Discrete Mathematics 312 (2012) 1062–1075

Contents lists available at SciVerse ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc

Linear and nonlinear constructions of DNA codes with Hamming distance d, constant GC-content and a reverse-complement constraint

Niema Aboluion, Derek H. Smith, Stephanie Perkins *

Division of Mathematics and Statistics, University of Glamorgan, Pontypridd, Mid Glamorgan, CF37 1DL, Wales, UK

ARTICLE INFO

Article history:
Received 24 September 2010
Received in revised form 22 November 2011
Accepted 23 November 2011
Available online 15 December 2011

Keywords:
DNA codes
Constant GC-content
Reverse-complement constraint
Linear and nonlinear codes

ABSTRACT

In a previous paper, the authors used cyclic and extended cyclic constructions to obtain codes over an alphabet {A,C,G,T} satisfying a Hamming distance constraint and a GC-content constraint. These codes are applicable to the design of synthetic DNA strands used in DNA microarrays, as DNA tags in chemical libraries and in DNA computing. The GC-content constraint specifies that a fixed number of positions are G or C in each codeword, which ensures uniform melting temperatures. The Hamming distance constraint is a step towards avoiding unwanted hybridizations. This approach extended the pioneering work of Gaborit and King. In the current paper, another constraint known as a reverse-complement constraint is added to further prevent unwanted hybridizations.

Many new best codes are obtained, and are reproducible from the information presented here. The reverse-complement constraint is handled by searching for an involution with 0 or 1 fixed points, as first done by Gaborit and King. Linear codes and additive codes over GF(4) and their cosets are considered, as well as shortenings of these codes. In the additive case, codes obtained from two different mappings from GF(4) to {A,C,G,T} are considered.

1. Introduction

The deoxyribonucleic acid (DNA) molecule consists of two complementary strands. Each strand is a sequence of four different nucleotide bases, called adenine (A), cytosine (C), guanine (G) and thymine (T). In coding theory terms, each strand can be regarded as a word constructed from the alphabet {A, C, G, T}. DNA can be synthesized, and there has been considerable interest in the construction of codes over this alphabet satisfying certain combinatorial constraints (see [8,5] and the references contained therein). The code is used to design the synthetic DNA strands (known as *oligonucleotides*) to control their hybridization, permitting desired hybridizations and deterring undesirable imperfect hybridizations.

Oligonucleotides can be used as probes in DNA microarray technologies. They can also be used as tags or bar codes in chemical libraries [2], exploiting their highly predictable hybridization chemistry. Tagged chemical libraries can help to automate the process of drug screening. In DNA computing, an important step is to construct an appropriate encoding of the problem in DNA oligonucleotide sequences. This must be done in such a way that hybridization finds the desired solution. In all these applications, imperfect hybridizations with an oligonucleotide which is close to, but somewhat different from the target can introduce errors and reduce efficiency. The library of words must always be large enough to represent the necessary information, suggesting that the code should be as large as possible.

The first three constraints specified below are the combinatorial constraints on a DNA code $\mathcal{C}$:

* Corresponding author.
E-mail addresses: niemaali@hotmail.com (N. Aboluion), dhsmith@glam.ac.uk (D.H. Smith), sperkins@glam.ac.uk (S. Perkins).

0012-365X/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.disc.2011.11.021

1. Let $H(x, y)$ denote the Hamming distance between two words (i.e. the number of positions in which they differ). The Hamming distance constraint is that $H(x, y) \geq d$ for all $x, y \in \mathcal{C}$ with $x \neq y$, for some prescribed minimum distance d.
2. The reverse-complement constraint is that $H(x^{RC}, y) \geq d$ for all $x, y \in \mathcal{C}$, where x^{RC} is the reverse-complement of x obtained by taking the reverse x^R of x and performing the symbol interchanges A $\leftrightarrow$ T, C $\leftrightarrow$ G (this is called taking Watson–Crick complements). Note that $x = y$ is included.
3. The GC-content constraint is that each codeword $x \in \mathcal{C}$ has the same GC-content. The GC-content of a *DNA* word is defined to be the number of positions in which the word has coordinate C or G.
4. A further constraint is used as an intermediate step in handling the reverse-complement constraint. The reverse constraint is that $H(x^R, y) \geq d$ for all $x, y \in \mathcal{C}$, where x^R is the reverse of a codeword x. As for the reverse-complement constraint, $x = y$ is included.

The purpose of the first two constraints is to make non-desirable hybridization unlikely to occur. The fixed GC-content constraint is used to ensure that similar melting temperatures are obtained, where DNA melting is the process by which double-stranded DNA unwinds and separates into single strands through the breaking of hydrogen bonding between the bases. Similar melting temperatures can be approximately achieved by ensuring that each word contains the same number of positions which are either G or C, referred to as constant GC-content [4]. Because CG base-pairing is generally stronger than AT base-pairing, this allows hybridization of multiple words to take place simultaneously [12].

Following [5], the maximum number of codewords of length n, minimum Hamming distance d, GC-content w satisfying the first three constraints is denoted $A_4^{GC,RC}(n, d, w)$. As the actual value of w is unimportant, the aim of the paper is to improve lower bounds for $\max_w(A_4^{GC,RC}(n, d, w))$. In [13] the algebraic approach of Gaborit and King [5] was extended for codes satisfying the Hamming distance and GC-content constraints. A comprehensive evaluation of linear cyclic and extended cyclic codes over GF(4) and $\mathbf{Z}_4$ as well as additive codes was undertaken for $n \leq 30$. Cosets of codes were considered, together with shortened and punctured codes. In the case of additive codes and codes over $\mathbf{Z}_4$, attention was also paid to the two distinct choices for the mapping from the field or ring to the alphabet {A, C, G, T}. As far as was computationally feasible, all possible choices of polynomial (or polynomials) were considered to generate the code. A GC-weight enumerator was calculated (see [5]) and the code was selected that gave the maximum number of codewords of some fixed GC-content.

This paper will follow the same approach, rather than using algorithmic methods [9–11]. The codes found in [13] will be used to construct codes that also satisfy the reverse-complement constraint. Many improvements to the best known values for $\max_w(A_4^{GC,RC}(n, d, w))$ are found, and tables of the best known codes are presented.

2. The reverse-complement constraint and involutions

The use of involutions to handle the reverse-complement constraint was pioneered by Gaborit and King [5], who stated and proved the following lemma:

Lemma 1. *Let $\mathcal{C}'$ be a code of length n such that:*

- *$n = 2k$ is even and $\mathcal{C}'$ has a fixed-point free involution in its permutation group (i.e. a permutation of the form $(a_1, a_2) \cdots (a_{2k-1}, a_{2k})$ which leaves no column unchanged); or*
- *$n = 2k + 1$ is odd and $\mathcal{C}'$ has a one-point fixed involution in its permutation group (i.e. a permutation of the form $(a_1, a_2) \cdots (a_{2k-1}, a_{2k})$ which leaves one column unchanged).*

Then $\mathcal{C}'$ is permutation equivalent to a code $\mathcal{C}$ that has the reverse permutation R in its permutation group.

The lemma is proved simply by considering the permutation that sends column a_{2i-1} to column i and column a_{2i} to column $n + 1 - i$, for $1 \leq i \leq k$ (and a_{2k+1} to $k + 1$ if n is odd). The code $\mathcal{C}$ can be written as a disjoint union $\mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1 \cup \mathcal{C}_2$, where $\mathcal{C}_0$ is the set of codewords fixed by R and $\mathcal{C}_1$, $\mathcal{C}_2$ are two sets that are interchanged by R. Either the set of codewords $\mathcal{C}_1$ or the set $\mathcal{C}_2$ can be chosen as a code that satisfies the reverse constraint and the Hamming distance constraint for the value of d prescribed for $\mathcal{C}'$. Gaborit and King also proved:

Lemma 2. *Let n be even. For the code $\mathcal{C}_1$ replace each of the first $n/2$ coordinates by its Watson–Crick complement ($A \leftrightarrow T$, $C \leftrightarrow G$). The code $\mathcal{C}_3$ obtained satisfies the reverse-complement constraint and the Hamming distance constraint for the value of d prescribed for $\mathcal{C}'$.*

Note that the GC-content of codewords is unaffected by this operation on the code. For odd n the situation is slightly more complicated, as the operation in Lemma 2 can reduce the Hamming distance between a codeword of $\mathcal{C}_1$ and the reverse-complement of a codeword of $\mathcal{C}_1$ by 1.

Lemma 3. *Let $n = 2k + 1$ be odd. For the code $\mathcal{C}_1$ replace each of the first $\lfloor n/2 \rfloor$ coordinates by its Watson–Crick complement ($A \leftrightarrow T$, $C \leftrightarrow G$). The code obtained consists of four subcodes $\mathcal{C}_A$, $\mathcal{C}_C$, $\mathcal{C}_G$, $\mathcal{C}_T$ in which the coordinate $k + 1$ is A,C,G and T respectively. Then the two subcodes $\mathcal{C}_3 = \mathcal{C}_A \cup \mathcal{C}_C$ or $\mathcal{C}_4 = \mathcal{C}_G \cup \mathcal{C}_T$ both satisfy the reverse-complement constraint and the Hamming distance constraint for the value of d prescribed for $\mathcal{C}'$.*

As $|\mathcal{C}_A| + |\mathcal{C}_C| + |\mathcal{C}_G| + |\mathcal{C}_T| = |\mathcal{C}_1|$, one of the two codes has at least half as many codewords as $\mathcal{C}_1$. If coordinate $k+1$ takes a constant value then one of the codes $\mathcal{C}_3$ or $\mathcal{C}_4$ has the same number of codewords as $\mathcal{C}_1$.

