SEQMAK Users' Manual

Version 0.12

Junghoon Kim
Department of Physics,
Sungkyunkwan University
jfkimberly@skku.edu

September 11, 2013

Contents

1 Prerequisites **1**

2 Introduction **3**

3 Running the Program **7**

GNU Free Documentation License **19**

1. APPLICABILITY AND DEFINITIONS 19
2. VERBATIM COPYING 21
3. COPYING IN QUANTITY 21
4. MODIFICATIONS 22
5. COMBINING DOCUMENTS 24
6. COLLECTIONS OF DOCUMENTS 24
7. AGGREGATION WITH INDEPENDENT WORKS 24
8. TRANSLATION 25
9. TERMINATION 25
10. FUTURE REVISIONS OF THIS LICENSE 26
11. RELICENSING 26

ADDENDUM: How to use this License for your documents 27

Bibliography **28**

Chapter 1

Prerequisites

SEQMAK is written in Python 2.7. In order to run the program you must have Python installed on your computer (preferably version 2.7x but version 2.4x should also work). This software has not been tested with Python 3.0 or higher, so I cannot guarantee it will work under those environments. If you do not have Python installed on your computer, you may download the latest version at http://www.python.org/download/. Python is GNU/Linux, Windows, and OS X compatible.

Chapter 2

Introduction

This is documentation for the program SEQMAK (pronounced SEC-mac), which stands for *sequence maker*. SEQMAK is a near clone of SEQUIN which was written by N. C. Seeman[1, 2]. The main differences are

- much of the repetitive nature of SEQUIN has been automated,
- exception handling has been much improved, and
- all features have been modularized, making the code much easier to read, learn, and maintain.

As is the case with SEQUIN, kinetic and thermodynamic factors are not taken into account when designing structures with SEQMAK . Only different combinations of DNA bases are considered. If you would like to consider other factors when designing your DNA structure, you may want to use a more complex program such as DNAdesign.[1]

The first step in designing any DNA structure is to draw up a blueprint of the structure itself. Once that has been done, the sequence of each strand must be determined. This is where SEQMAK comes in. By dividing the structure into different (topological) domains called arms and subarms, SEQMAK produces segments (called "critons") of random or user determined bases. Concatenating these critons produces the final strands which (with some luck) creates the intended structure. Since bases are produced at the criton level and then concatenated to form strands, strands of any significant length will inevitably have repeating sequence segments. Heuristically, this is thought to be undesirable as too many repeating segments may lead to crosstalk, *i.e.* hybridization between unintended domains of same/different strands. SEQMAK can minimize repeating sequence segments and hence crosstalk by calculating the frequency a certain criton occurs in all the strands and selecting critons which appear with the least frequency. After the sequence of all the critons have been determined, SEQMAK produces a text file with the sequence of the final strands.

[1] http://dna.caltech.edu/DNAdesign/

Below, two existing DNA structures are provided as pedagogical examples, the double-crossover (DX) tile[3] and 3-point star[4] (since the structural design has already been done, the focus henceforth is put on the use of SEQMAK). By design, both structures include Holliday junctions, which serve as fixtures, giving rigidity to the structures. Topologically, these junctions serve as indicators where the structure should be initially dissected into arms (see examples below and Fig. 2.1).

Among other things, DX tiles have been used to create DNA crystals[5], *i.e.* periodic arrays of DNA molecules, and molecular analogs of Wang tiles[6] to perform computations[7]. The DX tiles shown in Fig. 2.1 are of the **d**ouble-crossover **a**ntiparallel **e**ven (DAE) type. "Double-crossover" refers to the two junctions present in each DX tile, "antiparallel" is the relative orientation of the helical domains, and "even" refers to the number of half-turns between the two crossover points of a tile (sometimes you'll see DX tiles such as DAE-O where the extra "-O" stands for an odd number of half-turns in between crossover points of adjacent tiles). Once the structural design has been done,

