Introduction

The majority of high-throughput (HTP) Illumina™ NGS library construction workflows are built around a common series of enzymatic reactions – end repair, A-tailing, adapter ligation, and PCR amplification – and this approach remains the most flexible, cost-effective, and reliable solution available. While the basic library construction process is relatively simple, integrating this process into a tailor-made, application-specific pipeline can be a daunting and time-consuming challenge because of the many options and parameters to be considered. This guide is intended to help you to develop, improve, and validate your NGS production protocols by providing detailed technical information on a number of critical parameters relating to the reagents and protocols supplied with the KAPA HTP Library Preparation Kit.

To accommodate a wide variety of automated library construction scenarios, the protocols provided with the KAPA HTP Library Preparation Kit allow for flexibility in workflow and in the specifics of liquid handling. This Technical Guide is therefore intended to assist you in taking a creative and constructive approach in developing and optimizing your high-throughput library construction process.

We gratefully acknowledge the invaluable questions, answers, and insights through which many of you – our customers – have contributed to this guide. All of the information presented here was generated by scientists at Kapa Biosystems, and we have made every effort to ensure that it is both relevant and accurate. Nevertheless, we accept that your results may differ from ours for reasons that we have not anticipated, and we look forward to your critical feedback.

Please contact support@kapabiosystems.com for additional information, or to alert us to any corrections, improvements, or additions that you would like to see in future versions.
1. Automated Liquid Handling

The library construction protocols described in the KAPA HTP Library Preparation Kit Technical Data Sheet (TDS) can be carried out manually, and most protocol development and validation work is usually done this way. Although it is possible to achieve moderately high sample throughput by using multi-channel pipettes and 96-well plates, automated liquid handling is indispensable for the majority of HTP NGS production lines, and automating a validated manual library construction protocol can represent a significant challenge.

In addition to increased sample throughput, automation may be expected to provide additional advantages such as improved reproducibility and process control. Nevertheless, automation may result in slightly compromised yields and/or size distributions when compared with manual library construction performed by a skilled, experienced, and attentive technician. Most often, these discrepancies can be minimized through careful selection of appropriate hardware and plasticware, and by optimizing liquid handling parameters such as aspiration speeds and volumes in automation scripts.

Kapa Biosystems does not supply automated liquid handling equipment, and we are therefore unfortunately not in a position to provide detailed platform-specific protocols or technical support. However, we continue to work in partnership with a number of automation solution providers and customers to validate our reagents for a variety of protocols on their platforms.

2. Reaction Master Mixes

The master mix setups described in the KAPA HTP Library Preparation TDS are intended to accommodate a wide variety of automated liquid handling platforms and configurations.

- If a different final reaction volume is required for any enzymatic step, the recommended enzyme volume should be maintained, while the volume of concentrated reaction buffer should be scaled to achieve a 1X final concentration in the reaction.
- The volume of water required for each reaction is usually included in the reaction master mix, thereby minimizing the number of pipetting steps. Alternatively, for reasons related to specific automated liquid handling platforms and deck setups, it may be advantageous to add some or all of the required water separately during reaction setup.
- The total volume of water added to the end repair reaction can be adjusted to accommodate a range of input DNA volumes from 1 µl to 58 µl.

2.1. Master mix stability

The master mixes described in the TDS are stable for up to seven days at 4 °C or at -20 °C, and are unaffected by up to 3 freeze/thaw cycles, as shown in Figure 1. For each of the enzymatic reactions, we tested the stability of two master mix formulations: one comprised the reaction buffer and enzyme, plus the total volume of water required for that reaction, while the other consisted of the buffer and enzyme only (Table 1). Master mixes with intermediate amounts of water are therefore expected to show the same stability as these two extremes.

<table>
<thead>
<tr>
<th>Component</th>
<th>Setup 1</th>
<th>Setup 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>End Repair Master Mix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>8 µl</td>
<td>0 µl</td>
</tr>
<tr>
<td>10X KAPA End Repair Buffer</td>
<td>7 µl</td>
<td>7 µl</td>
</tr>
<tr>
<td>KAPA End Repair Enzyme</td>
<td>5 µl</td>
<td>5 µl</td>
</tr>
<tr>
<td>Total master mix volume</td>
<td>20 µl</td>
<td>12 µl</td>
</tr>
<tr>
<td>Final reaction setup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fragmented DNA</td>
<td>50 µl</td>
<td>50 µl</td>
</tr>
<tr>
<td>Balance of water required</td>
<td>0 µl</td>
<td>8 µl</td>
</tr>
<tr>
<td>End Repair Master Mix</td>
<td>20 µl</td>
<td>12 µl</td>
</tr>
<tr>
<td>Total reaction volume</td>
<td>70 µl</td>
<td>70 µl</td>
</tr>
<tr>
<td>A-Tailing Master Mix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>42 µl</td>
<td>0 µl</td>
</tr>
<tr>
<td>10X KAPA A-Tailing Buffer</td>
<td>5 µl</td>
<td>5 µl</td>
</tr>
<tr>
<td>KAPA A-Tailing Enzyme</td>
<td>3 µl</td>
<td>3 µl</td>
</tr>
<tr>
<td>Total master mix volume</td>
<td>50 µl</td>
<td>8 µl</td>
</tr>
<tr>
<td>Final reaction setup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beads with end repaired DNA</td>
<td>0 µl</td>
<td>0 µl</td>
</tr>
<tr>
<td>Balance of water required</td>
<td>0 µl</td>
<td>42 µl</td>
</tr>
<tr>
<td>A-Tailing Master Mix</td>
<td>50 µl</td>
<td>8 µl</td>
</tr>
<tr>
<td>Total reaction volume</td>
<td>50 µl</td>
<td>50 µl</td>
</tr>
<tr>
<td>Adapter Ligation Master Mix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>30 µl</td>
<td>0 µl</td>
</tr>
<tr>
<td>5X KAPA Ligation Buffer</td>
<td>10 µl</td>
<td>10 µl</td>
</tr>
<tr>
<td>KAPA T4 DNA Ligase</td>
<td>5 µl</td>
<td>5 µl</td>
</tr>
<tr>
<td>Total master mix volume</td>
<td>45 µl</td>
<td>15 µl</td>
</tr>
<tr>
<td>Final reaction setup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beads with A-tailed DNA</td>
<td>0 µl</td>
<td>0 µl</td>
</tr>
<tr>
<td>Adapter (10 µM)</td>
<td>5 µl</td>
<td>5 µl</td>
</tr>
<tr>
<td>Balance of water required</td>
<td>0 µl</td>
<td>30 µl</td>
</tr>
<tr>
<td>Adapter Ligation Master Mix</td>
<td>45 µl</td>
<td>15 µl</td>
</tr>
<tr>
<td>Total reaction volume</td>
<td>50 µl</td>
<td>50 µl</td>
</tr>
</tbody>
</table>
3. Input DNA