The operations in the lemmas can be applied to any linear or additive codes. Given the code $\mathcal{C}_3$ (or $\mathcal{C}_4$) the GC-weight enumerator can be calculated. The largest set of words of constant GC-weight then gives a code which is a candidate for $\max_w(A_4^{\mathrm{GC,RC}}(n, d, w))$. Different involutions can give different sizes for this set of words, either because of different sizes for $|\mathcal{C}_0|$, or from the nature of the construction when n is odd.

Note that there were no cases in [13] where a code gave a larger number of codeword using a linear code over the ring $\mathbf{Z}_4$ than could be obtained from a linear code over GF(4) or an additive code. Thus linear codes over $\mathbf{Z}_4$ are not considered further here. As in [13] all computations were carried out in Magma.[1]

3. Linear cyclic and extended cyclic codes over GF(4)

Ideally all generator polynomials (see [7]) should be examined (as in [13]), as well as all possible involutions (which are easily generated by Magma). The methods of the previous section are then applied to give a cyclic (or extended cyclic) code satisfying the Hamming distance constraint and the reverse-complement constraint. Then the largest set of codewords of fixed GC-content is selected as a candidate for $\max_w(A_4^{\mathrm{GC,RC}}(n, d, w))$. This was done for $n \leq 18$. For $19 \leq n \leq 30$ this did not prove feasible, and three restricted searches were used:

1. All polynomials were generated as in [13], but the involution search was truncated after a fixed period of time and the best result obtained in this period was used.
2. The polynomial that gave the best result for $\max_w(A_4^{\mathrm{GC}}(n, d, w))$ in [13] was used, together with a search of all possible involutions. This was possible when the permutation group of the code was small enough.
3. The polynomial that gave the best result for $\max_w(A_4^{\mathrm{GC}}(n, d, w))$ in [13] was also used when the permutation group was too large for a complete search. The Magma command to generate representatives of conjugacy classes was used. Either a suitable involution was found, or its nonexistence was shown (in which case no candidate code was generated).

4. Cosets of linear cyclic and extended cyclic codes over GF(4)

A coset of a cyclic or extended cyclic code satisfying the Hamming distance and reverse-complement constraints may have more codewords of fixed GC-content than the code itself. In order for the reverse-complement constraint to be satisfied the coset leader selected must be reversible (fixed by R). The code $\mathcal{C}_1$ is then replaced by the coset given by the selected coset leader, before Lemma 2 or Lemma 3 is applied. The coset then satisfies the Hamming distance and reverse-complement constraints and a GC-weight enumerator can be computed as before.

It only proved feasible to consider all generator polynomials and all involutions for $n \leq 18$. In these cases between 30 and 40 random reversible coset leaders were selected (using the method in [5] to ensure the coset leaders are reversible). For $19 \leq n \leq 30$ the polynomial that gave the best result for $\max_w(A_4^{\mathrm{GC}}(n, d, w))$ in [13] was used. For this single polynomial there were two methods that could be applied:

1. For smaller permutation groups all involutions were considered, together with between 5 and 10 random reversible coset leaders.
2. For larger permutation groups the Magma command to generate representatives of conjugacy classes was used. Either a suitable single involution was found, or its nonexistence was shown. For this involution 10 random reversible cosets were considered.

5. Additive cyclic and additive extended cyclic codes

Additive codes are additive subspaces of $(\mathrm{GF}(4))^n$. The following theorem describes all additive cyclic codes.

Theorem 4 (*[3]*)**.** *Let $\mathcal{C}$ be an $(n, 2^k)$ additive cyclic code of length n over* GF(4) *(with elements $0, 1, \omega, \omega^2$). Then $\mathcal{C} = \langle \omega p(x) + q(x), r(x)\rangle$ where $p(x)$, $r(x)$ are binary polynomials that divide $(x^n - 1)$mod 2, $r(x)$ divides $q(x)(x^n - 1)/p(x)$mod 2, and $k = 2n - \deg p - \deg r$.*

Note that, if $\langle \omega p(x) + q(x), r(x)\rangle$ and $\langle \omega p'(x) + q'(x), r'(x)\rangle$ are two representations of an additive cyclic code then $p'(x) = p(x)$, $r'(x) = r(x)$ and $q'(x) \equiv q(x) \bmod r(x)$.

As described in [13] there are two distinct mappings from $\{0, 1, \omega, \omega^2\}$ to $\{A, C, T, G\}$ that need to be considered for additive codes, pairing 0 with 1 for G, C or pairing 0 with ω for G, C. The larger number of codewords taken over the two cases is used.

Ideally all sets of three polynomials $p(x)$, $q(x)$, $r(x)$ should be examined (as in [13]), as well as all possible involutions (if any suitable involutions exist). The methods of the previous section can then be applied. This only proved feasible for

1 http://magma.maths.usyd.edu.au/magma/.

Table 1
Best lower bounds for $\max_w(A_4^{\mathrm{GC,RC}}(n,d,w))$ after shortening and puncturing operations $3 \le d \le 11$. Entries in bold and all entries for $n > 20$ are new best known results. Entries in italics are equal to the best known results.

n/d	3	4	5	6	7	8	9	10	11
4	$\mathit{6}_{ca}$	1_{ca}	-	-	-	-	-	-	-
5	4_{cf}	$\mathit{3}_{cf}$	-	-	-	-	-	-	-
6	27_{ca}	$\mathit{16}_{ca}$	-	1_{ca}	-	-	-	-	-
7	35_{ef}	22_{ea}	-	1_{ea}^{co}	-	-	-	-	-
8	280_{ca}^{co}	108_{ef}	22_{nb}^{T5}	$\mathit{12}_{ea}$	-	1_{ca}	-	-	-
9	252_{st}^{3}	126_{ca}	22_{ea}^{co}	8_{ca}	-	1_{ea}	-	-	-
10	3280_{st}^{6}	$\mathbf{860}_{cf}$	$\mathbf{210}_{ca}$	50_{ca}	5_{st}^{6}	4_{cf}	-	2_{nt}^{C6}	-
11	3656_{st}^{6}	920_{st}^{3}	226_{ea}	56_{nt}^{C2}	5_{st}^{4}	4_{ef}	-	$\mathit{2}_{nt}^{C5}$	-
12	$\mathbf{59136}_{ca}$	$\mathbf{14784}_{ca}$	$\mathbf{1848}_{ca}$	$\mathit{924}_{ca}$	84_{ca}	24_{ca}	6_{ca}	-	$\mathit{2}_{nt}^{C4}$
13	54752_{ea}	13520_{st}^{5}	1696_{ea}	880_{ea}^{co}	68_{st}^{3}	30_{st}^{6}	4_{ea}	3_{nt}^{A1}	-
14	$\mathbf{768768}_{st}^{8}$	$\mathit{192192}_{ca}$	7424_{ca}	$\mathbf{3712}_{nb}^{C8}$	$\mathbf{796}_{ca}^{co}$	$\mathbf{208}_{nt}^{A1}$	12_{st}^{6}	$\mathbf{21}_{ca}$	4_{nt}^{A1}
15	411560_{ea}	213584_{ea}	12835_{cf}	$\mathbf{6648}_{ea}^{co}$	802_{st}^{8}	$\mathbf{410}_{st}^{8}$	15_{nt}^{C2}	31_{cf}^{co}	3_{st1}^{5}
16	$\mathit{13174400}_{ef}$	$\mathit{3293600}_{ef}$	51608_{sb}^{9}	$\mathbf{55424}_{ea}^{co}$	$\mathbf{13856}_{ef}^{co}$	$\mathbf{3776}_{ef}^{co}$	194_{st}^{9}	20_{ca}^{co}	$\mathbf{68}_{ef}^{co}$
17	777640_{st}^{9}	3111120_{cf}	$\mathbf{97520}_{st}^{3}$	48550_{st}^{9}	12120_{cf}	$\mathit{6060}_{cf}$	196_{nt}^{A2}	28_{st}^{7}	5_{st}^{9}
18	$\mathbf{44933184}_{st}^{10}$	$\mathbf{11266112}_{sb}^{10}$	$\mathbf{699624}_{ca}$	$\mathbf{699624}_{ca}$	$\mathbf{43632}_{sb}^{10}$	$\mathbf{10908}_{st}^{10}$	$\mathbf{2691}_{ca}$	144_{ca}	28_{st}^{8}
19	11824384_{pt}^{10}	11266112_{st}^{10}	$\mathbf{1477796}_{st}^{10}$	$\mathbf{738772}_{ea}$	$\mathbf{92252}_{ea}$	$\mathbf{11542}_{ea}$	2893_{ea}^{co}	198_{st}^{10}	53_{st}^{5}
20	755728384_{st}^{11}	$\mathbf{189432064}_{cf}^{co}$	$\mathbf{11822368}_{sb}^{11}$	738520_{ca}	$\mathbf{369008}_{ca}$	$\mathit{184756}_{ef}$	6012_{ca}	1592_{pr}^{11}	400_{st}^{2}
21	90291264_{st}^{11}	188416000_{ef}	22573824_{cf}	1412068_{cf}^{co}	176772_{st}^{8}	45112_{pt}^{8}	11148_{pt}^{2}	2926_{st}^{2}	847_{st}^{4}
22	10602158336_{st}^{12}	2650495232_{st}^{12}	22607872_{st}^{12}	5643456_{pt}^{12}	1410864_{st}^{12}	353496_{pt}^{1}	88424_{sb}^{12}	22088_{st}^{11}	5522_{st}^{3}
23	1384513088_{st}^{12}	670222080_{st}^{12}	43264648_{pt}^{12}	10816624_{pt}^{12}	2703694_{st}^{12}	676312_{st}^{11}	169182_{st}^{6}	42968_{pt}^{3}	10701_{st}^{3}
24	177279886336_{st}^{13}	44319794176_{st}^{13}	346436544_{pt1}^{12}	43355616_{st}^{12}	21631400_{sb}^{13}	5406464_{pt}^{13}	1351616_{st}^{13}	338016_{st}^{13}	84964_{st}^{3}
25	21300369664_{ef}	11204500480_{st}^{2}	10399676_{pt1}^{13}	1299844_{st}^{13}	41600552_{pt}^{13}	10399676_{pt}^{13}	2599688_{st}^{13}	649922_{st}^{13}	162986_{st}^{3}
26	2532157069312_{st}^{14}	633038608384_{st}^{14}	326893568_{st}^{14}	618544192_{st}^{14}	38656528_{st}^{14}	83204800_{pt}^{14}	20801200_{sb}^{14}	5199376_{st}^{14}	1299844_{st}^{14}
27	-	158680788992_{st}^{14}	-	20057442_{ef}	-	300224_{st}^{14}	40114884_{pt}^{14}	10029150_{pt}^{14}	2506644_{st}^{14}
28	42061705248768_{st}^{15}	10515426312192_{st}^{15}	5154680832_{st}^{15}	10262347776_{st}^{15}	641396736_{st}^{15}	159987712_{st}^{15}	627776_{st}^{10}	80226336_{pr}^{15}	20056584_{sb}^{15}
29	-	2625500086272_{ef}	-	6691200_{ef}	-	3853632_{ef}	-	-	38777664_{cf}
30	609973884610560_{cf}	152493461268480_{cf}	80766566400_{cf}	149011451520_{cf}	9313176480_{cf}	2332609440_{cf}	9080016_{cf}	9110544_{cf}	708168_{cf}

