Electronic Supplementary Material (ESI) for Chemical Communications.
This journal is © The Royal Society of Chemistry 2015

Supplementary Information

Highly-efficient T4 DNA ligase-based SNP analysis using a ligation fragment containing a modified nucleobase at the end

Eui Kyoung Jang[a], Munhee Yang[b] and Seung Pil Pack[a,*]

[a] Department of Biotechnology and Bioinformatics, Korea University, Jochiwon, Sejong 339-700, Korea.

[b] Department of Psychology, Univ. of Texas, Austin, Texas 78712, U.S.A.

* Correspondence: Tel: +82-44-860-1419; Fax: +82-44-860-1598; E-mail: spack@korea.ac.kr

Supporting Information

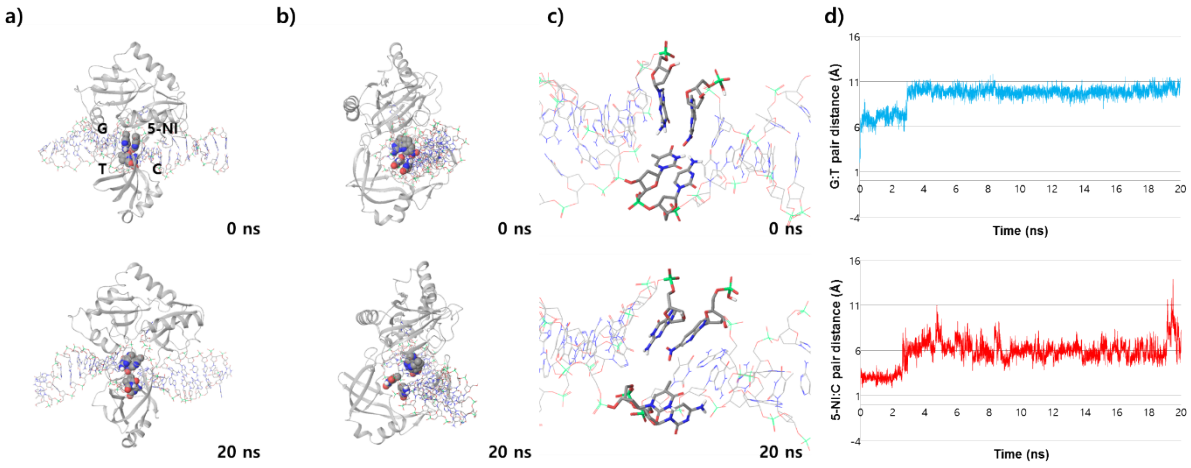

Fig. S1 Molecular dynamics simulation of base pairing at nick region on DNA ligase, when 5-nitroindole (5-NI)-end downstream ligation fragment was used. a) front view, b) side view, c) DNA structure, and d) G:T (SNP site) and 5-Ni:C pair distance during simulation time. G:T pair distance means the length from guanine (H1) to thymine (O4) and 5-NI:C pair distance means the length from 5-nitroindole (O1) to cytosine (H4)

