

8. Appendix

8.1. Lists of sequence words

8.1.1. List of sequence words after 5 rounds of minicircle selection

The sequence words represented in the selected sequences were extracted using the McSelex program. The words are sorted first according to their frequency and second alphabetically.

Motif	abs.	rel.									
G	250	0,397	GC	110	0,186	CA	27	0,046	AT	15	0,025
C	202	0,321	CG	107	0,181	AC	26	0,044	TA	12	0,020
T	100	0,159	GG	54	0,091	GA	25	0,042	AA	10	0,017
A	78	0,124	TG	50	0,085	TC	23	0,039	TT	8	0,014
			GT	48	0,081	AG	21	0,036			
			CC	34	0,058	CT	21	0,036			
Motif	rel.	abs.									
GCG	63	0,114	CGA	10	0,018	CCT	5	0,009	GTA	3	0,005
CGC	52	0,094	CTG	10	0,018	TAC	5	0,009	AAA	2	0,004
GTG	27	0,049	GTC	10	0,018	AAC	4	0,007	AAT	2	0,004
CGT	24	0,043	TGG	10	0,018	ACC	4	0,007	ATA	2	0,004
TGC	21	0,038	GGT	9	0,016	AGT	4	0,007	ATT	2	0,004
CGG	19	0,034	TCG	9	0,016	ATC	4	0,007	CAA	2	0,004
GCA	19	0,034	CAG	8	0,014	CAT	4	0,007	CTA	2	0,004
GGC	18	0,033	CCC	8	0,014	GTT	4	0,007	TAA	2	0,004
GGG	17	0,031	TCT	8	0,014	TCC	4	0,007	TTA	2	0,004
ACG	16	0,029	CTC	7	0,013	TGA	4	0,007	TTT	2	0,004
CCG	15	0,027	GAT	7	0,013	TTG	4	0,007	AAG	1	0,002
GCC	15	0,027	GCT	7	0,013	ACA	3	0,005	ACT	1	0,002
CAC	12	0,022	AGC	6	0,011	AGA	3	0,005	TAG	1	0,002
GAG	11	0,020	ATG	6	0,011	CCA	3	0,005	TAT	1	0,002
TGT	11	0,020	GGA	6	0,011	GAA	3	0,005			
			AGG	5	0,009	GAC	3	0,005			
Motif	rel.	abs.									
CGCG	37	0,072	ACGT	6	0,012	CCGT	4	0,008	GATC	3	0,006
GCGC	33	0,064	CGCC	6	0,012	CGAT	4	0,008	GCAT	3	0,006
CGTG	14	0,027	GCAG	6	0,012	CGCA	4	0,008	GCTC	3	0,006
GCGT	12	0,023	GGCA	6	0,012	CGGA	4	0,008	GGCC	3	0,006
CGGG	11	0,021	GTCT	6	0,012	GAGC	4	0,008	GGGT	3	0,006
GTGC	11	0,021	TGCA	6	0,012	GCTG	4	0,008	TACG	3	0,006
GCGG	10	0,019	TGGC	6	0,012	GGTG	4	0,008	TCGC	3	0,006
CACG	9	0,018	ATGC	5	0,010	GTGG	4	0,008	TCGG	3	0,006
ACGC	8	0,016	CCGC	5	0,010	TCTG	4	0,008	AATG	2	0,004
GCAC	8	0,016	CCGG	5	0,010	AACG	3	0,006	ACAC	2	0,004
TGCG	8	0,016	CGAG	5	0,010	AGGT	3	0,006	AGAG	2	0,004
TGTG	8	0,016	CGTC	5	0,010	ATCG	3	0,006	AGTC	2	0,004
GCGA	7	0,014	GCCC	5	0,010	CAGT	3	0,006	ATTG	2	0,004
GGCG	7	0,014	GCCG	5	0,010	CGGC	3	0,006	CAGA	2	0,004
GGGC	7	0,014	GGGG	5	0,010	CTGC	3	0,006	CATG	2	0,004
GTGT	7	0,014	AGCG	4	0,008	CTGG	3	0,006	CCAG	2	0,004
			CCCC	4	0,008	GAGG	3	0,006	CCCC	2	0,004

CAGCG	1	0,002	CTCGA	1	0,002	GCTCC	1	0,002	TACCA	1	0,002
CAGTG	1	0,002	CTCGG	1	0,002	GCTCG	1	0,002	TACGT	1	0,002
CATAC	1	0,002	CTGAC	1	0,002	GCTGC	1	0,002	TAGCG	1	0,002
CATCC	1	0,002	CTGCA	1	0,002	GCTGT	1	0,002	TATAA	1	0,002
CCACC	1	0,002	CTGCG	1	0,002	GGAAC	1	0,002	TCCCA	1	0,002
CCAGA	1	0,002	CTGGC	1	0,002	GGAAT	1	0,002	TCCGC	1	0,002
CCAGT	1	0,002	CTGGG	1	0,002	GGACG	1	0,002	TCCGT	1	0,002
CCCAG	1	0,002	CTGGT	1	0,002	GGAGC	1	0,002	TCGAT	1	0,002
CCCCC	1	0,002	CTGTC	1	0,002	GGCAT	1	0,002	TCGGT	1	0,002
CCCCG	1	0,002	CTGTG	1	0,002	GGCCA	1	0,002	TCTCG	1	0,002
CCCTC	1	0,002	GAATG	1	0,002	GGCCC	1	0,002	TCTGC	1	0,002
CCGAG	1	0,002	GACAC	1	0,002	GGCGG	1	0,002	TCTGG	1	0,002
CCGGG	1	0,002	GACCG	1	0,002	GGCTC	1	0,002	TCTGT	1	0,002
CCGTA	1	0,002	GACGT	1	0,002	GGCTG	1	0,002	TGACA	1	0,002
CCGTT	1	0,002	GAGAG	1	0,002	GGGCA	1	0,002	TGAGC	1	0,002
CCTCT	1	0,002	GAGCA	1	0,002	GGGCT	1	0,002	TGAGG	1	0,002
CCTGA	1	0,002	GAGCC	1	0,002	GGGGG	1	0,002	TGCAA	1	0,002
CCTGC	1	0,002	GAGGC	1	0,002	GGTAC	1	0,002	TGCAC	1	0,002
CGACC	1	0,002	GAGTT	1	0,002	GGTGC	1	0,002	TGCAG	1	0,002
CGAGC	1	0,002	GATGC	1	0,002	GGTGT	1	0,002	TGCCC	1	0,002
CGAGG	1	0,002	GCAGA	1	0,002	GGTTA	1	0,002	TGCCG	1	0,002
CGAGT	1	0,002	GCAGC	1	0,002	GGTTT	1	0,002	TGCGT	1	0,002
CGATG	1	0,002	GCAGG	1	0,002	GTAAA	1	0,002	TGGGT	1	0,002
CGCAG	1	0,002	GCATA	1	0,002	GTACA	1	0,002	TGGTG	1	0,002
CGCCC	1	0,002	GCATC	1	0,002	GTACG	1	0,002	TGGTT	1	0,002
CGCTC	1	0,002	GCATG	1	0,002	GTCCC	1	0,002	TGTCG	1	0,002
CGCTG	1	0,002	GCCAG	1	0,002	GTCCG	1	0,002	TGTGG	1	0,002
CGGAA	1	0,002	GCCCC	1	0,002	GTCGC	1	0,002	TGTGT	1	0,002
CGGAC	1	0,002	GCCCT	1	0,002	GTCGG	1	0,002	TTACC	1	0,002
CGGCA	1	0,002	GCCGA	1	0,002	GTGAG	1	0,002	TTGCG	1	0,002
CGGTA	1	0,002	GCCGC	1	0,002	GTGCT	1	0,002	TTTGC	1	0,002
CGTAA	1	0,002	GCCTG	1	0,002	GTGGT	1	0,002	TTTGT	1	0,002
CGTAC	1	0,002	GCGAC	1	0,002	GTTAC	1	0,002			
CGTCG	1	0,002	GCGCT	1	0,002	TAAAC	1	0,002			
CGTTA	1	0,002	GCGGC	1	0,002	TAACA	1	0,002			
CTCCT	1	0,002	GCGTA	1	0,002	TACAC	1	0,002			