Table 2

Best lower bounds for $\max_w(A_4^{\mathrm{GC,RC}}(n, d, w))$ after shortening and puncturing operations $12 \leq d \leq 30$. The entry in bold and all entries for $n > 20$ with $d = 12$ are new best known results. Entries in italics are equal to the best known results.

n/d	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
12	1^{7}_{stl}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	3_{cf}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	$\mathit{24}_{ef}$	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	26^{co}_{cf}	-	-	-	$\mathit{2}^{G1}_{nt}$	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	16_{ca}	-	-	-	$\mathit{2}^{A4}_{nt}$	-	-	-	-	-	-	-	-	-	-	-	-	-	-
19	28^{8}_{st}	-	-	-	2^{A6}_{nt}	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20	$\mathit{179}^{9}_{st}$	-	-	5^{1}_{pt}	$\mathbf{8}^{C8}_{nt}$	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	357^{5}_{st}	-	3_{cf}	-	6^{co}_{ef}	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22	2546^{12}_{sb}	4^{8}_{pt}	4^{12}_{pt1}	-	8^{A2}_{nt}	4^{7}_{pt1}	-	-	-	-	-	-	-	-	-	-	-	-	-
23	5312^{11}_{st}	-	-	-	2^{A1}_{nt}	-	-	-	-	-	-	-	-	-	-	-	-	-	-
24	42796^{13}_{sb}	-	-	4^{11}_{pt}	32^{G12}_{nt}	-	4^{13}_{pt}	-	-	-	-	-	-	-	-	-	-	-	-
25	81772^{3}_{st}	-	-	-	8^{C4}_{nt}	-	-	-	-	-	-	-	-	-	-	-	-	-	-
26	603734^{14}_{sb}	4_{cf}	-	-	80^{A8}_{nt}	4^{14}_{pr}	8^{4}_{st}	-	6^{14}_{pt}	-	-	-	-	-	-	-	-	-	-
27	1253105^{14}_{st}	-	3^{co}_{ef}	-	8^{A11}_{nt}	-	3_{cf}	-	-	-	-	-	-	-	-	-	-	-	-
28	10026576^{15}_{sb}	-	1104^{15}_{pt1}	-	320^{A4}_{nt}	-	28^{15}_{pt}	-	8^{15}_{pt}	-	4^{15}_{pt}	-	-	-	-	-	-	-	-
29	19388832_{cf}	-	20_{ef}	-	24^{co}_{ef}	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30	281928_{cf}	-	16384_{cf}	1024_{cf}	1112^{co}_{cf}	-	48_{cf}	-	58_{cf}	-	12_{cf}	-	6_{cf}	-	-	-	-	-	-

Table 3
New generator polynomials, coset leaders L and involutions for codes with at least 4 codewords, $27 < n \leq 30$.

(n,d)	new generator polynomials, coset leaders $\mathbf{L}$ and involutions
(30,3)	invol=(1, 18)(2, 25)(3, 16)(4, 19)(5, 12)(6, 23)(7, 30)(8, 21)(9, 24)(10, 17)(11, 28)(13, 26)(14, 29)(15, 22)(20, 27)
(30,4)	invol=(1, 28)(2, 17)(3, 6)(4, 25)(5, 14)(7, 22)(8, 11)(9, 30)(10, 19)(12, 27)(13, 16)(15, 24)(18, 21)(20, 29)(23, 26)
(30,5)	invol=(1, 16)(2, 17)(3, 18)(4, 19)(5, 20)(6, 21)(7, 22)(8, 23)(9, 24)(10, 25)(11, 26)(12, 27)(13, 28)(14, 29)(15, 30)
(30,6)	invol=(1, 28)(2, 17)(3, 6)(4, 25)(5, 14)(7, 22)(8, 11)(9, 30)(10, 19)(12, 27)(13, 16)(15, 24)(18, 21)(20, 29)(23, 26)
(30,7)	invol=(1, 28)(2, 17)(3, 6)(4, 25)(5, 14)(7, 22)(8, 11)(9, 30)(10, 19)(12, 27)(13, 16)(15, 24)(18, 21)(20, 29)(23, 26)
(30,8)	invol=(1, 10)(2, 29)(3, 18)(4, 7)(5, 26)(6, 15)(8, 23)(9, 12)(11, 20)(13, 28)(14, 17)(16, 25)(19, 22)(21, 30)(24, 27)
(30,9)	invol=(1, 28)(2, 17)(3, 6)(4, 25)(5, 14)(7, 22)(8, 11)(9, 30)(10, 19)(12, 27)(13, 16)(15, 24)(18, 21)(20, 29)(23, 26)
(30,10)	invol=(1, 28)(2, 17)(3, 6)(4, 25)(5, 14)(7, 22)(8, 11)(9, 30)(10, 19)(12, 27)(13, 16)(15, 24)(18, 21)(20, 29)(23, 26)
(30,11)	invol=(1, 28)(2, 17)(3, 6)(4, 25)(5, 14)(7, 22)(8, 11)(9, 30)(10, 19)(12, 27)(13, 16)(15, 24)(18, 21)(20, 29)(23, 26)
(30, 12)	$g(x) = x^{20} + \omega x^{16} + x^{15} + \omega x^{13} + x^{12} + \omega x^{11} + \omega x^{10} + x^8 + \omega^2 x^7 + x^6 + \omega^2 x^3 + \omega x^2 + x + \omega$
	invol=(1, 28)(2, 17)(3, 6)(4, 25)(5, 14)(7, 22)(8, 11)(9, 30)(10, 19)(12, 27)(13, 16)(15, 24)(18, 21)(20, 29)(23, 26)
(30,14)	invol=(1, 22)(2, 26)(3, 30)(4, 19)(5, 23)(6, 12)(7, 16)(8, 20)(9, 24)(10, 28)(11, 17)(13, 25)(14, 29)(15, 18)(21, 27)
(30,15)	invol=(1, 16)(2, 17)(3, 18)(4, 19)(5, 20)(6, 21)(7, 22)(8, 23)(9, 24)(10, 25)(11, 26)(12, 27)(13, 28)(14, 29)(15, 30)
(30,16)	invol=(1, 22)(2, 26)(3, 30)(4, 19)(5, 23)(6, 27)(7, 16)(8, 20)(9, 24)(10, 28)(11, 17)(12, 21)(13, 25)(14, 29)(15, 18)
(30,18)	invol=(1, 22)(2, 26)(3, 30)(4, 19)(5, 23)(6, 27)(7, 16)(8, 20)(9, 24)(10, 28)(11, 17)(12, 21)(13, 25)(14, 29)(15, 18)
(30,20)	invol=(1, 22)(2, 21)(3, 20)(4, 19)(5, 18)(6, 17)(7, 16)(8, 30)(9, 29)(10, 28)(11, 27)(12, 26)(13, 25)(14, 24)(15, 23)
(30,22)	invol=(1, 22)(2, 26)(3, 30)(4, 19)(5, 23)(6, 27)(7, 16)(8, 20)(9, 24)(10, 28)(11, 17)(12, 21)(13, 25)(14, 29)(15, 18)
(30,24)	invol=(1, 10)(2, 29)(3, 18)(4, 7)(5, 11)(6, 30)(8, 23)(9, 27)(12, 24)(13, 28)(14, 17)(15, 21)(16, 25)(19, 22)(20, 26)
(29, 4)	$g(x) = x^6 + x^5 + x^3 + 1$
	invol=(1, 2)(3, 10)(4, 18)(5, 21)(6, 27)(7, 19)(8, 23)(9, 22)(11, 25)(12, 28)(13, 20)(14, 26)(15, 16)(17, 24)
(29, 6)	$g(x) = x^{15} + x^{14} + x^{12} + x^7 + x^6 + x^4 + x^3 + x^2 + 1$
	invol=(1, 22)(2, 9)(3, 20)(4, 11)(5, 14)(6, 17)(7, 12)(8, 15)(10, 13)(16, 23)(18, 25)(19, 28)(21, 26)(24, 27)
(29, 8)	$g(x) = x^{16} + x^{15} + x^{14} + x^{13} + x^{12} + x^{11} + x^9 + x^6 + x^5 + x^4 + x + 1$
	invol=(1, 2)(3, 10)(4, 18)(5, 21)(6, 27)(7, 19)(8, 23)(9, 22)(11, 25)(12, 28)(13, 20)(14, 26)(15, 16)(17, 24)
(29, 11)	$g(x) = x^{14} + \omega^2 x^{13} + \omega^2 x^{11} + \omega x^{10} + x^9 + \omega x^8 + \omega^2 x^7 + \omega x^6 + x^5 + \omega x^4 + \omega^2 x^3 + \omega^2 x + 1$
	invol=(1, 2)(3, 29)(4, 28)(5, 27)(6, 26)(7, 25)(8, 24)(9, 23)(10, 22)(11, 21)(12, 20)(13, 19)(14, 18)(15, 17)
(29,12)	invol=(1, 2)(3, 29)(4, 28)(5, 27)(6, 26)(7, 25)(8, 24)(9, 23)(10, 22)(11, 21)(12, 20)(13, 19)(14, 18)(15, 17)
(29, 14)	$g(x) = x^{25} + x^{24} + x^{21} + x^{20} + x^{17} + x^{16} + x^{13} + x^{12} + x^9 + x^8 + x^5 + x^4 + x + 1$
	invol=(1, 7)(2, 8)(3, 5)(4, 6)(9, 11)(10, 12)(13, 15)(14, 16)(17, 19)(18, 20)(21, 23)(22, 24)(25, 27)(26, 28)
(29, 16)	$g(x) = x^{25} + x^{24} + x^{23} + x^{21} + x^{18} + x^{17} + x^{16} + x^{14} + x^{11} + x^{10} + x^9 + x^7 + x^4 + x^3 + x^2 + 1$
	$\mathbf{L} = (\omega, 0, \omega^2, \omega, \omega^2, 1, \omega, 1, \omega^2, \omega^2, 1, 0, 0, 1, \omega^2, 1, 0, 0, 1, \omega^2, \omega^2, 1, \omega, 1, \omega^2, \omega, \omega^2, 0, \omega)$
	invol=(1, 15)(2, 23)(3, 6)(4, 11)(5, 7)(8, 22)(9, 16)(10, 13)(12, 14)(17, 20)(18, 25)(19, 21)(24, 27)(26, 28)
(27, 6)	$g(x) = x^{12} + \omega^2 x^{10} + \omega x^6 + \omega^2 x^2 + 1$
	invol=(1, 26)(2, 9)(3, 20)(4, 17)(5, 14)(6, 25)(7, 8)(10, 15)(11, 22)(12, 23)(13, 16)(18, 21)(19, 24)

