18 h incubation at 37 offers the potential to produce even larger high fidelity sets than those proposed to be used with the cycled conditions presented in the main text, Table 1. While high GC overhangs do not appear to assemble inefficiently at this temperature, many more low-GC overhangs ligate less efficiently; these conditions are high fidelity but also highly biased. Supporting Table S10 enumerates high-fidelity Watson-Crick pair sets with up to 35 members, based on the ligation profile at 37°C 18 h; these sets are predicted to achieve the predicted fidelities with prolonged (overnight to ensure high efficiency ligation) static incubation at 37°C.

Supporting Table S1. Ten-fragment Golden Gate insert sequences

Insert name	Insert sequence
InsertA	TCGAAAGCAGACGTAATATATGAAGCCGCCTGGTATTTGGCTATGTACGGGACGA GGTTAATGTTGGGAGCGCTACTTAAGCCCTCATGAGAGTCGATTATTCTCGGCGC ATTCTTCGTGACGAAACGTAGGTAGTCCGCACTCGAAAGCACCGCAAAGTCAGAG CAACCTAAATAGGAACTGTGAATATTCTTCCATAAGACGGCTGTCTACCTCCCATA CCGGTGCTATCTATTGTTTGGGCTTTTCAGGCTCCATGGTAACGAGTTAGCGGGG ATAGCTCTTCCTCTTTTGCGCTCA
InsertB	GTTGCGCGCGGTTCGGGTCAGTTCCTTCTGCAAAGATCAGCCGACAAAAGGAAC AATTCGCCCCTGATGCTTTCTCCTAAGGTTTGTAGACTCGTTTTGCTCAAAGAGGT GAACATCGGGCAGCACTATTTAGGCCCCTTGAAGCTCGAAGGGCCATCCAGGCA AATTTGCTTGTCTTTCGGTAAGGCAGTAGTCTAACAACACTTGACACGTAATCAAC CCCGCTTAACCATACTACAGTCCGTGAGCACACTGAGCTCCACCCGACGGATATC TGACAAAAGGGTGTGGAAGTTAGAA
InsertC	CCCTAGTCACCTCCCACCAAGCGAGTGGGATCCCGGAACTTGTGTAAAACTTTTG TAGCATCCCAAACCCGTGACCCGATGCGTTGGTCAGTTTAGGCGGGAGCGGCTA CCCAAAACATACGAATACTACTAATATATTAAAGGGCTGAGGCTTGTCAACTGGGT CGATCTAAGTCGTATGATTGGCGCCTCCCCTTAAGAGGTACGAAGAAGCTCTCTC CCTAGACTCGTTCCATTCATTCCCCGGTAGGGACGGCTCAGGGAGGAATTCGACT TTGACATGATCGTCATGAACGCTGT
InsertD	ACTGGATGGGACGTCCATTGTCTGGCGCCGCCCGCTGAGGGGTCACGGAAATCT AGACGCATGACTGCTACCCTGGTGACCTTGTTATCATGTAAGTACACCAGTTTGG TGTTCGTGGAGTATGGATAACGTATATGATTGCCTCATTATACCGTAAACCATCAT GCCCGTCAGCTTCACAGGGGAACCAATCACTAGGTGGGTGCAGGCTATTACAAT CTCGCAATGAGCCGGCGCATCCGGACAAGTAGATGTTAGTGTACGATCCAGATAT CGTGGCTTCAGTAGATGCCCACATTT
InsertE	GGCCTCACAGATATCCAAAATAATGAGAGGGGCCAATTCCGCGGCAGCAATCTAG GTAAAGAATGGGTAAGTTGCTCGTGAGCGAATGGTGACGACCCTTGTGCTGCGG GTGACAGAGTGTCTTCTCTTCTGCGTAGCGACACATTTGCAATTCGGTAGCCATTA ATAAGCACCGACCGCGACCTATTGAAGCGCCGAACCCGTAACTACTGCGGAATA CCCCATTTTATCTTGAAGCGCGCAAGTGGCGCTACCGTGTTCTAACAGGATTCTA CAAGGAACATTTGATAGTTTTCTATC
InsertF	GTTCCTATGCTTGTCCTGTAAGCTTCAGACAACAAAGCTACCAGAACACAGTCGC GTCAATGAGCTACGCCAACTCTTACTGGTCACTTCCGATGTTTCATTTAGCACCCC GGCCTCAGATGTGCGGCCTTGAAGATACCGGTCCTGCGGTTGCGTGTCTAACCG GGCAAATCGCACCCCTAACCACGCTTCGTTACGGGATTGTTCTCGATTTGGATAA

	GTTGGCATGAGCCTGATGGCACACCATATTAGAGTAGGACAGAGTCGCACCAATA GGTCAGAGGATCGTAGAGGCAGGAT
InsertG	TTGCCTAGCAAGCGAAGATGCTGGACCCGTGTGTTTCTCCCCTGCACCAGACGAT CGCCGGTCGGACACGTCGCAGGTAGGGTATACGGACGACGTTTATTCCAATCAG TACCCGGAAAGGAGTTATGCTCGTTAAGCCCATGGATGCACCTAGTTACGCATTT GGCTTGTCAAACCTTTTGCAGGAGTGCCGTAAGAAAAGCCAATTTGATCGAGTCC TGATACATCCCACAAGCTATATGGACTTAGAAATCACTTGTATCATTACGCACACG GAAAACTCACACCCTTAATCGAACG
InsertH	ATAGGGTACTTTGAACAGCCTCCGCCGTCCTCGGTTCGTCATGATCATAAGTCTT CAGAAGCAGTAGCACCATCTTCCAAGAATGTCTGACGCAGGTGGGAGTTCAGTTG CACATTGATAATGTTAACCCATAACAGTAATGTCGGACGTGGCCTTTCAATATACG GAACCCCTGATACATATTAGCGGAGTTGTTCAAACTGGGTGAGGGTGGCACACAT CGGTTCTATACCTGCGACATGCCGGATTAGGTGACATAAAAGAAGGCGTATCCCA ATTAGCCATCCCAACTGTCCGCCC
InsertI	TACCCGCGCCGGTTTGAAGCATGGTAGTTCGTTCCATCGCAAGGGTCATTGGGAT TGCATCATGACCGTGCGTTGCGAGGTGTAGCGTCCCTCAGCTGAAAGGTCGCTC TATGGCGCCAGATTCAGGATTCAGGTCGCCGCTACCTTGACTAGCGGGCTGTGT GGAGAGGTGACGCACAGCCGCGGGAATTGAATCAGTGACTGCCTGCAACAAATG AGCTTTTGTTACTAGTTCGACGTTGCTACGAGAAGTCCTACAAATGCGCTTCTGTG ACTTACGCGCATTGAAGACAATGTTAT
InsertJ	TTAGACTCGGGGCCACGTAGCTCGCGTAGTCGAGTCCTAATCAGTTAATAATCCT ATCTGACCTCAATCAAGGGGCTCGAGCAAGTTCAAAGTTTCCAGACTCCGGAACA TAAATAGATGAGATAGTAGCGCCGGGAAACTATCGTTGTTTAGCGATGGCCATCT TCCCGGCTCTAAGCTTCTCATATGATCGGAGCCCCGGCTAACCGTGTCGAACGTG ATCTCACGGACCAGCAGCTACGCCTGATCCCGGCTCTACTCTCTACACTGGACCG ATAAACGAGGTACTGAGAGGGGTTT

Supporting Table S2. Ten-fragment Golden Gate assembly junction sequences [1]

Junction	High-fidelity set	Deletion-prone set	Failure-prone set	Low-fidelity set
1	AAGG	----	----	GCCC
1′	TTCC	----	----	CGGG
2	ACTC	----	----	GCCA
2′	TGAG	----	----	CGGT
3	AGGA	----	----	ACCC
3′	TCCT	----	----	TGGG
4	AGTC	----	----	AGCC
4′	TCAG	----	----	TCGG
5	ATCA	----	----	CGCC
5′	TAGT	----	----	GCGG
6	GCCG	----	----	AGCA
6′	CGGC	----	----	TCCT
7	CTGA	GCTG	TAAA	AGCG
7′	GACT	CGAC	ATTT	TCGC
8	GCGA	----	----	CGGC
8′	CGCT	----	----	GCCG
9	GGAA	----	----	AGGC
9′	CCTT	----	----	TCCG

[1] A notation of ---- / ---- indicates the junction pair used is identical to the HF set.