Motif	rel.	abs.									
			ACACGT	2	0,005	CGGCGC	2	0,005	TGCGCG	2	0,005
CGCGCG	17	0,039	ACGTCT	2	0,005	CGGGTG	2	0,005	TGCTGG	2	0,005
GCGCGC	14	0,032	AGCGCG	2	0,005	CGTCTG	2	0,005	TTGGCA	2	0,005
CGTGTG	4	0,009	AGGTCT	2	0,005	CGTGCG	2	0,005	AAACGC	1	0,002
GCACGC	4	0,009	AGTCTA	2	0,005	GAGGTC	2	0,005	AAATGC	1	0,002
GCGCGA	4	0,009	ATCGTG	2	0,005	GATCGT	2	0,005	AACAAG	1	0,002
GCGCGT	4	0,009	ATTGGC	2	0,005	GATTGG	2	0,005	AACGAG	1	0,002
GCGTGC	4	0,009	CACGCG	2	0,005	GCAGTC	2	0,005	AACGCC	1	0,002
TGCGGG	4	0,009	CACGTC	2	0,005	GCCCGT	2	0,005	AACGCG	1	0,002
CGCGAT	3	0,007	CAGTCT	2	0,005	GCCGGC	2	0,005	AAGGTT	1	0,002
CGCGCC	3	0,007	CCGCGC	2	0,005	GCGATC	2	0,005	AATGCC	1	0,002
CGGGCG	3	0,007	CCGGAT	2	0,005	GCGCGG	2	0,005	AATGGC	1	0,002
CGTGCA	3	0,007	CGATCG	2	0,005	GCGTGG	2	0,005	ACAAGG	1	0,002
GCGCAC	3	0,007	CGCACG	2	0,005	GGATTG	2	0,005	ACCACC	1	0,002
GCGGGC	3	0,007	CGCCGG	2	0,005	GGCACG	2	0,005	ACCGTG	1	0,002
GCGGGG	3	0,007	CGCGAG	2	0,005	GGCAGT	2	0,005	ACCTGA	1	0,002
GGGCGT	3	0,007	CGCGCA	2	0,005	GGCGTG	2	0,005	ACGCAG	1	0,002
GTGCGG	3	0,007	CGCGGG	2	0,005	GGGTGA	2	0,005	ACGCGA	1	0,002
GTGTGC	3	0,007	CGCGTC	2	0,005	GGTCTG	2	0,005	ACGCGC	1	0,002
TGGCAG	3	0,007	CGCGTG	2	0,005	TCGCGC	2	0,005	ACGCGG	1	0,002
			CGGATT	2	0,005	TCGTGT	2	0,005	ACGCGT	1	0,002

ACGCTC	1	0,002	CCTCTC	1	0,002	GCATAC	1	0,002	GGTTTG	1	0,002
ACGTCC	1	0,002	CCTGAC	1	0,002	GCATCC	1	0,002	GTAAAC	1	0,002
ACGTGC	1	0,002	CCTGCG	1	0,002	GCATGC	1	0,002	GTACAC	1	0,002
ACGTGT	1	0,002	CGACCG	1	0,002	GCCAGA	1	0,002	GTACGC	1	0,002
ACTCTC	1	0,002	CGAGCG	1	0,002	GCCCCC	1	0,002	GTCCCA	1	0,002
AGAGCG	1	0,002	CGAGGT	1	0,002	GCCC GC	1	0,002	GTCCGC	1	0,002
AGAGGT	1	0,002	CGAGTT	1	0,002	GCCCTC	1	0,002	GTCGCG	1	0,002
AGATCG	1	0,002	CGATGC	1	0,002	GCCGAG	1	0,002	GTCGGT	1	0,002
AGCACG	1	0,002	CGCAGG	1	0,002	GCCGCC	1	0,002	GTCTGC	1	0,002
AGCCGA	1	0,002	CGCCCG	1	0,002	GCCGGG	1	0,002	GTCTGG	1	0,002
AGCGGG	1	0,002	CGCCTG	1	0,002	GCCTGC	1	0,002	GTCTGT	1	0,002
AGCGTG	1	0,002	CGCGAC	1	0,002	GCGACC	1	0,002	GTGAGC	1	0,002
AGGCCA	1	0,002	CGCGGA	1	0,002	GCGAGC	1	0,002	GTGCAA	1	0,002
AGGTTT	1	0,002	CGCGTA	1	0,002	GCGAGT	1	0,002	GTGCAC	1	0,002
AGTGGC	1	0,002	CGCTGC	1	0,002	GCGCCC	1	0,002	GTGCAG	1	0,002
AGTTTG	1	0,002	CGGAAC	1	0,002	GCGCCG	1	0,002	GTGCAT	1	0,002
ATAACA	1	0,002	CGGACG	1	0,002	GCGCCT	1	0,002	GTGCGT	1	0,002
ATACGC	1	0,002	CGGCAT	1	0,002	GCGCTG	1	0,002	GTGCTG	1	0,002
ATCCGT	1	0,002	CGGGCC	1	0,002	GCGGAA	1	0,002	GTGGCA	1	0,002
ATCGCG	1	0,002	CGGGCT	1	0,002	GCGGAC	1	0,002	GTGGCG	1	0,002
ATGCC	1	0,002	CGGGGC	1	0,002	GCGGCG	1	0,002	GTGGTG	1	0,002
ATGCCG	1	0,002	CGGGGG	1	0,002	GCGGGT	1	0,002	GTGTGG	1	0,002
ATGCGC	1	0,002	CGGTAC	1	0,002	GCGTAC	1	0,002	GTGTGT	1	0,002
ATGCGG	1	0,002	CGTAAA	1	0,002	GCGTCC	1	0,002	GTTACC	1	0,002
ATGCTG	1	0,002	CGTACA	1	0,002	GCGTCG	1	0,002	GTTTGC	1	0,002
ATGGCG	1	0,002	CGTCCC	1	0,002	GCGTGT	1	0,002	GTTTGT	1	0,002
CAAGGT	1	0,002	CGTCCG	1	0,002	GCTCCT	1	0,002	TAAACG	1	0,002
CACGCT	1	0,002	CGTCGC	1	0,002	GCTCGG	1	0,002	TAACAA	1	0,002
CACGTG	1	0,002	CGTGGC	1	0,002	GCTGCA	1	0,002	TACACG	1	0,002
CAGAGG	1	0,002	CGTGGT	1	0,002	GCTGGC	1	0,002	TACCAC	1	0,002
CAGATC	1	0,002	CTCGAT	1	0,002	GCTGGT	1	0,002	TACGCA	1	0,002
CAGCGC	1	0,002	CTCGGG	1	0,002	GCTGTC	1	0,002	TACGCG	1	0,002
CAGTGG	1	0,002	CTCTCG	1	0,002	GGAACG	1	0,002	TACGTG	1	0,002
CATACG	1	0,002	CTGACA	1	0,002	GGAATG	1	0,002	TAGCGC	1	0,002
CATCCG	1	0,002	CTGCGC	1	0,002	GGACGT	1	0,002	TATAAC	1	0,002
CATGCG	1	0,002	CTGGCA	1	0,002	GGAGCC	1	0,002	TCCAG	1	0,002
CATGCT	1	0,002	CTGGGT	1	0,002	GGAGC	1	0,002	TCCGTT	1	0,002
CCAGAG	1	0,002	CTGGTT	1	0,002	GGATC	1	0,002	TCGATG	1	0,002
CCAGTG	1	0,002	CTGTCC	1	0,002	GGCCAG	1	0,002	TCGCGA	1	0,002
CCCAGT	1	0,002	GAACGA	1	0,002	GGCCCG	1	0,002	TCGGGC	1	0,002
CCCCCG	1	0,002	GAACGC	1	0,002	GGCGCG	1	0,002	TCGGGT	1	0,002
CCCCGC	1	0,002	GAATGG	1	0,002	GGCGCT	1	0,002	TCGGTA	1	0,002
CCC GCC	1	0,002	GACACG	1	0,002	GGCGGA	1	0,002	TCTCGA	1	0,002
CCC GCG	1	0,002	GACCGT	1	0,002	GGCTCC	1	0,002	TCTGGG	1	0,002
CCC GTA	1	0,002	GACGTC	1	0,002	GGCTGT	1	0,002	TCTGTG	1	0,002
CCC GTG	1	0,002	GAGAGC	1	0,002	GGGCAC	1	0,002	TGACAC	1	0,002
CCCTCT	1	0,002	GAGCAC	1	0,002	GGGCC	1	0,002	TGAGCA	1	0,002
CCGAGG	1	0,002	GAGCCG	1	0,002	GGGCTC	1	0,002	TGAGGC	1	0,002
CCGCCG	1	0,002	GAGCGG	1	0,002	GGGGCA	1	0,002	TGCATA	1	0,002
CCGCCT	1	0,002	GAGCGT	1	0,002	GGGGCC	1	0,002	TGCATG	1	0,002
CCGGCA	1	0,002	GAGGCC	1	0,002	GGGGGC	1	0,002	TGCCCT	1	0,002
CCGGCG	1	0,002	GAGTTT	1	0,002	GGGTGT	1	0,002	TGCCGG	1	0,002
CCGGGG	1	0,002	GATCGC	1	0,002	GGTACG	1	0,002	TGCGCA	1	0,002
CCGTAA	1	0,002	GATGCC	1	0,002	GGTGAG	1	0,002	TGCGTG	1	0,002
CCGTGC	1	0,002	GCACGT	1	0,002	GGTGCG	1	0,002	TGGCAC	1	0,002
CCGTGT	1	0,002	GCAGAT	1	0,002	GGTGTG	1	0,002	TGGCGC	1	0,002
CCGTTA	1	0,002	GCAGCG	1	0,002	GGTTAC	1	0,002	TGGCGG	1	0,002