$n \leq 10$ and in some cases with $11 \leq n \leq 20$, when the permutation group is small. For larger permutation groups just a small number of involutions were selected. In other cases with $11 \leq n \leq 20$ the set of polynomials $\{p(x), q(x), r(x)\}$ that gave the best result for $\max_w(A_4^{GC}(n, d, w))$ in [13] was used. Then a search of all possible involutions was carried out when the permutation group of the code was small enough; otherwise the Magma command to generate representatives of conjugacy classes can be used, and either a suitable involution can be found, or its nonexistence shown.

6. Cosets of additive cyclic and additive extended cyclic codes

The method here follows that in Section 4, this time with 20 randomly selected reversible coset leaders. Otherwise the searches are as in Section 5.

7. Shortening, puncturing and nonlinear shortening

Given an $[n, k]$ linear code $\mathcal{C}$ and given i with $1 \leq i \leq n$, the puncturing operation is to delete the i'th component from each codeword of $\mathcal{C}$. The shortening operation is to select all codewords with a 0 in the i'th component, and delete the i'th component from all of these codewords. In each case a linear code of length $n - 1$ is obtained, a GC-weight enumerator can be computed and a code of constant GC-weight can then be selected. However, this must be done in a way which ensures that a reverse-complement constraint is satisfied. If *any* position is used for shortening or puncturing then there may be no fixed point free involution (n even) or one-point fixed involution (n odd) in the permutation group of the code obtained. Long chains of shortened or punctured codes are not obtained and the benefits of a particularly good code are not inherited by shorter codes. A better approach is to proceed as follows. For a code $\mathcal{C}_3$ of even length, shorten or puncture in positions i and $n-i+1$ ($i \in \{1, 2, \ldots, n/2\}$). The pair $(i, n-i+1)$ is lost from the fixed point free involution of $\mathcal{C}_3$, but as the length of the code reduces by two a fixed point free involution remains. For a code $\mathcal{C}_3$ of odd length there are two options. Either shorten

Table 4
New generator polynomials, coset leaders L and involutions for codes with at least 4 codewords, $17 \le n \le 27$.