Input DNA quality, fragment size, and concentration are important determinants of library construction success, and much of the need for protocol optimization arises from potential variations in input DNA. It may therefore be necessary to develop separate protocols for distinct sample types. Similarly, even within the context of a specific application using well-defined samples, it is often beneficial to ensure uniformity of input DNA concentration, fragment size, and quality as far as possible.

Depending on sample characteristics and the intended application, there are many ways in which input DNA may be normalized. Methods range from relatively cheap and simple assays to more complicated, expensive, and thorough procedures, and may include:

- Normalization of concentration using spectrophotometric or fluorometric (e.g. PicoGreen®) assays
- Size selection and/or size determination techniques
- qPCR-based assays for determining DNA quality and/or concentration (e.g. KAPA Human Genomic DNA Quantification and QC Kit)

3.1. Buffer composition

The enzymes in the end repair reaction require free Mg^{2+}. Input DNA should therefore be in water or low-salt, weakly buffered solutions containing little or no metal ion chelating agents such as EDTA or EGTA. The composition of 1X KAPA End Repair Buffer is as follows:

- 50 mM Tris-HCl (pH 7.5)
- 10 mM MgCl₂
- 10 mM DTT
- 1 mM ATP
- 0.4 mM each of dATP, dCTP, dGTP, and dTTP

Since EDTA chelates Mg^{2+} in a 1:1 molar ratio, it should be possible to supplement the end repair reaction with an appropriate amount of MgCl₂ to account for any EDTA added along with the input DNA.

3.2. Fragmentation

Over and above application-specific considerations related to sequence assembly and/or analysis, input DNA should ideally be fragmented to achieve a size distribution that approximates the range that will be targeted in any size selection procedures that you intend to apply. Any method of size selection will necessarily result in significant loss of library material, but DNA recovery will be maximized in cases where the selected size range overlaps best with the size distribution of the input material.

The KAPA HTP Library Preparation Kit is compatible with all standard fragmentation techniques including ultrasonic, mechanical, and enzymatic methods. Kapa Biosystems does not currently supply DNA fragmentation products, and we are therefore not able to provide detailed recommendations or protocols for DNA fragmentation.

In some cases, it may be advantageous to perform fragmentation in 1X KAPA End Repair Buffer, or to recover fragmented DNA in 1X KAPA End Repair Buffer if post-fragmentation cleanup is required (see section 3.4 below).

3.3. Input amount

The KAPA HTP Library Preparation Kit reagents and protocol have been optimized and validated for DNA inputs from 100 ng – 5 µg. Lower amounts of DNA should work well if the sample represents sufficient copies to ensure the requisite coverage and complexity in the final library. Based on our experience and feedback from customers, 50 – 100 ng of good quality human genomic DNA (~8 000 – 16 000 diploid copies) represents the minimum sample input for robust HTP whole genome sequencing. While it is certainly possible to obtain useful data from considerably less material, we have not yet specifically validated lower inputs for this kit.

The kit and associated protocols have been validated for standard genomic DNA sequencing, but with appropriate adjustments the kit can be integrated into NGS library construction workflows for a wide variety of applications, including:

- Exome capture (e.g. Roche Nimblegen™, Agilent SureSelect®, IDT xGen Lockdown™ Probes, etc.)
- RNA-seq
- ChIP-seq
- Methyl-seq (using the KAPA HiFi Uracil+ Library Amplification ReadyMix)

We recommend that adapter concentrations be adjusted to suit the expected range of input DNA (see section 5 below).
3.4. Input volume

Our standard recommended end repair reaction setup accommodates up to 58 µl of fragmented input DNA in water or weak buffer solution (see section 3.1 above), and requires a final reaction volume of 70 µl. However, if input DNA is limiting and your workflow and fragmentation method (e.g. Covaris AFA™) allow for this, it may be advantageous to fragment the DNA in 1X KAPA End Repair Buffer.

Alternatively, if your chosen fragmentation protocol (or other upstream process, e.g. PCR) requires DNA recovery and/or cleanup using Agencourt® AMPure® XP beads, you may consider eluting the DNA directly into 1X KAPA End Repair Buffer. These options allow for the addition of up to 65 µl input DNA in 1X KAPA End Repair Buffer in the recommended 70 µl end repair reaction.

Another option is to elute fragmented DNA directly into end repair master mix, thus performing with-bead end repair; these beads may then be retained and re-used throughout the remainder of the protocol.