Supporting Table S3. Junction coordinates and sequences for high- and low-fidelity twelve fragment *lac* cassettes

Junction	High-fidelity set	Coordinates	Low-fidelity set	Coordinates
$\frac{1}{1'}$	$\frac{\text{GGAG}}{\text{CCTC}}$	1..4	$\frac{\text{GGAG}}{\text{CCTC}}$	1..4
$\frac{2}{2'}$	$\frac{\text{GGCA}}{\text{CCGT}}$	341..344	$\frac{\text{GGTC}}{\text{CCAG}}$	352..355
$\frac{3}{3'}$	$\frac{\text{TCGC}}{\text{AGCG}}$	762..765	$\frac{\text{AGCA}}{\text{TCGT}}$	747..750
$\frac{4}{4'}$	$\frac{\text{CAGT}}{\text{GTCA}}$	1116..1119	$\frac{\text{CAGT}}{\text{GTCA}}$	1116..1119
$\frac{5}{5'}$	$\frac{\text{TCCA}}{\text{AGGT}}$	1557..1560	$\frac{\text{GGTA}}{\text{CCAT}}$	1497..1500
$\frac{6}{6'}$	$\frac{\text{GAAT}}{\text{CTTA}}$	2023..2026	$\frac{\text{GAAT}}{\text{CTTA}}$	2023..2026
$\frac{7}{7'}$	$\frac{\text{AGTA}}{\text{TCAT}}$	2358..2361	$\frac{\text{GGTT}}{\text{CCAA}}$	2530..2533
$\frac{8}{8'}$	$\frac{\text{TCTT}}{\text{AGAA}}$	2955..2958	$\frac{\text{TCTT}}{\text{AGAA}}$	2955..2958
$\frac{9}{9'}$	$\frac{\text{CAAA}}{\text{GTTT}}$	3505..3508	$\frac{\text{GGTG}}{\text{CCAC}}$	3411..3414
$\frac{10}{10'}$	$\frac{\text{GCAC}}{\text{CGTG}}$	3898..3901	$\frac{\text{GCAC}}{\text{CGTG}}$	3898..3901
$\frac{11}{11'}$	$\frac{\text{AACG}}{\text{TTGC}}$	4209..4212	$\frac{\text{AGCG}}{\text{TCGC}}$	4204..4207
$\frac{12}{12'}$	$\frac{\text{GTCT}}{\text{CAGA}}$	4526..4529	$\frac{\text{GTCT}}{\text{CAGA}}$	4526..4529
$\frac{13}{13'}$	$\frac{\text{CCAT}}{\text{GGTA}}$	4848..4851	$\frac{\text{CCAT}}{\text{GGTA}}$	4848..4851

Supporting Table S4. Junction coordinates and sequences for high-fidelity twenty-four fragment *lac* cassette

Junction	Overhang	Coordinates
1 / 1′	GGAG / CCTC	1..4
2 / 2′	GATA / CTAT	119..122
3 / 3′	GGCA / CCGT	341..344
4 / 4′	GGTC / CCAG	563..566
5 / 5′	TCGC / AGCG	762..765
6 / 6′	GAGG / CTCC	899..902
7 / 7′	CAGT / GTCA	1116..1119
8 / 8′	GTAA / CATT	1313..1316
9 / 9′	TCCA / AGGT	1557..1560
10 / 10′	CACA / GTGT	1850..1853
11 / 11′	GAAT / CTTA	2023..2026
12 / 12′	ATAG / TATC	2150..2153
13 / 13′	AGTA / TCAT	2358..2361
14 / 14′	ATCA / TAGT	2645..2648
15 / 15′	TCTT / AGAA	2955..2958
16 / 16′	AGGT / TCCA	3246..3249
17 / 17′	CAAA / GTTT	3505..3508
18 / 18′	AAGC / TTCG	3681..3684
19 / 19′	GCAC / CGTG	3898..3901
20 / 20′	CAAC / GTTG	4033..4036
21 / 21′	AACG / TTGC	4209..4212
22 / 22′	CGAA / GCTT	4393..4396
23 / 23′	GTCT / CAGA	4526..4529
24 / 24′	TCAG / AGTC	4729..4732
25 / 25′	CCAT / GGTA	4848..4851

Supporting Table S5. Observations of blue/white colonies for 12- and 24-fragment *lac* cassette assemblies. 12HF = 12-fragment high-fidelity set; 12LF = 12-fragment low fidelity set; 24HF = 24-fragment high-fidelity set. Errors shown are one standard deviation from the mean.

	Transformation Colony Counts			
Assembly Tested	Blues	Whites	Totals	% Blues
12HF Replicate 1	626	9	635	99%
12HF Replicate 2	602	11	613	98%
12HF Replicate 3	834	7	841	99%
12HF Replicate 4	912	6	918	99%
12HF Replicate 5	325	0	325	100%
12HF Replicate 6	266	0	266	100%
12HF Replicate 7	353	4	357	99%
12HF Replicate 8	369	2	371	100%
12HF Average			540 ± 250	99.2% +/- 0.6%
12LF Replicate 1	106	123	229	46%
12LF Replicate 2	143	120	263	54%
12LF Replicate 3	135	203	338	40%
12LF Replicate 4	155	207	362	43%
12LF Replicate 5	239	262	501	48%
12LF Replicate 6	248	257	505	49%
12LF Replicate 7	256	362	618	41%
12LF Replicate 8	261	389	650	40%
12LF Average			430 ± 160	45% +/- 5%
24HF Replicate 1	32	6	38	84%
24HF Replicate 2	35	6	41	85%
24HF Replicate 3	95	14	109	87%
24HF Replicate 4	103	22	125	82%
24HF Replicate 5	58	17	75	77%
24HF Replicate 6	92	32	124	74%
24HF Replicate 7	32	8	40	81%
24HF Replicate 8	41	4	45	91%
24HF Replicate 9	76	9	85	90%
24HF Replicate 10	62	10	72	86%
24HF Average			75 ± 35	84% +/- 5%

Supporting Table S6. Validation of high-fidelity 12-fragment *lac* cassette assembly reactions by omission of each fragment. In the absence of any single one of the twelve insert fragments, no blue colonies are observed. Assembly reactions were performed as described in the Materials and Methods 37°C stages during the 30 cycles were 3 min at each temperature for expediency, and plating volumes of outgrowth for the single omission reactions were increased to 25 µl to increase detection sensitivity, and plating volumes for the positive control (complete assemblies, no omissions) were 2.5 µl outgrowth equivalents with counts multiplied by 10 to compare to the omission plates' outgrowth volume of 25 µl. Notes: 1 TMTC = too many to count; near lawn appearance. 2 na = not applicable.

