SUPPLEMENTAL DATA

Generation of single-chain LAGLIDADG homing endonucleases from native homodimeric precursor proteins

Hui Li[1,5], Stefan Pellenz[1,5], Umut Ulge[3,5], Barry L. Stoddard[4,5] and Raymond J. Monnat, Jr.[1,2,5,6]
[1]Departments of Pathology and [2]Genome Sciences and [3]the Molecular and Cellular Biology Program, University of Washington, Box 357705, Seattle, WA 98195; [4]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N. A3-025, Seattle, WA 98109; and the [5] Northwest Genome Engineering Consortium.

[6]Correspondence should be addressed to RJM, Jr.: Department of Pathology, Box 357705, University of WA, Seattle, WA 98195-7705. tel: 206.616.7392; fax: 206.543.3967; e-mail: monnat@u.washington.edu.

Supplemental Table 1. Amino acid sequences of linker regions in monomeric I-CreI and I-MsoI from *in vivo* selection.

Supplemental Figure 1. Distribution of linker lengths and linker amino acid composition in monomerized mCreI and mMsoI.

Supplemental Figure 2. DNA substrate distortion induced by I-MsoI and mMsoI binding.

26 Nov 2008/revised 30 Dec 2008

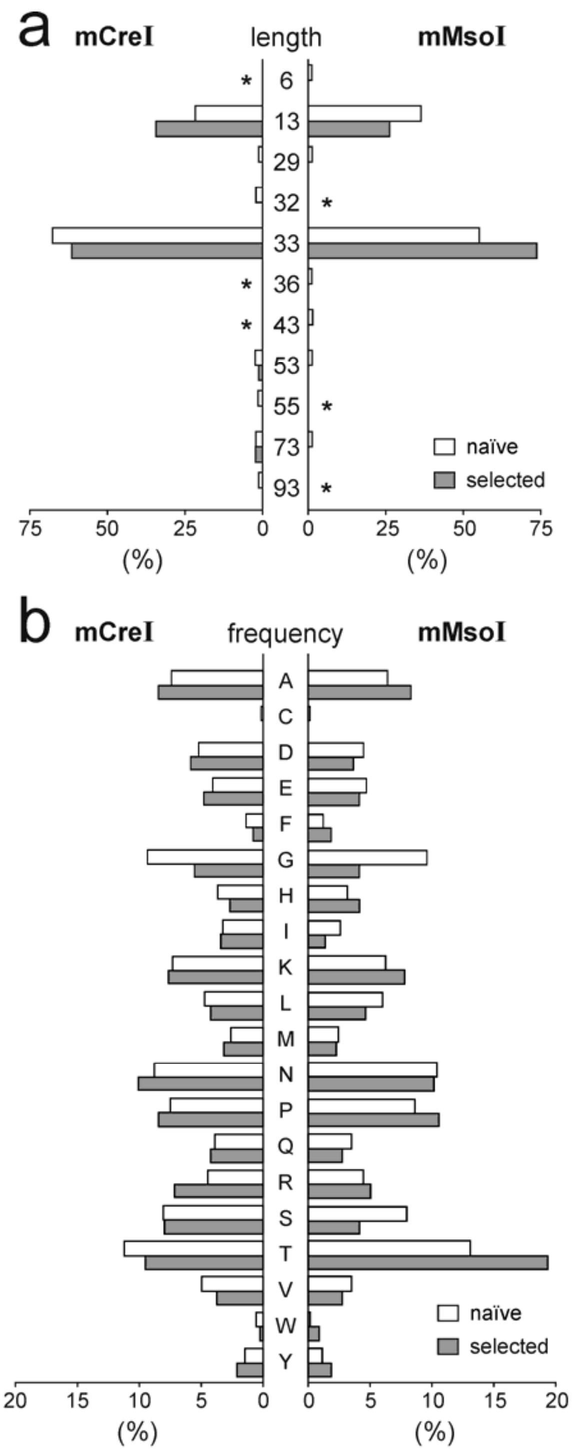
Supplemental Table 1. Amino acid sequences of linker regions in monomeric I-CreI and I-MsoI from *in vivo* selection.