Note that the end repair reaction setup should be adjusted to take account of any KAPA End Repair Buffer that is added along with input DNA, in order to ensure a 1X final concentration of DNA. Recommended adapter concentrations for various amounts of input DNA.

Table 2. Recommended adapter concentrations for various amounts of input DNA.

<table>
<thead>
<tr>
<th>Insert DNA per 50 µl ligation reaction</th>
<th>Adapter concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock</td>
<td>Final</td>
</tr>
<tr>
<td>3 - 5 µg</td>
<td>30 µM</td>
</tr>
<tr>
<td>1 µg</td>
<td>10 µM</td>
</tr>
<tr>
<td>500 ng</td>
<td>5 µM</td>
</tr>
<tr>
<td>100 ng</td>
<td>1 µM</td>
</tr>
</tbody>
</table>

The recommended final volume of the end repair reaction is 70 µl, in order to allow for the addition of 120 µl of KAPA PEG/NaCl SPRI (Solid Phase Reversible Immobilization) Solution – corresponding to ~1.7 volumes – without exceeding a standard well capacity of 200 µl during cleanup. A higher ratio of KAPA SPRI Solution may be used for cleanup by decreasing the total end repair reaction volume, and/or by increasing the volume of SPRI Solution, if this can be accommodated in the plate.

Once the DNA is bound to the beads during cleanup, the beads are washed with 80% ethanol to remove residual reaction components. Always use a freshly prepared ethanol solution, and store the solution properly while in use, as ethanol concentrations below 70% v/v will result in significant loss of bound DNA. While it is important to remove as much ethanol as possible before proceeding with subsequent reactions, avoid over-drying the beads, as this may also result in significant loss of DNA. Optimal drying times will vary, and depend on factors such as the volume of residual 80% ethanol left on the beads, the temperature, air-flow, and plasticware used.

5. Adapter Concentration and Post-Ligation Cleanups

In general, adapter ligation efficiency remains robust over a relatively wide range of adapter:insert molar ratios, making it unnecessary to adjust adapter concentrations for individual samples. It is nevertheless important to select an adapter concentration that is suitable for the range of DNA input amounts (Table 2). Excess adapter contributes to adapter-dimer formation and leads to increased carryover of adapter and adapter-dimer during the post-ligation cleanups. Conversely, too little adapter reduces ligation efficiency, and low adapter:insert molar ratios (approaching 2:1) result in catastrophic failure of library construction, because a significant proportion of insert molecules end up with an adapter ligated to only one end. Table 3 illustrates the effect of adapter concentration on the final yield of adapter-ligated library molecules for various amounts of input DNA.
Table 3. Library yields calculated by qPCR after two post-ligation cleanups for various adapter concentrations.

<table>
<thead>
<tr>
<th>Final adapter concentration</th>
<th>Yield of adapter-ligated fragments from various amounts of DNA input</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 µg</td>
</tr>
<tr>
<td>3 µM</td>
<td>1 189 ng</td>
</tr>
<tr>
<td>0.3 µM</td>
<td>463 ng</td>
</tr>
<tr>
<td>0.03 µM</td>
<td>N.D.</td>
</tr>
</tbody>
</table>

N.D. = Not detectable

Ultimately, the optimal adapter concentration for your specific workflow should be decided by considering trade-offs among the following factors:

- Cost
- Ligation efficiency
- Possible negative effects caused by adapter/adapter-dimer in your library
- The negative impacts on time, cost, and library yield of a second post-ligation cleanup that may be required to remove excess unligated adapter and/or adapter-dimer (see below)

The effects of adapter concentration on ligation efficiency and carryover of adapters and/or adapter-dimers is discussed in more detail in sections 5.1 to 5.3. For reference, Figure 2 shows unligated, sheared DNA and unligated Y-adapter as they appear on Bioanalyzer electropherograms generated with an Agilent High Sensitivity DNA Kit. Note that unligated Y-adapter gives rise to two distinct peaks on the electropherogram, presumably because of interactions between the single-stranded adapter arms.
5.1. One versus two consecutive cleanups

If your chosen library construction workflow includes a size-selection step between adapter ligation and library amplification (or cluster generation), it is likely that a single post-ligation cleanup using 1 volume of KAPA SPRI Solution will suffice, because residual adapter species are likely to be eliminated during size selection. Conversely, in the absence of post-ligation size selection, two consecutive cleanups may be required. Significant contamination of the library with adapters and/or adapter-dimers is likely to decrease the efficiency of library amplification and/or cluster generation/sequencing, and may produce unwanted artefacts.

Figure 3 shows DNA libraries prepared with the KAPA HTP Library Preparation Kit, starting with either 1 μg, 500 ng or 100 ng of sheared DNA, and using various final concentrations of adapter in the ligation reactions (as shown in Table 3). After ligation, samples were purified using either one or two 1X with-bead cleanups, as described in the KAPA HTP Library Preparation Kit TDS (i.e. 1 volume of KAPA SPRI Solution was added to 50 μl of adapter ligation reaction containing AMPure® XP beads).

While optimal adapter concentrations may render a second post-ligation cleanup unnecessary, two consecutive cleanups will reliably remove almost all detectable adapter species from most libraries. Excessively low adapter:insert ratios, approaching 2:1, do not support the ligation of adapters to both ends of most library molecules, and will result in unacceptably low library construction yields.