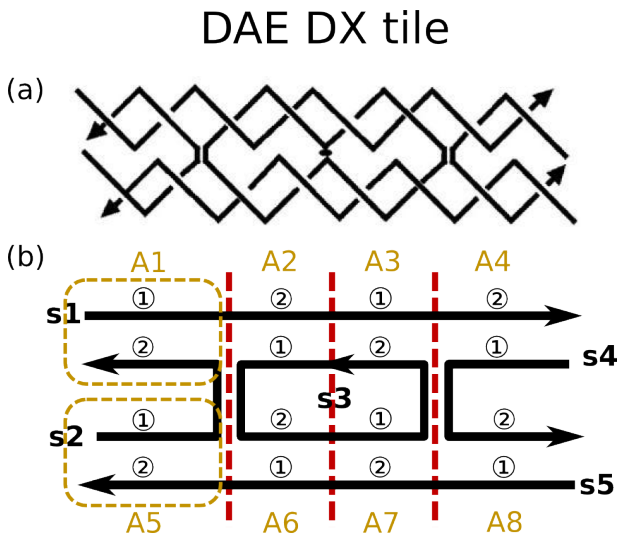

Figure 2.1: A double-crossover antiparallel even (DAE) type DX tile. (a) A schematic diagram of the DAE DX tile (taken from Ref. 3). (b) A DAE DX tile divided into strands (*e.g.* s1), arms (*e.g.* A1), and subarms (*e.g.* ①). The dashed red lines indicate the division of the structure into arm domains and should be made at every junction and in-between every junction. Each arm consists of two subarms (① and ②) encompassed by the dashed gold squares.

the following steps need to be carried out by hand before using SEQMAK :

1. index the strands
2. divide and index the strands into arms
3. divide and index the arms into subarms

First, the strands must be indexed. For the DAE DX tile shown in Fig. 2.1(b), they are s1, s2, . . . , s5. The order of the strands do not matter, *e.g.* s1 may be interchanged with s5. Next, divide the strands into arm domains. This is done by identifying the junctions and partitioning the structure along the junctions. The left and right dashed red lines of Fig. 2.1(b) illustrate these divisions. In addition, for any structure with more than two or more junctions, there needs to be a bisecting line in-between every junctions (which is the dashed red line down the center of the structure). This partitioning scheme separates the DAE DX tile into 4 arm domains, with each domain having 2 arms (*e.g.* the left-most arm domain consists of arms A1 and A5) for a total of eight arms, labeled A1, A2, . . . , A8. Furthermore, as can be seen from the figure, each arm consists of segments from two different strands. For instance, the A1 arm consists of segments from the s1 and s2 strands (likewise the A7 arm consists of segments from the s3 and s5 strands). These segments make up the subarms of the structure. The numbering scheme for the subarms is such that if the directionality ($5' \rightarrow 3'$) of the subarm is toward the junction, then it is labeled with ① and if it points away from the junction, then it is labeled with a ②. Now, we are in a position to run `SEQMAK` .

Chapter 3

Running the Program

As of version 0.12, SEQMAK consists of 3 files, `seqmak.py`, `commands.py`, and `functions.py`. Place all three files in the same directory. `seqmak.py` is the main file which is used to run the program. Double-click this file if you want to run it under Windows, or for *nix systems type

```
$ python seqmak.py
```

at the command line prompt (indicated by `$`). The following should appear

```
Enter a command (type 'help' for help):
```

At any point in the program, the user can type '`help`' (without the apostrophes) and a list of all the possible commands will be shown. Typing `help` gives out the following:

```
Possible commands are:
1. newarms (na)
2. show (s)
3. link (l)
4. crunch (c)
5. strandgen (sg)
6. repeatcheck (rp)
7. save (sv)
8. load (ld)
9. exit
Use '--help' for more information. e.g. 'na --help'
Enter a command (type 'help' for help):
```

With the exception of the `exit` command, there are 8 commands which are used to run the program. Each command may be executed by typing either the full name of the command (*e.g.* '`newarms`') or its abbreviated name indicated in parentheses (*e.g.* '`na`'). The first step is to create all the arms of the structure. Typing `newarms` asks for the number of arms the user would like:

```
How many arms do you want?
```

