Molecular Cell, *31*

Supplemental Data

A Biochemically Defined System for Coding Joint Formation in V(D)J Recombination

Haihui Lu, Noriko Shimazaki, Prafulla Raval, Jiafeng Gu, Go Watanabe, Klaus Schwarz, Patrick C. Swanson, and Michael R. Lieber

SUPPLEMENTARY RESULTS

DNA-PKcs Regulates the Activity of the RAG Complex and XRCC4:DNA Ligase IV

The RAG cleavage assay showed a small but consistent increase in both nicking and hairpin activity by DNA-PKcs (Suppl. Fig. 2). Although this would not be relevant to the first recombination event at a given chromosomal antigen receptor locus, an activated RAG complex would be relevant to subsequent recombination attempts, such as V to DJ recombination. The activity of XRCC4:DNA ligase IV is also affected by DNA-PKcs. DNA-PKcs greatly improves intermolecular ligation, also shown as an obvious decrease in unligated substrate (Suppl. Fig. 3A, lanes 2 & 3 vs lane 1). We confirmed phosphorylation of XRCC4 by DNA-PKcs (Leber et al., 1998) and found that autophosphorylation of DNA-PKcs greatly increased when XRCC4:DNA ligase IV was present (Suppl. Fig. 3B).

Coding Joint Sequences Showed Characteristics Similar to Junctional Sequences From in vivo V(D)J Recombination

After coding joint PCR products were run on gels, regions of coding joint products from 4 different sets of substrates were cut out and sequenced (Fig. 4). For reactions containing pol mu, pol lambda, and TdT, ~95% of the coding joints had features that corresponded to in vivo coding joints regarding P nucleotides, N nucleotide addition, and nucleolytic resection. The remaining ~5% were PCR products due to nonspecific annealing of the substrates. (Without polymerases, these events were more common because the coding joint formation was not as

efficient.) The substrates with fully compatible coding end overhangs yielded a disproportionate percentage of precise coding joints when polymerases mu, lambda, and TdT were not present. With the addition of these three polymerases, more deletion and addition were observed (Fig. 4A), similar to what is seen in vivo.

The substrates sharing 3 bp of homology have 1 nt gaps on the top and bottom strands, and these gaps are filled by pol mu or pol lambda in the most abundant coding joints generated in this system. The remainder of the products showed short deletions and additions (Fig. 4B). P nucleotides were found in 29% of the sequences, mostly 1 or 2 nt long, with one case of 4 nt P nucleotides. The 2 bp and 1 bp homology substrates yielded coding joints with higher diversity, as expected (Fig. 4C & D). P nucleotides were present in 26% of coding joints from the 2 bp microhomology substrate, and 67% of the coding joints for the 1 bp microhomology substrate. For the 1 bp substrate, one coding joint has 5 bp P nucleotides from the 12- coding end.

SUPPLEMENTARY DISCUSSION

Limitations of the Biochemically Defined System

In any type of PCR detection, one cannot determine the precise nature of the junctions at the time of the PCR. For example, one strand might be ligated rather than both. An unligated strand might include any of a number of configurations (gaps or flaps, for example).

We add XRCC4:DNA ligase IV after an initial 20 min incubation of all of the other proteins with the recombination substrates. The ligase complex is relatively quickly inactivated at 37°C, and therefore, there is no point to add it prior to any RAG cleavage events, and this is the reason for waiting 20 min. We add the ligase complex two subsequent times over a 2 hr incubation to permit increased ligation.

We have discussed the issue of Ku- and XLF-independence in the main text.

SUPPLEMENTARY METHODS

Oligonucleotides

45 bp double hairpin substrate: HL-157 (5'- TGATCGTGCT CGACTTCAGA GGCCTCTGAA GTCGAGCACG ATCAGCTCTC GAGACTGTAG TCTAGCATGC TAGACTACAG TCTCGAGAGC -3'). 86 bp double hairpin substrate: HL-150 (5'- GCTCGACCTC AGAGGCCTCT GAGGTCGAGC ACGATCAGCT TGAACTGTAG TCTAGCAC -3'), HL-151 (5'- CCGGGCTACC GTCAAGTAAG ATGCAGATAC GGAACACAGC TGGCCAGCTG TGTTCCGT -3'), and HL-152 (5'- ATCTGCATCT TACTTGACGG TAGCCCGGGT GCTAGACTAC AGTTCAAGCT GATCGT -3'). V(D)J recombination assay 12-substrate: HL-133 (5'- GGGTGCTAGA CTACAGTTCA AGCTGATCGT GCTCGACCTC AGAGGCACAG AGCTACAGAC TGGAACAAAA ACCCTGCAGT CGACTCTC[biotinTEG] -3'), and HL-134 is the complementary strand with the same biotinTEG at the 3' end. V(D)J recombination assay 23-substrate: HL-135 (5'- GGGCTACCGT CAAGTAAGAT GCAGATACGG AACACAGCTG GCACAGAGGT AGTACTCCAC TGTCTGGCTG TACAAAAACC CTCGGGATCC TCTCC[biotinTEG] -3'), and HL-136 is the complementary strand with the same biotinTEG at the 3' end. Coding joint PCR primers: HL-66 (5'-CTACCGTCAA GTAAGATG -3') and HL-68 (5'- CTAGACTACA GTTCAAGC -3'). The blue 12-substrate used in suppl. fig. 7: HL-163 (5'-GGGAAACAGC TATGACCATG ACGCCAAGCT TGGGCTGCAG GTCGGCACAG AGCTACAGAC TGGAACAAAA ACCCTGCAGT CGACTCTC [biotinTEG] -3'), and HL-164 is the complementary strand with the same biotinTEG at the 3' end. The coding joint primer for this blue 12-substrate (Suppl. Fig. 7) is HL-148 (5'- GGAAACAGCT ATGACCATG -3'). The V(D)J substrate for coding end exchange (Suppl. Fig. 8) is a purified PCR fragment from plasmid pHL-4 and pHL-6, both of which derived from plasmid pC139 (Sheehan and Lieber, 1993), with the 12-coding end last two base pairs changed from "AC" to "CC". pHL-6 was mutated 2 bp on each coding end to create two XhoI site, and the mutagenesis primer sequences are available upon request. The PCR primers used to amplify the substrates from the