Figure 3. Effect of post-ligation SPRI cleanups on adapter carryover and ligation efficiency. Libraries were constructed using 100 ng – 1 μg of sheared DNA with 3 μM (blue), 0.3 μM (green) or 0.03 μM (red) adapter (final concentrations), and analysed using an Agilent Bioanalyzer High Sensitivity DNA Kit. Electropherograms show the presence or absence of adapters after a single post-ligation cleanup using 1 volume of KAPA SPRI Solution, or after two successive cleanups using 1 volume of KAPA SPRI Solution each. Failed adapter ligation is evident in the distinctly smaller size distributions of libraries made using very low adapter concentrations.
5.2. A single cleanup with lower ratios of KAPA SPRI Solution

If you will not be performing size selection between ligation and library amplification and you prefer to avoid a second post-ligation cleanup, it may be possible to improve the elimination of adapter species using a single cleanup by employing a lower ratio of KAPA SPRI Solution. While this approach is less effective than the two consecutive 1X cleanups described above, it may be sufficient in some workflows when used in conjunction with lower adapter concentrations.

When optimizing your post-ligation cleanup strategy, note that 1X KAPA Ligation Buffer contains a significant concentration of PEG 6000, which dramatically affects DNA binding to AMPURE® XP beads.

We used a low molecular weight DNA ladder to determine the cleanup conditions necessary to retain library fragments of various sizes. Figure 4 shows the DNA ladder alternatively formulated in 10 mM Tris-Cl (pH 8.0) or in 1X KAPA Ligation Buffer, before being cleaned up using a range of volumes of KAPA SPRI Solution. Results were analyzed by agarose gel electrophoresis.

To determine the optimal cleanup conditions for removal of adapter-dimers produced from a standard ~60 bp Y-adapter, we performed with-bead adapter ligation in the absence of insert DNA, and then cleaned up the ligation reactions using various volumes of KAPA SPRI Solution. After eluting captured DNA from the beads in water, we determined the adapter-dimer concentrations using an adapter-dimer–specific qPCR assay. The relative concentration of adapter-dimer in each eluate is shown in Figure 5, which indicates that adapter-dimer elimination was most effective in cleanups using less than ~0.6 volumes of KAPA SPRI Solution.

5.3. Summarizing post-ligation cleanups

The results presented above demonstrate that two consecutive 1X cleanups are the most effective means of eliminating adapter species, while a single 1X cleanup results in carryover of a significant amount of adapter and/or adapter-dimer. The relatively poor performance of a single 1X cleanup is presumably largely due to the presence of PEG 6000 in the ligation reaction, but this effect cannot be entirely counteracted by reducing the volume of KAPA SPRI Solution used in the cleanup. For a single post-ligation cleanup, 0.4 – 0.6 volumes of KAPA SPRI Solution is likely to be optimal for reducing adapter species, while delivering relatively high yields of library DNA. These conclusions are illustrated in Figure 6.
using it with valuable samples.

When optimizing your size selection procedures, consider the following:

- Consult Figure 7 and Table 4 to determine the approximate SPRI volumes required to bind the desired fragment sizes.

- Long single-stranded arms of Y-adapters ligated to library fragments significantly affect size-dependent binding to SPRI beads, as well as the perceived fragment size determined using an Agilent Bioanalyzer High Sensitivity DNA Kit (see below).

- In many cases a single-sided size cut may suffice to remove larger fragments only, because unwanted small fragments may possibly be removed during a standard or modified reaction cleanup at another point in the protocol.

- For dual-SPRI size selections, we recommend that the second size cut should be performed with at least 0.2 volumes of AMPure® XP beads. DNA recovery is dramatically reduced if the difference between first and second cuts is less than ~0.2 volumes. To increase the amount of DNA recovered, use more than 0.2 volumes of AMPure® XP beads for the second cut, but note that this may result in the recovery of smaller library fragments and a broader size distribution.

- To raise the upper size limit of the selected fragments, decrease the volume of KAPA SPRI Solution added in the first cut. To decrease the upper size limit of the selected fragments, increase the volume of KAPA SPRI Solution used in the first cut.

- To raise the lower size limit of the selected fragments, reduce the volume of AMPure® XP beads added in the second cut. To decrease the lower size limit of the selected fragments, increase the volume of AMPure® XP beads added in the second cut.

Figure 7. The molecular weight of retained DNA fragments depends on the volume of KAPA SPRI Solution used during cleanup. Low molecular weight DNA ladder fragments were found to be bound completely to AMPure® XP beads (red), bound partially (blue), or not at all (green) using various ratios of KAPA SPRI Solution.
Table 4. Guidelines for dual-SPRI size selection.

<table>
<thead>
<tr>
<th>Approximate sizes selected† (2nd cut - 1st cut)</th>
<th>First cut (KAPA SPRI Solution)</th>
<th>Second cut (+ AMPure® XP beads)</th>
</tr>
</thead>
<tbody>
<tr>
<td>250 - 450 bp</td>
<td>0.6 vol.</td>
<td>+ 0.2 vol. = 0.8 vol.</td>
</tr>
<tr>
<td>300 - 750 bp</td>
<td>0.5 vol.</td>
<td>+ 0.2 vol. = 0.7 vol.</td>
</tr>
<tr>
<td>200 - 350 bp</td>
<td>0.7 vol.</td>
<td>+ 0.2 vol. = 0.9 vol.</td>
</tr>
</tbody>
</table>

† The stated sizes are actual total lengths of library fragments including ~60 bp Y-adapters ligated to both ends, and the volumes are appropriate for unamplified library fragments with single-stranded, forked ends (see 6.2 below). These parameters are provided as guidelines only, and are likely to require additional optimization for each specific implementation of the protocol.

6.2. Y-adapters affect size selection and size determination

The partially single-stranded DNA ends of unamplified library fragments carrying Y-adapters retard their migration during electrophoresis using the Agilent Bioanalyzer High Sensitivity DNA Kit. For this reason, the actual fragment sizes of unamplified libraries may be significantly smaller than those determined using the Bioanalyzer. Moreover, the single-stranded DNA ends produced by ligation of Y-adapters appear to reduce the efficiency of binding to AMPure® XP beads, which may lead to the loss of smaller, desired fragments during SPRI-based size selection if SPRI volumes are not increased to compensate for this effect.

