SUPPLEMENTARY INFORMATION

Table S1: Non-redundant list of selected sequences. Sequences marked by ** were deleted from the all analyses because they contained additional TATA boxes. Sequences marked by * are those that were further deleted initially, for the same reasons, but that their deletion did not change the observed pattern (see the main text for details).

a. MLP sequences selected by thermodynamic criteria

MLP t1	TGCCCTTGTT	TATAAAAG	GTTTAAAAAG
MLP t2	CCTGGGATAT	TATAAAAG	GGCTATTCGC
MLP t3	GCTACATG	TATAAAAG	GTTTTCCTTT
MLP t5	AAAGAGCTTG	TATAAAAG	TTGACGTGGG
MLP t6	AACCGCTTTC	TATAAAAG	TCGGGAAATT
MLP t7	GTAAGCGGGC	TATAAAAG	ATTCTAGCTT
MLP t8	GAGCGTGATC	TATAAAAG	CCAGAACACC
MLP t10	CCATTATTTG	TATAAAAG	GTACCGTCGG
MLP t11	GTATAATCGT	TATAAAAG	TGTAGGTCTT
MLP t12	ATCGGAATGA	TATAAAAG	AGGTGATCTA
MLP t13	TAATTCTGAG	TATAAAAG	CTGGCGATAA
MLP t14	GGTGGCGTT	TATAAAAG	TCTTAAACTT
MLP t16	TGGCAAACAA	TATAAAAG	CGTGACGAGG
MLP t17	GCACCTTCGG	TATAAAAG	CACGGGAGTA
MLP t18	GTGCAATCTG	TATAAAAG	CGTAAATTTA
MLP t19	AGAATCTCCG	TATAAAAG	CAGTTGATCC
MLP t21	CGTACCTTGA	TATAAAAG	ACTAATTAAG
MLP t22 **	TCGTGGTTTA	TATAAAAG	ATGGGCTCGA
MLP t23 **	ACGTCCTCGT	TATAAAAG	TATATATGGA
MLP t29 **	GGAGACACTA	TATAAAAG	ATTGGGGGCT
MLP t32	GGCGGTATAC	TATAAAAG	TTAGTGTTGT
MLP t33	AGATTTGGCG	TATAAAAG	CGCGAATGGC
MLP t34	TGCTTGCGGG	TATAAAAG	GCTCGTATAC
MLP t37	TTTAGACTT	TATAAAAG	GAAAGCGCGC
MLP t38	TGATGGGCCT	TATAAAAG	CCTTTTAAGC
MLP t39	GTCAGTAGGT	TATAAAAG	GTATACGTCT
MLP t40 **	AAGTCTACTA	TATAAAAG	GAGATAGGGC
MLP t42	GTCTAAATAC	TATAAAAG	CTGTCCCAGT
MLP t51	ATCACTATCA	TATAAAAG	GCATAAGATT
MLP t52	TGTGTAACTG	TATAAAAG	TCGGCTTGTT
MLP t53	GTCTGTGGTT	TATAAAAG	TAATTACCAA
MLP t5	TAAAGCCTGT	TATAAAAG	TAACCCTTGC
MLP t55	GTGGTGGATG	TATAAAAG	GCGATTGAGT
MLP t56	GCGTGGCCAG	TATAAAAG	AGCCGAACAC
MLP t58	AGGGTGAAGG	TATAAAAG	GTTAGCAA
MLP t59 **	TGGCAACTTT	TATAAAAG	TGGAGAACTC
MLP t63	AAAGGAGAGA	TATAAAAG	GGCGGGGAAT
MLP t65	CTTGGTTCGA	TATAAAAG	GCGACGGTAA
MLP t67	AGCTTCAAGT	TATAAAAG	CCTGGTGCGC
MLP t68	GTACCCAGGT	TATAAAAG	TCGCTAGACT
MLP t69	TTCAAAGGTC	TATAAAAG	CGGAATTGGT
MLP t70	ATAGTTTAAT	TATAAAAG	GTCCCTTCAC
MLP t71	GATGGGGAGG	TATAAAAG	TGGGATAAGG

