

# Explaining DNA Structure

V. Manca<sup>a,\*</sup>, G. Scollo<sup>b</sup>

<sup>a</sup>University of Verona, Department of Computer Science, Strada Le Grazie, 15 – 37134 Verona, Italy

<sup>b</sup>University of Catania, Department of Mathematics and Computer Science, Via Santa Sofia, 64 – 95123 Catania, Italy

---

## Abstract

The logic of DNA molecules is explained as a consequence of their main functions: efficiently duplicating polymers conveying enormous quantities of biological information with minimal spatial occupancy. An abstract model is provided, where monomers are represented by triangles, that was materialized by means of a 3D printer.

*Keywords:* Biopolymers, Duplication, Bilinearity, Complementarity, Antiparallelism, Compactness

---

## 1. Introduction

Informational (bio)polymers are molecules composed of strands of monomers, each of which brings a minimal information represented by a letter of some finite alphabet. The crucial role of duplication for life was already emphasized in 1944 by Ervin Schrödinger, almost ten years before the discovery of DNA, by using very general arguments [21]. According to the famous physicist, heredity has to be interpreted as the conveyance of biological information from parents to their descendants, and this transfer has to be based on a specific kind of molecules, with a sort of regular structure similar to crystals but of *aperiodic* nature (otherwise it would be informationally poor). In 1953 the famous model of Watson and Crick was discovered, based on the essential findings of Rosalind Franklin. DNA molecules have a typical double helix form, which soon became popular and revolutionized all fields of biology [1]. The question addressed in the present paper is: “Why DNA molecules have this so specific and particular structure?”. The answer we propose is obtained through an analysis that links the double helix shape to the twofold role of DNA, viz. to represent information and to duplicate it efficiently. The former requires robustness of the whole structure, while the latter entails the necessity of separating its two strands, to accomplish template driven duplication. This occurs either as entire molecule duplication or as piece transcription into RNA strands that activate biological information by means of molecules that exert biological functions.

Along this way, we present an abstract model of DNA, which is independent of any biochemical aspect, but emerges directly from some general principles that DNA structure follows, in order to carry out its functions.

The DNA model given in this paper comes from a first attempt in [10] and several revisions in subsequent works [15, 12, 11, 6, 13, 14]. However, the present interaction between the authors, aimed at making a 3D-printable model out of it, led to correction and improvement of a few motivations and explanations of DNA structural principles. As a matter of fact, the 3D reconstruction in terms of triangle modules imposed a logical refinement of some important aspects of the previous versions of the original model.

The structure of a monomer (see Figure 1) consists of four parts: *head*, *tail*, *body*, *variable*. The head and the tail are extremal parts where their concatenation bonds are placed. A sequence of concatenated monomers builds up a directed strand. The monomer that has an unlinked tail is considered to be the first

---

\*Corresponding author

*Email addresses:* vincenzo.manca@univr.it (V. Manca), giuseppe.scollo@unict.it (G. Scollo)

monomer of the strand, while that one which has an unlinked head is considered to be the last element of the strand; any other monomer in the sequence concatenates its tail with the head of the monomer preceding it and its head with the tail of the monomer following it. Letters are a conventional way of designating different kinds of the variable, informational part of monomers.

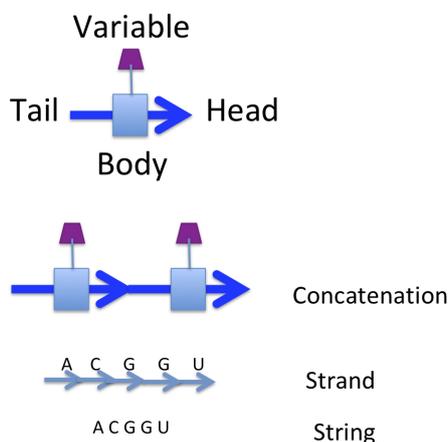


Figure 1: The abstract structure of a monomer and of a linear biopolymer.

In this way, the whole strand has a structure that is similar to that of a single monomer, where the tail of the first monomer is the tail of the entire strand and the head of the last monomer is the head of the entire strand, while the sequence of the linked monomers (including the bodies of the first and of the last element) is the body of the entire strand, and the sequence of letters of the monomers represents the information that is associated to the whole strand.