TGGGTG	1	0,002	TGTCGG	1	0,002	TGTGCT	1	0,002	TTGCGC	1	0,002
TGGTGC	1	0,002	TGTGCA	1	0,002	TGTGTG	1	0,002	TTTGCG	1	0,002
TGGTTA	1	0,002	TGTGCG	1	0,002	TTACCA	1	0,002			
Motif	rel.	abs.	ACCTGAC	1	0,003	CCCTCTC	1	0,003	CGTCTGC	1	0,003
GCGCGCG	11	0,028	ACGCAGG	1	0,003	CCGAGGT	1	0,003	CGTCTGG	1	0,003
CGCGCGC	8	0,020	ACGCGAG	1	0,003	CCGCCGG	1	0,003	CGTGCAA	1	0,003
CGCGCGA	4	0,010	ACGCGCG	1	0,003	CCGCCTG	1	0,003	CGTGCAC	1	0,003
CGCGCGT	3	0,008	ACGCGGA	1	0,003	CCGCGCA	1	0,003	CGTGCAT	1	0,003
CGGGCGT	3	0,008	ACGCGTA	1	0,003	CCGCGCG	1	0,003	CGTGGCA	1	0,003
GCGCGAT	3	0,008	ACGTCCC	1	0,003	CCGGCAT	1	0,003	CGTGGTG	1	0,003
GTGCGGG	3	0,008	ACGTGCA	1	0,003	CCGGCGC	1	0,003	CGTGTGC	1	0,003
ACGTCTG	2	0,005	ACTCTCG	1	0,003	CCGTAAA	1	0,003	CGTGTGG	1	0,003
AGCGCGC	2	0,005	AGAGCGG	1	0,003	CCGTGCA	1	0,003	CGTGTGT	1	0,003
AGGTCTG	2	0,005	AGAGGTC	1	0,003	CCGTGTG	1	0,003	CTCGATG	1	0,003
ATCGTGT	2	0,005	AGATCGC	1	0,003	CCTGACA	1	0,003	CTCGGGC	1	0,003
ATTGGCA	2	0,005	AGCACGC	1	0,003	CCTGCGC	1	0,003	CTCTCGA	1	0,003
CAGTCTA	2	0,005	AGCCGAG	1	0,003	CGACCGT	1	0,003	CTGACAC	1	0,003
CCGGATT	2	0,005	AGCGGGC	1	0,003	CGAGCGT	1	0,003	CTGCGCG	1	0,003
CGATCGT	2	0,005	AGGCCAG	1	0,003	CGAGGTC	1	0,003	CTGGCAC	1	0,003
CGCGATC	2	0,005	AGTTTG	1	0,003	CGAGTTT	1	0,003	CTGGGTG	1	0,003
CGCGCAC	2	0,005	AGTGGCG	1	0,003	CGATGCC	1	0,003	CTGGTTA	1	0,003
CGCGTGC	2	0,005	AGTTTGC	1	0,003	CGCACGC	1	0,003	CTGTCGG	1	0,003
CGGATTG	2	0,005	ATAACAA	1	0,003	CGCCCGT	1	0,003	GAACGAG	1	0,003
CGTGCGG	2	0,005	ATACGCG	1	0,003	CGCCGGC	1	0,003	GAACGCG	1	0,003
GAGGTCT	2	0,005	ATCCGTT	1	0,003	CGCCGGG	1	0,003	GAATGGC	1	0,003
GATCGTG	2	0,005	ATCGCGC	1	0,003	CGCCTGC	1	0,003	GACACGT	1	0,003
GATTGGC	2	0,005	ATGCCCT	1	0,003	CGCGACC	1	0,003	GACCGTG	1	0,003
GCACGCG	2	0,005	ATGCCGG	1	0,003	CGCGAGC	1	0,003	GACGTCT	1	0,003
GCAGTCT	2	0,005	ATGCGCA	1	0,003	CGCGCCC	1	0,003	GAGAGCG	1	0,003
GCGATCG	2	0,005	ATGCGGG	1	0,003	CGCGCCG	1	0,003	GAGCACG	1	0,003
GCGCACG	2	0,005	ATGCTGG	1	0,003	CGCGCCT	1	0,003	GAGCCGA	1	0,003
GCGCGCC	2	0,005	ATGGCGG	1	0,003	CGCGCGG	1	0,003	GAGCGGG	1	0,003
GCGCGGG	2	0,005	CAAGGTT	1	0,003	CGCGGAC	1	0,003	GAGCGTG	1	0,003
GCGCGTC	2	0,005	CACGCGC	1	0,003	CGCGGGC	1	0,003	GAGCCA	1	0,003
GCGCGTG	2	0,005	CACGCGG	1	0,003	CGCGGGG	1	0,003	GAGTTTG	1	0,003
GCGGGCG	2	0,005	CACGCTC	1	0,003	CGCGTAC	1	0,003	GATCGCG	1	0,003
GCGTGCG	2	0,005	CACGTCC	1	0,003	CGCGTCC	1	0,003	GATGCC	1	0,003
GGATTGG	2	0,005	CACGTCT	1	0,003	CGCGTCG	1	0,003	GCACGCT	1	0,003
GGCAGTC	2	0,005	CACGTGT	1	0,003	CGCTGCA	1	0,003	GCAGATC	1	0,003
GGGCGTG	2	0,005	CAGAGGT	1	0,003	CGGAACG	1	0,003	GCAGCGC	1	0,003
TCGCGCG	2	0,005	CAGATCG	1	0,003	CGGACGT	1	0,003	GCATACG	1	0,003
TCGTGTG	2	0,005	CAGCGCG	1	0,003	CGGCATC	1	0,003	GCATCCG	1	0,003
TGCGGGG	2	0,005	CAGTGGC	1	0,003	CGGCGCG	1	0,003	GCATGCG	1	0,003
TGGCAGT	2	0,005	CATACGC	1	0,003	CGGCGCT	1	0,003	GCCAGAG	1	0,003
TTGGCAG	2	0,005	CATCCGT	1	0,003	CGGGCCC	1	0,003	GCCCCCG	1	0,003
AAACGCC	1	0,003	CATGCGC	1	0,003	CGGGCTC	1	0,003	GCCCCGC	1	0,003
AAATGCC	1	0,003	CATGCTG	1	0,003	CGGGGCC	1	0,003	GCCCCGTA	1	0,003
AACAAGG	1	0,003	CCAGAGG	1	0,003	CGGGGGC	1	0,003	GCCCCGTG	1	0,003
AACGCGT	1	0,003	CCAGTGG	1	0,003	CGGGTGA	1	0,003	GCCCCTCT	1	0,003
AAGGTTT	1	0,003	CCCAGTG	1	0,003	CGGGTGT	1	0,003	GCCGAGG	1	0,003
AATGCCG	1	0,003	CCCCCGC	1	0,003	CGGTACG	1	0,003	GCCGCCG	1	0,003
AATGGCG	1	0,003	CCCCGCG	1	0,003	CGTAAAC	1	0,003	GCCGGCA	1	0,003
ACAAGGT	1	0,003	CCCGCCT	1	0,003	CGTACAC	1	0,003	GCCGGCG	1	0,003
ACACGTC	1	0,003	CCCGCGC	1	0,003	CGTCCCA	1	0,003	GCCGGGG	1	0,003
ACACGTG	1	0,003	CCCGTAA	1	0,003	CGTCCGC	1	0,003	GCCTGCG	1	0,003
ACCGTGC	1	0,003	CCCGTGT	1	0,003	CGTCGCG	1	0,003	GCGACCG	1	0,003

GCGAGCG	1	0,003	GGCAGCG	1	0,003	GTCTGGG	1	0,003	TCTGGGT	1	0,003
GCGAGTT	1	0,003	GGCATCC	1	0,003	GTCTGTG	1	0,003	TGACACG	1	0,003
GCGCCCG	1	0,003	GGCCAGA	1	0,003	GTGAGCA	1	0,003	TGAGCAC	1	0,003
GCGCCGG	1	0,003	GGCCGT	1	0,003	GTGCATG	1	0,003	TGAGGCC	1	0,003
GCGCGAC	1	0,003	GGCGCGG	1	0,003	GTGCGTG	1	0,003	TGCATAC	1	0,003
GCGCGCA	1	0,003	GGCGCTG	1	0,003	GTGCTGG	1	0,003	TGCATGC	1	0,003
GCGCTGC	1	0,003	GGCGGAA	1	0,003	GTGGCAG	1	0,003	TGCCCTC	1	0,003
GCGGAAC	1	0,003	GGCGTGC	1	0,003	GTGGCGC	1	0,003	TGCCGGC	1	0,003
GCGGACG	1	0,003	GGCGTGT	1	0,003	GTGGTGC	1	0,003	TGCCCAC	1	0,003
GCGGGCG	1	0,003	GGCTCCT	1	0,003	GTGTGCA	1	0,003	TGCGCGC	1	0,003
GCGGGCT	1	0,003	GGCTGTC	1	0,003	GTGTGCG	1	0,003	TGCGCGT	1	0,003
GCGGGGC	1	0,003	GGGCACG	1	0,003	GTGTGTG	1	0,003	TGCGGGC	1	0,003
GCGGGGG	1	0,003	GGGCCCG	1	0,003	GTTACCA	1	0,003	TGCGGGT	1	0,003
GCGGGTG	1	0,003	GGGCTCC	1	0,003	GTTTGCG	1	0,003	TGCGTGC	1	0,003
GCGTACA	1	0,003	GGGGCAC	1	0,003	TAAACGC	1	0,003	TGCTGGC	1	0,003
GCGTCCG	1	0,003	GGGGGCA	1	0,003	TAACAAG	1	0,003	TGCTGGT	1	0,003
GCGTCCG	1	0,003	GGGTGAG	1	0,003	TACACGT	1	0,003	TGGCACG	1	0,003
GCGTGCA	1	0,003	GGGTGTG	1	0,003	TACCACC	1	0,003	TGGCAGC	1	0,003
GCGTGGC	1	0,003	GGTACGC	1	0,003	TACGCAG	1	0,003	TGGCGGA	1	0,003
GCGTGGT	1	0,003	GGTCTGT	1	0,003	TACGCGA	1	0,003	TGGGTGA	1	0,003
GCGTGTG	1	0,003	GGTGAGC	1	0,003	TACGTGC	1	0,003	TGGTGC	1	0,003
GCTCGGG	1	0,003	GGTGCGG	1	0,003	TAGCGCG	1	0,003	TGGTTAC	1	0,003
GCTGGCA	1	0,003	GGTGTGC	1	0,003	TATAACA	1	0,003	TGTCGGT	1	0,003
GCTGGTT	1	0,003	GGTTACC	1	0,003	TCCCAGT	1	0,003	TGTGCAG	1	0,003
GCTGTCG	1	0,003	GGTTTGT	1	0,003	TCCGTTA	1	0,003	TGTGCGT	1	0,003
GGAACGA	1	0,003	GTAAACG	1	0,003	TCGATGC	1	0,003	TGTGCTG	1	0,003
GGAATGG	1	0,003	GTACACG	1	0,003	TCGCGAG	1	0,003	TGTGTGC	1	0,003
GGACGTC	1	0,003	GTACGCA	1	0,003	TCGGGCG	1	0,003	TTACCAC	1	0,003
GGAGCCG	1	0,003	GTCCCAG	1	0,003	TCGGGTG	1	0,003	TTGCGCG	1	0,003
GGCACGC	1	0,003	GTCGCGA	1	0,003	TCGGTAC	1	0,003	TTTGCGC	1	0,003
GGCACGT	1	0,003	GTCGGTA	1	0,003	TCTCGAT	1	0,003			