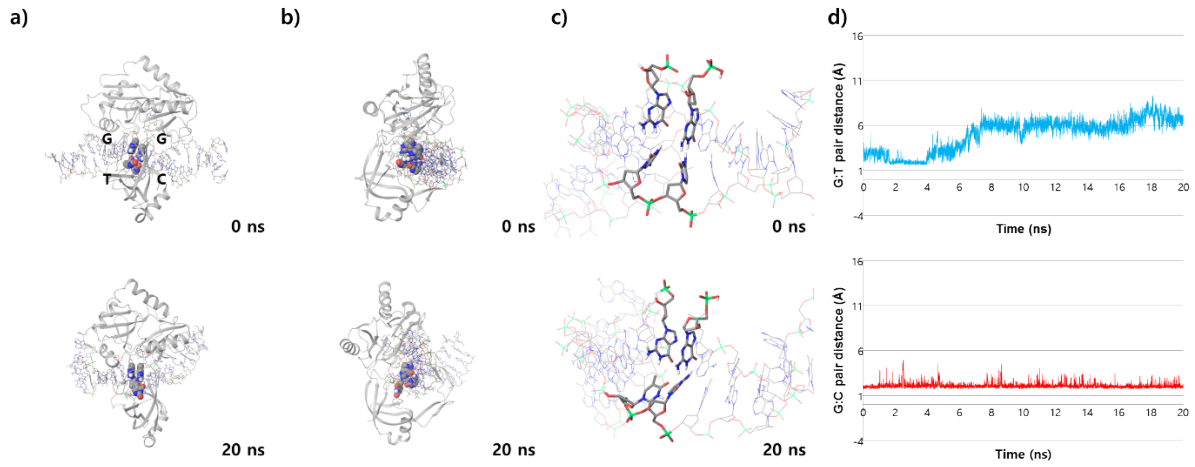

Fig. S2 Molecular dynamics simulation of base pairing at nick region on DNA ligase, when normal base-end downstream ligation fragment was used. a) front view, b) side view, c) DNA structure, and d) G:T (SNP site) and G:C pair distance during simulation time. G:T pair distance means the length from guanine (H1) to thymine (O4) and G:C pair distance means the length from 5-nitroindole (O1) to cytosine (H4)

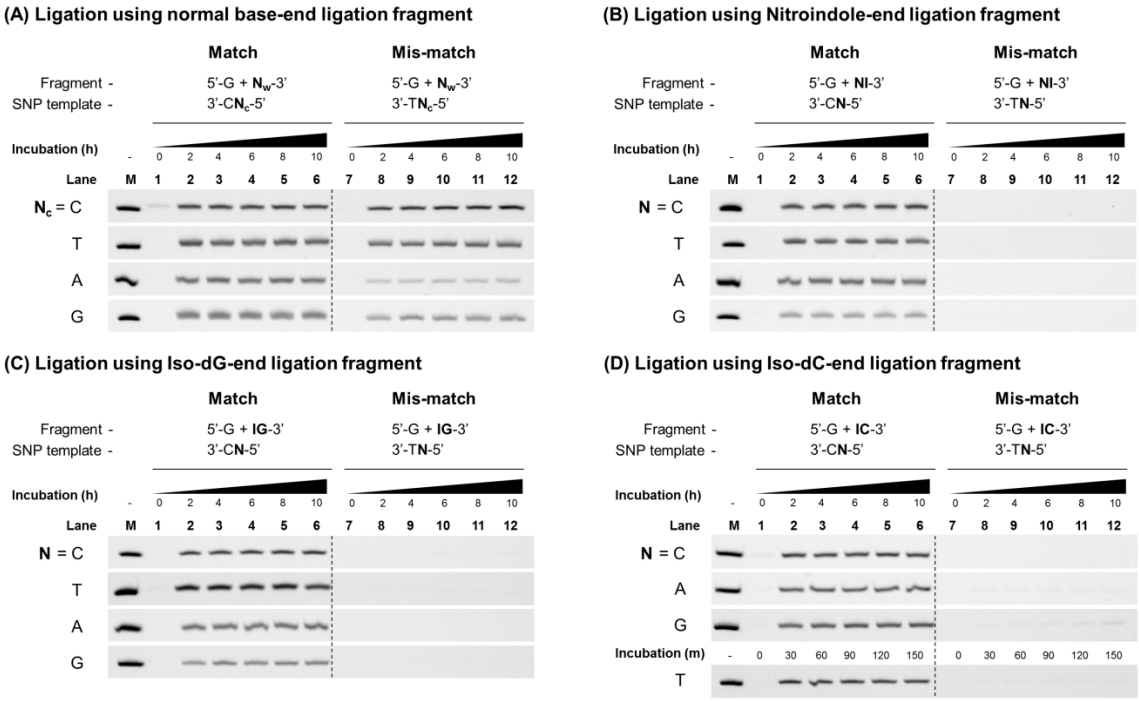

Fig. S3 (A) Time-dependent assays were performed with each of the three modified nucleobase-end downstream ligation fragments; 5-Nitroindole- (B), Iso dG- (C), Iso dC-end fragment (D), by comparing with four normal base-end fragments (A). Each modified nucleotide is indicated by their respective abbreviation in the figure. 5-Nitroindole, NI; Iso-dG, IG; Iso-dC, IC