(n,d)		new generator polynomials, coset leaders **L** and involutions
$(27,14)$	$g(x) =$	$x^{24}+x^{22}+x^{20}+x^{18}+x^{16}+x^{14}+x^{12}+x^{10}+x^{8}+x^{6}+x^{4}+x^{2}+1$
	$\mathbf{L} =$	$(\omega,\omega^2,1,\omega^2,1,1,0,\omega^2,1,\omega,0,\omega,1,0,1,\omega,0,\omega,1,\omega^2,0,1,1,\omega^2,1,\omega^2,\omega)$
	invol $=$	(1, 4)(2, 3)(5, 6)(7, 8)(9, 10)(11, 12)(13, 14)(15, 16)(17, 18)(19, 20)(21, 22)(23, 24)(25, 26)
$(27,18)$	$g(x) =$	$x^{25}+x^{24}+x^{22}+x^{21}+x^{19}+x^{18}+x^{16}+x^{15}+x^{13}+x^{12}+x^{10}+x^{9}+x^{7}+x^{6}+x^{4}+x^{3}$
		$+x+1$
	invol $=$	(1, 2)(3, 6)(4, 5)(7, 8)(9, 12)(10, 11)(13, 14)(15, 18)(16, 17)(19, 20)(21, 24)(22, 23)(25, 26)
$(26,13)$	$g(x) =$	$x^{24}+x^{22}+x^{20}+x^{18}+x^{16}+x^{14}+x^{12}+x^{10}+x^{8}+x^{6}+x^{4}+x^{2}+1$
	invol $=$	(1, 14)(2, 15)(3, 16)(4, 17)(5, 18)(6, 19)(7, 20)(8, 21)(9, 22)(10, 23)(11, 24)(12, 25)(13, 26)
$(25,3)$	$g(x) =$	$x^5+\omega x^4+\omega x+\omega^2$
	invol $=$	(1, 7)(2, 14)(3, 21)(4, 16)(5, 11)(6, 18)(8, 20)(9, 15)(10, 22)(12, 24)(13, 19)(17, 23)
$(21,4)$	$g(x) =$	$x^5+\omega^2x^4+\omega x^3+\omega x^2+\omega^2x+1$
	invol $=$	(1, 6)(2, 5)(3, 4)(7, 20)(8, 19)(9, 18)(10, 17)(11, 16)(12, 15)(13, 14)
$(21,5)$	$g(x) =$	$x^7+\omega^2x^6+x^4+x^3+\omega^2x+1$
	invol $=$	(1, 2)(3, 21)(4, 20)(5, 19)(6, 18)(7, 17)(8, 16)(9, 15)(10, 14)(11, 13)
$(21,6)$	$g(x) =$	$x^9+\omega^2x^8+\omega^2x^7+\omega^2x^6+\omega^2x^3+\omega^2x^2+\omega^2x+1$
	$\mathbf{L} =$	$(\omega,0,0,\omega,0,\omega,0,0,\omega^2,\omega,1,\omega,\omega^2,0,0,\omega,0,\omega,0,0,\omega)$
	invol $=$	(1, 8)(2, 7)(3, 6)(4, 5)(9, 21)(10, 20)(11, 19)(12, 18)(13, 17)(14, 16)
$(21,14)$	$g(x) =$	$x^{19}+x^{18}+x^{16}+x^{15}+x^{13}+x^{12}+x^{10}+x^{9}+x^{7}+x^{6}+x^{4}+x^{3}+x+1$
	invol $=$	(1, 10)(2, 3)(4, 13)(5, 6)(7, 16)(8, 9)(11, 12)(14, 15)(17, 18)(20, 21)
$(21,16)$	$g(x) =$	$x^{18}+\omega^2x^{17}+\omega^2x^{16}+x^{15}+x^{13}+\omega^2x^{12}+\omega^2x^{11}+x^{10}+x^{8}+\omega^2x^{7}+\omega^2x^{6}$
		$+x^5+x^3+\omega^2x^2+\omega^2x+1$
	$\mathbf{L} =$	$(\omega,0,0,\omega^2,1,0,\omega,1,\omega,\omega,0,\omega,\omega,1,\omega,0,1,\omega^2,0,0,\omega)$
	invol $=$	(1, 11)(2, 10)(3, 9)(4, 8)(5, 7)(6, 16)(12, 15)(13, 14)(17, 20)(18, 19)
(20,4)	$\mathbf{L} =$	$(\omega^2,0,\omega^2,0,0,\omega,\omega^2,0,\omega,1,1,\omega,0,\omega^2,\omega,0,0,\omega^2,0,\omega^2)$
	invol $=$	(1, 7)(2, 16)(3, 5)(4, 14)(6, 12)(8, 10)(9, 19)(11, 17)(13, 15)(18, 20)
$(20,6)$	$\omega p(x)+q(x) =$	$x^{12}+x^9+x^7+x^6+x^5+x^4+x^3+x^2+\omega^2x+\omega^2$
	$r(x) =$	$x^{17}+x^{16}+x^{13}+x^{12}+x^9+x^8+x^5+x^4+x+1$
	invol $=$	(1, 11)(2, 12)(3, 13)(4, 14)(5, 15)(6, 16)(7, 17)(8, 18)(9, 19)(10, 20)
$(20,7)$	$\omega p(x)+q(x) =$	$x^{12}+x^{10}+x^7+x^4+x^3+x^2+\omega^2x+\omega^2$
	$r(x) =$	$x^{18}+x^{16}+x^{14}+x^{12}+x^{10}+x^8+x^6+x^4+x^2+1$
	invol $=$	(1, 11)(2, 12)(3, 13)(4, 14)(5, 15)(6, 16)(7, 17)(8, 18)(9, 19)(10, 20)
(20,8)	invol $=$	(1, 18)(2, 14)(3, 20)(4, 11)(5, 7)(6, 12)(8, 16)(9, 17)(10, 15)(13, 19)
(20,9)	invol $=$	(1, 11)(2, 12)(3, 13)(4, 14)(5, 15)(6, 16)(7, 17)(8, 18)(9, 19)(10, 20)
(19,6)	invol $=$	(1, 10)(2, 11)(3, 12)(4, 13)(5, 14)(6, 15)(7, 16)(8, 17)(9, 18)
$(19,7)$	$\omega p(x)+q(x) =$	$x^{13}+x^{11}+x^{10}+x^6+x^5+x^4+x^3+x^2+\omega^2x+\omega^2$
	$r(x) =$	$x^{16}+x^{14}+x^{12}+x^{10}+x^8+x^6+x^4+x^2+1$
	invol $=$	(1, 10)(2, 11)(3, 12)(4, 13)(5, 14)(6, 15)(7, 16)(8, 17)(9, 18)
$(19,8)$	$\omega p(x)+q(x) =$	$x^{17}+x^{13}+x^{12}+x^{10}+x^8+x^6+x^5+x^4+x^3+\omega^2x^2+\omega^2x+\omega^2$
	$r(x) =$	$x^{18}+1$
	invol $=$	(1, 10)(2, 11)(3, 12)(4, 13)(5, 14)(6, 15)(7, 16)(8, 17)(9, 18)
$(19,9)$	$\omega p(x)+q(x) =$	$x^{17}+x^{16}+x^{11}+x^7+\omega x^4+\omega^2x^2+x+\omega^2$
	$r(x) =$	$x^{18}+1$
	$\mathbf{L} =$	$(0,1,\omega^2,1,0,1,\omega,0,1,1,1,0,\omega,1,0,1,\omega^2,1,0)$
	invol $=$	(1, 10)(2, 11)(3, 12)(4, 13)(5, 14)(6, 15)(7, 16)(8, 17)(9, 18)
(18,5)	invol $=$	(1, 10)(2, 11)(3, 12)(4, 13)(5, 14)(6, 15)(7, 16)(8, 17)(9, 18)
(18,6)	invol $=$	(1, 10)(2, 11)(3, 12)(4, 13)(5, 14)(6, 15)(7, 16)(8, 17)(9, 18)
(18,9)	invol $=$	(1, 10)(2, 11)(3, 12)(4, 13)(5, 14)(6, 15)(7, 16)(8, 17)(9, 18)
(18,10)	invol $=$	(1, 10)(2, 11)(3, 12)(4, 13)(5, 14)(6, 15)(7, 16)(8, 17)(9, 18)
$(18,12)$	$\omega p(x)+q(x) =$	$x^{15}+\omega x^{14}+\omega^2x^{13}+\omega^2x^{12}+x^9+\omega x^8+\omega^2x^7+\omega^2x^6+x^3+\omega x^2+\omega^2x+\omega^2$
	$r(x) =$	$x^{16}+x^{15}+x^{13}+x^{12}+x^{10}+x^9+x^7+x^6+x^4+x^3+x+1$
	invol $=$	(1, 4)(2, 5)(3, 6)(7, 10)(8, 11)(9, 12)(13, 16)(14, 17)(15, 18)
$(17,4)$	$g(x) =$	$x^4+x^3+\omega x^2+x+1$
	invol $=$	(1, 2)(3, 17)(4, 16)(5, 15)(6, 14)(7, 13)(8, 12)(9, 11)
$(17,7)$	$g(x) =$	$x^8+\omega^2x^7+x^6+x^5+\omega x^4+x^3+x^2+\omega^2x+1$
	invol $=$	(1, 2)(3, 17)(4, 16)(5, 15)(6, 14)(7, 13)(8, 12)(9, 11)
$(17,8)$	$g(x) =$	$x^9+\omega x^8+\omega x^7+\omega^2x^5+\omega^2x^4+\omega x^2+\omega x+1$
	invol $=$	(1, 2)(3, 17)(4, 16)(5, 15)(6, 14)(7, 13)(8, 12)(9, 11)

or puncture once in position $\lfloor (n+1)/2 \rfloor$ or shorten or puncture twice in positions i and $n-i+1$ ($i \in \{1, 2, \ldots, \lfloor (n-1)/2 \rfloor\}$). In both cases a suitable involution remains.

As was noted in [13], a nonlinear shortening operation sometimes gives more codewords. Given a code $\mathcal{C}$ of constant GC-content over {A, C, G, T}, compute the frequency of each letter A, C, G, T in each column i of the matrix of codewords. Choose the letter and column of the most frequent occurrence and select all codewords with the chosen letter in the chosen column. Delete this component from all selected codewords and a (normally nonlinear) code of constant GC-content is obtained. As the operation in nonlinear, it is only feasible for smaller codes. Again, as for linear shortening, in order to

Table 5
New generator polynomials, coset leaders L and involutions for codes with at least 4 codewords, $13 \le n \le 17$.