When the single stranded Y-adapter arms are resolved to complementary double-stranded DNA during library amplification, DNA fragments once again bind to AMPure® XP beads as expected, and their sizes can be accurately determined using a Bioanalyzer.

To investigate the discrepancy in Bioanalyzer size determination caused by single-stranded Y-adapter ends, we used a discrete 470 bp insert fragment that was ligated to standard ~60 bp Y-adapters, and then amplified by PCR. Unligated insert (470 bp), insert ligated to Y-adapter (~590 bp, with single-stranded ends), and PCR-amplified adapter-ligated fragment (~590 bp, fully double-stranded) were analyzed using a Bioanalyzer High Sensitivity DNA Kit, and by standard agarose gel electrophoresis (Figure 8).

Figure 8. Comparison of apparent fragment size by gel electrophoresis and the Agilent Bioanalyzer High Sensitivity DNA assay. Unligated 470 bp insert (top), adapter-ligated insert (middle) and amplified adapter-ligated insert (bottom) were visualised by gel electrophoresis (left) or with a Bioanalyzer High Sensitivity DNA Kit (right).

An apparently disproportionate size shift caused by Y-adapters is also seen when performing dual-SPRI size selection with a low molecular weight lambda DNA ladder. We digested lambda DNA with multiple restriction enzymes to generate a dense, low molecular weight ladder. The ladder was then used as input in our standard library construction protocol to add ~60 bp Y-adapters. Unligated and adapter-ligated ladder samples were subjected to dual-SPRI size selection targeting a variety of size distributions. The recovered DNA fragments were assayed using a Bioanalyzer High Sensitivity DNA Kit, which showed that the size-selected ladder fragments carrying Y-adapters appeared to be significantly larger than the equivalent unligated ladder fragments (Figure 9).
Figure 9. Size distributions of ligated and unligated low molecular weight DNA fragments using the Agilent Bioanalyzer. Unligated (left panel) and adapter-ligated (right panel) samples were subjected to identical dual-SPRI cleanup procedures prior to analysis with the Bioanalyzer High Sensitivity DNA Kit. Fragments ligated to Y-adapters resulted in significantly larger size distributions than could be expected from the small difference in molecular weight between the ligated and unligated DNA species.

To confirm whether the apparent shift in size distribution was entirely due to misrepresentation by the Bioanalyzer, we PCR-amplified adapter-ligated ladder DNA that had been size-selected, and then compared the size distributions of unamplified and amplified DNA (data not shown). Figure 10 summarises the results obtained when adapter-ligated ladder DNA is size-selected either before or after PCR amplification. Adapter-ligated DNA that is size selected before amplification shows a size distribution that is ~50 bp larger than that of adapter-ligated DNA that is size-selected after PCR-amplification, indicating that single-stranded Y-adapter ends do not fully contribute to size-dependent binding to AMPure® XP beads. Unamplified libraries carrying Y-adapters therefore require slightly higher PEG/NaCl concentrations during dual-SPRI size selection to achieve similar size distributions as those seen for fully double-stranded DNA such as PCR-amplified libraries. Note that the guidelines provided in Table 4 take all of the above factors into account.

Figure 10. Effects of size selection at different stages during library preparation on the apparent size distributions of fragmented DNA. Samples were analysed using the Agilent Bioanalyzer High Sensitivity Kit. Apparent size distributions are shown for digested lambda DNA (1), digested DNA which was size selected using dual selection with 0.6 and 0.8 volumes of SPRI solution (2), digested DNA which was adapter ligated and then size selected (3), digested DNA which was adapter-ligated, size selected and PCR amplified (4) as well as digested DNA which was adapter ligated, PCR amplified and then size-selected (5).

The effect that Y-adapters have on size selection is illustrated in Figure 11, where the adapter-ligated 470 bp fragment and the amplified adapter-ligated fragment were cleaned up using various ratios of KAPA SPRI Solution. The amplified fragment binds more strongly to the beads than the unamplified adapter-ligated fragment, presumably because of the single-stranded DNA at the ends of the unamplified fragment.

Figure 11. Binding efficiency of adapter-ligated or PCR-amplified DNA fragments to AMPure® XP beads. A 470 bp DNA fragment was ligated to adapters and then either not amplified (blue) or PCR amplified (red), prior to cleanup with a range of volumes of KAPA SPRI Solution. Unamplified, adapter-ligated fragments are bound less efficiently to AMPure® XP beads.
7. Library Amplification with KAPA HiFi HotStart ReadyMix

Most library construction workflows include one or more library PCR amplification steps to increase the proportion and total amount of adapter-carrying DNA fragments, and/or to add functional sequences (e.g. multiplexing indices or primer sequences) not provided by the adapters. Despite the very low amplification bias and high fidelity of KAPA HiFi HotStart DNA Polymerase, excessive library amplification should be avoided to minimize the following adverse effects:

- Increased duplicate reads
- Uneven coverage depth and sequence dropout
- Chimeric library inserts
- Nucleotide substitutions
- Heteroduplex formation due to thermocycling after substrate depletion, which interferes with library QC and/or quantification using electrophoretic or fluorometric assays (e.g. Bioanalyzer, PicoGreen™)
- High final library concentrations requiring large dilutions for qPCR-based library quantification

For these reasons, we recommend that the number of library amplification cycles be optimized, and minimized without jeopardizing a high success rate for your specific workflow. For most cases in which amplified library material is used directly for cluster amplification and sequencing, it should not be necessary to produce more than ~500 ng of amplified library DNA (~10 ng/µl in a 50 µl reaction). For libraries with an average fragment size of 350 bp, this equates to a molar concentration of ~50 nM, which is more than sufficient for library QC and cluster amplification purposes (typically, 2 nM library stock is required for most standard Illumina™ cluster amplification protocols). Nevertheless, some applications (e.g. targeted capture) may require larger amounts of amplified library material.