	Transformation colony counts			
Assembly tested	**Blues**	**Whites**	**Totals**	**% Blues**
minus LacZ fragment # 1	0	3	3	0.0%
minus LacZ fragment # 2	0	4	4	0.0%
minus LacZ fragment # 3	0	3	3	0.0%
minus LacZ fragment # 4	0	31	31	0.0%
minus LacZ fragment # 5	0	37	37	0.0%
minus LacZ fragment # 6	0	25	25	0.0%
minus LacZ fragment # 7	0	11	11	0.0%
minus LacZ fragment # 8	0	51	51	0.0%
minus LacZ fragment # 9	0	17	17	0.0%
minus LacZ fragment #10	0	45	45	0.0%
minus LacZ fragment #11	0	19	19	0.0%
minus LacZ fragment #12	0	2	2	0.0%
minus pGGA destination plasmid	0	0	0	0.0%
minus BsaI-HFv2 restriction enzyme	0	TMTC[1]	TMTC	na[2]
minus T4 DNA ligase	0	0	0	0.0%
Positive Control; no omissions	4170	11	4181	99.7%
Positive Control; no omissions	4250	10	4260	99.8%

Supporting Table S7. Validation of high-fidelity 24-fragment *lac* cassette assembly reactions by omission of each fragment. In the absence of any single one of the twenty-four insert fragments, no blue colonies are observed. Assembly reactions were performed as described in the Materials and Methods, except enzyme levels were doubled to increase sensitivity. Notes: [1] TMTC = too many to count; near lawn appearance. [2] na = not applicable.

	Transformation colony counts			
Assembly tested	**Blues**	**Whites**	**Totals**	**% Blues**
minus LacZ fragment #1	0	5	5	0%
minus LacZ fragment #2	0	10	10	0%
minus LacZ fragment #3	0	60	60	0%
minus LacZ fragment #4	0	35	35	0%
minus LacZ fragment #5	0	25	25	0%
minus LacZ fragment #6	0	20	20	0%
minus LacZ fragment #7	0	31	31	0%
minus LacZ fragment #8	0	26	26	0%
minus LacZ fragment #9	0	30	30	0%
minus LacZ fragment #10	0	51	51	0%
minus LacZ fragment #11	0	33	33	0%
minus LacZ fragment #12	0	30	30	0%
minus LacZ fragment #13	0	41	41	0%
minus LacZ fragment #14	0	55	55	0%
minus LacZ fragment #15	0	57	57	0%
minus LacZ fragment #16	0	80	80	0%
minus LacZ fragment #17	0	22	22	0%
minus LacZ fragment #18	0	21	21	0%
minus LacZ fragment #19	0	15	15	0%
minus LacZ fragment #20	0	26	26	0%
minus LacZ fragment #21	0	18	18	0%
minus LacZ fragment #22	0	114	114	0%
minus LacZ fragment #23	0	21	21	0%
minus LacZ fragment #24	0	17	17	0%
minus pGGA destination plasmid	0	0	0	0%
minus BsaI-HFv2 restriction enzyme	TMTC[1]	TMTC	TMTC	na[2]
minus T4 DNA ligase	0	6	6	0%
Positive Control; no omissions	219	27	246	89%
Positive Control; no omissions	195	14	209	93%
Positive Control; no omissions	148	10	158	94%

Supporting Table S8. Predicted high fidelity four-base overhang sets for use with Golden Gate assembly methods (based on 18 h at 37°C).

Set	Number of overhangs	Estimated fidelity	Overhang sequences
1	10	100%	CTTA, CTCC, ACTA, GGTA, TCCA, CGAA, AATG, AGCG, ATGG, AGAT
2	15	99.8%	AGAG, ACAT, GACA, AGCA, AATC, GGTA, CAAA, CCAA, AACG, CTGA, CCTC, ACGG, TCCA, CAGC, ACTA
3	20	99.3%	GACA, ACTA, CGGA, ATTA, AGAG, AACG, CCAA, GGTA, CTGA, AGGA, CAGC, ACGG, CAAA, GAAC, AGAT, CCTC, CTAC, AGCA, AATC, ATGA
4	25	98.5%	AGCA, GACA, GTAA, CAGC, AATC, ATAG, GAAC, ATGA, AACT, CAAA, CTTC, CGTA, ATTA, CTGA, TCCA, ACTC, AATG, GCGA, ACAA, AGGG, CTCA, ACCG, CCAA, GGTA, AGAT
5	30	97.2%	GTAA, AAGT, ATAG, GAAA, CCAG, AATC, ATGA, GCAC, GGTA, CGTC, ACCG, ACAA, GCCA, AGGG, AATG, CAAC, AACT, CACA, AGCA, ATTA, CGAA, GAGA, CTTA, CCGA, ACGC, AGAT, CTCC, CAGA, CCTA, TCCA
6	35	94.1%	TCCA, AACT, CGTA, GTAA, AAGC, CCGA, GGGA, GCAA, ATAG, ATCC, AAGA, CCAG, ATGA, AATC, AGAA, ACAT, CAGA, CTCA, CCTA, ACGA, GACA, ATTA, AGAC, CAAA, GGTA, CGAA, CCAC, GAAC, AGGG, AATG, ACTA, CTTC, ACCG, ACTC, AGCA

Supporting Table S9. Ligation fidelity substrate and ligation library sequences

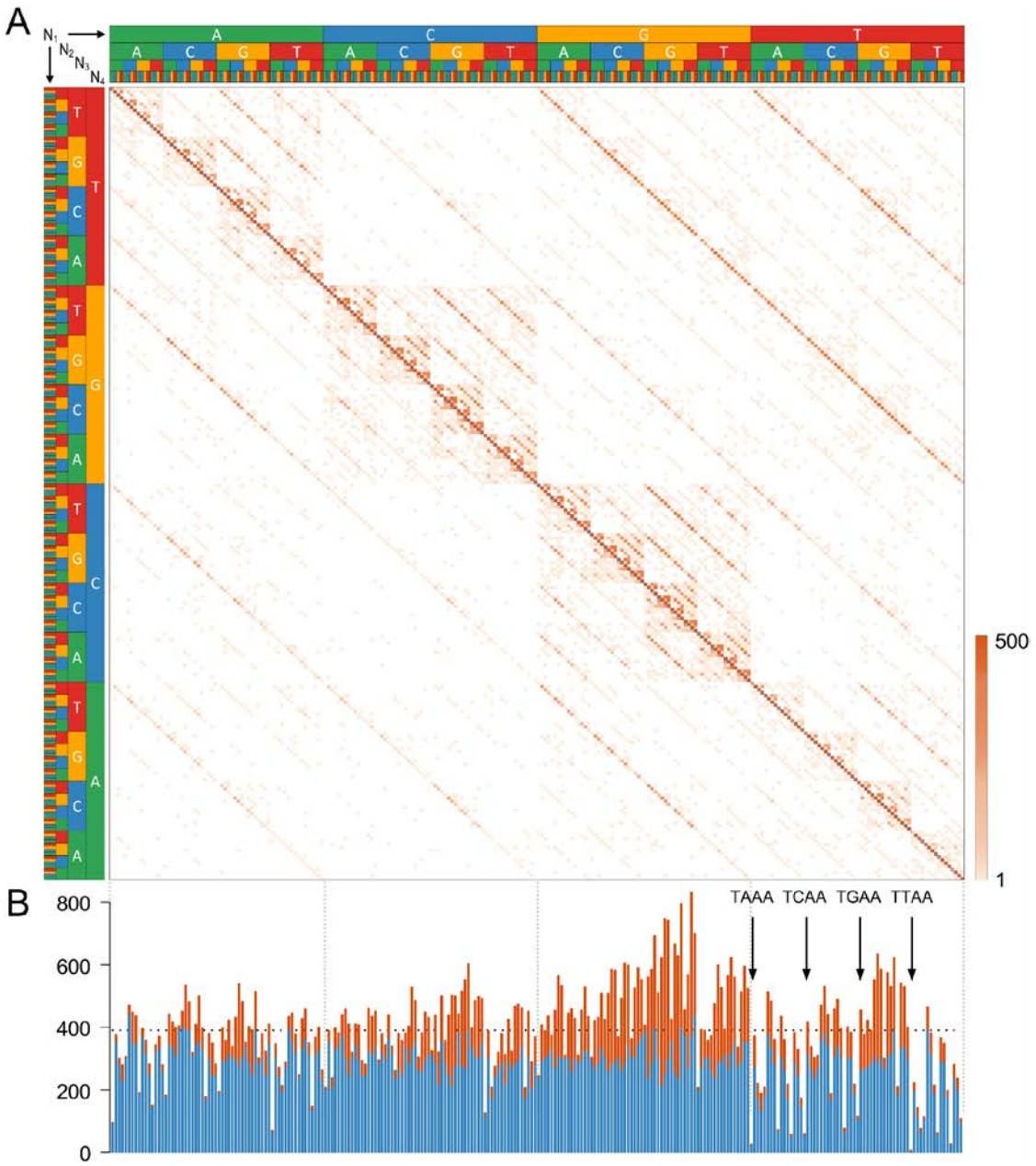
Substrate	Sequence [1]
Precursor oligonucleotide	TCACGTNNNNG**GAGACC**TGCGATCCAGTGCGCCGTCCATTGATC AACGNNNNNNCAAATCTCTCTCTTTTCCTCCTCCTCCGTTGTTGTT GTTGAGAGAG
Ligation profile substrate	pNNNNG**GAGACC**TGCGATCCAGTGCGCCGTCCATTGATCAACGN NNNNNCAAATCTCTCTCTTTTCCTCCTCCTCCGTTGTTGTTGTTGA GAGAGATTTGNNNNNNNCGTTGATCAATGGACGGCGCACTGGATC GCA**GGTCTC**C
Expected insert [2]	TTGNNNNNNCGTTGATCAATGGACGGCGCACTGGATCGCAGGTC TCCNNNNGGAGACCTGCGATCCAGTGCGCCGTCCATTGATCAAC GNNNNNNCAA