plasmids are: HL-146 (5'- [biotinTEG]CAGGAAACAG CTATGACCAT GA -3') and HL-147 (5'- [biotinTEG]GCTTCCTTAG CTCCTGAAAA TCTC -3')

RAG Cleavage Assay and Ligation Assay

V(D)J recombination assay substrates HL-133/134 and HL-135/136 were used in the RAG cleavage assay, and the 12-substrate radioactively was labeled with P^{32} at the 5' end of HL-133. 1 pmol each of the 12- and 23- substrate was incubated with 1.5 pmol (+) or 3 pmol (++) GST-core RAG1/2, 5 pmol of HMGB1 and 0.5 pmol of DNA-PKcs in a total volume of 10 µl of reaction buffer (25 mM HEPES, pH 7.5, 60 mM KCl, 5 mM $MgCl_2$, 250 uM ATP, 1 mM DTT) at 37ºC for 1.5 hr, 0.8 pmol of Ku was added where indicated. The reactions were deproteinated with 0.5% SDS and 1mg/ml proteinase K at 55ºC overnight, extracted with phenol/chloroform and precipitated with ethanol. The products were resolved on 7 M urea-10% PAGE. The ligation assay has been described (Lu et al., 2007).

Test for Coding End Exchange

The V(D)J recombination assay to test for exchange of the coding ends (Suppl. Fig. 8) has the following modifications from the assays described in earlier sections. The DNA substrate is incubated first with the RAG complex, HMGB1, Artemis and DNA-PKcs at 37°C for 1.5 hr to allow RSS synapsis and cleavage. Then XRCC4:DNA ligase IV are added for another 20 min incubation at 37°C for coding end ligation. For the reactions with both substrates, each substrate was individually incubated in a separate tube for the first 1.5 hr for RSS synapsis and cleavage. Then the two reactions were mixed gently for the 20 min ligation. After the V(D)J recombination, the reaction was digested with ClaI which cuts between the two signals to remove most unreacted substrates. The PCR and denaturing PAGE was carried out the same way as described in main text.

REFERENCES

Leber, R., Wise, T.W., Mizuta, R., and Meek, K. (1998). The XRCC4 gene product is a target for and interacts with the DNA-dependent protein kinase. J Biol Chem *273*, 1794-1801.
Lu, H., Pannicke, U., Schwarz, K., and Lieber, M.R. (2007). Length-dependent binding of human XLF to DNA and stimulation of XRCC4.DNA ligase IV activity. J Biol Chem *282*, 11155-11162.
Ma, Y., Lu, H., Tippin, B., Goodman, M.F., Shimazaki, N., Koiwai, O., Hsieh, C.-L., Schwarz, K., and Lieber, M.R. (2004). A Biochemically Defined System for Mammalian Nonhomologous DNA End Joining. Molecular Cell *16*, 701-713.
Raval, P., Kriatchko, A.N., Kumar, S., and Swanson, P.C. (2008). Evidence for Ku70/Ku80 association with full-length RAG1. Nucleic Acids Res. (in press)
Sheehan, K.M., and Lieber, M.R. (1993). V(D)J recombination: Signal and coding joint resolution are uncoupled and depend on parallel synapsis of the sites. Mol Cell Biol *13*, 1363-1370.

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. Coomassie Staining of Purified V(D)J Recombination Proteins

Purified proteins were fractionated by SDS-PAGE (6% for DNA-PKcs and 8% for all the other gels). The gels were stained with Coomassie Blue R250. The name of each protein and the sizes of the protein standards (in kDa) are labeled on the side. Each set of 2 or 3 lanes of the specified protein covers a range of concentrations. The purity of Ku70/Ku86, polymerase mu and lambda were shown previously (Ma et al., 2004), as were the purity of MBP tagged full-length and core RAG proteins (Raval et al., 2008) and XLF (Lu et al., 2007).

Supplementary Figure 2. The Nicking and Hairpinning Activity of the RAG Complex are Increased by DNA-PKcs.

A. The configuration of the12-substrate, nicking and hairpinning products are illustrated on the right, next to the corresponding bands of the gel. Two concentrations of the RAG complex 150 nM (+) and 300 nM (++) are used for the left and right panel respectively. The nicking and hairpin opening products are both increased in the reactions where DNA-PKcs is added, and this increase is observed with both concentrations of RAG complex. We do not know the mechanism of the DNA-PKcs stimulation of the RAG activity, and it may not require phosphorylation.

B. The nicking and hairpin efficiencies of the RAG complex without DNA-PKcs (Panel A, lanes 5 & 8), and with DNA-PKcs (Panel A, lanes 3, 4 and 6, 7) are quantified and plotted. Addition of DNA-PKcs increases the nicking and hairpin efficiencies by 1.6 ~ 2.6 fold.