7.1. Library amplification yields

To provide some context for the purposes of comparison and troubleshooting during method development, we investigated library amplification using a range of input amounts and various numbers of PCR cycles. Note that size selection and other factors related to your particular workflow will determine the actual amount of adapter-ligated library DNA available for PCR amplification. The input amounts referenced in this section relate to the actual template input (measured by qPCR) used for library amplification, and should not be confused with total fragmented DNA input used for library construction.

Table 5 and Figure 12 provide typical library yields and concentrations following library amplification for various numbers of PCR cycles using 25 ng, 50 ng, 100 ng or 250 ng of template library DNA (quantified by qPCR using the KAPA Library Quantification Kit). Reaction setup and cycling parameters for the amplification reactions were as recommended in the KAPA HTP Library Preparation Kit TDS. Following a single cleanup with 1 volume of AMPure® XP beads to remove dNTPs and primers, amplified libraries were quantified by qPCR, NanoDrop™, Bioanalyzer and PicoGreen™.

Yields calculated by qPCR and NanoDrop™ correlate well across the full range of PCR cycles, reaching a plateau around 8 – 10 µg total yield per 50 µl reaction (Figure 12). Conversely, library concentrations determined using PicoGreen® or Bioanalyzer assays show a different pattern and appear to be less reliable for libraries amplified for greater numbers of PCR cycles.

In the case of NGS library amplification the PCR template material is typically extremely heterogeneous, comprising a large number of diverse library fragments. During early PCR cycles, each successive round of denaturation, primer annealing, and strand extension produces roughly double the number of full-length, double-stranded library fragments.

As thermocycling continues, primers and dNTPs are depleted until one or both become limiting, at which point denatured, single-stranded library molecules are no longer converted to double-stranded DNA via primer annealing and extension. In NGS library amplification reactions set up according to our recommended protocol, primers are typically depleted before dNTPs. When DNA synthesis can no longer take place because of substrate depletion, subsequent rounds of DNA denaturation and annealing result in the separation of complementary DNA strands followed by imperfect annealing to non-complementary partners. This presumably results in the formation of so-called “daisy-chains” (tangled knots), comprising large assemblies of improperly annealed, partially double-stranded, heteroduplex DNA.

Table 5. Library amplification yields for various template input amounts, using different numbers of PCR cycles.

<table>
<thead>
<tr>
<th>PCR cycles</th>
<th>Recovered yield per 50 µl reaction using various input amounts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25 ng</td>
</tr>
<tr>
<td>6</td>
<td>650 ng</td>
</tr>
<tr>
<td>8</td>
<td>2 200 ng</td>
</tr>
<tr>
<td>10</td>
<td>4 800 ng</td>
</tr>
<tr>
<td>14</td>
<td>8 500 ng</td>
</tr>
<tr>
<td>22</td>
<td>9 000 ng</td>
</tr>
</tbody>
</table>

Library amplification reactions were cleaned up using one volume of AMPure® XP beads and eluted in 50 µl. Concentrations were determined by qPCR using the KAPA Library Quantification Kit, and molar concentrations were calculated using an average fragment size of 550 bp.
The PicoGreen® assay is specific for double-stranded DNA and thus the single-stranded component of "daisy-chain" molecules in over-amplified samples leads to under-estimation of the DNA concentration (Figure 12). Quantification using a double-stranded DNA-specific Bioanalyzer assay (e.g. the Agilent High Sensitivity DNA Kit; see Figure 13) is also problematic for such samples, for the following possible reasons:

• As in the case of PicoGreen® and other dye-assisted fluorometric assays, it is unlikely that the equivalent masses of single- and double-stranded DNA produce the same amount of fluorescence.

• The tangled assemblies of heteroduplex library fragments migrate slowly during electrophoresis and are represented as high molecular weight species, leading to difficulties in setting baselines and identifying peaks (Figure 13 c - e). In extreme cases, DNA may be assembled into such large complexes that it apparently falls outside the range of the assay, and is not visible in the chromatogram (Figure 13 e).

• If fragment size is over-estimated then molar fragment concentrations calculated by the instrument will be correspondingly under-estimated, even assuming that the determined mass concentrations are accurate.

Apart from the discrepancies attributed to different quantification methods, there may be several reasons for suboptimal yields from library amplification reactions performed with KAPA HiFi HotStart ReadyMix. These include, but are not limited to, the following:

• **Suboptimal primer concentration.** Primer concentrations may be reduced to save costs, but it is important to use a primer concentration that is able to support exponential amplification throughout the required number of cycles. The typical yields given in Table 5 should be used as a guideline during optimization of primer concentration. Note that a number of artefacts may result if primer concentration becomes limiting during library amplification (see above).

• **Primer quality.** Primers stored in water are prone to degradation, especially with repeated freezing and thawing. This leads to a lower effective primer concentration in the reaction, which may not support exponential amplification throughout the desired number of cycles. Degraded primers also promote mispriming and spurious amplification products which lead to an overall reduction in reaction efficiency. We therefore recommend that primers should be stored in 10 mM Tris-Cl (pH 8.0 – 8.5).