[1] The type IIS restriction enzyme recognition site is indicated in bold. SMRT adapter region is underlined.

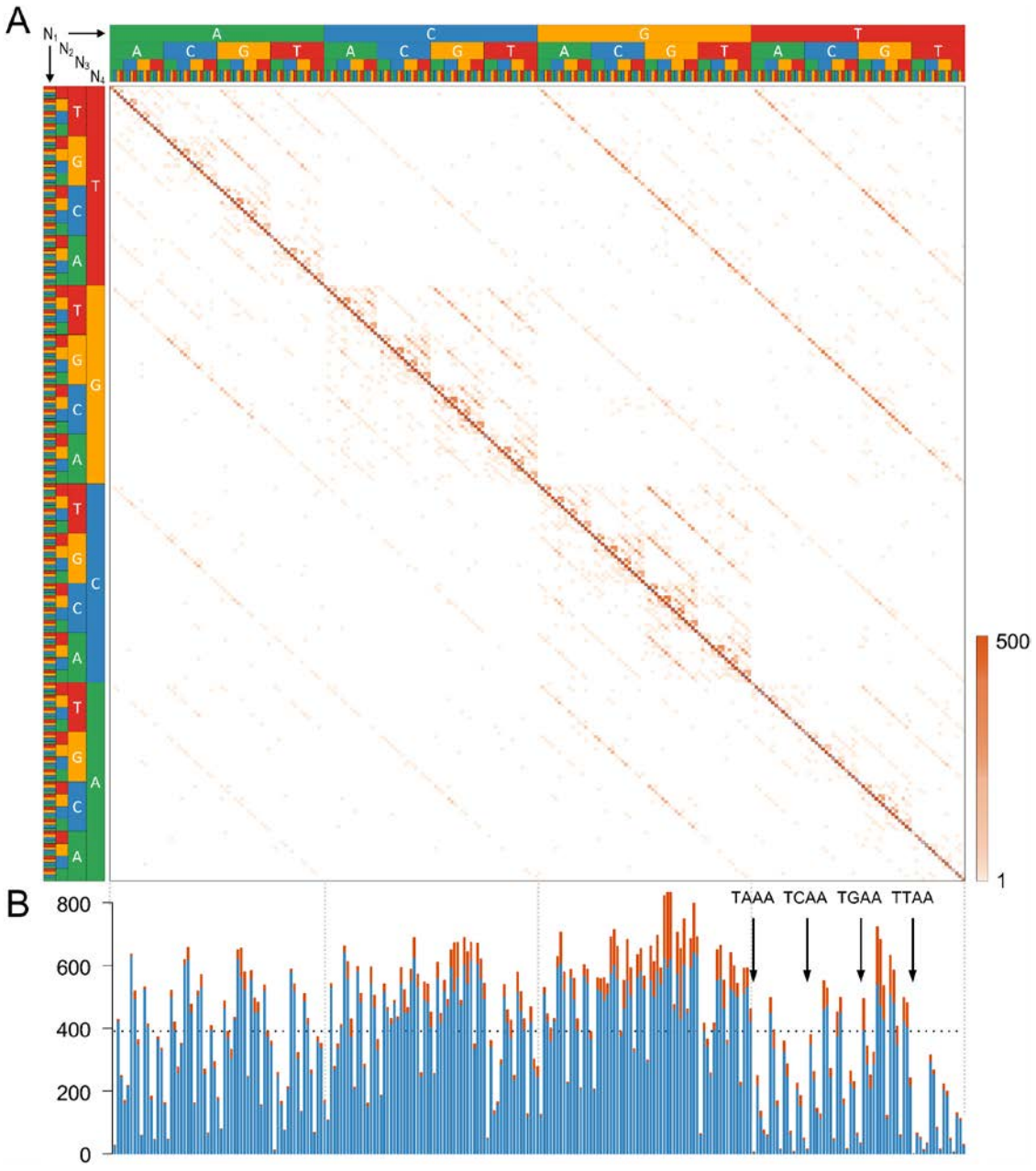
[2] The expected insert length is 99nt. The location of three-base overhang is in position 48..51, 3'-randomized region is in position 4..9, and 5'-randomized region is position 90..95.

Supporting Table S10. Number of ligation events per experiment.

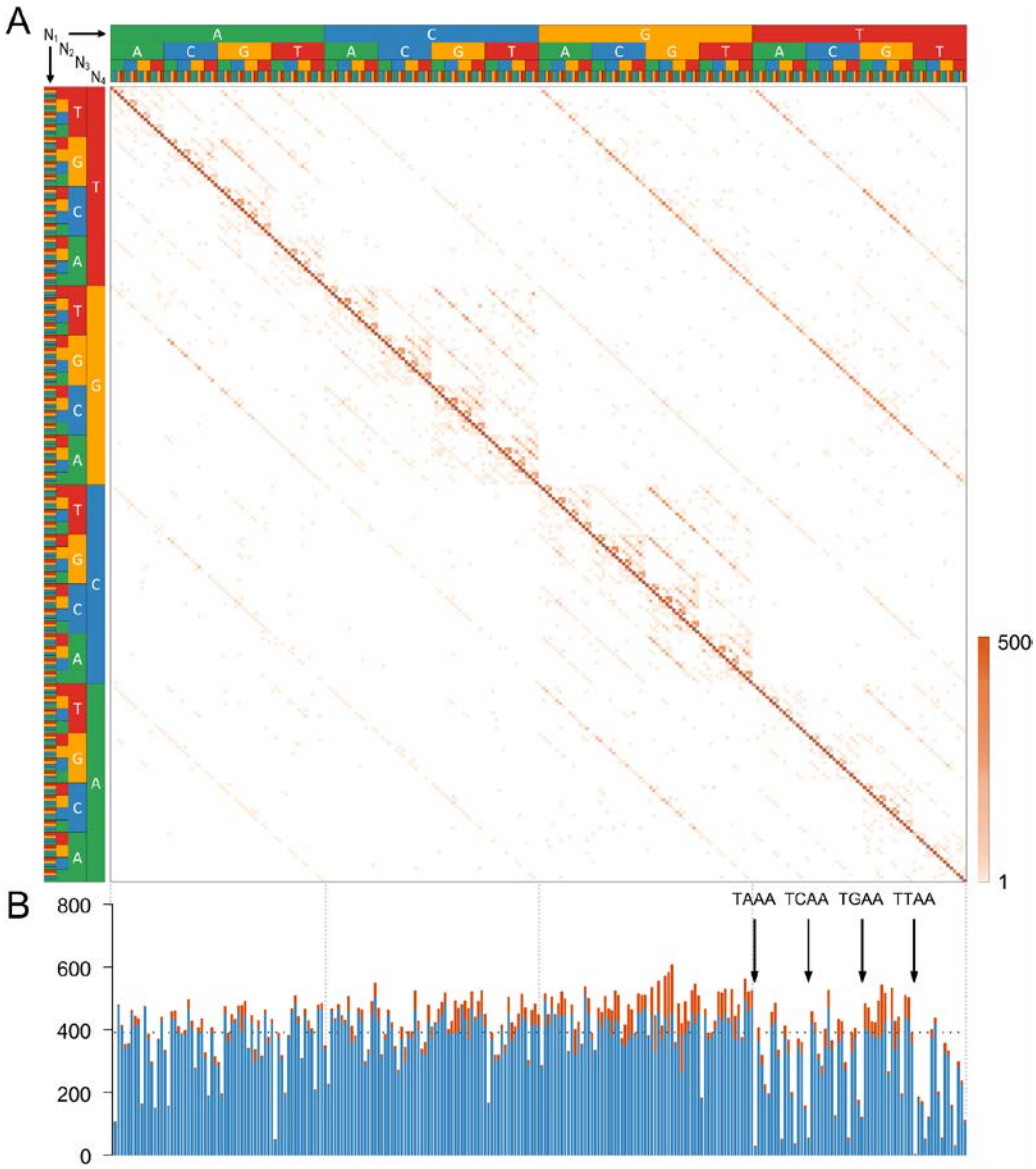
Time, h	Temperature, °C	Replicate	Number of runs	Number of ligation events	
				Per replicate	Per experiment
1	25	1	12	431,424	
1	25	2	2	83,954	
1	25	3	2	74,660	590,038
1	37	1	4	161,703	
1	37	2	2	87,857	
1	37	3	2	91,515	341,075
18	25	1	12	573,991	
18	25	2	2	84,322	
18	25	3	2	61,234	719,547
18	37	1	4	203,878	
18	37	2	2	86,171	
18	37	3	2	97,734	387,783