Supplementary Figure 3. DNA-PKcs Phosphorylates XRCC4 and Stimulates Ligation by XRCC4:DNA Ligase IV, Which then Improves the Autophosphorylation of DNA-PKcs.

A. The DNA substrate for the ligation reactions has two identical double-strand ends with 4 bp overhangs that are fully compatible. Intramolecular ligation results in formation of a circular monomer that migrates slightly slower than the substrate. Intermolecular ligation products are dimer and multimers. DNA-PKcs increased overall ligation efficiency, and especially favored intermolecular ligation.

B. DNA-PKcs with or without XRCC4:DNA ligase IV was incubated in the presence of pseudo Y structure DNA and [γ-P^{32}] ATP to allow detection of protein phosphorylation. The same pseudo Y structure DNA as in Figure 1 was used to stimulate the kinase activity of DNA-PKcs (linear duplex DNA showed similar result). Strong phosphorylation of XRCC4 was detected, and the autophosphorylation of DNA-PKcs also increased about 4-fold.

Supplementary Figure 4. Coding Joint Ligation with Addition of XLF and Additional Ku Protein. V(D)J recombination assay was carried out with the indicated substrates. The RAG complex, HMGB1, DNA-PKcs, Artemis, polymerase mu, lambda and TdT are included in every reaction. XRCC4:DNA ligase IV, Ku and XLF are added as indicated above the lanes. In reaction 1 where 50 nM Ku was added to the reaction with 200 nM XRCC4:DNA ligase IV, the amount of coding joint was reduced. In the other reactions, addition of Ku and XLF in different concentration and ratios did not change the coding joint efficiency for any of the three substrates.

Supplementary Figure 5. Ku is not Required for Coding Joint Formation. MBP tagged core RAG1 and core RAG2 were coexpressed in 293T cells and purified using more stringent conditions with 500 mM NaCl, where Ku association is not detectable. V(D)J recombination assays were carried out using the MBP tagged core RAG complex, in the same way as described in Supplementary Figure 4. Addition of 5 to 10 nM of Ku (about 3% of the amount of RAG) showed slight inhibition of the coding joint formation. In reactions 4 to 6, the polymerases were omitted, but Ku still showed no stimulation of coding joint ligation.

Supplementary Figure 6. Addition of Klenow in the Biochemically Defined System Changes the Pattern of Coding Joint Products. Klenow was added in addition to the pol X family polymerases and the resulting products were much shorter in length.

Supplementary Figure 7. Coding Joint Formation Between Two Different 12-Coding Ends Indicates Release of Coding Ends by the RAG Complex.

A. Two different 12-RSS substrates (blue and red) with the same signal end sequence but different coding end sequences were added to the same V(D)J recombination assay in equal amounts. The 23-RSS substrate was present at a 4:1 ratio relative to the individual 12-RSS substrates. Every reaction includes the RAG complex, HMGB1, Artemis and DNA-PKcs. Reaction a is a no-ligase control, and reactions b and c are duplicates. Reactions d and e do not have the 23-RSS substrate as a test for 12/23 rule. The red 12-RSS substrate and the 23-RSS substrate are the same as the ones in Figure 2. The 4 bp microhomology between the 23-coding end and either 12- coding end, as well as between the two 12- coding ends are the same, and so the ligation between different ends should have similar efficiency.

B. Three different PCR primers specific for the different coding ends are used to amplify joining products between the 12(red)-12(blue) coding ends (84 bp, panel A&D), the 12 (blue)-23 coding ends (82 bp, panel B), and the 12(red)-23 coding ends (78 bp, panel C). Ligation of identical 12-coding ends (red to red) cannot be efficiently detected due to the inefficiency of amplification of palindromic DNA.

The precise coding joints are indicated by arrowheads. Duplicate PCR for reactions b, c and e indicate reproducible PCR products. Panel A and B use the same radioactively labeled primer 3, specific for the blue 12-coding end. If there is equal synapsis and cleavage efficiency between the 23-RSS and each of the two 12-RSSs, the cleaved coding ends of 12(red):12(blue):23 should be 1:1:2. The amount of 12(red)-12(blue) coding joint observed is comparable to the 12(blue)-23 coding joint (compare panel A and B), indicating random ligation between the coding ends released by the RAG complex. Control reactions with 12(red) and 12(blue) substrates in the absence of 23 substrates showed no detectable coding joint formation in panel D. Therefore the *in vitro* coding joint formation conforms to the 12/23 rule.

C. The full-length RAG1/core RAG 2 complex (FLR1/CR2) and CR1/FLR2 are tested for the formation of 12(red)-12(blue) coding joints (panel F). Duplicate reactions g and h with FLR1/CR2 and duplicate PCRs indicate reproducible coding joint detection. Similar to the core RAG proteins, the coding ends are released by the full-length RAG proteins and ligated randomly.

Supplementary Figure 8. Coding End Exchange Between Two V(D)J Recombination Substrates that Each Have a cis 12- and 23- RSS Configuration.

A. In contrast to the substrates used earlier in this study, the substrates here have a 12-RSS and a 23-RSS on the same linear substrate, with 138 bp between the RAG cut sites. The biotin groups on the ends are bound by streptavidin during the reactions. Substrate 1 and 2 have almost the same sequence except for 2 bp on each coding end, resulting in an XhoI digestion site on each coding end of substrate 2. The four possible ligation products resulting from mixed coding ends are illustrated on the left. After PCR, only the radioactive strands (indicated by *) are detectable. Coding joint (CJ) 2-2 is 2 bp longer than CJ 1-1, whereas CJ 1-2 and 2-1 are 1 bp longer than CJ 1-1. Only CJ 1-1 is resistant to XhoI digestion.