protein	***linker amino acid sequence***	***frequency***
mCreI.2	TGSGSGSTNMKPPVRAFEPTGVRSRGSGSGSGT	6/84
*mCreI	**TGSGSGSKSQAVAHPTDGQRDFGAKGSGSGSGT**	5/84
*mCreI.6	TGSGSGSKPAGGDAPRLMQGVNRIDGSGSGSGT	19/84
*mCreI.7	TGSGSGSGSGSGT	29/84
*mCreI.14	TGSGSGSNPRNSPNSKTSMPIDVNNGSAYSMQSNRGYVKEEYLHRGSGSGSGT	1/84
*mCreI.15	TGSGSGSKTKNMSPKANIERTPENKGSGSGSGT	7/84
*mCreI.19	TGSGSGSSTKERTNLKDNMTIDKPRGSGSGSGT	1/84
mCreI.45	TGSGSGSKDVTQANRTYIPRENASRGSGSGSGT	1/84
*mCreI.48	TGSGSGSTDQAGHDPGAKTAKPMLGGSGSGSGT	1/84
*mCreI.53	TGSGSGSNYAAKPIPSAGQLETSHNGSGSGSGT	3/84
*mCreI.56	TGSGSGSIPQTQFHLVLGAAATRDNGSGISETNPRDPTQVSDKNIGSTVTGQVVRTDSLEENKANGSGSGSGT	2/84
mCreI.65	TGSGSGSKTKNMSPSANIERTPDNKGSGSGSGT	1/84
mCreI.81	TGSGSGSKYEGKAILSAGQLDTSYKGSGSGSGT	1/84
*mCreI.90	TGSGSGSNNKSSHPQGDVEQKHQHSGSGSGSGT	1/84
mCreI.102	TGSGSGSTSARLYPQTTATMNDSTMGSGSGSGT	1/84
*mCreI.119	TGSGSGSNPAMLADPKNTGLATGAIGSGSGSGT	1/84
*mCreI.121	TGSGSGSNDTEMSSWTAERRTPRPTGSGSGSGT	1/84
*mCreI.124	TGSGSGSNPGVRSPRNNDLPDHRLIGSGSGSGT	1/84
mCreI.125	TGSGSGSNAGNLPSRENNTSKHSAEGSGSGSGT	2/84
*mMsoI.3	TGSGSGSTAAKPPVRTTDGMESTFMGSGSGSGT	1/57
*mMsoI.5	TGSGSGSGSGSGT	15/57
*mMsoI.14	TGSGSGSAYTTTTDEAPTLVKPRHNGSGSGSGT	1/57
*mMsoI.15	TGSGSGSKPTALNPWNIDRTTIPAKGSGSGSGT	6/57
*mMsoI	**TGSGSGSKHPTLTLPTTTSQENLPNGSGSGSGT**	3/57
mMsoI.25	TGSGSGSRFAGESHVNNTTKTTKLEGSGSGSGT	9/57
*mMsoI.27	TGSGSGSKTKNPHPENPGQSMTQAKGSGSGSGT	1/57
mMsoI.28	TGSGSGSRFAGESHVNNTTKTTKLEGSGSGSGT	3/57
mMsoI.29	TGSGSGSTHTTRHNRTPTAPNYRPIGSGSGSGT	1/57
mMsoI.43	TGSGSGSGFANKYNVDHNPLSNMNSGSGSGSGT	1/57
mMsoI.55	TGSGSGSKTKNPHPWNPDRSTTPAKGSGSGSGT	1/57
*mMsoI.70	TGSGSGSTTQAPPTMTYTRGVATTDGSGSGSGT	1/57
*mMsoI.96	TGSGSGSNLGAENAQSASQKDDALRGSGSGSGT	1/57