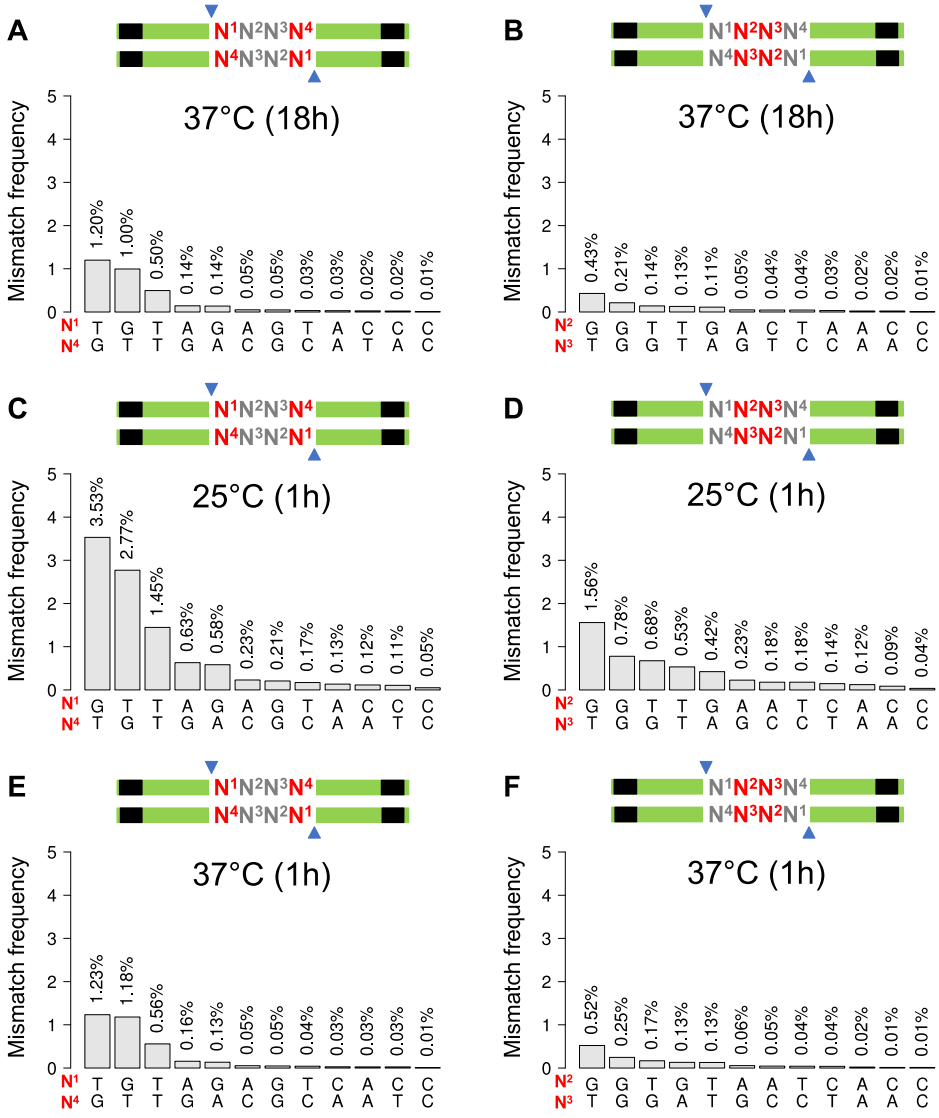
Supporting Figure S1. Ligation fidelity of T4 DNA Ligase for four-base overhangs (1 h at 25°C). SMRT sequencing results for ligating 100 nM of the multiplexed four-base overhang substrate 1 h at 25°C, with 1.75 µM T4 DNA ligase in standard ligation buffer. Observations have been normalized to 100,000 ligation events (see Supporting Information Data files for actual observation totals). (A) Frequency heat map of all ligation events (log-scaled). Overhangs are listed alphabetically left to right (AAAA, AAAC…TTTG, TTTT) and bottom to top such that the Watson-Crick pairings are shown on the diagonal. (B) Stacked bar plot showing the frequency of ligation products containing each overhang, corresponding to each column in the heat map in (A). Fully Watson-Crick paired ligation results are indicated in blue, and ligation products containing one or more mismatches are in orange. The dashed line indicates the median number of ligation events. The dashed line indicates the expected level of ligation if all overhangs appeared in equal frequency.