B. The V(D)J recombination assays are carried out with substrate 1 only, 2 only and both substrates (the exchange reactions are shown in lanes 5, 6, 11 &12). For the duplicate exchange reactions, the two different substrates were kept separate during synapsis and cleavage, and after gentle mixing, ligation was initiated by addition of DNA ligase IV:XRCC4. Coding joint 1-1 is completely resistant to XhoI digestion (lanes 2 vs 8), whereas coding joint 2-2 is entirely digested (lanes 4 vs 10). Because coding joint formation from substrate 1 seems to be more efficient than substrate 2 (lanes 1 vs 2), the expected coding joint 1-1 would be at least 50% if there were coding joint retention. However, in the exchange reactions, the XhoI-resistant coding joint 1-1 is much less than 50% (lanes 5 & 6 vs 11 & 12), indicating the two cis coding ends are not preferentially joined.

Supplementary Figure 9. Efficiency of Coding Joint Formation in the Biochemically Defined System.

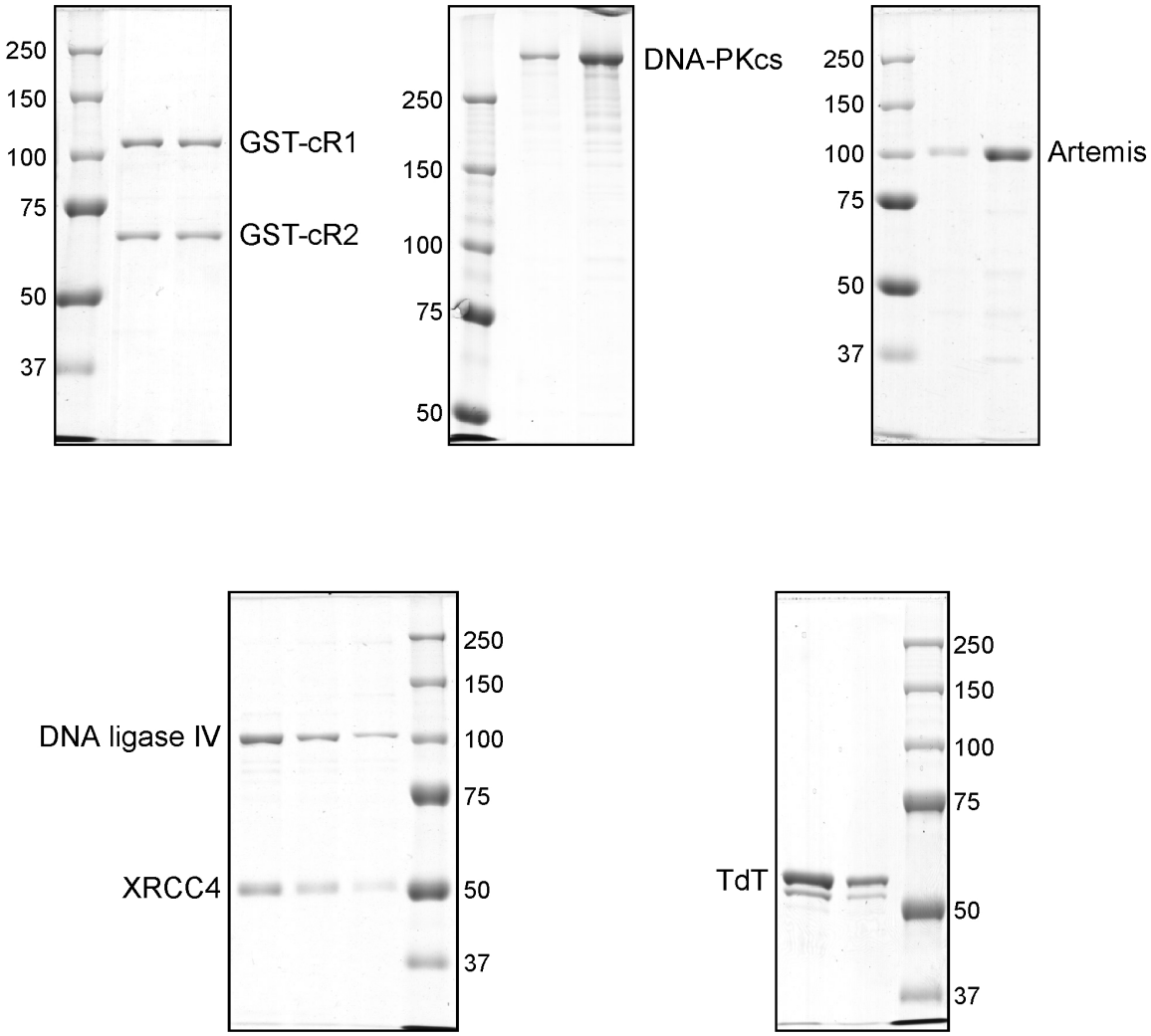
A. Estimation of coding joint efficiency from individual steps. The efficiency of each enzymatic step of coding joint formation is estimated from individual activity assays (Fig. 1, Suppl. Fig. 2 & 3, and data not shown). The overall efficiency is estimated by multiplying the efficiency of individual steps. When polymerases μ or TdT add template-independent nucleotides, the resulting coding ends are not always compatible, which then reduce ligation efficiency. Therefore, the calculations here do not include the efficiency of polymerase additions.