(n,d)	new generator polynomials, coset leaders $\mathbf{L}$ and involutions		
(17,12)	$\mathbf{L}$	=	$(0,\omega^2,\omega^2,0,1,1,0,\omega^2,\omega,\omega^2,0,1,1,0,\omega^2,\omega^2,0)$
	invol	=	(1, 9)(2, 8)(3, 7)(4, 6)(10, 17)(11, 16)(12, 15)(13, 14)
(16,3)	invol	=	(1, 3)(2, 8)(4, 12)(5, 15)(6, 9)(7, 11)(10, 16)(13, 14)
(16,4)	invol	=	(1, 3)(2, 11)(4, 14)(5, 6)(7, 15)(8, 12)(9, 16)(10, 13)
(16,6)	$\omega p(x)+q(x)$	=	$x^6+\omega^2x^4+x^3+x^2+\omega^2x+\omega^2$
	$r(x)$	=	$x^8+x^7+x^6+x^4+1$
	$\mathbf{L}$	=	$(1,0,\omega^2,1,1,\omega^2,1,0,0,1,\omega^2,1,1,\omega^2,0,1)$
	invol	=	(1, 9)(2, 6)(3, 16)(4, 7)(5, 11)(8, 13)(10, 12)(14, 15)
(16,7)	$\mathbf{L}$	=	$(0,0,1,0,\omega^2,\omega^2,\omega^2,\omega,\omega,\omega^2,\omega^2,\omega^2,0,1,0,0)$
	invol	=	(1, 3)(2, 11)(4, 14)(5, 6)(7, 15)(8, 12)(9, 16)(10, 13)
(16,8)	$g(x)$	=	$x^8+\omega^2x^7+\omega x^6+\omega x^5+\omega^2x^4+x^3+x^2+\omega x+1$
	$\mathbf{L}$	=	$(0,\omega,\omega,\omega,1,0,\omega,0,0,\omega,0,1,\omega,\omega,\omega,0)$
	invol	=	(1, 3)(2, 11)(4, 14)(5, 6)(7, 15)(8, 12)(9, 16)(10, 13)
(16,10)	$\omega p(x)+q(x)$	=	$x^{13}+\omega x^{11}+\omega^2x^{10}+\omega^2x^9+\omega^2x^8+x^5+\omega x^3+\omega^2x^2+\omega^2x+\omega^2$
	$r(x)$	=	$x^{15}+x^{14}+x^{13}+x^{12}+x^{11}+x^{10}+x^9+x^8+x^7+x^6+x^5+x^4+x^3+x^2+x+1$
	$\mathbf{L}$	=	$(1,\omega^2,\omega,\omega,\omega^2,1,\omega,\omega,\omega,\omega,1,\omega^2,\omega,\omega,\omega^2,1)$
	invol	=	(1, 13)(2, 14)(3, 15)(4, 16)(5, 9)(6, 10)(7, 11)(8, 12)
(16,11)	$\mathbf{L}$	=	$(\omega^2,0,1,0,\omega,1,\omega,\omega,\omega,\omega,1,\omega,0,1,0,\omega^2)$
	invol	=	(1, 3)(2, 11)(4, 14)(5, 6)(7, 15)(8, 12)(9, 16)(10, 13)
(16,12)	$g(x)$	=	$x^{12}+\omega x^{10}+\omega^2x^9+\omega^2x^8+\omega^2x^6+\omega x^5+x^4+x^3+\omega^2x^2+x+\omega$
	invol	=	(1, 3)(2, 8)(4, 12)(5, 15)(6, 9)(7, 11)(10, 16)(13, 14)
(15,3)	$\omega p(x)+q(x)$	=	$x^5+x^4+x^3+x^2+\omega^2x+\omega^2$
	$r(x)$	=	x^6+x^2+1
	invol	=	(1, 4)(2, 13)(3, 6)(5, 8)(7, 10)(9, 12)(11, 14)
(15,4)	$\omega p(x)+q(x)$	=	$x^6+x^4+x^3+\omega^2x+\omega^2$
	$r(x)$	=	$x^7+x^6+x^3+x^2+x+1$
	invol	=	(1, 6)(2, 3)(4, 11)(5, 12)(7, 14)(8, 13)(9, 10)
(15,5)	$g(x)$	=	$x^6+x^5+\omega x^4+x^3+\omega x^2+x+1$
	invol	=	(1, 2)(3, 15)(4, 14)(5, 13)(6, 12)(7, 11)(8, 10)
(15,6)	$\omega p(x)+q(x)$	=	$x^9+x^8+x^7+x^5+x^4+\omega x^3+\omega^2x^2+x+\omega^2$
	$r(x)$	=	$x^{10}+x^8+x^7+x^3+x+1$
	$\mathbf{L}$	=	$(1,1,1,0,\omega^2,1,\omega,\omega,\omega,1,\omega^2,0,1,1,1)$
	invol	=	(1, 8)(2, 9)(3, 10)(4, 11)(5, 12)(6, 13)(7, 14)
(15,10)	$g(x)$	=	$x^{11}+\omega x^{10}+\omega^2x^8+\omega^2x^7+x^6+x^5+\omega^2x^4+\omega^2x^3+\omega x+1$
	$\mathbf{L}$	=	$(\omega,\omega,0,\omega,\omega,\omega^2,\omega,\omega,\omega,\omega^2,\omega,\omega,0,\omega,\omega)$
	invol	=	(1, 2)(3, 15)(4, 14)(5, 13)(6, 12)(7, 11)(8, 10)
(14,4)	$\omega p(x)+q(x)$	=	$x^6+x^4+x^3+\omega^2x+\omega^2$
	$r(x)$	=	$x^7+x^6+x^3+x^2+x+1$
	invol	=	(1, 8)(2, 5)(3, 10)(4, 13)(6, 11)(7, 14)(9, 12)
(14,5)	$\omega p(x)+q(x)$	=	$x^8+x^7+x^6+x^5+x^4+\omega^2x^3+\omega^2x^2+x+\omega^2$
	$r(x)$	=	$x^9+x^8+x^6+x^5+x^4+x^3+1$
	invol	=	(1, 8)(2, 9)(3, 10)(4, 11)(5, 12)(6, 13)(7, 14)
(14,7)	$\omega p(x)+q(x)$	=	$x^{10}+\omega^2x^5+\omega^2x^4+\omega x^3+x^2+x+\omega^2$
	$r(x)$	=	$x^{11}+x^{10}+x^9+x^7+x^4+x^3+x^2+1$
	$\mathbf{L}$	=	$(1,\omega^2,\omega^2,\omega^2,\omega^2,\omega,\omega,\omega,\omega,\omega^2,\omega^2,\omega^2,\omega^2,1)$
	invol	=	(1, 8)(2, 9)(3, 10)(4, 11)(5, 12)(6, 13)(7, 14)
(14,10)	invol	=	(1, 8)(2, 9)(3, 10)(4, 11)(5, 12)(6, 13)(7, 14)
(13,3)	$\omega p(x)+q(x)$	=	$x^3+\omega^2x+\omega^2$
	$r(x)$	=	$x^5+x^3+x^2+1$
	invol	=	(1, 7)(2, 8)(3, 9)(4, 10)(5, 11)(6, 12)
(13,5)	$\omega p(x)+q(x)$	=	$x^9+x^7+x^6+x^5+x^4+x^3+x^2+\omega^2x+\omega^2$
	$r(x)$	=	$x^{10}+x^8+x^6+x^4+x^2+1$
	invol	=	(1, 7)(2, 8)(3, 9)(4, 10)(5, 11)(6, 12)
(13,6)	$\omega p(x)+q(x)$	=	$x^8+x^6+\omega^2x^2+\omega^2x+\omega^2$
	$r(x)$	=	$x^{10}+x^9+x^7+x^6+x^4+x^3+x+1$
	$\mathbf{L}$	=	$(\omega,0,\omega^2,1,1,0,1,0,1,1,\omega^2,0,\omega)$
	invol	=	(1, 7)(2, 8)(3, 9)(4, 10)(5, 11)(6, 12)

preserve the reverse-complement constraint it is necessary to select position $n/2$ (n odd) or to apply the operation twice in the positions i and $n-i+1$ ($i \in \{1, 2, \ldots, \lfloor n/2\rfloor\}$). If the operation is applied twice, the same letter must be selected in both positions.

Normally shortening gives an unchanged Hamming distance, and puncturing reduces the minimum distance by 1. Sometimes, however, the minimum distance is greater than is anticipated. Thus it is necessary to assess shortenings and puncturings of codes with two consecutive values of d (one position) or three consecutive values of d (two positions).

Table 6
New generator polynomials, coset leaders L and involutions for codes with at least 4 codewords, $4 \le n \le 13$.

(n,d)			new generator polynomials, coset leaders $\mathbf{L}$ and involutions
$(13,9)$	$\omega p(x)+q(x)$	$=$	$x^{10}+x^9+\omega^2x^8+x^6+x^5+\omega^2x^4+x^2+x+\omega^2$
	$r(x)$	$=$	$x^{12}+1$
	invol	$=$	(1, 11)(2, 12)(3, 9)(4, 10)(5, 7)(6, 8)
$(12,3)$	$\omega p(x)+q(x)$	$=$	$x^3+\omega^2x+\omega^2$
	$r(x)$	$=$	$x^5+x^3+x^2+1$
	invol	$=$	(1, 7)(2, 8)(3, 9)(4, 10)(5, 11)(6, 12)
(12,4)	invol	$=$	(1, 11)(2, 4)(3, 9)(5, 7)(6, 12)(8, 10)
(12,5)	invol	$=$	(1, 7)(2, 8)(3, 9)(4, 10)(5, 11)(6, 12)
$(12,6)$	$\omega p(x)+q(x)$	$=$	$x^8+x^6+x^4+x^3+x^2+\omega^2x+\omega^2$
	$r(x)$	$=$	$x^{11}+x^{10}+x^9+x^8+x^7+x^6+x^5+x^4+x^3+x^2+x+1$
	invol	$=$	(1, 7)(2, 8)(3, 9)(4, 10)(5, 11)(6, 12)
(12,7)	invol	$=$	(1, 11)(2, 4)(3, 9)(5, 7)(6, 12)(8, 10)
$(12,8)$	$\omega p(x)+q(x)$	$=$	$x^{10}+x^8+\omega^2x^7+\omega x^5+\omega x^4+\omega x^3+\omega x^2+x+\omega^2$
	$r(x)$	$=$	$x^{11}+x^{10}+x^9+x^8+x^7+x^6+x^5+x^4+x^3+x^2+x+1$
	invol	$=$	(1, 7)(2, 8)(3, 9)(4, 10)(5, 11)(6, 12)
(12,9)	invol	$=$	(1, 11)(2, 12)(3, 9)(4, 10)(5, 7)(6, 8)
$(11,5)$	$\omega p(x)+q(x)$	$=$	$x^7+x^5+x^4+x^3+x^2+\omega^2x+\omega^2$
	$r(x)$	$=$	$x^9+x^8+x^7+x^6+x^5+x^4+x^3+x^2+x+1$
	invol	$=$	(1, 6)(2, 7)(3, 8)(4, 9)(5, 10)
(11,8)	invol	$=$	(1, 2)(3, 5)(4, 9)(6, 7)(8, 10)
$(10,4)$	$g(x)$	$=$	$x^4+\omega x^3+\omega x+1$
	invol	$=$	(1, 8)(2, 7)(3, 6)(4, 5)(9, 10)
$(10,5)$	$\omega p(x)+q(x)$	$=$	$x^7+x^6+x^5+x^4+\omega^2x^2+\omega^2$
	$r(x)$	$=$	$x^8+x^6+x^4+x^2+1$
	invol	$=$	(1, 6)(2, 7)(3, 8)(4, 9)(5, 10)
$(10,6)$	$\omega p(x)+q(x)$	$=$	$x^6+x^5+x^4+\omega^2x^2+x+\omega^2$
	$r(x)$	$=$	$x^{10}+1$
	invol	$=$	(1, 6)(2, 7)(3, 8)(4, 9)(5, 10)
(10,8)	invol	$=$	(1, 7)(2, 6)(3, 10)(4, 9)(5, 8)
(9,4)	invol	$=$	(1, 6)(2, 5)(3, 4)(7, 9)
$(9,5)$	$\omega p(x)+q(x)$	$=$	$x^5+\omega x^3+\omega^2x^2+\omega^2x+\omega^2$
	$r(x)$	$=$	$x^7+x^6+x^5+x^4+x^3+x^2+x+1$
	$\mathbf{L}$	$=$	$(0,\omega,0,0,\omega,0,0,\omega,0)$
	invol	$=$	(1, 5)(2, 6)(3, 7)(4, 8)
$(9,6)$	$\omega p(x)+q(x)$	$=$	$x^8+x^4+\omega^2x^3+x^2+x+\omega^2$
	$r(x)$	$=$	x^9+1
	invol	$=$	(1, 6)(2, 5)(3, 4)(7, 9)
(8,3)	$\mathbf{L}$	$=$	$(\omega^2,1,1,0,0,1,1,\omega^2)$
	invol	$=$	(1, 5)(2, 6)(3, 7)(4, 8)
$(8,4)$	$g(x)$	$=$	x^3+x+1
	invol	$=$	(1, 4)(2, 3)(5, 8)(6, 7)
(8,6)	invol	$=$	(1, 5)(2, 7)(3, 4)(6, 8)
$(7,3)$	$g(x)$	$=$	$x^2+\omega^2$
	invol	$=$	(1, 6)(2, 3)(4, 5)
$(7,4)$	$\omega p(x)+q(x)$	$=$	$x^3+\omega x^2+\omega^2x+\omega^2$
	$r(x)$	$=$	x^4+x^3+x+1
	invol	$=$	(1, 4)(2, 5)(3, 6)
(6,3)	invol	$=$	(1, 4)(2, 5)(3, 6)
$(6,4)$	$\omega p(x)+q(x)$	$=$	$x^3+\omega x^2+\omega^2x+\omega^2$
	$r(x)$	$=$	x^4+x^3+x+1
	invol	$=$	(1, 4)(2, 5)(3, 6)
$(5,3)$	$g(x)$	$=$	$x^2+\omega x+1$
	invol	$=$	(1, 4)(2, 3)
$(4,3)$	$\omega p(x)+q(x)$	$=$	$x^2+\omega^2x+\omega^2$
	$r(x)$	$=$	x^3+x^2+x+1
	invol	$=$	(1, 3)(2, 4)