• **Primer design.** Not all adapters contain all of the motifs required for (multiplexed) sequencing, and some of these are added during library amplification. This necessitates the use of long amplification primers with overhangs that are not complementary to the template. Whilst this is unavoidable in some workflows, it is recommended that the shortest possible primers with the highest degree of complementarity be used during library amplification. Long primers are more prone to intra- and inter-molecular interactions which compete with specific priming, reduce overall reaction efficiency, and may lead to excessive primer dimer formation.

• **PCR volume.** Some workflows require relatively large amounts of library DNA, and it is tempting to scale up PCR volumes to achieve higher yields of amplified library DNA. This approach often fails, because the thermal transfer (and consequently reaction kinetics and efficiency) in larger volumes differ from those in the standard 50 µl reaction volume, particularly if the level of liquid in the tube or plate is close to or above the level of the thermocycler block. A better strategy is to ensure that the reaction is fully optimized, and to perform multiple 50 µl reactions per sample if necessary.

• **Suboptimal annealing temperature.** An annealing temperature of 60 °C is optimal for most typical Illumina™ TruSeq® adapter/primer combinations, but the annealing temperature may have to be optimized for different adapter/primer designs. The best way to determine the optimal primer annealing temperature is to perform an annealing temperature gradient PCR (55 – 70 °C).

### 7.2. With-bead library amplification

While most PCR reagents commonly used for NGS library amplification are completely inhibited by AMPure® XP beads, KAPA HiFi Hotstart ReadyMix produces only slightly lower yields in the presence of beads (data not shown). Although with-bead library amplification is efficient and has been used successfully by some of our customers, we have not fully investigated the impact that SPRI beads might have on library amplification. We therefore recommend that you validate with-bead library amplification to ensure that there are no negative effects such as amplification biases or reduced fidelity, prior to implementing it in your process.

While AMPure® XP beads themselves do not significantly inhibit KAPA HiFi Hotstart DNA Polymerase, the associated PEG/NaCl solution and preservatives may do so. In typical with-bead library construction protocols, the beads will have been thoroughly washed prior to library amplification, in which case this should not be a concern.
Figure 12. Yield of PCR amplified libraries as determined by four quantification methods. Library amplification was performed with different numbers of PCR cycles according to the recommended protocol, using 25 ng (a), 50 ng (b), 100 ng (c), or 250 ng (d) of template library DNA. Reactions were cleaned up using one volume of AMPure® XP beads and eluted in 50 µl. Amplified library concentrations were determined by qPCR (KAPA Library Quantification Kit, red), NanoDrop™ (blue), PicoGreen® (purple), or Bioanalyzer (green).

Figure 13. Excessive library amplification leads to high molecular weight DNA heteroduplexes, or “daisy-chains”. Library amplification reactions using 25 ng of template DNA were cleaned up using one volume of AMPure® XP beads after library amplification for 6 cycles (a), 8 cycles (b), 12 cycles (c), 14 cycles (d) or 22 cycles (e). The amplified and cleaned up libraries were analysed using an Agilent Bioanalyzer DNA Kit.
When contemplating a new high-throughput sequencing pipeline, it is important to consider the quality and the amount of DNA available for each sample, and whether it will be possible to produce libraries of sufficient diversity to satisfy the requirements of the intended application. Later, during process development, optimization, or troubleshooting, it may be useful to track library yields and recoveries through various points in the workflow, in order to identify those steps that may require special attention.

In this section we have attempted to provide guidelines for assessing yields and recoveries at various steps in a generic workflow (Table 6).

Table 6. Typical recoveries throughout the library construction process.

<table>
<thead>
<tr>
<th>Step</th>
<th>Estimated loss (% of input into this step)</th>
<th>Estimated recovery (as a % of input into end repair)</th>
</tr>
</thead>
<tbody>
<tr>
<td>End repair and cleanup</td>
<td>0 – 20% (mass)</td>
<td>80 – 100% (mass)</td>
</tr>
<tr>
<td>A-tailing and cleanup</td>
<td>10 – 20% (mass)</td>
<td>70 – 90% (mass)</td>
</tr>
<tr>
<td>Adapter ligation and 1st post-ligation cleanup</td>
<td>Some mass gain is possible due to the addition of adapters. 40 – 60% of input molecules do not end up with both adapters.</td>
<td>70 – &gt;100% (mass) 25 – 50% (molecules)</td>
</tr>
<tr>
<td>2nd post-ligation cleanup</td>
<td>~5% (molecules)</td>
<td>20 – 45% (molecules)</td>
</tr>
<tr>
<td>Size-selection</td>
<td>Variable</td>
<td>N/A</td>
</tr>
<tr>
<td>Post-amplification cleanup</td>
<td>Variable</td>
<td>N/A</td>
</tr>
</tbody>
</table>

When troubleshooting, or evaluating the efficiency of your library construction process, consider the following points:

- Many customer-specific factors – both known and unknown – affect library construction efficiency. In particular, differences in sample quality and fragment size distribution can have dramatic impacts on final library yields.
- Library construction yield ultimately depends on two critical factors – the retention of input DNA through multiple cleanup and size-selection steps, and the combined efficiency of the end-repair, A-tailing, and adapter ligation reactions. While the simple retention of DNA mass may be monitored using spectrophotometry, fluorometry or electrophoresis (gels or chips), all of these methods are susceptible to inherent inaccuracies, in addition to those caused by molecular artefacts and/or interference from contaminants. In contrast, the combined efficiency of the enzymatic reactions is best determined by qPCR using adapter-specific primers, because only library fragments carrying adapters at both ends are quantified, and qPCR offers QC metrics that can be used to evaluate the reliability of the data. In this regard, note the following:
  - KAPA Library Quantification Kits can be used to quantify any type of Illumina™ library that is ready for cluster amplification. If the flow cell sequences are not yet present after adapter ligation (i.e. are only added during library amplification), it may still be possible to use the DNA Standards and qPCR MasterMix provided in the kit in combination with customized primers. Please contact support@kapabiosystems.com for more information in this regard.
  - Quantification methods that are not specific for adapter-carrying library fragments (e.g. those employing a NanoDrop™, a Bioanalyzer, or PicoGreen® dye) will return much higher post-ligation recoveries than adapter-specific qPCR. Using such quantification methods, ~100% of input DNA mass may be observed to be retained through “with-bead” end repair, A-tailing, adapter ligation, and the intervening cleanups. However, using adapter-specific qPCR, we estimate that 25 – 50% of input DNA fragments are typically converted to adapter-ligated library molecules.
  - While a conversion rate of 25 – 50% of input DNA to library molecules may seem low, this represents a significant improvement over the traditional approach. In our experience, when end repair, A-tailing and adapter ligation are performed in separate tubes and cleanups are performed using fresh beads after each step, an estimated 10 – 20% of input DNA fragments are recovered with adapters at both ends. The with-bead strategy therefore facilitates the construction of more diverse ("complex") libraries from lower amounts of input DNA.
• Library construction yields are significantly affected by the fragment size distribution of the input DNA, cleanups, and size selection:
  - Small DNA fragments are preferentially lost even in standard cleanups using AMPure® XP beads, while very large library fragments are unlikely to amplify efficiently during qPCR or cluster amplification. If your input DNA has a wider-than-normal size distribution, with long tails at either end, you will experience reduced yields (see Sections 4, 5 and 6).
  - A large discrepancy between input and selected size distributions inevitably results in a dramatic loss (>>90%) of input molecules. Similarly, narrower size cuts result in larger losses than broader size cuts.
  - Different methods of size selection usually provide different levels of stringency, as well as very different yields.
• When tailoring a library construction process for a particular application, it may be helpful to “work backward” from the amount of material required for sequencing:
  - Notwithstanding standard protocols for cluster amplification with particular Illumina™ instruments, special procedures for cluster amplification using low concentrations of library DNA may be available, particularly with respect to so-called “no-PCR” or “PCR-free” sequencing protocols.
  - If you are performing any process between library construction and sequencing (e.g. capture), consider the input requirements for that process to determine the amount of DNA required for library construction.
  - As a point of reference, the KAPA HTP Library Preparation Kit is used routinely for whole-genome re-sequencing from 100 ng of high quality human gDNA, corresponding to ~16 000 diploid copies. For whole-genome sequencing of smaller genomes, or for samples of lower complexity (e.g. ChIP-seq), successful library construction is possible from considerably lower amounts of input DNA.

9. Useful Publications and Resources


An extremely valuable NGS community forum can be found at: http://www.seqanswers.com

While we are not in a position to make specific recommendations for particular sequencing projects, we hope that this guide has provided you with some of the information needed to make those decisions. Your feedback on the performance of our kits and protocol in different types of sequencing projects, or protocol improvements that you are willing to share with others in the community is always greatly appreciated!
Restrictions and Liabilities

This Technical Guide is provided “as is” and Kapa Biosystems assumes no responsibility for any typographical, technical, or other inaccuracies. The document is subject to being changed, without notice, in future editions.

To the maximum extent permitted by applicable law, Kapa Biosystems disclaims all warranties, either express or implied, with regard to this guide and any information contained herein, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. Kapa Biosystems shall not be liable for errors or for incidental or consequential damages in connection with the furnishing, use, or performance of this document or of any information contained herein.

This document might contain references to third party sources of information, hardware or software, products, or services and/or third party web sites (collectively the “Third-Party Information”). Kapa Biosystems does not control, and is not responsible for, any Third-Party Information. The inclusion of Third-Party Information in this document does not imply endorsement by Kapa Biosystems of the Third-Party Information or the third party in any way.

Kapa Biosystems does not in any way guarantee or represent that you will obtain satisfactory results from using Kapa Biosystems products as described herein. You assume all risk in connection with your use of Kapa Biosystems products.

Kapa Biosystems is not responsible nor will be liable in any way for your use of any software or equipment that is not supplied by Kapa Biosystems in connection with your use of Kapa Biosystems products.

Note to Purchaser: Limited License

KAPA Library Preparation Kits are developed, designed and sold exclusively for research purposes and in vitro use. Neither the product, nor any individual component, has been tested for use in diagnostics or for drug development, nor is it suitable for administration to humans or animals. Please refer to the MSDS, which is available on request.

Certain applications of this product are covered by patents issued to parties other than Kapa Biosystems and applicable in certain countries. Purchase of this product does not include a license to perform any such applications. Users of this product may therefore be required to obtain a patent license depending upon the particular application and country in which the product is used.

Licensed under U.S. Patent nos. 5,338,671 and 5,587,287 and corresponding patents in other countries.

PicoGreen® is a registered trademark of Life Technologies. NanoDrop™ is a trademark of Thermo Fisher Scientific. Illumina™ and TruSeq® are trademarks of Illumina. Nimblegen™ is a registered trademark of Roche. SureSelect® is a registered trademark of Agilent. Lockdown™ is a trademark of IDT. AMPure® is a registered trademarks of Beckman Coulter. AFA™ is a trademark of Covaris.

For support please contact: support@kapabiosystems.com

Boston, Massachusetts, United States
Int. Tel: +1 781 497 2933 Tel: +1 855 KAPA BIO
Fax: +1 781 497 2934
Email: sales@kapabiosystems.com

Cape Town, South Africa
Tel: +27 21 448 8200 Fax: +27 21 448 6503
Email: sales@kapabiosystems.com