From Fig. 2.1(b) we see that a DAE DX tile has 8 arms, so type `8` and press enter.

```
How many arms do you want?
8
What is the length of the subarms of these arms?
```

Next the length (number of bases) of the subarms of each arm needs to be entered. For a DAE DX tile, every arm consists of 2 subarms each of which are 8 bases long, so type `8` and press enter.

```
How many arms do you want?
8
What is the length of the subarms of these arms?
8
Any more arms? (y/n)
```

The next question that comes up is whether the user would like additional arms. The user can type '`y`' if they would like additional arms or if the structure requires arms of different lengths, *e.g.* 4 arms consisting of subarms of length 8 and 4 arms consisting of subarms of length 4. Since the length of all the subarms of the DAE DX tile are the same, type '`n`'. To check the newly created arms type '`show`'.

```
Enter a command (type 'help' for help):
show

arm1
xxxxx xxx
xxxxx xxx

arm2
xxxxx xxx
xxxxx xxx

arm3
xxxxx xxx
xxxxx xxx

arm4
xxxxx xxx
xxxxx xxx

arm5
xxxxx xxx
xxxxx xxx

arm6
xxxxx xxx
```

```
xxxxx xxx

arm7
xxxxx xxx
xxxxx xxx

arm8
xxxxx xxx
xxxxx xxx
Enter a command (type 'help' for help):
```

The '`show`' command lists all the arms along with the sequences of their subarms below each arm. Since the sequences of the subarms have yet to be determined, all the bases are represented by `x`'s, and each subarm sequence is space separated every 5 bases for easy viewing. Next, we must link the arms to form strands. This can be done by making the following table. Not only is each strand a concatenation of subarms, but the subarm

strands	$3'$-linked arm ① → ②	$5'$-linked arm ② → ①
s1	1,2 / 3,4	2,3
s2	5,1	
s3	4,8	
s4	8,7 / 6,5	7,6
s5	2,6 / 7,3	6,7 / 3,2

Table 3.1: Linking table.

indices of each strand alternates between ① and ② (Fig. 2.1(b)). Hence the entries under the $3'$-linked arm lists all the arm numbers of a given strand in the $5'$ to $3'$ direction. For instance, strand s1=A1-①→A2-②→A3①→A4②. The arms in the $5'$ to $3'$ direction (①→②) are A1, A2 and A3, A4 (not A2, A3). Keeping just the arm numbers, we write 1,2 / 3,4 in the intersecting entry of the s1 row and $3'$-linked arm column. Conversely, under $5'$-linked arm, we write down the arm numbers going in the opposite direction ($3'$ to $5'$ or ②→①), which for s1 is A2, A3. In this manner, the information shown in this table allows us to link all the arms to reconstruct the strands. Hence, typing `link` brings up the following prompt

```
Enter the arm and it's 3'-linked arm:
```

under which the user should type all the numbers in the $3'$-linked arm column separated by a comma.

```
Enter the arm and it's 3'-linked arm:
1,2,3,4,5,1,4,8,8,7,6,5,2,6,7,3
Enter the arm and it's 5'-linked arm:
```

and the same can be analogously done for the $5'$-linked arm.

```
Enter the arm and it's 3'-linked arm:
1,2,3,4,5,1,4,8,8,7,6,5,2,6,7,3
Enter the arm and it's 5'-linked arm:
2,3,7,6,6,7,3,2
```

At this point, we can generate the strands with the 'strandgen' command.

```
Enter a command (type 'help' for help):
strandgen

strand 1
xxxxx xxxxx xxxxx xxxxx xxxxx xxxxx xx

strand 2
xxxxx xxxxx xxxxx x

strand 3
xxxxx xxxxx xxxxx x

strand 4
xxxxx xxxxx xxxxx xxxxx xxxxx xxxxx xx

strand 5
xxxxx xxxxx xxxxx xxxxx xxxxx xxxxx xxxxx xxxxx
xxxxx xxxxx xxxxx xxxxx xxxx

Enter a command (type 'help' for help):
```