8. Results

The results obtained by the methods described in Sections 2–7 are given in Tables 1 and 2. The labels in the tables have the following meaning:

Table 7
Involutions and shortening positions for codes obtained from best known linear codes.

(n,d)	involution or shortening positions
(22, 14)	invol= $(1,10)(2,14)(3,18)(4,9)(5,8)(6,7)(11,15)(12,17)(13,16)(19,22)(20,21)$
(24, 13)	invol= $(1,5)(2,6)(3,7)(4,8)(9,13)(10,14)(11,15)(12,16)(17,21)(18,22)(19,23)(20,24)$
(24, 16)	invol= $(1,5)(2,17)(3,22)(4,7)(6,9)(8,10)(11,16)(12,14)(13,15)(18,19)(20,23)(21,24)$
(25, 16)	invol= $(1,7)(2,6)(3,5)(8,13)(9,12)(10,11)(14,25)(15,24)(16,23)(17,22)(18,21)(19,20)$
(28, 14)	shortening of (30,14) in positions 10 and 21
(28, 15)	shortening of (30,15) in positions 10 and 21
(30, 14)	invol= $(1,6)(2,7)(3,8)(4,9)(5,10)(11,16)(12,17)(13,18)(14,19)(15,20)(21,26)(22,27)(23,28)(24,29)(25,30)$
(30, 15)	invol= $(1,6)(2,7)(3,8)(4,9)(5,10)(11,16)(12,17)(13,18)(14,19)(15,20)(21,26)(22,27)(23,28)(24,29)(25,30)$

Subscripts	
cf	cyclic linear code over GF(4)
ef	extended cyclic linear code over GF(4)
ca	cyclic additive code over GF(4)
ea	extended cyclic additive code over GF(4)
sb	linear shortening of the code below
st	linear shortening twice of the code two positions below
pt	linear puncturing twice of the code two positions below and two positions to the right
*pt*1	linear puncturing twice of the code two positions below and one position to the right
pr	linear puncturing of the code below and one position to the right
nb	nonlinear shortening of the code below
nt	nonlinear shortening in two positions of the code two positions below
*st*1	linear shortening twice of the code two positions below and one position to the left

Superscripts	
co	coset of code
i	position (or first position) for shortening or puncturing
Xi	letter and position (or first position) for nonlinear shortening

Entries that improve results in the online table of Gaborit and King[2] for $n \le 20$ are marked in bold; entries that equal these results are marked in italic. For $n \ge 21$ and $d \le 12$ the entries are all new bests and are not marked. Files of codewords for these codes (when the code is best known and the number of codewords does not exceed 50 000) are maintained at two web sites.[3]

The involutions used, the generator polynomials (where these differ from those used in the same case in [13]) and the coset leaders for cosets of linear codes are given in Tables 3–6.

9. Codes with an all-ones codeword in the dual

Gaborit and King [5] recommend the use of linear codes with an all-ones vector in the dual. If this is the case the code has only even GC-weights. This means that the set of all codewords is spread over a smaller number of possible weights. Thus, a larger number of codewords for a give choice of GC-weight is likely.

Lemma 5 ([5]). *Let* C *be a code over* GF(4). *If the all-ones vector belongs to* $C^{\perp}$, *then the GC-weight enumerator of C is even (has all even weights).*

It is possible to apply this approach to all the codes constructed in the current paper by applying the following lemma to the code $\mathcal{C}'$.

Lemma 6 ([5]). *Let* C *be a linear code over* GF(4) *of length n. Suppose* $C^{\perp}$ *has a vector* $c = (c_1, \dots, c_n)$ *with no 0s. Then* C *is equivalent to a code that has the all-ones vector in its dual.*

The equivalent code is obtained as follows. First select the codewords in $\mathcal{C}'^{\perp}$ that have no 0s. For every vector, each entry that is ω is replaced by 1 and the corresponding column of the generator matrix of the code $\mathcal{C}'$ is multiplied by ω to leave the inner product unchanged. Similarly, each entry that is ω^2 is replaced by 1 and the corresponding column of the generator matrix of the code $\mathcal{C}'$ is multiplied by ω^2. Thus $\mathcal{C}'$ is replaced by an equivalent code that has the all-ones vector in its dual, which therefore has an even GC-weight enumerator.

2 http://llama.med.harvard.edu/~king/dnacodes.html.

3 http://data.research.glam.ac.uk/projects/ ; http://www.idsia.ch/~roberto/DNA10.zip.

Table 8
Best known lower bounds for $\max_w(A_4^{\mathrm{GC,RC}}(n, d, w))$ $3 \le d \le 11$.

n/d	3	4	5	6	7	8	9	10	11
4	6	2_{gk}	-	-	-	-	-	-	-
5	15_{gk}	3	1_{gk}	-	-	-	-	-	-
6	44_{gk}	16	4_{gk}	2_{gk}	-	-	-	-	-
7	135_{gk}	36_{gk}	11_{gk}	2_{gk}	1_{gk}	-	-	-	-
8	528_{gk}	128_{gk}	28_{gk}	12	2_{gk}	2_{gk}	-	-	-
9	1354_{gk}	275_{gk}	67_{gk}	21_{m3}	8_{gk}	2_{gk}	1_{gk}	-	-
10	4542_{gk}	860	210	54_{gk}	17_{m3}	8_{gk}	2_{gk}	2_{gk}	-
11	14405_{gk}	2457_{gk}	477_{gk}	117_{gk}	37_{m1}	14_{m1}	5_{gk}	2	1_{gk}
12	59136	14784	1848	924	87_{m1}	29_{m1}	12_{m3}	4_{gk}	2
13	167263_{gk}	27376_{gk}	3974_{m1}	924_{gk}	206_{m1}	62_{m1}	23_{m3}	10_{m3}	4_{gk}
14	768768	192192	11878_{gk}	3712	796	208	49_{m3}	21	8_{m1}
15	1646240_{gk}	411821_{gk}	25670_{gk}	6648	1600_{gk}	410	109_{m3}	37_{m1}	18_{gk}
16	13174400	3293600	55376_{gk}	55424	13856	3776	243_{m3}	83_{m3}	68
17	26355520_{gk}	6587200_{gk}	97520	97450_{gk}	12864_{gk}	6060	579_{m1}	175_{m3}	68_{A}
18	44933184	11266112	699624	699624	43632	10908	2691	407_{m1}	133_{m3}
19	47102080_{gk}	23647760_{gk}	1477796	738772	92252	11542	3678_{m3}	960_{m1}	285_{m1}
20	756760576_{gk}	189432064	11822368	11806240_{gk}	369008	184756	11452_{gk}	2868_{gk}	766_{gk}
21	756760576_{A}	189432064_{A}	22573824	11806240_{A}	369008_{A}	184756_{A}	11452_{A}	2926	847
22	10602158336	2650495232	22607872	11806240_{A}	1410864	353496	88424	22088	5522
23	10602158336_{A}	2650495232_{A}	43264648	11806240_{A}	2703694	676312	169182	42968	10701
24	177279886336	44319794176	346436544	43355616	21631400	5406464	1351616	338016	84964
25	177279886336_{A}	44319794176_{A}	346436544_{A}	43355616_{A}	41600552	10399676	2599688	649922	162986
26	2532157069312	633038608384	346436544_{A}	618544192	41600552_{A}	83204800	20801200	5199376	1299844
27	2532157069312_{A}	633038608384_{A}	346436544_{A}	618544192_{A}	41600552_{A}	83204800_{A}	40114884	10029150	2506644
28	42061705248768	10515426312192	5154680832	10262347776	641396736	159987712	40114884_{A}	80226336	20056584
29	42061705248768_{A}	10515426312192_{A}	5154680832_{A}	10262347776_{A}	641396736_{A}	159987712_{A}	40114884_{A}	80226336_{A}	38777664
30	609973884610560	152493461268480	80766566400	149011451520	9313176480	2332609440	40114884_{A}	80226336_{A}	38777664_{A}

Table 9
Best known lower bounds for $\max_w(A_4^{\mathrm{GC,RC}}(n, d, w))$ $12 \leq d \leq 30$.