Motif	rel.	abs.									
C	6	0,017	GCGTGC	2	0,006	AGAGTCT	1	0,003	CAGCGCG	1	0,003
G	4	0,011	GGATTGG	2	0,006	AGATCGG	1	0,003	CAGTGGG	1	0,003
C	4	0,011	GGCAGTCT	2	0,006	AGCCGAG	1	0,003	CATACGG	1	0,003
G	3	0,008	GTGCGGG	2	0,006	AGCGGGC	1	0,003	CATCCGT	1	0,003
C	2	0,006	TGGCAGT	2	0,006	AGGCCAGA	1	0,003	CATGCGCA	1	0,003
A	2	0,006	TTGGCAGT	2	0,006	AGGTCTGT	1	0,003	CATGCTGG	1	0,003
A	2	0,006	AAATGCCG	1	0,003	AGGTTTGT	1	0,003	CCAGAGGT	1	0,003
A	2	0,006	AACAAGGT	1	0,003	AGTGCGC	1	0,003	CCAGTGG	1	0,003
C	2	0,006	AACGCGTA	1	0,003	AGTTTGC	1	0,003	CCCAGTGG	1	0,003
C	2	0,006	AAGTTT	1	0,003	ATAACAAG	1	0,003	CCCCCGG	1	0,003
C	2	0,006	AATGCCG	1	0,003	ATACGCGA	1	0,003	CCCCGCG	1	0,003
C	2	0,006	AATGGCG	1	0,003	ATCCGTTA	1	0,003	CCCGCTG	1	0,003
C	2	0,006	ACAAGGTT	1	0,003	ATCGCGC	1	0,003	CCCGCGCA	1	0,003
C	2	0,006	ACACGTCT	1	0,003	ATGCCCTC	1	0,003	CCCGTAA	1	0,003
C	2	0,006	ACACGTGT	1	0,003	ATGCCGGC	1	0,003	CCCGTGT	1	0,003
C	2	0,006	ACCGTGCA	1	0,003	ATGCGCAC	1	0,003	CCGAGGTC	1	0,003
G	2	0,006	ACCTGACA	1	0,003	ATGCGGGT	1	0,003	CCGCCGG	1	0,003
G	2	0,006	ACGCGAGC	1	0,003	ATGCTGGC	1	0,003	CCGCCTG	1	0,003
G	2	0,006	ACGCGGAC	1	0,003	ATGGCGGA	1	0,003	CCGCGCAC	1	0,003
G	2	0,006	ACGCGTAC	1	0,003	CAAGTTT	1	0,003	CCGCGCG	1	0,003
G	2	0,006	ACGTCCA	1	0,003	CACGCGC	1	0,003	CCGGCATC	1	0,003
G	2	0,006	ACGTCTGC	1	0,003	CACGCGGA	1	0,003	CCGGCGCT	1	0,003
G	2	0,006	ACGTCTGG	1	0,003	CACGTCCC	1	0,003	CCGTAAAC	1	0,003
G	2	0,006	ACGTGCAT	1	0,003	CACGTCTG	1	0,003	CCGTGCAA	1	0,003
G	2	0,006	ACTCTCGA	1	0,003	CAGAGGTC	1	0,003	CCGTGTGT	1	0,003
G	2	0,006	AGAGCGGG	1	0,003	CAGATCGC	1	0,003	CCTGACAC	1	0,003

CCTGCGCG	1	0,003	GAACGCGT	1	0,003	GCGTACAC	1	0,003	GTGTGCAG	1	0,003
CGACCGTG	1	0,003	GAATGGCG	1	0,003	GCGTCCGC	1	0,003	GTGTGCGT	1	0,003
CGAGCGTG	1	0,003	GACACGTC	1	0,003	GCGTCGCG	1	0,003	GTGTGTGC	1	0,003
CGAGGTCT	1	0,003	GACCGTGC	1	0,003	GCGTGCAC	1	0,003	GTTACCAC	1	0,003
CGAGTTTG	1	0,003	GACGTCTG	1	0,003	GCGTGCCA	1	0,003	GTTTGCGC	1	0,003
CGATGCCC	1	0,003	GAGAGCGG	1	0,003	GCGTGGTG	1	0,003	TAAACGCC	1	0,003
CGCACGCT	1	0,003	GAGCACGC	1	0,003	GCGTGTGC	1	0,003	TAACAAGG	1	0,003
CGCCCGTG	1	0,003	GAGCCGAG	1	0,003	GCTCGGGC	1	0,003	TACACGTG	1	0,003
CGCCGGCA	1	0,003	GAGCGGGC	1	0,003	GCTGGCAC	1	0,003	TACGCAGG	1	0,003
CGCCGGGG	1	0,003	GAGGCCAG	1	0,003	GCTGGTTA	1	0,003	TACGCGAG	1	0,003
CGCCTGCG	1	0,003	GAGTTTGC	1	0,003	GCTGTGCG	1	0,003	TACGTGCA	1	0,003
CGCGACCG	1	0,003	GATCGCGC	1	0,003	GGAACGAG	1	0,003	TAGCGCGC	1	0,003
CGCGAGCG	1	0,003	GATGCCCT	1	0,003	GGAATGGC	1	0,003	TATAACAA	1	0,003
CGCGCACG	1	0,003	GCACGCGC	1	0,003	GGACGTCT	1	0,003	TCCCAGTG	1	0,003
CGCGCCCG	1	0,003	GCACGCGG	1	0,003	GGAGCCGA	1	0,003	TCGATGCC	1	0,003
CGCGCCCG	1	0,003	GCACGCTC	1	0,003	GGCACGCG	1	0,003	TCGCGCGC	1	0,003
CGCGCGAC	1	0,003	GCAGATCG	1	0,003	GGCAGCGC	1	0,003	TCGCGCGT	1	0,003
CGCGCGGG	1	0,003	GCAGCGCG	1	0,003	GGCATCCG	1	0,003	TCGGGCGT	1	0,003
CGCGCGTG	1	0,003	GCATACGC	1	0,003	GGCCAGAG	1	0,003	TCGGGTGT	1	0,003
CGCGGACG	1	0,003	GCATCCGT	1	0,003	GGCCCGTA	1	0,003	TCGGTACG	1	0,003
CGCGGGCT	1	0,003	GCATGCGC	1	0,003	GGCGCGGG	1	0,003	TCGTGTGG	1	0,003
CGCGGGGG	1	0,003	GCCAGAGG	1	0,003	GGCGCTGC	1	0,003	TCTCGATG	1	0,003
CGCGTACA	1	0,003	GCCCCGCG	1	0,003	GGCGAAC	1	0,003	TCTGGGTG	1	0,003
CGCGTCCG	1	0,003	GCCCGCCT	1	0,003	GGCGTCCG	1	0,003	TGACACGT	1	0,003
CGCGTCGC	1	0,003	GCCCGTAA	1	0,003	GGCGTGTG	1	0,003	TGAGCACG	1	0,003
CGCGTGCG	1	0,003	GCCCGTGT	1	0,003	GGCTGTCTG	1	0,003	TGAGGCCA	1	0,003
CGGAACGA	1	0,003	GCCCTCTC	1	0,003	GGGCACGT	1	0,003	TGCATACG	1	0,003
CGGACGTC	1	0,003	GCCGAGGT	1	0,003	GGGCCCGT	1	0,003	TGCATGCG	1	0,003
CGGCATCC	1	0,003	GCCGCCGG	1	0,003	GGGCGTGC	1	0,003	TGCCCTCT	1	0,003
CGGCGCGG	1	0,003	GCCGGCAT	1	0,003	GGGCGTGT	1	0,003	TGCCGGCG	1	0,003
CGGCGCTG	1	0,003	GCCGGCGC	1	0,003	GGGCTCCT	1	0,003	TGCGCACG	1	0,003
CGGGCCCG	1	0,003	GCCTGCGC	1	0,003	GGGGCACG	1	0,003	TGCGCGCA	1	0,003
CGGGCTCC	1	0,003	GCGACCGT	1	0,003	GGGGGCAC	1	0,003	TGCGCGTG	1	0,003
CGGGGGCA	1	0,003	GCGAGCGT	1	0,003	GGGTGAGC	1	0,003	TGCGGGCG	1	0,003
CGGGTGAG	1	0,003	GCGAGTTT	1	0,003	GGGTGTGC	1	0,003	TGCGGGGC	1	0,003
CGGGTGTG	1	0,003	GCGCACGC	1	0,003	GGTACGCA	1	0,003	TGCGGGTG	1	0,003
CGGTACGC	1	0,003	GCGCCCGT	1	0,003	GGTCTGTG	1	0,003	TGCGTGCA	1	0,003
CGTAAACG	1	0,003	GCGCCGGG	1	0,003	GGTGAGCA	1	0,003	TGCTGGCA	1	0,003
CGTACACG	1	0,003	GCGCGACC	1	0,003	GGTGCGGG	1	0,003	TGCTGGTT	1	0,003
CGTCCCAG	1	0,003	GCGCGCAC	1	0,003	GGTGTGCG	1	0,003	TGGCACGC	1	0,003
CGTCGCGA	1	0,003	GCGCGCCG	1	0,003	GGTTACCA	1	0,003	TGGCAGCG	1	0,003
CGTCTGGG	1	0,003	GCGCGCCT	1	0,003	GTAACGCG	1	0,003	TGGCGGAA	1	0,003
CGTGCATG	1	0,003	GCGCGCGG	1	0,003	GTACACGT	1	0,003	TGGTGCGG	1	0,003
CGTGGCAG	1	0,003	GCGCGGGC	1	0,003	GTACGCAG	1	0,003	TGGTTACC	1	0,003
CGTGGTGC	1	0,003	GCGCGGGG	1	0,003	GTCCAGT	1	0,003	TGTCGGTA	1	0,003
CGTGTGTG	1	0,003	GCGCGTCC	1	0,003	GTCGCGAG	1	0,003	TGTGCGTG	1	0,003
CTCGATGC	1	0,003	GCGCGTCTG	1	0,003	GTCGGTAC	1	0,003	TGTGCTGG	1	0,003
CTCGGGCG	1	0,003	GCGCTGCA	1	0,003	GTCTGGGT	1	0,003	TGTGTGCA	1	0,003
CTCTCGAT	1	0,003	GCGGAACG	1	0,003	GTGAGCAC	1	0,003	TTACCACC	1	0,003
CTGACACG	1	0,003	GCGGACGT	1	0,003	GTGCATGC	1	0,003	TTGCGCGC	1	0,003
CTGCGCGT	1	0,003	GCGGCGCG	1	0,003	GTGCGGGC	1	0,003	TTTGCGCG	1	0,003
CTGGCACG	1	0,003	GCGGGCTC	1	0,003	GTGCGTGC	1	0,003			
CTGGGTGA	1	0,003	GCGGGGCC	1	0,003	GTGCTGGT	1	0,003			
CTGGTTAC	1	0,003	GCGGGGGC	1	0,003	GTGGCAGC	1	0,003			
CTGTGCGT	1	0,003	GCGGGTGA	1	0,003	GTGGTGCG	1	0,003			