B. Experimentally determined overall coding joint formation efficiency. An oligonucleotide of length 86 nt was synthesized and used as control template. It has the same sequence as the precise coding joint product from the 4 bp compatible substrates. In the control PCRs, the amounts of the control template are $1x10^{-3}$, $5x10^{-3}$ and $10x10^{-3}$ of the amount of substrate, respectively. The 4 bp compatible substrates were used for the joining reactions, and the RAG complex, HMGB1, DNA-PKcs and Artemis were the proteins present. XRCC4:DNA ligase IV was added where indicated. The efficiency of coding joint formation is between 0.1% ~ 0.5%.

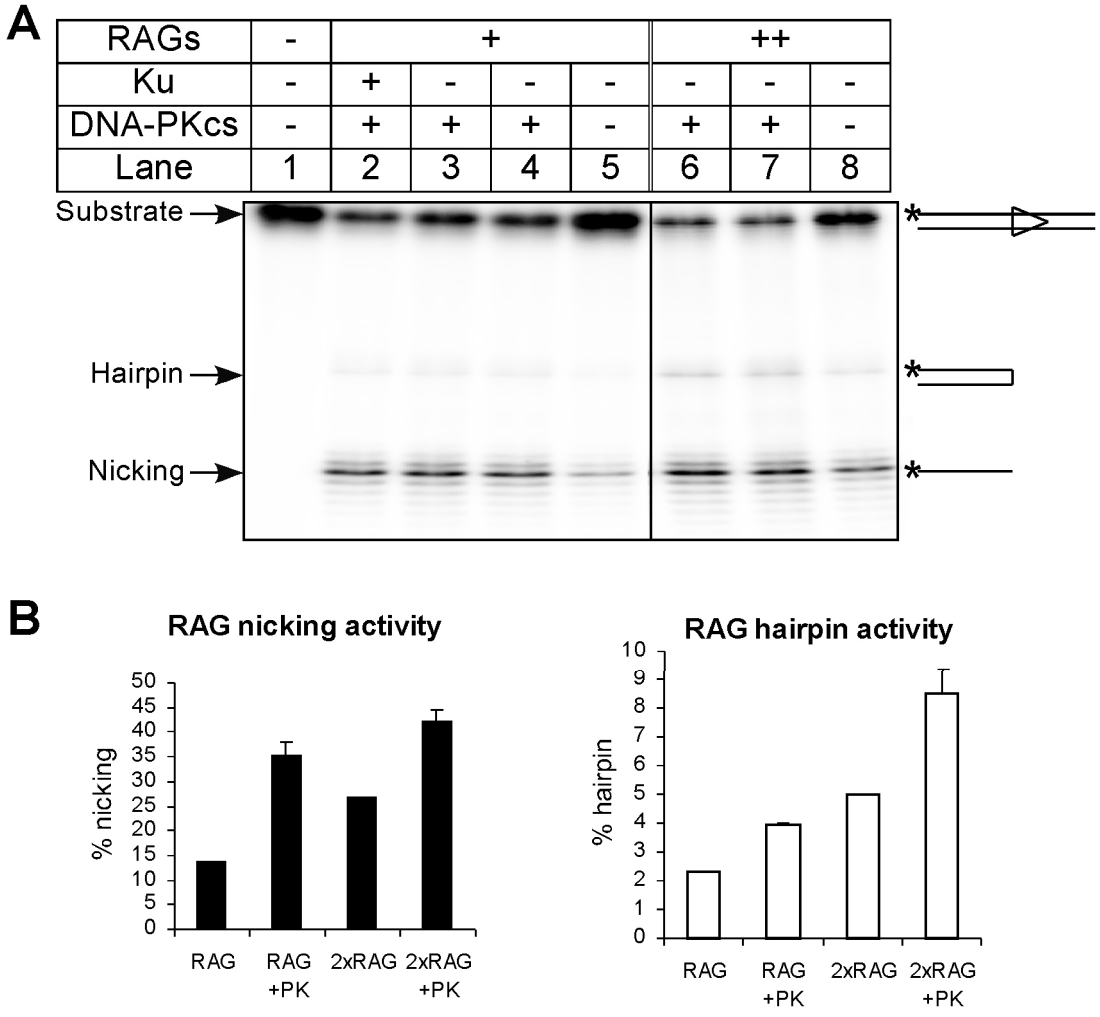
Supplementary Table 1. Protein Purification of the V(D)J Recombination Factors.

Protein	*Origin*	*Tag*	*Cell*	*Purification*	*Reference*
Core RAG1/ core RAG2	Mouse	GST	293T	Glutathione agarose	Yu, et al., 2002
		MBP	293T	Amylose agarose	Bergeron, et al., 2006
FLRAG1/ core RAG2	Mouse	MBP	293T	Amylose agarose	Bergeron, et al., 2006
Core RAG1/ FLRAG2					
HMGB1	Human	His	E.coli	Ni-NTA→MonoQ	Bianchi, et al., 1992
Artemis	Human	Myc-His	293T	Anti-myc protein G sepharose	Ma, et al., 2002
		His	Insect	Ni-NTA→MonoQ	
DNA-PKcs	Human	Endo-genous	HeLa	Ion exchange and affinity DNA chromatography	Chan et al., 1996
Ku70/Ku80	Human	His	Insect	Ni-NTA→MonoQ	Yaneva et al., 1997
DNA ligase IV:XRCC4	Human	His	Insect	Ni-NTA→MonoQ	Nick McElhinny et al., 2000
XLF	Human	Myc-His	293T	Ni-NTA→MonoQ	Lu et al., 2007
Pol mu	Human	No tag	E.coli	Ion exchange chromatography	Dominguez et al., 2000
Pol lambda	human	His	E.coli	Ni-NTA→ heparin sepharose	Shimazaki et al., 2002
TdT	human		Insect	Ion exchange chromatography	Chang et al., 1988