MLP t72	CAGGTTTATC	TATAAAAG	GTTGGGGAAC
MLP t73	AGGGCAAGAA	TATAAAAG	TGTTGGCGAG
MLP t74	TAGCGATGCG	TATAAAAG	GCTTCGTAAT
MLP t66	ATGAGTAACA	TATAAAAG	GGAAGTAGTG

b. MLP sequences selected by kinetic criteria

MLP k1	GGCCTCGGTG	TATAAAAG	ACTAGCACAA
MLP k4	TACATGTACT	TATAAAAG	GTTATATGT
MLP k9 **	TCAGACTCT	TATAAAAG	TAGATTTCTG
MLP k15	ATTGATAGCG	TATAAAAG	GCTAGCGGTT
MLP k20	CAGCGAATGC	TATAAAAG	GTCTTGCTGC
MLP k21	GGTTGGACCC	TATAAAAG	TACCCTGAAT
MLP k22	GTCTGATTGA	TATAAAAG	TTCGAGATGC
MLP k23	TCATGGTGGA	TATAAAAG	CTAGGAACGC
MLP k27	TCGTGCTTGG	TATAAAAG	TCACTGCTCG
MLP k30	TAGAATGGGA	TATAAAAG	CTCGGGTGAA
MLP k36	ATGTTGTACG	TATAAAAG	GATAGTGATA
MLP k40	GATAATGCGA	TATAAAAG	GAGATGGGGG
MLP k17	AGCATTAAGT	TATAAAAG	GCGGGTTGTC
MLP k51	TTGTTTGCGT	TATAAAAG	AACACAGGTG
MLP k52	CGGTTGACGG	TATAAAAG	TGCCTAGCCC
MLP k54	GTTCGCCTCA	TATAAAAG	TGGTGTTGGG
MLP k55	ATACGACGAG	TATAAAAG	CGGGCCGGGC
MLP k57	ATCGTTCGTT	TATAAAAG	TTAGTCGTAG
MLP k60	TGTGTCATCA	TATAAAAG	TTTTAATTTT
MLP k61	TTTTACGCGC	TATAAAAG	GTTGGTAGTA
MLP k62	TAAGTTGGCG	TATAAAAG	CGCGCGTAAA
MLP k63	GGGGCGTGTT	TATAAAAG	GGGATGTGTG
MLP k64	GATGCTGGGG	TATAAAAG	GTCGGGTACA
MLP k66	GGTGGGTCTC	TATAAAAG	CCTGGGGTGT
MLP k67	CAGATTGGCC	TATAAAAG	TACCTCGGGC
MLP k68	TTCATGCTGG	TATAAAAG	AGTGGTATTC
MLP k70 **	TGTTTGTTT	TATAAAAG	GGAAAGCTAT
MLP k76	AGTGGTAGG	TATAAAAG	TTTAGGGGTG
MLP k77	AGCATTCGAT	TATAAAAG	TTACGCACAG
MLP k80	CGGTTTGGGA	TATAAAAG	CTTGCTACTT
MLP k82	CCCCGTTCCA	TATAAAAG	TAAGACTAGT
MLP k83	GAGTTTGGCG	TATAAAAG	TGTAGGGCCG
MLP k84	TATCGTGGGG	TATAAAAG	GTGCCTGAGC
MLP k85	TATTCCATCA	TATAAAAG	CGGAGGCTCT
MLP k87	CCGGTTGATC	TATAAAAG	TTTTAGGTGG
MLP k88	GATATTCGTC	TATAAAAG	GGCGTGAGGC
MLP k93	TTGACCTCGG	TATAAAAG	GGCGCTGGCC
MLP k94	TCCATTGAAA	TATAAAAG	ATAGATCTTA
MLP k95	TGTACAGTAA	TATAAAAG	TTGCGATTGT
MLP k96	GTTATACAAG	TATAAAAG	TCAGCGGAGG
MLP k97	CTGATTTGCA	TATAAAAG	CGGGCTCATT
MLP k98	CAAGTGGGCT	TATAAAAG	GTAACTGATC
MLP k99 **	CAACGCCATA	TATAAAAG	AGGACTTCGA