n/d	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
12	2_{gk}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	2_{gk}	1_{gk}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	4_{gk}	2_{gk}	2_{gk}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	6_{gk}	3_{gk}	2_{gk}	1_{gk}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	24	5_{gk}	2_{gk}	2_{gk}	2_{gk}	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	30_{gk}	12_{m3}	4_{gk}	2_{gk}	2	1_{gk}	-	-	-	-	-	-	-	-	-	-	-	-	-
18	49_{m3}	21_{m3}	10_{m3}	4_{gk}	2	2_{gk}	2_{gk}	–	-	-	-	-	-	-	-	-	-	-	-
19	99_{m3}	39_{m3}	18_{m3}	8_{m3}	4_{gk}	2_{gk}	2_{gk}	1_{gk}	-	-	-	-	-	-	-	-	-	-	-
20	179	77_{m3}	33_{m3}	15_{m3}	8	4_{gk}	2_{gk}	2_{gk}	2_{gk}	-	-	-	-	-	-	-	-	-	-
21	357	88_{m3}	43_{m3}	22_{m3}	11_{m3}	6_{m3}	3_{m3}	2_{m3}	2_{m3}	1_{m3}	-	-	-	-	-	-	-	-	-
22	2546	174_{m3}	192_{L1}	36_{m3}	20_{m3}	10_{m3}	6_{m3}	2_{m3}	2_{m3}	2_{m3}	2_{m3}	-	-	-	-	-	-	-	-
23	5312	336_{m3}	126_{m3}	57_{m3}	31_{m3}	16_{m3}	8_{m3}	4_{m3}	2_{m3}	2_{m3}	2_{m3}	1_{m3}	-	-	-	-	-	-	-
24	42796	2744_{L1}	244_{m3}	102_{m3}	320_{L1}	27_{m3}	14_{m3}	7_{m3}	4_{m3}	2_{m3}	2_{m3}	2_{m3}	2_{m3}	-	-	-	-	-	-
25	81772	1402_{m3}	480_{m3}	190_{m3}	320_{A}	65_{m3}	23_{m3}	12_{m3}	6_{m3}	4_{m3}	2_{m3}	2_{m3}	2_{m3}	1_{m3}	-	-	-	-	-
26	603734	2974_{m3}	977_{m3}	351_{m3}	320_{A}	67_{m3}	38_{m3}	20_{m3}	11_{m3}	6_{m3}	4_{m3}	2_{m3}	2_{m3}	2_{m3}	2_{m3}	-	-	-	-
27	1253105	6308_{m3}	1927_{m3}	655_{m3}	320_{A}	114_{m3}	56_{m3}	32_{m3}	18_{m3}	9_{m3}	5_{m3}	4_{m3}	2_{m3}	2_{m3}	2_{m3}	1_{m3}	-	-	-
28	10026576	13688_{m3}	9790_{L2}	2420_{L2}	459_{m3}	194_{m3}	93_{m3}	50_{m3}	28_{m3}	15_{m3}	8_{m3}	4_{m3}	4_{m3}	2_{m3}	2_{m3}	2_{m3}	2_{m3}	-	-
29	19388832	29292_{m3}	9790_{A}	2599_{m3}	898_{m3}	353_{m3}	155_{m3}	77_{m3}	42_{m3}	24_{m3}	12_{m3}	7_{m3}	4_{m3}	3_{m3}	2_{m3}	2_{m3}	2_{m3}	1_{m3}	-
30	19388832_{A}	61270_{m3}	141810_{L1}	35480_{L1}	1767_{m3}	546_{m3}	266_{m3}	127_{m3}	65_{m3}	36_{m3}	20_{m3}	11_{m3}	7_{m3}	4_{m3}	3_{m3}	2_{m3}	2_{m3}	2_{m3}	2_{m3}

The approach is capable of generating some improvements to the results in [13] when finding lower bounds for $\max_w(A_4^{\mathrm{GC}}(n,d,w))$ [1]. However, when the reverse-complement constraint was added, as in the current paper, it was found that either $\mathcal{C}'^{\perp}$ had no words with no 0s, or the permutation group of the equivalent form of $\mathcal{C}'$ had no fixed point free involution (n even) or no one-point fixed involution (n odd). Thus no improvements to $\max_w(A_4^{\mathrm{GC,RC}}(n,d,w))$ were obtained beyond those available in Tables 1 and 2.

10. The Magma database of best linear codes

Magma contains a database of best linear codes, based on results in a web site previously maintained by Brouwer and now superseded by a web site maintained by Grassl [6]. Specifically, for a given length n and minimum Hamming distance d, Magma is able to return a linear code over GF(4) with the largest dimension known to be possible. Although not guaranteed to give a good result, this is clearly a good candidate for the code $\mathcal{C}'$ provided a suitable involution exists. If an involution does exist the techniques presented in this paper can be used directly.

10.1. Results

Six new best results for $A_4^{\mathrm{GC,RC}}(n,d,w)$ are given in Table 9 and marked with a subscript $L1$. Two further new best results (marked $L2$) are given by shortening the codes for $n=30$, $d=14$ and $n=30$, $d=15$ in positions 10 and 21. The involutions used and shortening positions are given in Table 7.

11. The best known codes

The best known lower bounds for $\max_w(A_4^{\mathrm{GC,RC}}(n,d,w))$ are reported in Tables 8 and 9. Unmarked entries are those given in Tables 1 and 2. In the case of entries in italics in Tables 1 and 2, the values have been published previously (often without a specific construction) in [5] or elsewhere. Entries marked with a subscript gk are taken from [5] or the authors' updated online table[4] where details of the individual constructions used can be found. Entries with subscript $m1$ are from [9] and entries with subscript $m3$ are from [11]. Entries with subscripts $L1$ and $L2$ are taken from Section 10. Files of codewords for the codes marked $m1$ and $m3$ can be found at the authors' web sites.[5]

Note that in some cases in Tables 1 and 2 a code of length n has more codewords than the code of length $n+1$ with the same d, or no code of length $n+1$ exists for this d. In such cases a code of length $n+1$ can be constructed by a process of central extension. A new position with coordinate $\lfloor (n+3)/2 \rfloor$ is inserted, with the same value for every codeword. If n is even then a code of length $n+1$ with the same number of codewords as the code of length n is obtained, with the GC-content, Hamming distance and reverse-complement constraints still satisfied. The same is true if n is odd and the original code of length n was constructed using Lemmas 1 and 3. Entries in Tables 8 and 9 obtained from the entry above in this way are shown with a subscript A.

12. Conclusion

Detailed attention to the options for constructing DNA codes from cyclic and extended cyclic codes, in combination with the use of involutions pioneered by Gaborit and King, has produced many new best codes. These results complement the results for $\max_w(A_4^{\mathrm{GC}}(n,d,w))$ (without a reverse-complement constraint) given in [13]. All of the codes in Tables 1 and 2 should be reproducible from the information given here. The entries marked $L1$ and $L2$ in Tables 8 and 9. should be reproducible from the information given here unless the entry in the Magma database of best linear codes changes.

Note that entries in the tables are given for the actual values of n, d. Thus a better code can sometimes be obtained corresponding to a value given in the table when d is increased.

References

[1] N. Aboluion, The construction of DNA codes using a computer algebra system, Ph.D. Thesis, University of Glamorgan, Wales, UK, 2011.
[2] S. Brenner, R.A. Lerner, Encoded combinatorial chemistry, Proc. Natl. Acad. Sci. USA 89 (1992) 5381–5383.
[3] A.R. Calderbank, E.M. Rains, P.W. Shor, N.J.A. Sloane, Quantum error correction via codes over GF(4), IEEE Trans. Inform. Theory 4 (1998) 1369–1387.
[4] A.G. Frutos, Q. Liu, A.J. Thiel, A.M.W. Sanner, A.E. Condon, L.M. Smith, R.M. Corn, Demonstration of a word design strategy for DNA computing on surfaces, Nucleic Acids Res. 25 (1997) 4748–4757.
[5] P. Gaborit, O.D. King, Linear constructions for DNA codes, Theoret. Comput. Sci. 334 (2005) 99–113.
[6] M. Grassl, Code tables: bounds on the parameters of various types of codes. http://www.codetables.de/ (accessed 06.09.10).
[7] F.J. MacWilliams, N.J.A. Sloane, The Theory of Error-Correcting Codes, North-Holland, Amsterdam, The Netherlands, 1977.
[8] A. Marathe, A.E. Condon, R.M. Corn, On combinatorial DNA word design, J. Comput. Biol. 8 (3) (2001) 201–219.

4 http://llama.med.harvard.edu/king/dnacodes.html.
5 http://data.research.glam.ac.uk/projects/ ; http://www.idsia.ch/~roberto/DNA08.zip ; http://www.idsia.ch/~roberto/DNA09.zip.

[9] R. Montemanni, D.H. Smith, Construction of constant GC-content DNA codes via a variable neighbourhood search algorithm, J. Math. Model. Algorithms 7 (2008) 311–326.
[10] R. Montemanni, D.H. Smith, Metaheuristics for the construction of constant GC-content DNA codes, MIC 2009: The VIII Metaheuristics International Conference, Hamburg, July 13–16 2009.
[11] R. Montemanni, D.H. Smith, N. Koul, Three metaheuristics for the construction of constant GC-content DNA codes, in: S. Voss, M. Caserta (Eds.), Metaheuristics: Intelligent Decision Making (Operations Research / Computer Science Interface Series), Springer, 2011.
[12] D. Shoemaker, D.A. Lashkari, D. Morris, M. Mittmann, R.W. Davis, Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy, Nat. Genet. 16 (1996) 450–456.
[13] D.H. Smith, N. Aboluion, R. Montemanni, S. Perkins, Linear and nonlinear constructions of DNA codes with Hamming distance *d* and constant GC-content, Discrete Math. 311 (2011) 1207–1219.