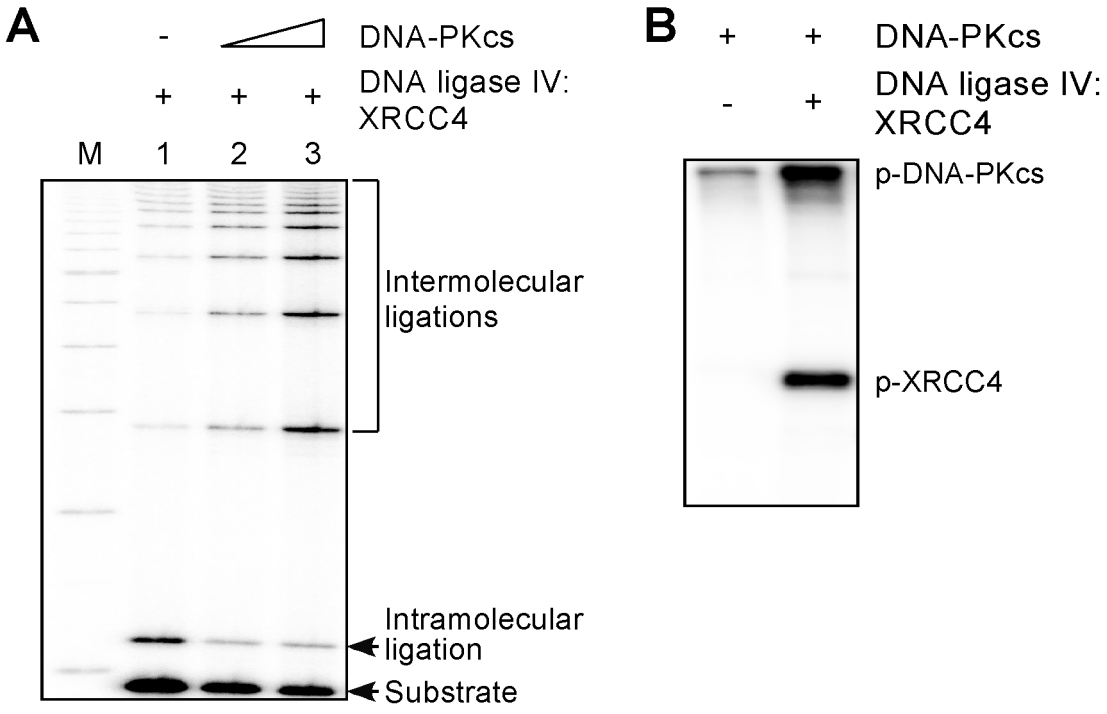
Supplementary Figure 1



Supplementary Figure 2

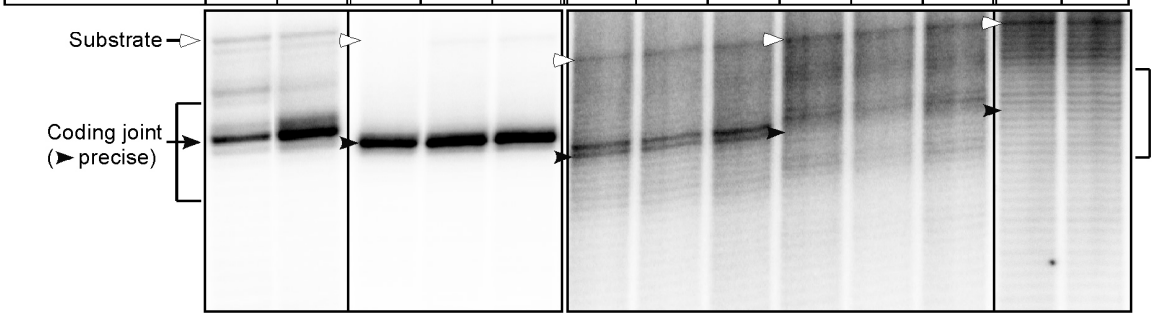


Supplementary Figure 3



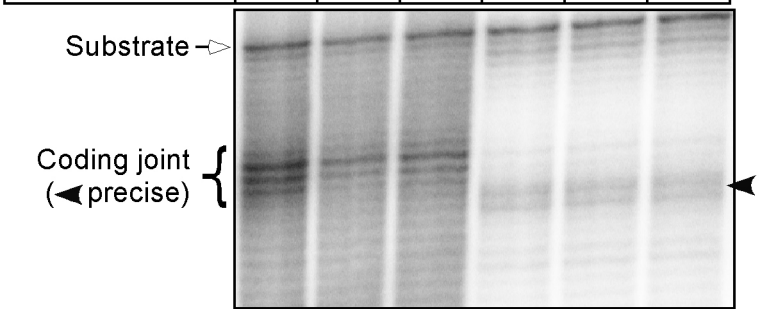
Supplementary Figure 4

CE Homology	4 bp					3 bp			1 bp			1 bp	
LIV:X4 (nM)	200	200	200	200	200	100	100	100	100	100	100	50	50
Ku (nM)	50	-	5	-	-	-	5	-	-	5	-	-	5
XLF (nM)	-	-	-	67	-	-	-	30	-	-	30	-	16
Lanes	1	2	3	4	5	6	7	8	9	10	11	12	13