With reference to informational biopolymers of RNA (ribonucleic acid) and DNA (deoxyribonucleic acid), the respective alphabets of variables are [1]:

$\{A, U, C, G\}$ ,  
 $\{A, T, C, G\}$ .

RNA monomers are *ribo-nucleotides*, DNA monomers are *nucleotides*. Polymers are often designated by the sequences of letters of their monomers.

## 2. DNA structural principles

RNA and DNA polymers are informational, while polypeptides are more properly conformational or functional (RNA molecules can be informational and conformational as well). Moreover, DNA polymers are bilinear in that they are double strands paired by complementarity, where  $A$  and  $T$  are each other complementary and also  $C$  and  $G$  are each other complementary. Bilinear monomer strands are already present in RNA (see Figure 2), albeit only temporarily in template duplication, but they assume a stable and structural character in DNA molecules.

In the schema of Figure 2 three principles appear: *Bilinearity*, *Complementarity*, *Antiparallelism*, that is, 1) two lines of monomers are paired, 2) paired monomers are complementary, and 3) paired lines exhibit opposite directions.

- Bilinearity is the basis of *template driven duplication*, which is carried out by splitting two paired strands and then letting these act as templates. That is, each single strand builds up a copy of the original bilinear structure by pairing each of its monomers to a complementary monomer, floating in the environment, and

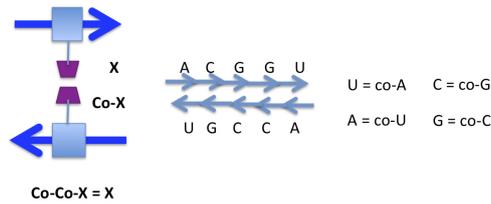


Figure 2: Pairing of complementary monomers.

then concatenating all the newly paired monomers in the given order. This duplication has a computational complexity that is linear in the length of the template, viz. the minimum achievable complexity, whereas other string duplication algorithms usually have quadratic complexity in the string length.

- Complementarity makes pairing bonds sufficiently weak to allow separation of the two paired lines (for reading, partially copying, or duplicating the molecule). In reality, these bonds are hydrogen bonds. If paired monomers were identical rather than complementary ones, then the typical bonds between equal parts would be stronger, covalent ones, whence separation of paired strands would be harder. However, the weakness of pairing need be contrasted when DNA molecules should keep a safe repository of biological information. As we are going to see, the helix shape solves the satisfaction of these opposite requirements in a very brilliant way.

- Pairing occurs along two specular axes, thus in opposite directions, so that paired monomers cannot be further paired with other monomers—pairing of three monomers along the same direction is drawn at the bottom of Figure 3. The exactly binary pairing is coherent with template driven duplication of DNA.

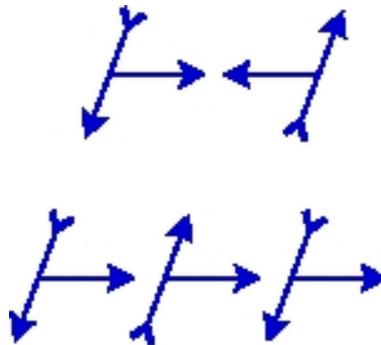


Figure 3: Top: Pairing of two monomers has to be specular. In this way, when two monomers are paired, then their pairing is *closed*, that is, neither of them may pair with another monomer. Bottom: if pairing is not specular, more than two monomers may bind.

An object is *asymmetric along a direction* (a straight line), if no plane orthogonal to that direction splits the object into two specular parts.

Directed concatenation and specularity of pairing require that monomers be asymmetric with respect to the concatenation direction, y-axis, and with respect to the pairing direction, x-axis. However, when the double-stranded sequence of paired monomers is considered to live in 3D-space, as it actually does, a third, linearly independent direction ought to be taken into account, that is, a z-axis.