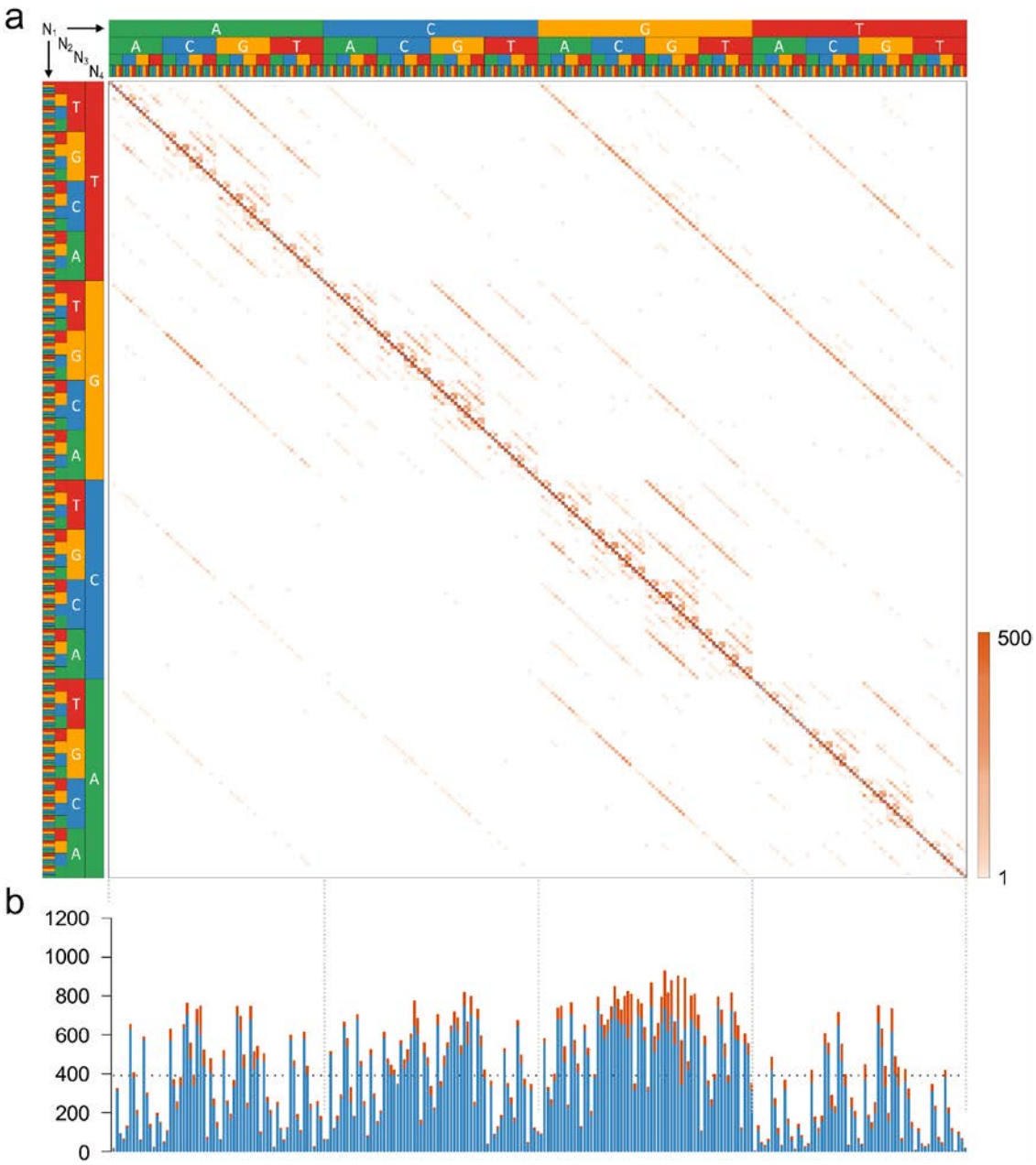
Supporting Figure S2. Ligation fidelity of T4 DNA Ligase for four-base overhangs (1 h at 37°C). SMRT sequencing results for ligating 100 nM of the multiplexed four-base overhang substrate 1 h at 37°C, with 1.75 µM T4 DNA ligase in standard ligation buffer. Observations have been normalized to 100,000 ligation events (see Supporting Information Data files for actual observation totals). (A) Frequency heat map of all ligation events (log-scaled). Overhangs are listed alphabetically left to right (AAAA, AAAC…TTTG, TTTT) and bottom to top such that the Watson-Crick pairings are shown on the diagonal. (B) Stacked bar plot showing the frequency of ligation products containing each overhang, corresponding to each column in the heat map in (A). Fully Watson-Crick paired ligation results are indicated in blue, and ligation products containing one or more mismatches are in orange. The dashed line indicates the median number of ligation events. The dashed line indicates the expected level of ligation if all overhangs appeared in equal frequency.



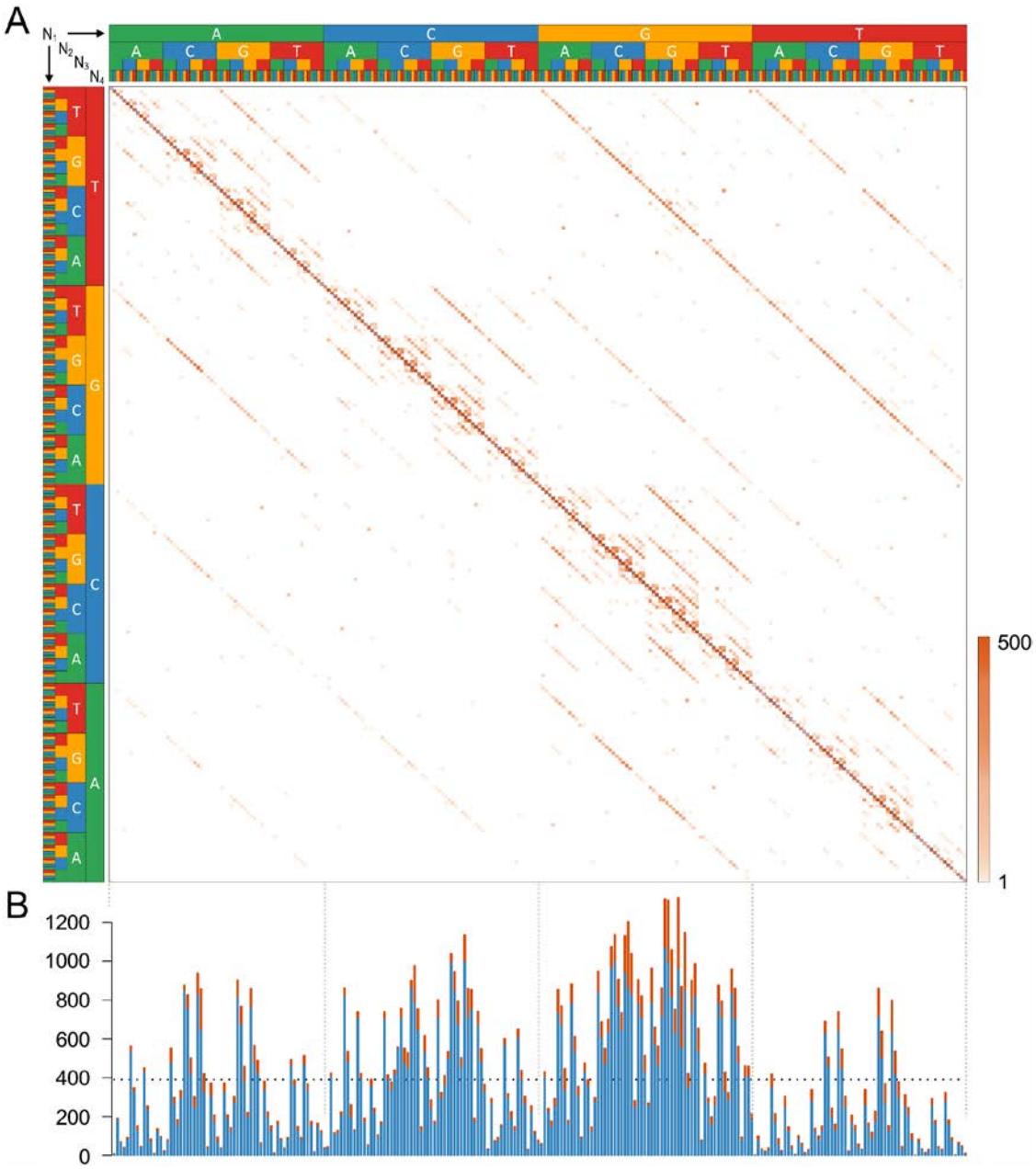
Supporting Figure S3. Ligation fidelity of T4 DNA Ligase for four-base overhangs (18 h at 37°C). SMRT sequencing results for ligating 100 nM of the multiplexed four-base overhang substrate 18 h at 37°C, with 1.75 μM T4 DNA ligase in standard ligation buffer. Observations have been normalized to 100,000 ligation events (see Supporting Information Data files for actual observation totals). (A) Frequency heat map of all ligation events (log-scaled). Overhangs are listed alphabetically left to right (AAAA, AAAC…TTTG, TTTT) and bottom to top such that the Watson-Crick pairings are shown on the diagonal. (B) Stacked bar plot showing the frequency of ligation products containing each overhang, corresponding to each column in the heat map in (A). Fully Watson-Crick paired ligation results are indicated in blue, and ligation products containing one or more mismatches are in orange. The dashed line indicates the median number of ligation events. The dashed line indicates the expected level of ligation if all overhangs appeared in equal frequency.