5 strands have been produced by concatenating arms according to the directionality given by Tab. 3.1. A couple of caveats are that the numbering of the strands produced by 'strandgen' most likely will not match the numbering of the original strand drawn (*e.g.* strand 5, which is the longest strand above is actually strand s3 in Fig. 2.1(b)). This is not a problem since the order of the strands hold no meaning and are relative to each other. The important point is that all the strands are correctly produced. Also, since no specific bases have been determined yet, all the strands consist of x's at this point which will change once the user starts to determine the sequences. Now we can start filling in the bases with the crunch command.

```
Enter a command (type 'help' for help):
crunch
strands {'strand5':
'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx',
'strand4': 'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx',
'strand3': 'xxxxxxxxxxxxxxx',
'strand2': 'xxxxxxxxxxxxxxx',
```

```
'strand1': 'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx'}
Please enter the following
arm #, starting base, end base, CRITON size, # of repeats (default: None)
```

The user needs to input 4 digits. The first is the arm number. The second and third are the starting and end base numbers, respectively. The fourth is the CRITON size (which is the desired length of the unit segments that SEQMAK produces). An optional fifth digit is the maximum number of repeats this criton is allowed. If this fifth digit is omitted, then no repeats of this criton are allowed. As an example type 1,1,4,4 and press enter.

```
Please enter the following
arm #, starting base, end base, CRITON size, # of repeats (default: None)
1,1,4,4
critkey crit4
GATT
(a)ccept or (r)eject or (s)et
```

A random sequence of criton size 4 has been created, namely GATT. We can choose to accept this segment, reject this segment and create a new random segment of criton size 4, or manually set a user determined sequence of bases of criton size 4. Let's reject this choice by typing 'r'.

```
Please enter the following
arm #, starting base, end base, CRITON size, # of repeats (default: None)
1,1,4,4
critkey crit4
GATT
(a)ccept or (r)eject or (s)et
r
ACTT
(a)ccept or (r)eject or (s)et
```

A new sequence, ACTT, has been created. Let's accept this choice by typing 'a'.

```
Please enter the following
arm #, starting base, end base, CRITON size, # of repeats (default: None)
1,1,4,4
critkey crit4
GATT
(a)ccept or (r)eject or (s)et
r
ACTT
(a)ccept or (r)eject or (s)et
a

strand 1
ACTTx xxxxx xxxxx xxxxx xxxxx xxxxx xx
```

```
strand 2
xxxxx xxxxx xxAAG T

strand 3
xxxxx xxxxx xxxxx x

strand 4
xxxxx xxxxx xxxxx xxxxx xxxxx xxxxx xx

strand 5
xxxxx xxxxx xxxxx xxxxx xxxxx xxxxx xxxxx xxxxx xxxxx xxxxx xxxxx xxxxx xxxx

['ACTT', 'TGAA']
```

Not only has the first four bases of A1 has been created (`ACTT`), but its complementary segment (`TGAA`) has also been automatically created and placed in strand 2. These two segments are now stored in memory (`['ACTT', 'TGAA']`) and will not be used again as long as repeats of this criton size are not allowed (which is the default). The user can also check which arms have already been determined by typing '`show`' at any time.

```
Enter a command (type 'help' for help):
show

arm1
ACTTx xxx
TGAAx xxx

arm2
xxxxx xxx
xxxxx xxx

arm3
xxxxx xxx
xxxxx xxx

arm4
xxxxx xxx
xxxxx xxx

arm5
xxxxx xxx
xxxxx xxx

arm6
xxxxx xxx
```

```
xxxxx xxx

arm7
xxxxx xxx
xxxxx xxx

arm8
xxxxx xxx
xxxxx xxx
Enter a command (type 'help' for help):
```