Any winding of a line is essentially a spiral. If one considers all the ways DNA double strands can be arranged in a spiral form, one easily concludes that these spirals cannot develop in a plane. For, in any planar spiral the concatenation angles will vary along the spiral. This means that for a double strand of millions of monomers the concatenation angles between distant points would vary very much. In terms of geometric structures of monomers, an unbounded number of different kinds of them would have to lie along



Figure 4: The two structures above are paired in a parallel way, but they have different handedness (the person on the left has her heart on the left, while her specular copy has the heart on the right). On the contrary, if two paired monomers have the same handedness, then they need be paired with opposite concatenation directions. With kind permission by G. M.

a spiral, that would not even be similar. This fact is intolerable if one considers that monomers have similar biochemical structures and all concatenation (phosphodiester) bonds occur on similar kinds of molecules.

On the contrary, the winding of DNA according to a helicoidal shape provides a spiral form that develops along to the z-axis, where concatenation angles remain almost the same in any pair of consecutive monomers, see Figures 9, 10.

Monomers are chiral, that is, as the hands of a human body, they are 3D objects different from their specular images. In fact, the word chiral comes from a Greek root meaning hand, and the difference between chiral objects that are specular images of each other is expressed as their left or right *handedness*, see Figure 4. Clearly, the specular relation is an involution on handedness, that is to say, a specular image of a specular image of object X has the same handedness as X. Chiral objects that have the same handedness are said to be *homochiral*.

According to the historical definition due to Lord Kelvin [25], a geometrical figure is *chiral*, and said to have chirality, if its mirror image cannot be brought to coincide with itself. This is made more accurate by adding “by rotations and translations alone”. For example, a right shoe is different from a left shoe, and clockwise is different from anticlockwise in the Euclidean plane.

Please note that the definition is independent of the mirror position and direction. The mirror is a plane in 3D, while it is a straight line in 2D, whereby it is immediately recognized the chirality of the previously introduced 2D structure of monomers, since it cannot be brought to coincide with its mirror image, with respect to any mirror line, by any rototranslation in the 2D plane. However, more can be said about the chirality of paired monomers in the 2D plane, as they are represented by the oriented T-shapes in Figure 3.

**Proposition 1.** *Paired 2D monomers are homochiral iff they pair in an antiparallel way.*

**Proof.** The requirement of specular pairing (see Figure 3) entails that, if paired monomers would concatenate in the same direction, then they would be specular images of each other, hence would have different handedness; by contraposition, homochirality entails antiparallelism. Conversely, specular pairing and antiparallelism together amount to a double specular reflection between paired monomer structures; since specularity is involutory on handedness, antiparallelism entails homochirality.  $\square$

We are going to argue the necessity of chirality of the 3D monomer structure as well. An example of 3D achiral object is a bi-glove, that is a glove that, indifferently, a right hand or a left hand can wear (by rotating it). In fact, such a glove is identical to itself after a rotation around the axis going from the wrist to the middle finger (see Figure 5).

A more general definition of chirality is given in [18], that is based on group theory and permits to define chirality in metric spaces which are not Euclidean, while proving equivalent to the usual one in the case of



Figure 5: Surgery gloves where the palm and the back are indistinguishable.

Euclidean spaces. The usual definition in 3D Euclidean space is adequate for the present purpose, where a few words may be useful about helix chirality and chirality of reference systems (orthogonal ones, for the sake of simplicity).

Helix chirality is usually characterized as right-handed when its clockwise rotational sequence leads in the screw tightening direction. On the other hand, a right-handed orthogonal reference system  $(x,y,z)$  assigns the upward direction to the  $z$  axis when the  $x$ -to- $y$  rotation around it is anticlockwise. There is no conflict between these conventions, since the  $x$ -to- $y$  rotation is viewed as anticlockwise when looking downwards, thus in the direction that is opposite to the one assigned to the  $z$ -axis, but it becomes clockwise when looking in the latter direction, which is the upward direction of screw tightening in Figure 6. Clearly, the two helices which form the DNA double strand must be homochiral, otherwise they would cross repeatedly on the cylinder surface. The most stable form of DNA is the B form, which is a right-handed double helix.