c. E4 sequences selected by thermodynamic criteria

E4 t1	ACTTCGGCAG	TATATATA	GCGTCGTAAG
E4 t3	ACCGGGAGGT	TATATATA	CTGCCACGCA
E4 t4	GCTGGATCCG	TATATATA	GCGGGATGCA
E4 t6	CCCTGCGTGA	TATATATA	TCGGACGCTC
E4 t8	CGCACCCAAA	TATATATA	AGTAACTCAT
E4 t10	GAAACCAGGG	TATATATA	GCAGGGTTGG
E4 t11	CGGCTCAGCC	TATATATA	GCACAATGCG
E4 t13	CGGGACATGA	TATATATA	GGTATTTGAC
E4 t14	GGGGTCAGAG	TATATATA	ACTCGGGATG
E4 t15	GACGTCCACA	TATATATA	CATCGAACTC
E4 t16	CGTGTGCCGT	TATATATA	GCGGTTCCTA
E4 t17	GGCGGTAACC	TATATATA	ACCCGTGCCT
E4 t20 **	TGCTAACGTT	TATATATA	ACGGATCGAC
E4 t21 *	TGGGGTAGGG	TATATATA	ATTGTGATAA
E4 t24	ACCCTGTTCC	TATATATA	TTTGTCCCTG
E4 t26	TGCGGGATCC	TATATATA	GGGCTATTCG
E4 t27	TGTCAACACG	TATATATA	ACTCCCTAA
E4 t28	CAAGTGCCGG	TATATATA	GCGGCGTTTT
E4 t29	AGGTCGCGGA	TATATATA	CAAACTACAG
E4 t31 *	ACGCCGTTAG	TATATATA	ATGTCCAATA
E4 t32 **	GAGCGCACCC	TATATATA	TACCGTCGCG
E4 t39	CCTGCCTGTG	TATATATA	ACGGACACTT
E4 t42	AACCCCAATC	TATATATA	CCAGAAAACC
E4 t45	GGGGCAAGCC	TATATATA	GGGTGGTTGG
E4 t51	CGTGCTCCGG	TATATATA	ACCGCCAATG
E4 t52	TCGCCGCCTG	TATATATA	ACGGACGCAG
E4 t53	AGGGGCAGAC	TATATATA	TCTGATCTCC
E4 t58	GCGCCGACTG	TATATATA	GACCAGACCT
E4 t60	TATATCACGC	TATATATA	CCCGAACCAG
E4 t61 **	ACGCCCGCAC	TATATATA	AACATACAAT
E4 t62	ACCGCCAACT	TATATATA	CTGCTTGATG
E4 t64	TCCGCCAACC	TATATATA	CCAGATGCTT
E4 t65 **	CGCGGCCGTA	TATATATA	TATGCACG
E4 t66	TGTGACAACA	TATATATA	GTCGAACAAC
E4 t67 *	AGGTGCGTAT	TATATATA	ATTGAGAGCG
E4 t68 *	AACTTCACAT	TATATATA	ATCCGCTCC
E4 t69	CTAGCCAAGG	TATATATA	CCTCATCTCC
E4 t70 *	TAGCTCAACA	TATATATA	ATATCTTCAA
E4 t72	TACGTTAGCA	TATATATA	GGTGCAGCCT
E4 t73	CGACGCAGCG	TATATATA	CCCATGCGAC
E4 t74	TGAGGGAGGG	TATATATA	GGCCGTTGGC
E4 t81 *	CGTGGTAGAT	TATATATA	ACGGGGAGTA
E4 t82	ACCCACTCCC	TATATATA	CCACCAGAGC
E4 T83	TACATCTTCT	TATATATA	TCTATTAGAC
E4 t85	GCAGAGCCAG	TATATATA	ACGCACGTGC
E4 t87	GCGCTGCCCT	TATATATA	CGCCATCAGC
E4 t91	CGCCACCCCA	TATATATA	CAAGCCCACC
E4 t92	CCGGTCTACT	TATATATA	CAATGCATGC
E4 t95	GATGTAAATG	TATATATA	ACGAGGAAGA