8.1.2. List of sequence words after 8 rounds of minicircle selection

The sequence words represented in the selected sequences were extracted using the McSelex program. The words are sorted first according to their frequency and second alphabetically.

Motif	abs.	rel.									
GC	55	0,242	GT	14	0,062	TC	7	0,031	GA	4	0,018
CG	51	0,225	CT	10	0,044	TT	7	0,031	AA	3	0,013
GG	19	0,084	CA	9	0,040	TA	6	0,026			
TG	15	0,066	CC	9	0,040	AG	5	0,022			
			AC	8	0,035	AT	5	0,022			
Motif	abs.	rel.									
GCG	39	0,184	ATG	3	0,014	GAT	2	0,009	ATC	1	0,005
CGC	30	0,142	CCC	3	0,014	GGT	2	0,009	ATT	1	0,005
TGC	11	0,052	CCG	3	0,014	GTA	2	0,009	CAA	1	0,005
GTG	10	0,047	CGA	3	0,014	GTC	2	0,009	CAG	1	0,005
CGG	8	0,038	GCT	3	0,014	TGT	2	0,009	CCA	1	0,005
GGG	8	0,038	TCG	3	0,014	TTC	2	0,009	CTA	1	0,005
CGT	6	0,028	TCT	3	0,014	AAA	1	0,005	TAA	1	0,005
GCA	6	0,028	TTA	3	0,014	AAC	1	0,005	TAT	1	0,005
GGC	6	0,028	AGG	2	0,009	AAG	1	0,005	TCA	1	0,005
CTT	5	0,024	CAC	2	0,009	ACA	1	0,005	TGG	1	0,005
GCC	5	0,024	CAT	2	0,009	ACC	1	0,005	TTT	1	0,005
ACG	4	0,019	CCT	2	0,009	ACT	1	0,005			
TAC	4	0,019	CTC	2	0,009	AGA	1	0,005			
			CTG	2	0,009	AGC	1	0,005			
			GAG	2	0,009	AGT	1	0,005			
Motif	abs.	rel.									
CGCG	22	0,112	CTTA	2	0,010	ATTA	1	0,005	GCCG	1	0,005
TGCG	9	0,046	GCAC	2	0,010	CAAA	1	0,005	GCTC	1	0,005
GTGC	7	0,036	GCAT	2	0,010	CACG	1	0,005	GGCA	1	0,005
CGGG	6	0,030	GCCC	2	0,010	CACT	1	0,005	GGCC	1	0,005
GCGG	6	0,030	GCCT	2	0,010	CAGG	1	0,005	GGGT	1	0,005
CGTG	5	0,025	GCTT	2	0,010	CCCA	1	0,005	GGTA	1	0,005
GCGT	5	0,025	GGGG	2	0,010	CCGG	1	0,005	GGTC	1	0,005
GGGC	4	0,020	GTGT	2	0,010	CCTG	1	0,005	GTAA	1	0,005
ATGC	3	0,015	TCGC	2	0,010	CCTT	1	0,005	GTAC	1	0,005
CGCA	3	0,015	TCTT	2	0,010	CGAG	1	0,005	GTCA	1	0,005
CGCC	3	0,015	TGTG	2	0,010	CGGC	1	0,005	GTCT	1	0,005
GCGA	3	0,015	TTCT	2	0,010	CGTA	1	0,005	TAAG	1	0,005
GGCG	3	0,015	AAAC	1	0,005	CTAT	1	0,005	TACA	1	0,005
TTAC	3	0,015	AACG	1	0,005	CTGC	1	0,005	TACC	1	0,005
ACGC	2	0,010	AAGG	1	0,005	CTGG	1	0,005	TACG	1	0,005
CATG	2	0,010	ACCC	1	0,005	CTTC	1	0,005	TATT	1	0,005
CCCG	2	0,010	ACGG	1	0,005	CTTT	1	0,005	TCGT	1	0,005
CCGC	2	0,010	ACTA	1	0,005	GAGC	1	0,005	TCTG	1	0,005
CGAT	2	0,010	AGAG	1	0,005	GAGT	1	0,005	TGCA	1	0,005
CGCT	2	0,010	AGCT	1	0,005	GATC	1	0,005	TGGT	1	0,005
CTCG	2	0,010	AGGC	1	0,005	GATG	1	0,005	TTTC	1	0,005
			AGTG	1	0,005	GCAA	1	0,005			
			ATCG	1	0,005	GCAG	1	0,005			
Motif	abs.	rel.									
CGCGC	15	0,082	GCGCG	15	0,082	CGTGC	4	0,022	GCGTG	4	0,022
			GTGCG	5	0,027	GCGGG	4	0,022	ATGCG	3	0,016

CGGGC	3	0,016	ACTAT	1	0,005	CTTCT	1	0,005	GGTAC	1	0,005
GCGCC	3	0,016	AGAGC	1	0,005	CTTTC	1	0,005	GGTCA	1	0,005
TGCGC	3	0,016	AGCTC	1	0,005	GAGCT	1	0,005	GTAAG	1	0,005
ACGCG	2	0,011	AGGCG	1	0,005	GAGTG	1	0,005	GTACA	1	0,005
CATGC	2	0,011	AGTGT	1	0,005	GATCG	1	0,005	GTCTT	1	0,005
CCCGC	2	0,011	ATCGT	1	0,005	GATGC	1	0,005	GTGCA	1	0,005
CCGCG	2	0,011	ATTAC	1	0,005	GCAAA	1	0,005	TAAGG	1	0,005
CGCCT	2	0,011	CAAAC	1	0,005	GCACG	1	0,005	TACCC	1	0,005
CGCGA	2	0,011	CACGC	1	0,005	GCACT	1	0,005	TATTA	1	0,005
CGCGG	2	0,011	CACTA	1	0,005	GCAGG	1	0,005	TCGCA	1	0,005
CGCGT	2	0,011	CAGGC	1	0,005	GCCGG	1	0,005	TCGCG	1	0,005
CGCTT	2	0,011	CCGGG	1	0,005	GCCTG	1	0,005	TCGTG	1	0,005
CGGGG	2	0,011	CCTGC	1	0,005	GCCTT	1	0,005	TCTGG	1	0,005
CTCGC	2	0,011	CCTTA	1	0,005	GCGAG	1	0,005	TCTTC	1	0,005
CTTAC	2	0,011	CGAGT	1	0,005	GCGGC	1	0,005	TCTTT	1	0,005
GCATG	2	0,011	CGATC	1	0,005	GCGTA	1	0,005	TGCAT	1	0,005
GCCCG	2	0,011	CGATG	1	0,005	GCTCG	1	0,005	TGCGA	1	0,005
GCGAT	2	0,011	CGCAC	1	0,005	GCTTA	1	0,005	TGCGG	1	0,005
GCGCA	2	0,011	CGCAG	1	0,005	GGCAA	1	0,005	TGGTC	1	0,005
GCGCT	2	0,011	CGCAT	1	0,005	GGCCC	1	0,005	TGTGC	1	0,005
GGCGG	2	0,011	CGCCC	1	0,005	GGCGC	1	0,005	TTACC	1	0,005
GTGTG	2	0,011	CGGCG	1	0,005	GGGCA	1	0,005	TTACG	1	0,005
TGCGT	2	0,011	CGTAA	1	0,005	GGGCC	1	0,005	TTCTG	1	0,005
AAACG	1	0,005	CGTGT	1	0,005	GGGCG	1	0,005	TTCTT	1	0,005
AACGC	1	0,005	CTATT	1	0,005	GGGGC	1	0,005	TTTCT	1	0,005
ACCCA	1	0,005	CTGCG	1	0,005	GGGGT	1	0,005			
ACGGG	1	0,005	CTGGT	1	0,005	GGGTA	1	0,005			