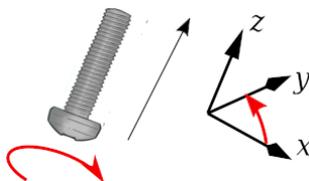


Figure 6: Right-handed screw and 3D reference system conventions.

**Proposition 2.** *Chirality of the 3D-structure of monomers is necessary to the correct formation of bilinear DNA strands.*

**Proof.** If monomers were achiral, any of them could be brought to coincide with its mirror image by rototranslations in 3D space, with respect to any mirror plane. Since template-driven duplication is a process where pairing bonds are formed before concatenation bonds, the free motion of monomers getting close to the template could lead to pairing both kinds of their mirror images with respect to the mirror plane that is orthogonal to the concatenation direction, without affecting orientedness of the pairing direction. However, neither head-to-head nor tail-to-tail concatenation bonds are allowed to form the DNA double strands.  $\square$

In conclusion, monomers are chiral and, as we are going to argue, they all have the same chirality, that is: 1) the tail concatenation extreme point T, 2) the head concatenation extreme point H, and 3) the pairing extreme point P form a triangle THP, called *monomeric triangle*, see Figure 7, where the axis orthogonal to the THP plane is oriented in the direction for which the rotation of H around T that leads H toward P is anticlockwise, according to the right-hand rule.

The next analysis deals with the relationship between the “global” chirality of the helix-shaped double strand and the “local” chirality of paired monomers, whereby a geometrical basis of antiparallelism is found.

First, consider how a linear double strand may wrap around a cylinder surface while keeping the same relative pairing directions of paired monomers. Please note that, because of antiparallel wrapping, paired

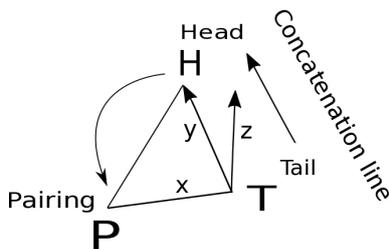


Figure 7: The monomeric triangle.

2D-monomers lie on different planes, see Figure 8. The intersection of those planes is their *pairing axis*, while the segment that joins each monomer's concatenation bonds along its single strand is its *concatenation side*. Finally, the *phase angle* of monomer pairing is defined as the angle  $\widehat{MCN}$ , not wider than  $\pi$ , on a plane orthogonal to the cylinder axis, say the  $xy$ -plane, where  $C$  is the intersection of the cylinder axis with that plane while  $MN$  is the projection of the pairing axis on it.

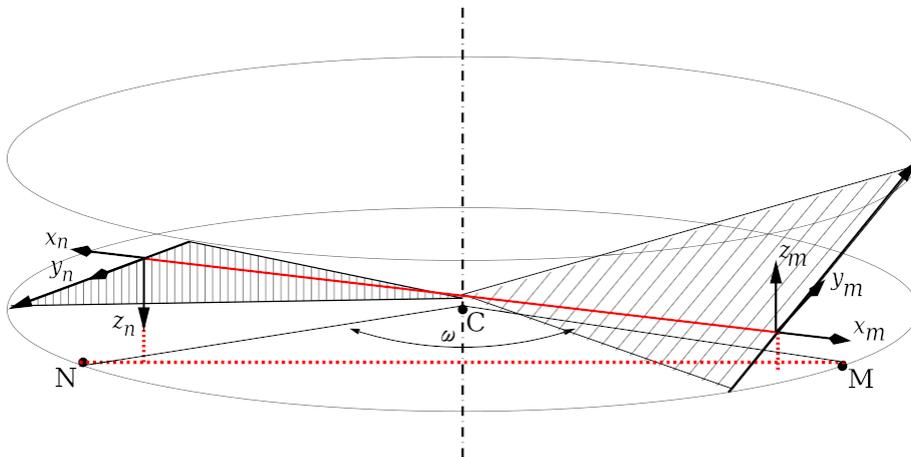


Figure 8: Monomer pairing in 3D.