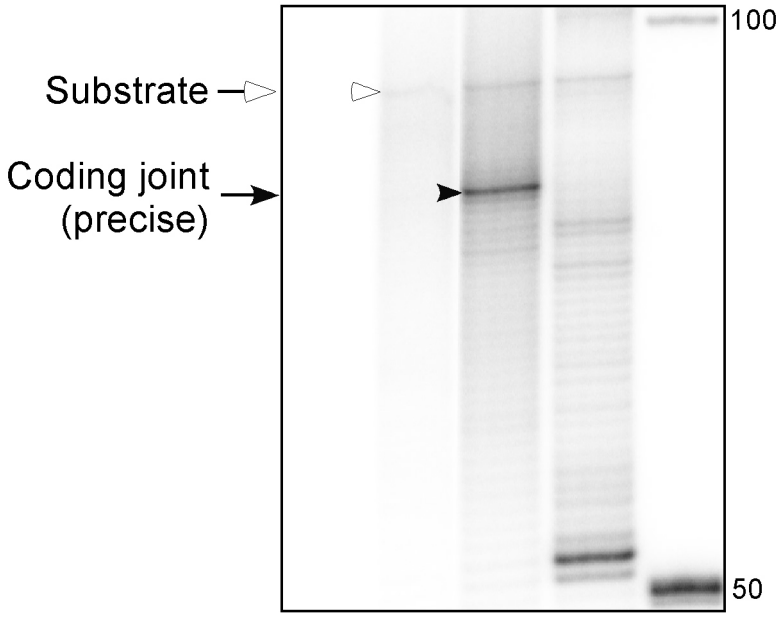
Supplementary Figure 5

CE Homology	1 bp					
LIV:X4 (nM)	100					
Ku (nM)	-	5	10	-	5	10
Polymerases	+	+	+	-	-	-
Lanes	1	2	3	4	5	6



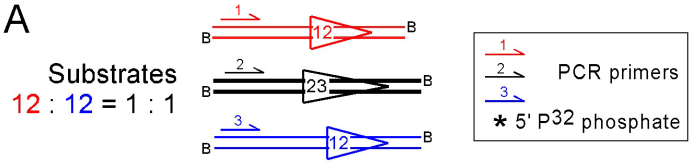
Supplementary Figure 6

Klenow		-	-	+	
Pol μ, λ, TdT		+	+	+	
LIV:X4		-	+	+	
Lanes	Ctl	1	2	3	M



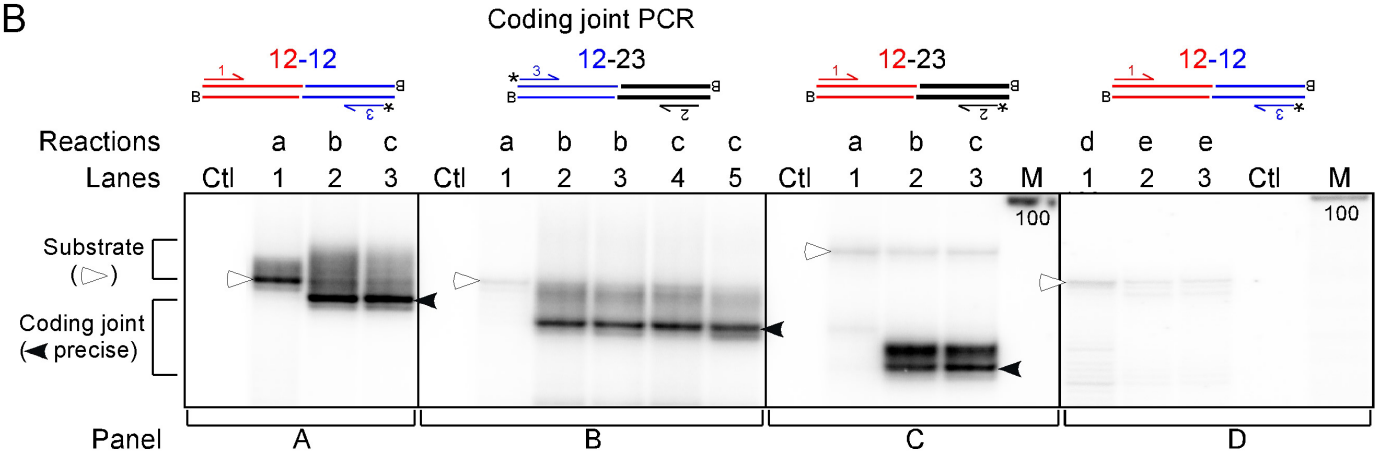
Supplementary Figure 7

A

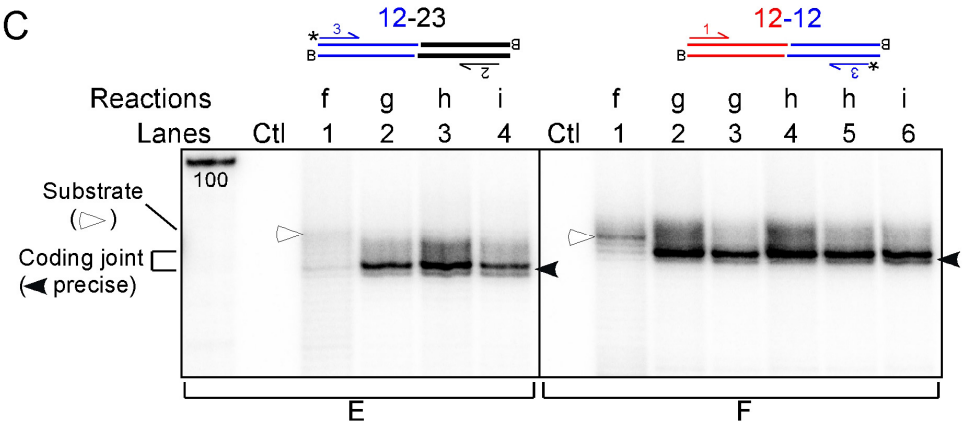


Reaction	a	b	c	d	e
12+12 substrates	+	+	+	+	+
23 - substrate	+	+	+	-	-
XRCC4:DNA ligase IV	-	+	+	-	+