Modeling Method

Protein preparation

The initial setup of DNA ligase for calculations was performed using Schrödinger' "Protein Preparation Wizard" starting from the X-ray crystal structure (PDB ID: 2Q2T) of the DNA ligase and DNA complex[1-4]. We selected Chlorella virus DNA ligase because of lack of T4 DNA ligase crystal structure. ATP-dependent DNA ligase reveal a common tertiary structure of the catalytic core and a nucleotide binding pocket[5]. Protein residue bond orders were assigned and hydrogen atoms added, with the initial assignment of protonation states for basic and acidic residues, and tautomeric states based on residue pK_as at their normal pH 7.0. Subsequent optimization of hydroxyl, histidine protonation states and C/N atom "flips," and sidechain O/N atom "flips" of Asn and Gln was based on optimizing hydrogen bonding patterns, so that the final assignments were checked on visual inspection of the protein. In particular, all final His residues were assigned as neutral, either in a HIE or HID state[6]. Finally an "Impref" minimization of the DNA ligase and DNA complex was performed using the OPLA-AA(2005) force field to remove steric clashes and bad contacts[7, 8]. At the end of the minimization, the root-mean-square deviation (RMSD) of all heavy atoms was within 0.3 Å of the crystallographic positions. Using computationally prepared structure, guanine base at the 5′ end of the downstream fragment is modified to 5-nitroindole. We used Maestro's edit tool of Schrödniger to modify part of guanine and its pair. Then, Macro model performed to minimize the structures.

Molecular dynamics simulations

MD simulations in the NPT ensemble were performed using the program Desmond[9, 10]. According to the software, the OPLS 2005 force field was used to build aqueous biological systems, and the TIP3P model was used to simulate water molecules[11, 12]. The orthorhombic periodic boundary conditions were set up to specify the shape and size of the repeating unit. In order to get an electrically neutral system, the minimum number of sodium and chloride ions needed to balance the system charge was placed randomly in the solvated system, and 0.15 mol/L sodium and chloride were then added. Energy of prepared systems for MD simulation was minimized up to maximum 2000 steps using steepest descent method until a gradient threshold (25 kcal/mol/Å) is reached. The systems were equilibrated with the default protocol provided in Desmond. For this, the program ran six steps composed of minimizations and short (12 and 24 ps) molecular dynamics simulations to relax the model system before performing the final long

simulations. After that, a 20 ns MD simulation was performed in an NPT ensemble at 300 K and 1 atm using a reversible system propagation algorithm (RESPA), a Nosé-Hoover thermostat and a Martyna-Tobias-Klein barostat. Energy and atomic coordinate trajectory data were recorded every 4.8 ps

References

1. *Protein Preparation Wizard 2015-2; Epik version 2.4, Schrödinger, LLC, New York, NY, 2015; Impact version 5.9, Schrödinger, LLC, New York, NY, 2015; Prime version 3.2, Schrödinger, LLC, New York, NY, 2015.*
2. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Research*, 2000, **28**, 235-242.
3. P. A. Nair, J. Nandakumar, P. Smith, M. Odell, C. D. Lima and S. Shuman, *Nature structural & molecular biology*, 2007, **14**, 770-778.
4. G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *Journal of computer-aided molecular design*, 2013, **27**, 221-234.
5. M. Odell, V. Sriskanda, S. Shuman and D. B. Nikolov, *Mol Cell*, 2000, **6**, 1183-1193.
6. M. H. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, *Journal of Chemical Theory and Computation*, 2011, **7**, 525-537.
7. G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *The Journal of Physical Chemistry B*, 2001, **105**, 6474-6487.
8. W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *Journal of the American Chemical Society*, 1996, **118**, 11225-11236.
9. K. J. Bowers, E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes and F. D. Sacerdoti, 2006.
10. *Desmond Molecular Dynamics System, version 4.2, D. E. Shaw Research, New York, NY, 2015. Maestro-Desmond Interoperability Tools, version 4.2, Schrödinger, New York, NY, 2015.*
11. W. L. Jorgensen, *Encyclopedia of computational chemistry*, 1998.
12. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of chemical physics*, 1983, **79**, 926-935.