How deep should the pairing axis get across the inside of the cylinder? The answer is simple: the more, the better, according to the aforementioned DNA function of keeping a safe repository of its biological information. This means that a pretty wide **phase angle is to be expected**. Let the **cylinder axis** be the  $z$ -axis of a right-handed reference system associated to the (double) helix. When placed on the cylinder, each monomer  $m$  may be endowed with a local, **nonorthogonal** reference system with  $z_m$ -axis parallel to the global  $z$ -axis,  $x_m$ -axis along the pairing axis and  $y_m$ -axis along the concatenation direction.

Consider two paired, complementary monomers, as drawn in Figure 8. Please note that the direction of local  $x$ -axes is reverted with respect to the pairing arrows in Figure 3, for consistency with the usual convention on right-handedness of reference systems as drawn in Figure 6. Now, in order to contribute to form a given helix, the concatenation side of either monomer must exhibit a fixed *rise angle* over the  $xy$ -plane. For monomer  $m$ , this is the angle between its concatenation side and the  $xy$ -plane. Clearly, the sequence direction of a linear strand is turned to the common  $z$  direction of the cylinder axis and the double helix geometry lifts the 2D-antiparallelism of the bilinear double strand to the 3D-antiparallelism of its wrapping around a cylinder, whereby the  $z_m$  and  $z_n$  directions of paired monomers  $m, n$  are opposite.

There is a close relationship between antiparallelism and chirality. According to Proposition 1, antiparallel pairing of monomer structures in the 2D bilinear sequence agrees with their 2D-homochirality, which

means that the same monomer may be indifferently located in either strand without coming out of the plane. Homochirality of monomers must also hold in their 3D geometry, for otherwise two variants of each monomer kind would be needed, the right-handed one and its specular, left-handed mate. Having the same handedness is a very useful property, because it entails that any monomer can be posed indifferently in either strand, and this homogeneity is a real advantage in many processes of DNA recombination that are fundamental for natural selection. However, Proposition 1 only holds for 2D monomers, in 3D space an additional reason for antiparallelism of DNA bilinear strands is found. *Local distinguishability* of paired strands at distant sites is the possibility to assess whether two monomers, separated by several monomer pairs in the bilinear strand, do or do not belong to the same single strand, without a sequential travel keeping track of the whole sequence that separates the subject monomers. The rationale for antiparallelism of bilinear strands in 3D space is based on the following statement.

**Proposition 3.** *Bilinear DNA strands of homochiral monomers can be locally distinguished if they pair in antiparallel way.*

**Proof.** The scan of monomers at distant sites in a DNA double strand may determine whether they do or do not belong to the same single strand by comparing their concatenation directions relative to the cylinder axis. A monomer concatenation direction may be taken as either its projection along the (vertical) cylinder axis or its rotational direction around it. Now, parallel paired monomers would have the same vertical concatenation direction. Specularity of pairing then entails that homochiral monomers would also have the same rotational direction. Therefore, same strand membership of monomers at distant sites cannot be assessed from either of their concatenation direction components. On the contrary, antiparallel pairing of homochiral monomers entails concatenation with opposite vertical as well as rotational directions in the paired strands, thus enabling local distinguishability of paired strands at distant sites.  $\square$

Biological motivations for local distinguishability of DNA strands at distant sites are found in the DNA dynamics, that is, in the way certain specific searches by biological agents need be performed in order to effectuate corresponding biological actions. For example, restriction enzymes scan DNA while looking for specific subsequences of monomers that activate their cleavage of the DNA double strand. The search proceeds by linear diffusion combined with hopping [19], therefore it is not necessarily a strictly sequential one, and certain enzymes have recognition elements at distance of tens to thousands of monomer pairs [24].

*Discontinuous reading* of DNA subsequences, that skips intermediate parts of the DNA sequence, is another factor that highlights the relevance of recognition of subsequence elements at distant sites, for it yields genes that are formed by noncontiguous parts, whose recognition as belonging to the same single strand is necessary to enable specific biological actions.

The *convention* of right-handed chirality of the 2D monomer structure, entails right-handed chirality of the double helix whenever the rise angle is positive, as it is the case with the B form of DNA. Note, however, that the same, right-handed monomer structure may lead to a left-handed double helix development, viz. by having a negative rise angle, as it is the case with the Z form of DNA.