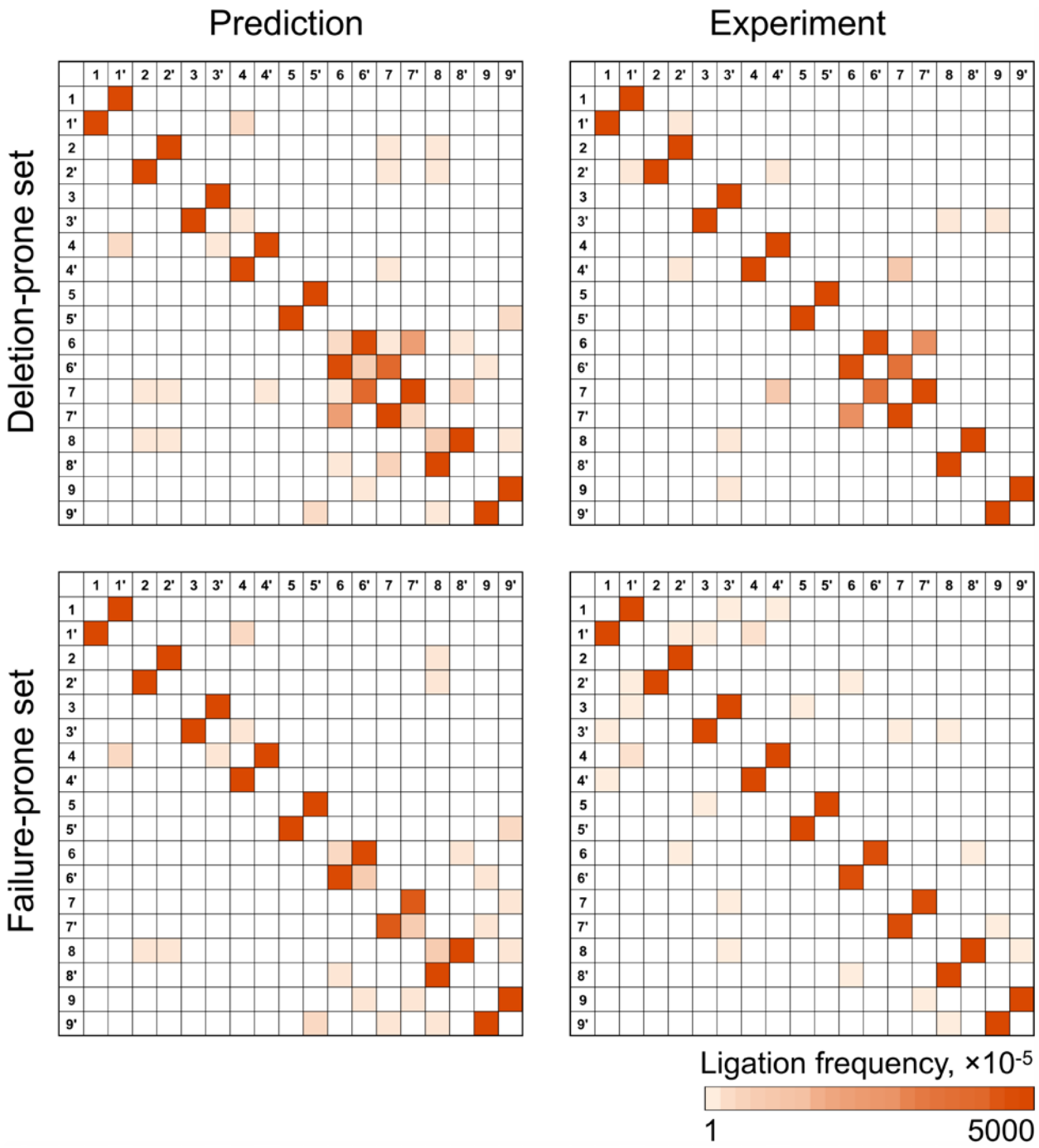
Supporting Figure S4. Frequency of specific base pair mismatches by position, additional ligation conditions. Incidence of each possible mismatched base pair observed during ligation of four-base overhangs, with 100 nM of the multiplexed substrate, 1.75 µM T4 DNA ligase, at varied temperatures and incubation times in standard ligation buffer. This figure was generated from the same data as shown in Figures S1, S2, and S3. (A, C, and E) show the results for the edge position (N1:N4′); (B, D, and F) for the middle position (N2:N3′).



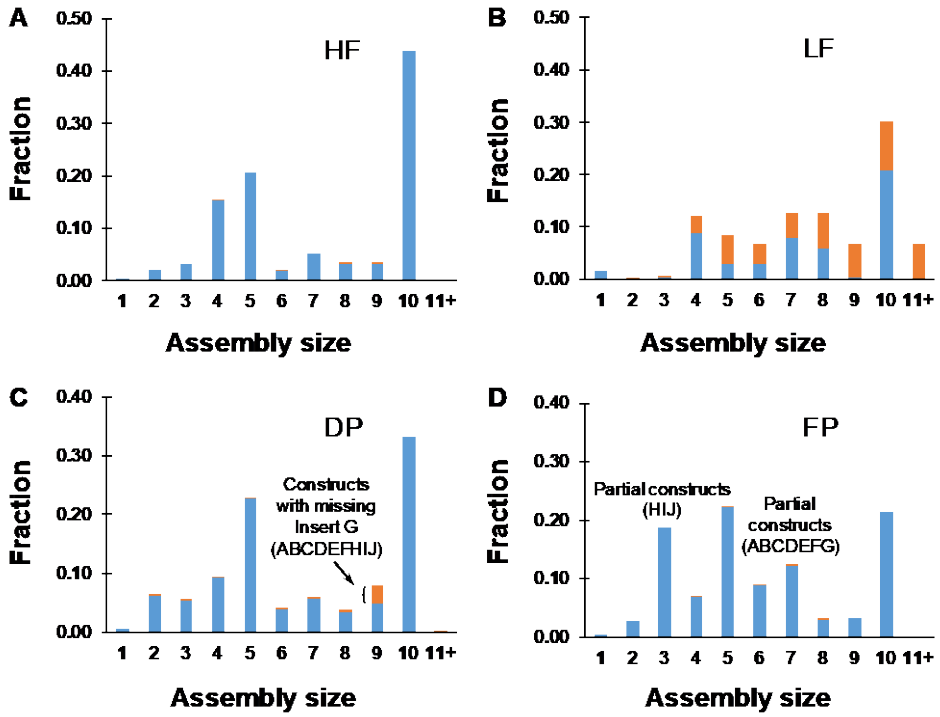
Supporting Figure S5. Ligation fidelity of T7 DNA Ligase for four-base overhangs (18 h at 25°C). SMRT sequencing results for ligating 100 nM of the multiplexed four-base overhang substrate 1 h at 37°C, with 1.75 µM T7 DNA ligase in standard ligation buffer. Observations have been normalized to 100,000 ligation events (see Supporting Information Data files for actual observation totals). (A) Frequency heat map of all ligation events (log-scaled). Overhangs are listed alphabetically left to right (AAAA, AAAC…TTTG, TTTT) and bottom to top such that the Watson-Crick pairings are shown on the diagonal. (B) Stacked bar plot showing the frequency of ligation products containing each overhang, corresponding to each column in the heat map in (A). Fully Watson-Crick paired ligation results are indicated in blue, and ligation products containing one or more mismatches are in orange. The dashed line indicates the median number of ligation events. The dashed line indicates the expected level of ligation if all overhangs appeared in equal frequency.