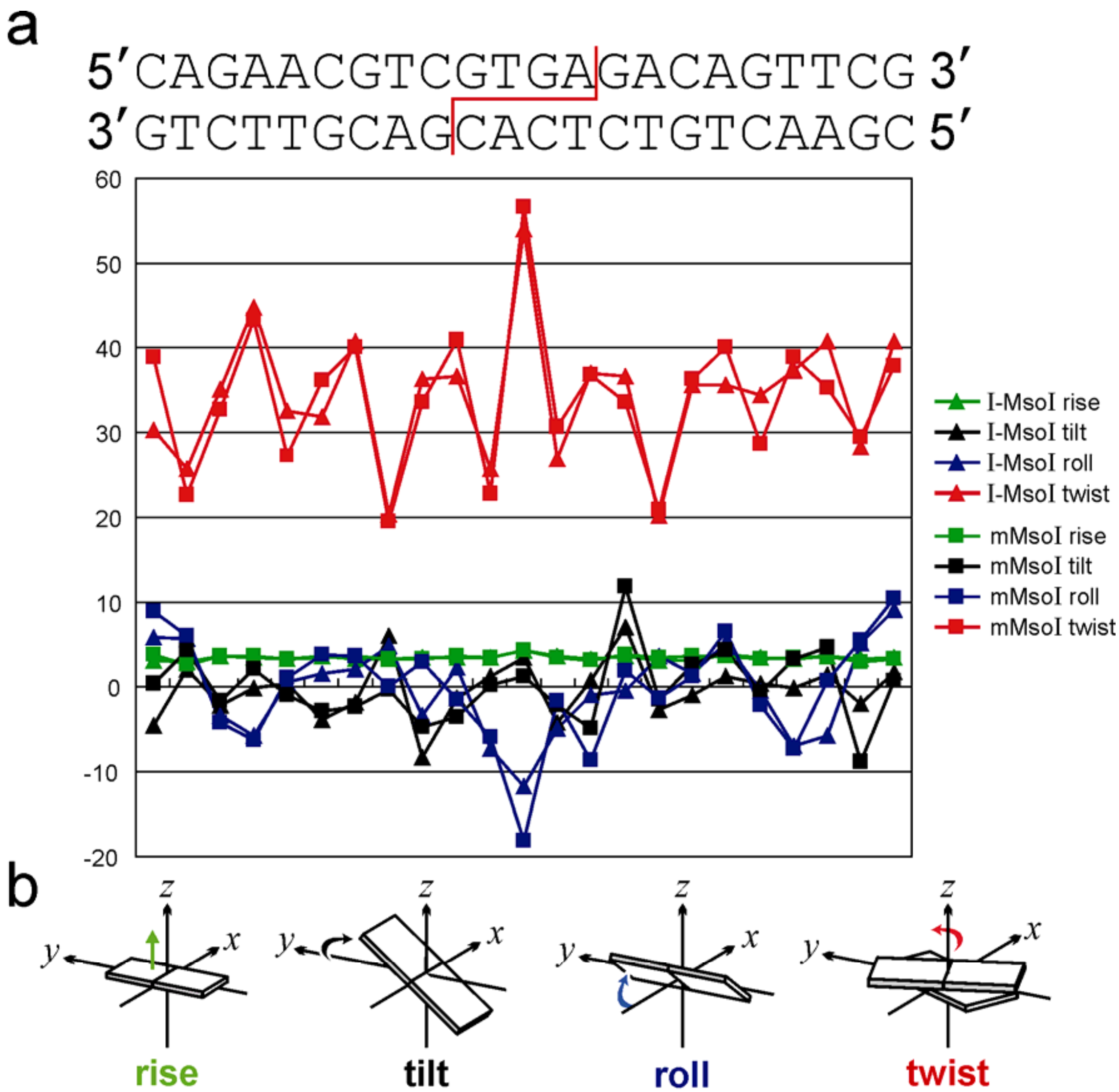
Supplemental Table 1 notes: The common adapter sequences at both ends of linker are shown with the random linker regions underlined. The linker sequences of mCreI and mMsoI that were characterized in detail are shown in **bold**. Proteins marked with an asterisk (*) were purified and their site-specific cleavage activity analyzed *in vitro* as shown in Figure 2. Frequency refers to the number of times each linker was identified when sequencing pENDO plasmids recovered from two rounds of selection in *E. coli*. A total of 141 plasmid DNA's were sequenced, 84 colonies for mCreI and 57 colonies for mMsoI.

Supplemental Figure 1. Distribution of linker lengths and linker amino acid composition in monomerized mCreI and mMsoI. The random linker regions of randomly chosen unselected and of the 141 selected clones shown in Supplemental Table 1 were sequenced to determine the predicted amino acid sequences of random linker regions. (a,b) The linker length distributions are shown as stem plots for mCreI (left stem) and mMsoI (right panel; a) and linker amino acid compositions in b. Linker data from unselected (naïve) clones are shown as open boxes, whereas selected clone data (selected) are shown as gray boxes. * = no linker of the specified length was identified in either library of the indicated protein.

Supplemental Figure 2. DNA substrate distortion induced by I-MsoI and mMsoI binding. Four different distortions in DNA base pair geometry were calculated from I-MsoI and mMsoI co-crystal data using 'Readout' (http://gibk26.bse.kyutech.ac.jp/jouhou/ readout/) (34) and plotted versus the native I-MsoI homing site shown at top with the cleavage site indicated by the staggered line. Base pair roll, tilt, twist and rise are plotted as % deviations from standard coordinate B-form double-stranded DNA.



Supplemental Figure 1



Supplemental Figure 2