Motif	abs.	rel.									
GCGGCG	10	0,060	AGAGCT	1	0,006	CGGGGC	1	0,006	GCGTAA	1	0,006
CGCGCG	9	0,054	AGCTCG	1	0,006	CGGGGT	1	0,006	GCTCGC	1	0,006
GCGTGC	4	0,024	AGGCGG	1	0,006	CGTAAG	1	0,006	GCTTAC	1	0,006
CGCGCC	3	0,018	AGTGTG	1	0,006	CGTGTG	1	0,006	GGCAAA	1	0,006
CGTGCG	3	0,018	ATCGTG	1	0,006	CTATTA	1	0,006	GGCCCG	1	0,006
TGCGCG	3	0,018	ATGCGG	1	0,006	CTCGCA	1	0,006	GGCGCA	1	0,006
ATGCGC	2	0,012	ATTACC	1	0,006	CTCGCG	1	0,006	GGCGGG	1	0,006
CATGCG	2	0,012	CAAACG	1	0,006	CTGGTC	1	0,006	GGGCAA	1	0,006
CCCGCG	2	0,012	CACGCG	1	0,006	CTTACG	1	0,006	GGGCC	1	0,006
CCGCGC	2	0,012	CACTAT	1	0,006	CTTCTG	1	0,006	GGGCGC	1	0,006
CGCGAT	2	0,012	CAGGCG	1	0,006	CTTTCT	1	0,006	GGGGCA	1	0,006
CGCGCT	2	0,012	CCGGGG	1	0,006	GAGCTC	1	0,006	GGGGTA	1	0,006
CGCGGG	2	0,012	CCTGCG	1	0,006	GAGTGT	1	0,006	GGGTAC	1	0,006
GCATGC	2	0,012	CCTTAC	1	0,006	GATCGT	1	0,006	GGTACA	1	0,006
GCCCGC	2	0,012	CGAGTG	1	0,006	GATGCG	1	0,006	GTAAGG	1	0,006
GCGCCT	2	0,012	CGATCG	1	0,006	GCAAAC	1	0,006	GTCTTT	1	0,006
GCGCGA	2	0,012	CGATGC	1	0,006	GCACGC	1	0,006	GTGCAT	1	0,006
GCGCGT	2	0,012	CGCACT	1	0,006	GCACTA	1	0,006	GTGCGA	1	0,006
GCGCTT	2	0,012	CGCAGG	1	0,006	GCAGGC	1	0,006	GTGCGC	1	0,006
GCGGGC	2	0,012	CGCATG	1	0,006	GCCGGG	1	0,006	GTGTGC	1	0,006
GTGCGT	2	0,012	CGCCCG	1	0,006	GCCTGC	1	0,006	TACCCA	1	0,006
TGCGTG	2	0,012	CGCCTG	1	0,006	GCCTTA	1	0,006	TATTAC	1	0,006
AAACGC	1	0,006	CGCCTT	1	0,006	GCGAGT	1	0,006	TCGCAC	1	0,006
AACGCG	1	0,006	CGCGCA	1	0,006	GCGATC	1	0,006	TCGCGG	1	0,006
ACGCGC	1	0,006	CGCGTA	1	0,006	GCGATG	1	0,006	TCGTGT	1	0,006
ACGCGG	1	0,006	CGCGTG	1	0,006	GCGCAG	1	0,006	TCTGGT	1	0,006
ACGGGC	1	0,006	CGCTTA	1	0,006	GCGCAT	1	0,006	TCTTCT	1	0,006
ACTATT	1	0,006	CGGCGG	1	0,006	GCGCCC	1	0,006	TCTTTC	1	0,006
			CGGGCC	1	0,006	GCGGCG	1	0,006	TGCATG	1	0,006
			CGGGCG	1	0,006	GCGGGG	1	0,006	TGCGAG	1	0,006

Motif	abs.	rel.	Motif	abs.	rel.	Motif	abs.	rel.	Motif	abs.	rel.
TGCGGC	1	0,006	TGTGCG	1	0,006	TTCTGG	1	0,006	TTTCTT	1	0,006
TGGTCA	1	0,006	TTACCC	1	0,006	TTCTTC	1	0,006			
			CATGCGC	1	0,007	CTCGCGG	1	0,007	GGCAAAC	1	0,007
			CATGCGG	1	0,007	CTGGTCA	1	0,007	GGCCCGC	1	0,007
			CCGCGCG	1	0,007	CTTCTGG	1	0,007	GGCGCAG	1	0,007
			CCGCGCT	1	0,007	CTTTCTT	1	0,007	GGCGGGC	1	0,007
			CCGGGGC	1	0,007	GAGCTCG	1	0,007	GGGCAAA	1	0,007
			CCTTACG	1	0,007	GAGTGTG	1	0,007	GGGCCCG	1	0,007
			CGAGTGT	1	0,007	GATCGTG	1	0,007	GGGCGCA	1	0,007
			CGATCGT	1	0,007	GATGCGC	1	0,007	GGGGCAA	1	0,007
			CGATGCG	1	0,007	GCAAACG	1	0,007	GGGGTAC	1	0,007
			CGACTA	1	0,007	GCACGCG	1	0,007	GGGTACA	1	0,007
			CGCAGGC	1	0,007	GCACTAT	1	0,007	GTCTTTC	1	0,007
			CGCATGC	1	0,007	GCAGGCG	1	0,007	GTGCATG	1	0,007
			CGCCCGC	1	0,007	GCCGGGG	1	0,007	GTGCGAG	1	0,007
			CGCCTGC	1	0,007	GCCTGCG	1	0,007	GTGCGCG	1	0,007
			CGCCTTA	1	0,007	GCCTTAC	1	0,007	GTGTGCG	1	0,007
			CGCGATC	1	0,007	GCGAGTG	1	0,007	TATTACC	1	0,007
			CGCGATG	1	0,007	GCGATCG	1	0,007	TCGCACT	1	0,007
			CGCGCAT	1	0,007	GCGATGC	1	0,007	TCGCGGG	1	0,007
			CGCGCCC	1	0,007	GCGCAGG	1	0,007	TCGTGTG	1	0,007
			CGCGCGT	1	0,007	GCGCATG	1	0,007	TCTGGTC	1	0,007
			CGCGGGG	1	0,007	GCGCCCG	1	0,007	TCTTCTG	1	0,007
			CGCGTAA	1	0,007	GCGCCTG	1	0,007	TCTTTCT	1	0,007
			CGCGTGC	1	0,007	GCGCCTT	1	0,007	TGCATGC	1	0,007
			CGCTTAC	1	0,007	GCGCGCA	1	0,007	TGCGAGT	1	0,007
			CGGGCCC	1	0,007	GCGCGCT	1	0,007	TGCGCGT	1	0,007
			CGGGCGC	1	0,007	GCGCGTA	1	0,007	TGCGGCG	1	0,007
			CGGGGCA	1	0,007	GCGCGTG	1	0,007	TGTGCGT	1	0,007
			CGGGGTA	1	0,007	GCGCTTA	1	0,007	TTACCCA	1	0,007
			CGTAAGG	1	0,007	GCGGCGG	1	0,007	TTCTGGT	1	0,007
			CGTGCGC	1	0,007	GCGGGCC	1	0,007	TTCTTCT	1	0,007
			CGTGCGT	1	0,007	GCGGGGT	1	0,007	TTTCTTC	1	0,007
			CTATTAC	1	0,007	GCGTAAG	1	0,007			
			CTCGCAC	1	0,007	GCTCGCG	1	0,007			
			ATTACCCA	1	0,007	CGCCTTAC	1	0,007	CTATTACC	1	0,007
			CAAACGCG	1	0,007	CGCGATCG	1	0,007	CTCGCACT	1	0,007
			CACGCGCG	1	0,007	CGCGATGC	1	0,007	CTCGCGG	1	0,007
			CACTATTA	1	0,007	CGCGCATG	1	0,007	CTTCTGGT	1	0,007
			CAGGCGGG	1	0,007	CGCGCCCG	1	0,007	CTTTCTTC	1	0,007
			CATGCGCG	1	0,007	CGCGCCTG	1	0,007	GAGCTCGC	1	0,007
			CATGCGGC	1	0,007	CGCGCCTT	1	0,007	GAGTGTGC	1	0,007
			CCC GCGCG	1	0,007	CGCGCGCC	1	0,007	GATCGTGT	1	0,007
			CCC GCGCT	1	0,007	CGCGCGCT	1	0,007	GATGCGCG	1	0,007
			CCGCGCGC	1	0,007	CGCGCGTG	1	0,007	GCAAACGC	1	0,007
			CCGCGCTT	1	0,007	CGCGCTTA	1	0,007	GCACGCGC	1	0,007
			CCGGGGCA	1	0,007	CGCGGGGT	1	0,007	GCACTATT	1	0,007
			CGAGTGTG	1	0,007	CGCGTAAG	1	0,007	GCAGGCGG	1	0,007
			CGATCGTG	1	0,007	CGCGTGCG	1	0,007	GCATGCGC	1	0,007
			CGATGCGC	1	0,007	CGGGCCCG	1	0,007	GCATGCGG	1	0,007
			CGCACTAT	1	0,007	CGGGCGCA	1	0,007	GCCGGGGC	1	0,007
			CGCAGGCG	1	0,007	CGGGGCAA	1	0,007	GCCTTACG	1	0,007
			CGCATGCG	1	0,007	CGGGGTAC	1	0,007	GCGAGTGT	1	0,007
			CGCCCGCG	1	0,007	CGTGCGCG	1	0,007	GCGATCGT	1	0,007
			CGCCTGCG	1	0,007	CGTGCGTG	1	0,007	GCGATGCG	1	0,007