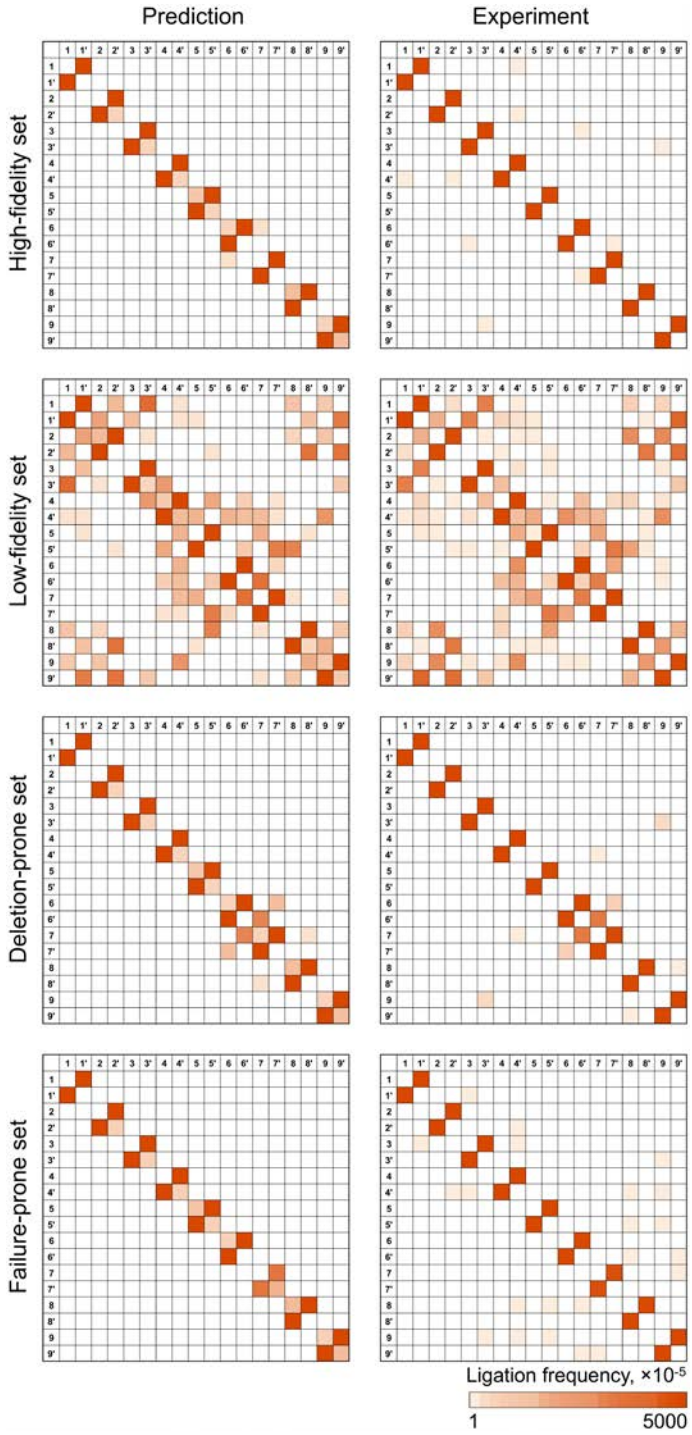
Supporting Figure S6. Ligation fidelity of T7 DNA Ligase for four-base overhangs (18 h at 37°C). SMRT sequencing results for ligating 100 nM of the multiplexed four-base overhang substrate 1 h at 37°C, with 1.75 µM T7 DNA ligase in standard ligation buffer. Observations have been normalized to 100,000 ligation events (see Supporting Information Data files for actual observation totals). (A) Frequency heat map of all ligation events (log-scaled). Overhangs are listed alphabetically left to right (AAAA, AAAC…TTTG, TTTT) and bottom to top such that the Watson-Crick pairings are shown on the diagonal. (B) Stacked bar plot showing the frequency of ligation products containing each overhang, corresponding to each column in the heat map in (A). Fully Watson-Crick paired ligation results are indicated in blue, and ligation products containing one or more mismatches are in orange. The dashed line indicates the median number of ligation events. The dashed line indicates the expected level of ligation if all overhangs appeared in equal frequency.



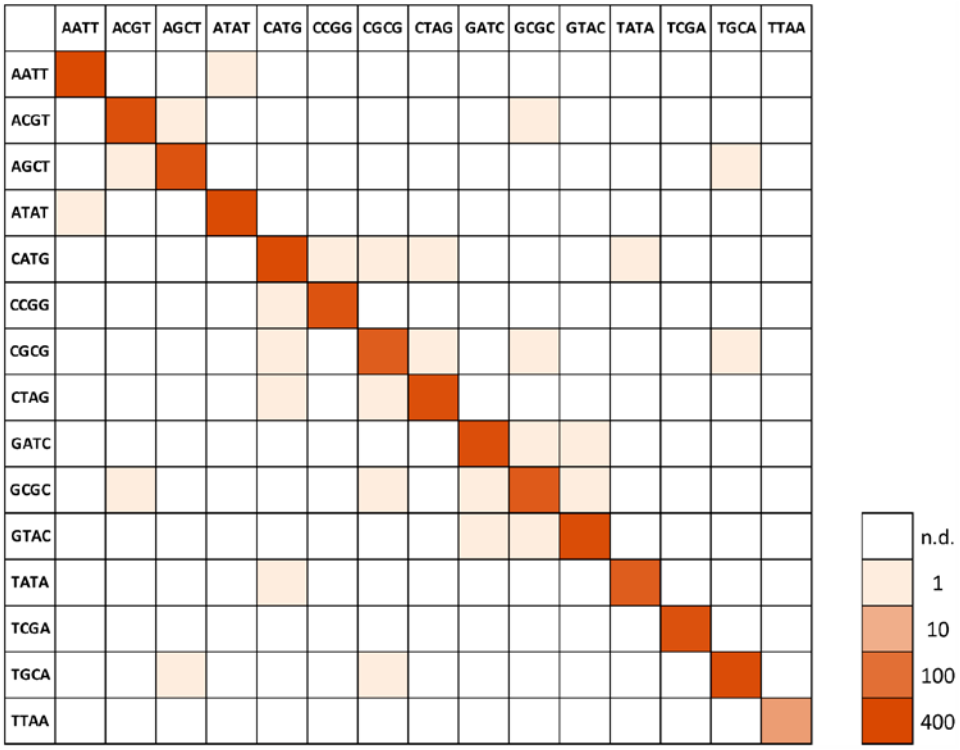
Supporting Figure S7. Predicted versus observed fragment linkages in a multi-fragment (Golden Gate) assembly for the DP and FP 10-fragment assemblies. Predicted frequencies of junctions are based on the fidelity library data generated for the four-base overhang substrate ligated with T4 DNA ligase at 25°C for 18 h. The experimental observations shown are for assembly of the 10-fragment HF and LF sets with Golden Gate Assembly mix, 37°C 5 min/16°C 5 min, 30 cycles.



Supporting Figure S8. Distribution of assembly sizes for the 10-fragment Golden Gate assemblies (18 h at 37°C).



Supporting Figure S9. Predicted versus observed fragment linkages in a multi-fragment (Golden Gate) assembly for the HF, LF, DP and FP 10-fragment assemblies. Predicted frequencies of junctions are based on the fidelity library data generated for the four-base overhang substrate ligated with T4 DNA ligase at 37°C for 18 h. The experimental observations shown are for assembly of the 10-fragment HF, LF, DP, and FP sets with Golden Gate Assembly mix (18 h at 37°C).



Supporting Figure S10. Predicted mismatch ligation potential between all palindromic overhangs. Palindromic overhangs generated by Type IIP restriction enzymes are used in traditional restriction enzyme cloning methods. The fidelity profile (18 h at 25°C) was used to predict any likely mismatch ligation between these overhangs. No significant cross talk is predicted, but the TATA and TTAA overhangs are expected to be low efficiency in ligation compared to all other palindromes.