By repeating this process over all the arms, we can obtain all the segments to produce strands such as the following:

```
...
Enter a command (type 'help' for help):
show

arm1
ACTTA TCA
TGAAT AGT

arm2
CCTGA CAT
GGACT GTA

arm3
AGGAC TGG
TCCTG ACC

arm4
TGGCA CGA
ACCGT GCT

arm5
CCGAG TTT
GGCTC AAA

arm6
CGCCC CTT
GCGGG GAA

arm7
AATCC GTT
TTAGG CAA
```

```
arm8
CCACT GCG
GGTGA CGC
Enter a command (type 'help' for help):
strandgen

strand 1
ACTTA TCAAT GTCAG GAGGA CTGGT CGTGC CA

strand 2
CCGAG TTTTG ATAAG T

strand 3
TGGCA CGACG CAGTG G

strand 4
CCACT GCGAA CGGAT TCGCC CCTTA AACTC GG

strand 5
CCTGA CATAA GGGGC GAATC CGTTC CAGTC CTCCT GACAT AAGGG GCGAA
TCCGT TCCAG TCCT

Enter a command (type 'help' for help):
```

With all the strands completed, we need to check the strands for repeating segments so as to minimize any crosstalk. This is accomplished with the '`repeatcheck`' command. Typing '`repeatcheck`', we get

```
Enter a command (type 'help' for help):
repeatcheck
Enter min. CRITON size, max. CRITON size, min. # of repeats, max. # of repeats
```

Four input parameters are needed. The first is the minimum criton size. This is the minimum segment length that the user would like to check over all the strands. For example if '`min. CRITON size`' = 4, then all segments of length 4 bases or more are checked for repeats over all the strands. The second parameter is the maximum criton size the user would like to check. This sets an upper-bound on the maximum segment length to check. Hence, if '`min. CRITON`size+' = 4 and '`max. CRITON size`' = 6, then criton sizes of 4, 5, and 6 are analyzed over all the strands for repeats. The third and fourth parameters are the minimum and maximum number of repeats the user would like to check. If `min. # of repeats` = 3 and `max. # of repeats` = 6, then all segments of criton sizes 4, 5, and 6 which appear between 3 and 6 times throughout the strands are displayed. As an example, type `4,6,3,6`

```
Enter min. CRITON size, max. CRITON size, min. # of repeats, max. # of repeats
4,6,3,6
'ATAA' has 3 repeats
```

```
strand # => base position
strand5 => 7
strand5 => 39
strand2 => 11

'TAAG' has 3 repeats
strand # => base position
strand5 => 8
strand5 => 40
strand2 => 12

'GCGA' has 3 repeats
strand # => base position
strand5 => 14
strand5 => 46
strand4 => 6

'CGAA' has 3 repeats
strand # => base position
strand5 => 15
strand5 => 47
strand4 => 7

'CAGT' has 3 repeats
strand # => base position
strand5 => 26
strand5 => 58
strand3 => 11

'TCCT' has 3 repeats
strand # => base position
strand5 => 29
strand5 => 32
strand5 => 61

'ATAAG' has 3 repeats
strand # => base position
strand5 => 7
strand5 => 39
strand2 => 11

'GCGAA' has 3 repeats
strand # => base position
strand5 => 14
strand5 => 46
```

```
strand4 => 6

Enter a command (type 'help' for help):
```

The results show that all segments of criton sizes 4, 5, and 6 have a maximum of 3 repeats, which are displayed along with the strand number the criton lies in and its starting base position in the strand. If the user decides everything is satisfactory, then the strands can be saved to a file by typing '`save`'.

```
Enter a command (type 'help' for help):
save
Type the name of the output file:
```

Choose the name of the output file and press enter.

```
Enter a command (type 'help' for help):
save
Type the name of the output file:
strands
Save to file (y/n)?
y
Enter a command (type 'help' for help):
```