GCGCAGGC	1	0,007	GCGCGTGC	1	0,007	GGGCGCAG	1	0,007	TCTTTCTT	1	0,007
GCGCATGC	1	0,007	GCGCTTAC	1	0,007	GGGGCAA	1	0,007	TGCATGCG	1	0,007
GCGCCCGC	1	0,007	GCGGGCCC	1	0,007	GGGGTACA	1	0,007	TGCGAGTG	1	0,007
GCGCCTGC	1	0,007	GCGGGGTA	1	0,007	GTCTTCT	1	0,007	TGCGCGCC	1	0,007
GCGCCTTA	1	0,007	GCGTAAGG	1	0,007	GTGCATGC	1	0,007	TGCGCGCG	1	0,007
GCGCGATC	1	0,007	GCGTGCGC	1	0,007	GTGCGAGT	1	0,007	TGCGCGTA	1	0,007
GCGCGATG	1	0,007	GCGTGCCT	1	0,007	GTGCGCGC	1	0,007	TGCGGCGG	1	0,007
GCGCGCAT	1	0,007	GCTCGCGG	1	0,007	GTGTGCGT	1	0,007	TGCGTGCG	1	0,007
GCGCGCCC	1	0,007	GGCAAACG	1	0,007	TATTACCC	1	0,007	TGTGCGTG	1	0,007
GCGCGCGA	1	0,007	GGCCCGCG	1	0,007	TCGCACTA	1	0,007	TTCTGGTC	1	0,007
GCGCGCGT	1	0,007	GGCGCAGG	1	0,007	TCGCGGGG	1	0,007	TTCTTCTG	1	0,007
GCGCGCTT	1	0,007	GGGCAAAC	1	0,007	TCTGGTCA	1	0,007	TTTCTTCT	1	0,007
GCGCGTAA	1	0,007	GGGCCCGC	1	0,007	TCTTCTGG	1	0,007			

8.2. Predicting the propensity of a DNA sequence to adopt the Z-DNA conformation

The propensity of DNA to flip to the Z-DNA conformation is strongly dependent on the base sequence. A few simple rules have been formulated in order to predict whether a given sequence is likely to adopt the Z-conformation. First of all, alternating purine and pyrimidine bases are required. Second, the particular dinucleotides have to be considered. Whereas CG/GC dinucleotides are most conducive and TG/CA dinucleotides favorable for the formation of Z-DNA, the presence of TA/AT dinucleotides are generally thought not to form Z-DNA well (Sinden 1996).

Since it is not feasible to measure the Z-DNA propensity of large numbers of sequences, algorithms have been developed that use simple rules for the computation of values indicative of a sequences likelihood to form Z-DNA. (Konopka et al 1985; Ho et al. 1986) One algorithm implemented in the Z-Hunt programs uses a model based on statistical mechanics. In short, empirically derived energetic parameters for the individual basepairs are used to calculate the propensity of the sequence to adopt the Z-DNA conformation relative to that of a random sequence. Limitations of the program are the treatment of sequences as a set of dinucleotides and the fact that since not all energetic parameters are known, some of the parameters had to be estimated. Hence the quality of the prediction is limited and some of the predicted Z-DNA sites are not consistent with experimental data (Schroth et al. 1992).

Recently a different model has been proposed in which the sequence is no longer treated as a series of dinucleotides. Starting from experimental data free energies needed to induce the transition from B- to Z-DNA were defined for the individual basepairs. The free energies depend on which base is present in the anti or syn

conformation (Eichman et al. 1999). The propensity of a sequence to adopt the Z-DNA conformation is the sum of the free energies of the individual basepairs.

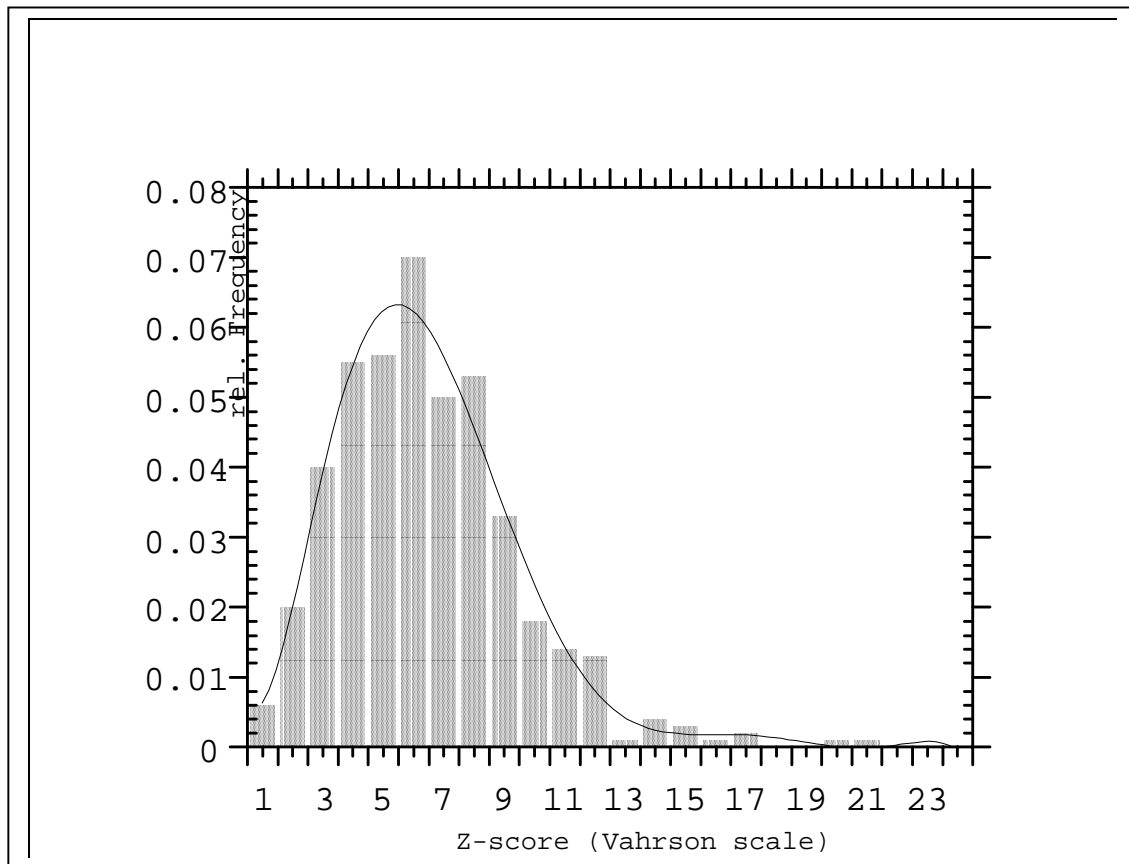


Figure 8.1 Distribution of Z-scores in randomized sequences. The relative frequencies of the Z-scores are shown in bars. A smoothed curve has been added.

The algorithm used here originates in the work of W. Vahrson and uses a simple scheme of assigning different values to the different dinucleotides (CG/GC : 2; TG/AC : 1; TA/AT : 0). Breaks in the sequence of alternating purine and pyrimidine bases are assigned a penalty of -1 (Wölfl et al. 1995). Application of this algorithm to a particular sequence results in a number, which I refer to as the Z-score on the Vahrson scale. But what is the significance of the values obtained ? In other words : what value indicates a good Z-DNA forming sequence ? In order to investigate this question, an attempt was made to calculate the distribution of Z-scores for randomly chosen sequences. Using the computer program 'Shuffle', a collection of 1000 sequences of 16 random nucleotides enclosed on either side by the adjoining 6 nucleotides present in the minicircle DNA was assembled. In order to mimic the situation of the selection experiment, the frequencies of the individual bases were adjusted to the ratio A:C:G:T = 20:30:30:20 . The Z-scores were calculated and the frequencies of individual Z-scores determined. A graph of the results is shown in

Figure 8.1 . It should be noted that the bars do not add up to a frequency of 1 . Since the algorithm used, does not assign a Z-score to sequences containing less than 5 consecutive bases of alternating purines and pyrimidines. Likewise, any negative Z-score is disregarded. Therefore, the vast majority of Z-scores are not represented in the graph. It is of interest to compare the calculated Z-scores of sequences with experimental data. Both the plasmids pME.C and pME.E are often used as sources for Z-DNA. The (TG)₃₀ insert present in pME.C has a Z-score of 59 and the (CG)₁₁ insert present in pME.E has a Z-score of 42 . The motifs Z1, Z2 and Z3 present in the promoter of the human c-myc gene and shown to adopt the Z-DNA conformation *in vivo* have Z-scores of 17, 20/18 and 30 respectively (Vahrson 1991). The plasmid pME.A is often used as a plasmid that does not contain any sequences likely to adopt the Z-DNA conformation. The highest Z-scores found in the pME.A plasmid are 12 (once), 11 (5 times) and 10 (twice). One of the sites with a Z-score of 11 was eliminated in the plasmid pME.JA. In the minicircles the highest Z-score outside the randomized region has a value of 8.