E4 t97	CACCCCACAG	TATATATA	TGGGCCCCAA
E4 t98	GGTGGGAACG	TATATATA	ATGTGGATGG

d. E4 sequences selected by kinetic criteria

E4 k2 *	AGAACATAAT	TATATATA	GGGTTGGGGG
E4 k4	GTCTCGCGC	TATATATA	GTGCCGTGGG
E4 k5 **	CAGGTCGGTT	TATATATA	TGGGGGGGC
E4 k7	GGTGCGGCGT	TATATATA	AGTGGGGGTT
E4 k8	CGCAGGGTGA	TATATATA	GGCGCGACCA
E4 k9	GGCGTAGAAC	TATATATA	TGAATGCCGC
E4 k11	TAAGGCCCGT	TATATATA	CAGGGCGTTC
E4 k13	AATGGGGCGG	TATATATA	CGCGGCGGGA
E4 k15	CCTCCCCCGC	TATATATA	TGGCGGCTAA
E4 k17 **	CCGTTGCTTT	TATATATA	GTGCGCCAGG
E4 k18	CCGACGTTG	TATATATA	AGGGGGCTTA
E4 k19	TTCTGTGGTG	TATATATA	GAGGGGTCAT
E4 k20 *	GTATCCGTTT	TATATATA	ATCCGGCGGG
E4 k21	GCGGGTGACG	TATATATA	GTGGGGAGC
E4 k27	TAGTGGCTGC	TATATATA	CGCCCCCAGC
E4 k28	TCGCTCCTAG	TATATATA	CTGAGTGGAG
E4 k30	GCACTAGCGC	TATATATA	TGGTCTGGGG
E4 k31	AATCCACCGC	TATATATA	CACGACACGT
E4 k32	GCGTTTTGCC	TATATATA	GGGGTACGGT
E4 k33	CCCATCCAGC	TATATATA	CTGATCCCCC
E4 k34	AGCTTGGGTG	TATATATA	GTTGCGGGCC
E4 k36	CTGCTGGTGC	TATATATA	GACTGGCGGG
E4 k38	GTCTAGTAAG	TATATATA	AGCAGGCCTC
E4 k51	AACTCCAGCC	TATATATA	TGCGGCTACA
E4 k52	TCCATTCGAC	TATATATA	CGCCGGTCAA
E4 k53	CGCTTGAACC	TATATATA	CGCAGCCCAT
E4 k55	TTTAGGAAGC	TATATATA	CACCCCAACG
E4 k56	TCTTGGGTCT	TATATATA	GGGCGTTCTG
E4 k58	CAGGTAGGGC	TATATATA	CAAAGGGGGG
E4 k59	GAGGTTTAGG	TATATATA	TTGACGCACT
E4 k60	TCGATTGGGG	TATATATA	GTGTGGCAAG
E4 k61	CTAGCCCCGG	TATATATA	CAGCGCCGCC
E4 k63	CGAAAGAGTC	TATATATA	CGCGGGGTGT
E4 k64	AGGGCGTGTC	TATATATA	AGAGGGACGT
E4 k67	CGGAGGTCGA	TATATATA	CGCCGTGTGA
E4 k68	GGGCAGAAGC	TATATATA	CACTTGTGGC
E4 k71	GGAAGGTCGT	TATATATA	TCCGGCCGGC
E4 k74	TCGTTCTTGT	TATATATA	GTGGCGACCG
E4 k76	GGTTAAGTGC	TATATATA	CACGCGTGTC
E4 k78	GTACCAGGTG	TATATATA	CGCGATGGGG
E4 k80 *	TTGCGCTCGC	TATATATA	ATGCCTGGGA
E4 k82	GCGTACAGGC	TATATATA	CAGCGCTGTA