The sequence map of all the produced strands along with the arm information will be saved to output file. For the case at hand, the output file is `strands`:

```
arm1
ACTTATCA
TGAATAGT

arm2
CCTGACAT
GGACTGTA

arm3
AGGACTGG
TCCTGACC

arm4
TGGCACGA
ACCGTGCT

arm5
CCGAGTTT
GGCTCAAA

arm6
```

```
CGCCCCTT
GCGGGGAA

arm7
AATCCGTT
TTAGGCAA

arm8
CCACTGCG
GGTGACGC

5' -> 3'
strand 1
ACTTA TCAAT GTCAG GAGGA CTGGT CGTGC CA

strand 2
CCGAG TTTTG ATAAG T

strand 3
TGGCA CGACG CAGTG G

strand 4
CCACT GCGAA CGGAT TCGCC CCTTA AACTC GG

strand 5
CCTGA CATAA GGGGC GAATC CGTTC CAGTC CTCCT GACAT AAGGG
GCGAA TCCGT TCCAG TCCT
```

Now the user may exit the program with the 'exit' command and order the strands!

SEQMAK also allows for the user to check repeat segments of predefined user sequences. The 'load' command can be used to read in strand sequences from a file and check for repeat segments. First, write out the sequences of the strands in a file, one strand per line, *e.g.*,

```
ACGT TGCA TTATA
ATAGTAAAGGA   TTA
  CGCGTGAGTGCAGA
atagtaaaggatta
AGCATGAG agatTTCGGAA
```

The sequences can be written out in either upper or lower case and the whitespaces in-between or at either ends of the sequences don't matter. All that matters is that the user writes one strand per line. Give a name to the file, *e.g.* 'strands.txt', and save it. Run SEQMAK and type 'load' or 'ld'. The output should be the following:

```
Enter a command (type 'help' or 'h' for help):
ld
Enter the name of the file (e.g. strands.txt) or 'q' to exit command:
```

Type 'strands.txt' to load the saved file and an output of all the predefined strands will appear on the screen (reformatted to uppercase and with whitespaces removed),

```
Enter the name of the file (e.g., strands.txt) or 'q' to exit command:
strands.txt

strand1: ACGTTGCATTATA

strand2: ATAGTAAAGGATTA

strand3: CGCGTGAGTGCAGA

strand4: ATAGTAAAGGATTA

strand5: AGCATGAGAGATTTCGGAA

Enter a command (type 'help' or 'h' for help):
```

The sequences of the strands have now been loaded in and the 'repeatcheck' command can be run as described above.

TODO

- add other structures
- include .pdb graphics
- fix bugs
- revise documentation
- other?

GNU Free Documentation License

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "**Document**", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed

as "**you**". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "**Modified Version**" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "**Secondary Section**" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "**Invariant Sections**" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "**Cover Texts**" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "**Transparent**" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "**Opaque**".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "**Title Page**" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page"

means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "**publisher**" means any person or entity that distributes copies of the Document to the public.

A section "**Entitled XYZ**" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "**Acknowledgements**", "**Dedications**", "**Endorsements**", or "**History**".) To "**Preserve the Title**" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified

Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with … Texts." line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

Bibliography

[1] N. C. Seeman and N. R. Kallenbach, "Design of immobile Nucleic Acid Junctions", *Biophysical Journal* **44**, 201-209 (1983)

[2] N. C. Seeman, "De Novo Design of Sequences for Nucleic Acid Structural Engineering", *Journal of Biomolecular Structure Dynamics* **8**, 573-581 (1990)

[3] T. Fu and N. C. Seeman, "DNA Double-Crossover Molecules", *Biochemistry* **32**, 3211-3220 (1993)

[4] Y. He, Y. Chen, H. Liu, A. E. Ribbe, and C. Mao, "Self-Assembly of Hexagonal DNA Two-Dimensional (2D) Arrays", *J. Am. Chem. Soc.* **127**, 12202-12203 (2005)

[5] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, "Design and self-assembly of two-dimensional DNA crystals", *Nature* **394**, 539-544 (1998)

[6] Wang, H. Dominoes and the AEA case of the decision problem. *Proc. Symp. Math. Theory of Automata*, Fox, J. Eds.; Polytechnic Press, Brooklyn, NY, 23-55 (1963)

[7] Rothemund, P. W. K.; Papadakis, N.; Winfree, E. "Algorithmic Self-Assembly of DNA Sierpinski Triangles" *PLoS Biol.* **2**, 2041-2053 (2004)