The logic of DNA bilinear geometry is abstractly, but completely, described by representing monomers by means of triangles [10]. The bilinear arrangement geometry turns out to be, for many aspects, independent of the biochemical materialization of those triangles.

The argument developed so far implies that the requirements of homochirality and homogeneity of monomers can be added to the three principles, of DNA bilinearity, complementarity and antiparallelism.

Now we argue that not only does the helicoidal shape wind the strands while maintaining constant concatenation angles, but it also ensures DNA bistability, that is, the property of DNA of assuming open and closed configurations.

When DNA is closed, the biological information stored in its strands is potential, that is stored and protected in a very stable way, but not available for being read or processed. When DNA is open, then the two strands can be easily unpaired and reading and duplication are available. In biochemical terms this is obtained thanks to the possibility of making pairing bonds strong (in the closed configuration) or weak (in the open configuration). In other words, helix winding/unwinding realizes the closed/open switch between stored information and information that acts inside cells.

Information reading by biological agents is a complex process that is carried out with help by suitable enzymes. To this end, first a portion of DNA helix (delimited by specific factors) is unwound from its tightly coiled shape, in such a way that the monomers of the double-stranded portion are “unzipped”. In this conformation, they are sequentially scanned by the reading agent, according to suitable docking mechanisms that establish a correspondence between the geometry of the monomer and the internal geometry of the agent. This enables specific monomer recognition as well as assessment of which strand does it belong to, according to Proposition 3.

When the size of DNA double strands grows, the problem of space occupancy of such long polymers becomes important. The increased duplication reliability of DNA with respect to RNA is due to the substitution of the  $U$  monomer with the  $T$  monomer and to the shift from ribo-nucleotides, which have a pentose with five Carbon atoms and five water molecules, to nucleotides, where one of the water molecules is deprived of an Oxygen atom. This second aspect yields a greater chemical stability, while the  $U/T$  substitution, coupled with bilinearity, affords a better error checking mechanism (due to specific biochemical properties).

A biological process, called *transcription*, transforms DNA polymers into corresponding RNA polymers, where a portion of single-stranded DNA delivers an RNA strand where every  $T$  monomer is replaced by a  $U$  monomer [1] (the substitution of  $U$  by  $T$  confers more stability to DNA strands). Another fundamental biological process, called *translation*, transforms a RNA polymer into a Polypeptide called *Protein*, where each of 61 RNA triplets, among all possible 64 triples over 4 letters, is replaced by a monomer having a letter of the Polypeptide alphabet (according to a code, called *genetic code*). RNA polymers are a sort of bridge between DNA and Proteins [8]. DNA polymers are biological memories encoding the information transmitted along the generation when membranes become cells [17]. Proteins are molecules that perform biological functions encoded in DNA (enzymatic, kinetic, structural, energetic ones, and so on).

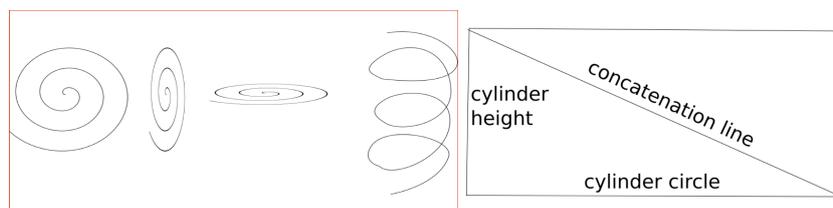


Figure 9: Left: Possible DNA winding in spiral forms. Planar spirals are not allowed, because concatenation angles vary with the distance from the origin. In the helicoidal shape concatenation angles are constant. Right: spatial compression along a run of  $2\pi$  radians, computed by a triangle where the basis is equal to the length of the cylinder circle, the height is the distance between two monomers placed on the same vertical line, and the hypotenuse is the length of the spiral, along the cylinder surface, between the two monomers. Helix is the first compression level of DNA (with a compression rate around 2.17 in the B form of DNA), which confers sufficient robustness for further windings (nucleosomes, loops, fibers, chromosomes).