8.3. Statistics of randomized libraries

The aim of the *in vitro* selection experiments was the selection of DNA sequences that bind to the Z α peptide. Of interest was not only whether the selected sequences favored formation of Z-DNA, but also whether, within the sequence-space of potentially Z-DNA forming sequences, certain sequences were bound better than others. Since the binding sequence was expected to be shorter than the length of the randomized sequence employed, analysis of the data looked for sequence words of a certain but initially unknown length. One question that has to be addressed is whether the occurrence of certain words at more than average frequency could be due to statistical chance. In the following I therefore want to present a calculation of the odds of finding an arbitrary sequence-word in the selected sequences by chance.

First the total number of possible different sequences has to be computed. In the case of the (YR)₈ library each of the 16 positions is twofold degenerate, i.e. there are $2^{16} = 65,536$ different sequences.

How many different sequences will then contain a specific sequence-word of a given length ? As an example a sequence length of 5 nucleotides is chosen here. If out of the 16 total nucleotides 5 are fixed by the chosen word, the other 11 nucleotides are

still random . Since each nucleotide is twofold degenerate, there are 2^{11} different sequences. However, the block of 5 nucleotides containing the chosen sequence can be placed at 12 different positions within the 16 nucleotide sequence, therefore a total of $12 \times 2^{11} = 24,576$ different sequences contain the chosen word. Plasmid DNA is a double stranded molecule. A sequence-motif bound by a ligand can occur in either of the two strands. Therefore the number of possible sequences containing a chosen word has to be multiplied twofold, yielding : $2 \times 12 \times 2^{11} = 49,152$ different sequences.

In the case of the selection experiment performed in this work it is useful to take into account that the randomized sequence does not exist in isolation but is surrounded by fixed plasmid sequences. The sequences were designed such that the randomized alternating purine and pyrimidine bases were surrounded by a 5´G and a 3´C . Since these two bases fit with the expectation that selected sequences should be rich in alternating GC dinucleotides this ´bordereffect´ should be taken into account. The border sequence accounts for another 2^{12} possible sequences containing the chosen motif if it has a 5´G and a 3´C . Therefore, the total number of different sequences containing the chosen word comes to $2^{11} \times 2 \times 12 + 2^{12} = 53,248$. The chance of finding a word is given by the number of different sequences containing the word divided by the number of different sequences contained in the library. In the chosen example of a word 5 nucleotides long this computes to $49,152 / 65,536 = 0.75$ or if the border effect is included in the calculation $53,248 / 65,536 = 0.81$. Hence, in a $(YR)_8$ library, a given 5mer sequence will appear in 81 % of all sequences by chance. Table 8.1 lists the chances of finding a specific word in a given sequence calculated for different word-lengths :

length of word	statistical likelihood	incl. bordereffect
4	1	1
5	0.75	0.81
6	0.20	0.38
7	0.094	0.17
8	0.043	0.078

Table 8.1 Chance of a sequence word being represented in any sequence of the $(YR)_8$ library

It can be concluded that using a $(YR)_8$ library only sequence motifs of at least 6 nucleotides length can be detected. If shorter binding motifs are to be detected, the randomized sequence obviously has to be shortened. In the particular case of the

selection of Z-DNA binding proteins shortening the randomized sequence is not an option since the ease of formation of Z-DNA increases with the length of the stretch of alternating purines and pyrimidines. The selection experiments in this work always were a combination of two selection processes : on the one hand those sequences were selected that most easily formed Z-DNA. on the other hand those, that were best bound by the $Z\alpha$ protein. In the case that $Z\alpha$ does not just recognize the Z-DNA conformation the experiment has to be designed such that the selection for the Z-DNA conformation does not become dominant.

With the above considerations in mind experiments were done in which a minicircle library containing a stretch of 16 totally randomized nucleotides was used. Disregarding any border effects, the statistics are as follows. The total number of possible sequences amounts to $4^{16} = 4.3 \times 10^9$. The number of possible sequences containing a sequence word of length n can be calculated using the formula : $2 \times 17 - n \times 4^{16-n}$. The statistical likelihood to find a given sequence word by chance in such a library is listed in Table 8.2 for different lengths of the randomized region.

length of word	statistical likelihood
3	0.44
4	0.10
5	0.023
6	0.0053
7	0.0012
8	0.00028

Table 8.2 Chance of a sequence word to be represented in any sequence of a N_{16} library

It is obvious that the situation encountered with a N_{16} library is quite different from that of a $(YR)_8$ library. In the case of a N_{16} library sequence motifs as short as 4 nucleotides and possibly just 3 nucleotides long can be detected.

8.4. Computer programs

All computer programs were written in the computer language C and compiled using the Think C compiler on an Apple Macintosh computer.

8.4.1. The McSelex program

The McSelex program was used to collect the sequence word contained in the selected sequences

```
#include <stdio.h>
#include <string.h>

#define NUMBERSEQUENCE 9      /* Number of sequences to analyze */
#define MAXMOTIVES 1024
#define MOTIVLENGTH 4        /* Length of the sequence words */
#define TRUE 1
#define FALSE 0
#define OUTPUT "MOLCH:unselex_4" /* Filename for output */

/* To make things easy the sequences are coded in the program itself by
initializing the *sequence[] definition. The sequences have to entered in
the following fashion :{"ACGT","TGCA"}; */
/* Likewise the output file has to be included in the source code using
the OUTPUT definition */

static char *sequence[] = { }; /* Enter sequences here ! */

main()
{
/* A data structure motivsatz is defined to hold the detected motives */

struct motivsatz {
    char motiv[MOTIVLENGTH + 1];
    int anzahl;
};

struct motivsatz motive[MAXMOTIVES];
int runner1, runner2, runner3;
int numbermotives = 0; /* keeps track of the list of motives */
int flag; /* motiv already in list ? */
int tester;
char currentmotiv[MOTIVLENGTH];
FILE *filep;

/* main loop */

for (runner1 = 0; runner1 < NUMBERSEQUENCE; runner1++)
{
    for (runner2 = 0; runner2 < strlen(sequence[runner1]) - MOTIVLENGTH +
1; runner2++)
    {
        /* Copy that part of the sequence to be checked : */

        for (runner3 = 0; runner3 < MOTIVLENGTH; runner3++)
```



```

        currentmotiv[runner3] = sequence[runner1][runner2 +
            runner3];
currentmotiv[runner3] = '\0';

/* check the current motive against all existing motives */

for (runner3 = 0, flag = FALSE; runner3 < numbermotives;
runner3 ++)
    {
        if (!strcmp(currentmotiv, motive[runner3].motiv))
            {
                motive[runner3].anzahl++;
                flag = TRUE;
            }
    }
if (!flag)
    {
        strcpy(motive[runner3].motiv, currentmotiv);
        motive[runner3].anzahl = 1;
        numbermotives++;
    }
}

/* Output to the screen */

for (runner1 = 0; runner1 < numbermotives; runner1++)
    {
        printf("%s %d \n",&motive[runner1].motiv,
            motive[runner1].anzahl);
    }

/* Results are saved in a file */

filep = fopen(OUTPUT,"w");

for (runner1 = 0; runner1 < numbermotives; runner1++)
    {
        fprintf(filep,"%s %d \n",&motive[runner1].motiv,
            motive[runner1].anzahl);
    }
fclose (filep);
}

```

8.4.2. The Shuffle program

The program Shuffle was used to generate randomized sequences.

```

#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#define RIGHT_END "GAAGGG"
#define FILENAME "Molch:mischung2" /* filename for output */
#define ANZ_SEQUENZEN 1000 /* Number of sequences to be generated*/

FILE *outputfile;

char sequence[30] ;
char left_end [] = "CGCTCAG";
float random1;

```

```
int runner, runner2;

main()
{
    outputfile = fopen(FILENAME,"w");
    for (runner2 = 0; runner2 < ANZ_SEQUENZEN; runner2++)
    {
        strcpy(sequence, left_end);
        for (runner = 1; runner < 17; runner++)
        {
            random1 = rand();
            random1 = random1/RAND_MAX;
            if (random1 < 0.2) strcat(sequence,"A");
            else if (random1 < 0.4) strcat(sequence,"T");
            else if (random1 < 0.7) strcat(sequence,"C");
            else strcat(sequence, "G");
        }
        strcat(sequence, RIGHT_END);
        printf("%s",sequence);fprintf(outputfile,"%s\n",sequence);
    }
    fclose(outputfile);
}
```