B



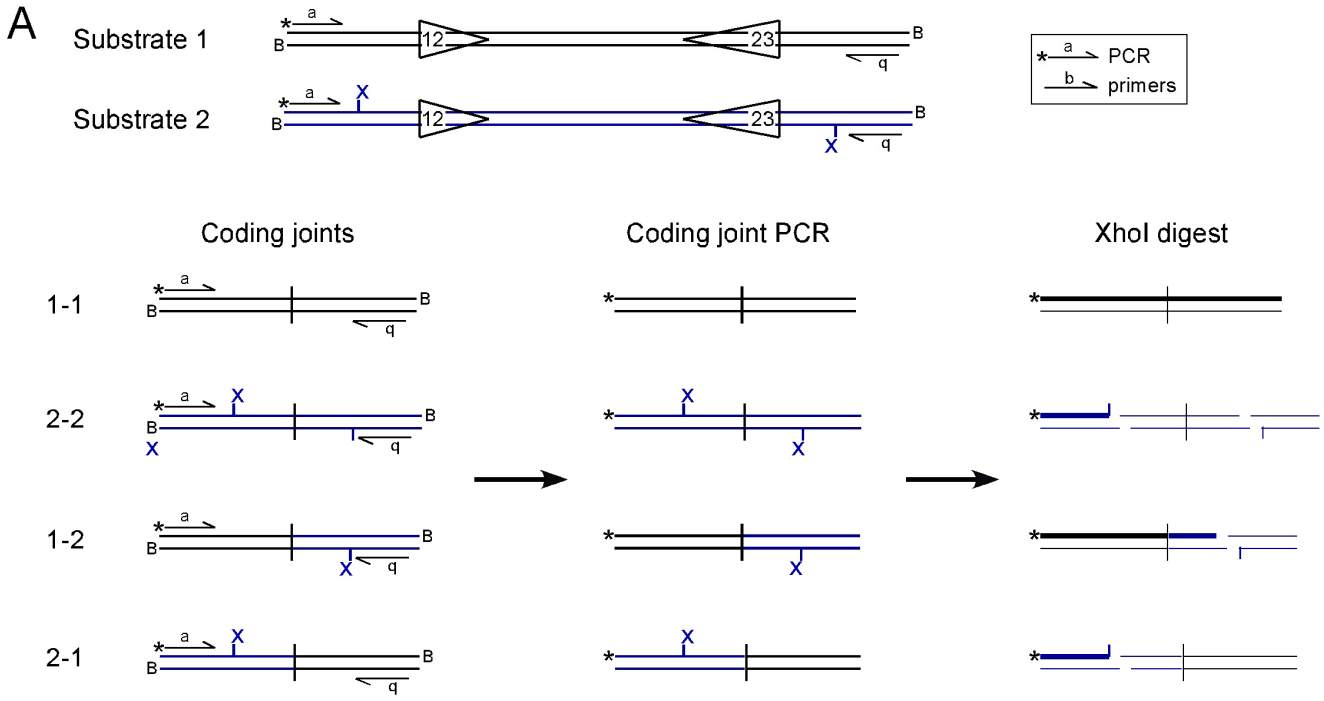
C



Reaction	f	g	h	i
12+12+23	+	+	+	+
RAG1	FL*			C
RAG2	C*			FL
X4:LIV	-	+	+	+

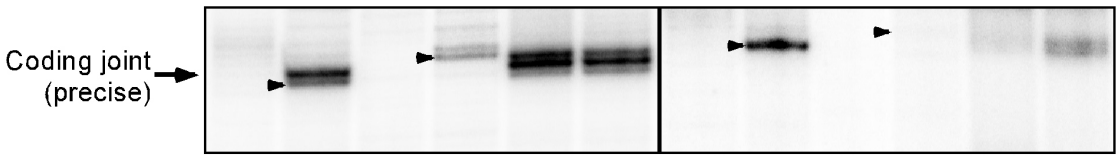
* FL: full-length, C: core

Supplementary Figure 8



B

XhoI	-						+					
Substrate	1		2		1 + 2		1		2		1 + 2	
Ligase	-	+	-	+	+	+	-	+	-	+	+	+
Lanes	1	2	3	4	5	6	7	8	9	10	11	12



Supplementary Figure 9

A

Protein	Reaction	Efficiency
RAG1/RAG2	Hairpinning	10%
Artemis:DNA-PKcs	Hairpin opening	2~10%
Polymerase mu, lambda and TdT	Nucleotide additions	ND
XRCC4:DNA ligase IV	Ligation	< 50%
Overall efficiency		Less than 0.1~0.5%

ND: not determined

B

PCR cycles	X 14					X 20					
X4:LIV	+	-				+	-				
Control template (x 10^{-3})			1	5	10	/	/	1	5	10	
Lanes	1	2	3	4	5	6	7	8	9	10	Ctl