Information stored in the DNA bilinear structure is only potential; namely, only when it is transcribed into RNA does it become active biological information. Moreover, bilinear strands afford a mechanism of error control and repair; they can be spatially compressed by winding them around the ideal cylinder of a double helix, where the length compression rate is given by the length of the double helix divided by the height of the cylinder (see Figure 9). This compression is the basis of further windings, which yield a huge reduction of spatial occupancy; this is essential when DNA size dramatically grows in complex organisms. Consider that all DNA strands inside the  $10^{14}$  cells of a human body, when chained together along a straight line, reach 600 times the Earth-Sun distance [2].

In conclusion, the only spiral arrangement of a DNA double strand has to wind over an ideal cylinder as depicted in Figures 9, 10. In this arrangement monomeric triangles remain *reasonably* similar.

The double helix winding of paired monomeric triangles is completely defined by four parameters:

- The *radius*  $r$  of the cylinder where helices are placed (in reality, 1 nm);
- The *twist angle*  $\varphi$  between the x-axes of two concatenated triangles, projected on a plane orthogonal to the cylinder axis (in reality, around  $36^\circ$ );

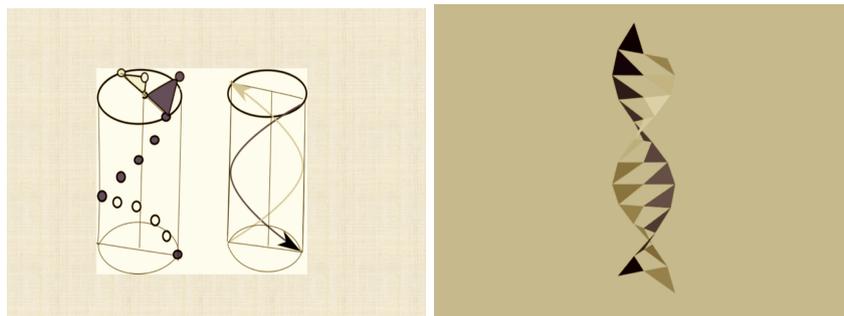


Figure 10: DNA cylinder and monomeric triangles inside it.

- The *axial rise angle*  $\chi$  of the x-axis of a triangle over a plane orthogonal to the cylinder axis (in reality, around  $\pm 30^\circ$ , where the sign discriminates the opposite z-components of the concatenation directions of paired triangles);

- The *phase angle*  $\omega$ , that is the rotation angle around the cylinder axis between two paired triangles, according to its previous definition. This angle is smaller than  $\pi$  because the pairing axes do not cross the cylinder axis. Therefore the cylinder volume is partitioned in two grooves of different width, that are separated by the monomer strands along the cylinder axis direction, see Figure 11. Correspondingly, the projection of the pairing axis (the broken red line in Figure 11) on a plane orthogonal to the cylinder axis partitions the cylinder circle according to the ratio  $\omega : 2\pi - \omega$ . In reality  $\omega$  is around  $127^\circ$ , near to  $3.5\varphi = 126^\circ$ .

The pictures in Figure 11 come from the 3D model described in the next section. The top view of the model, see picture on the lower left side of the figure, shows the internal hole drawn by pairing bonds. The front view on the right side shows the major groove (labelled M, upper part) and the minor groove (labelled m, lower part) in the DNA double helix (the broken red line on the bottom is the lowest pairing axis).

A summary of twelve structural principles can be extracted from the analysis developed so far. These principles are by no means mutually independent ones; on the contrary, the present analysis aimed at uncovering some of their interdependencies while proposing a rationale for their existence.

1. Directed linearity (of concatenation)
2. Specularity of pairing direction (no more than two monomers may pair)
3. Complementarity (weak pairing)
4. Monomeric chirality (unique **concatenation** direction)
5. Homogeneity (monomeric homochirality)
6. Antiparallelism (of paired monomer strands)
7. Discontinuous reading (of substrands)
8. Strand distinguishability (local at distant sites)
9. Uniform angular winding (helix)
10. Internal pairing (information protection)
11. Bistability (winding-unwinding)
12. Asymmetry of grooves (phase angle  $< \pi$ )

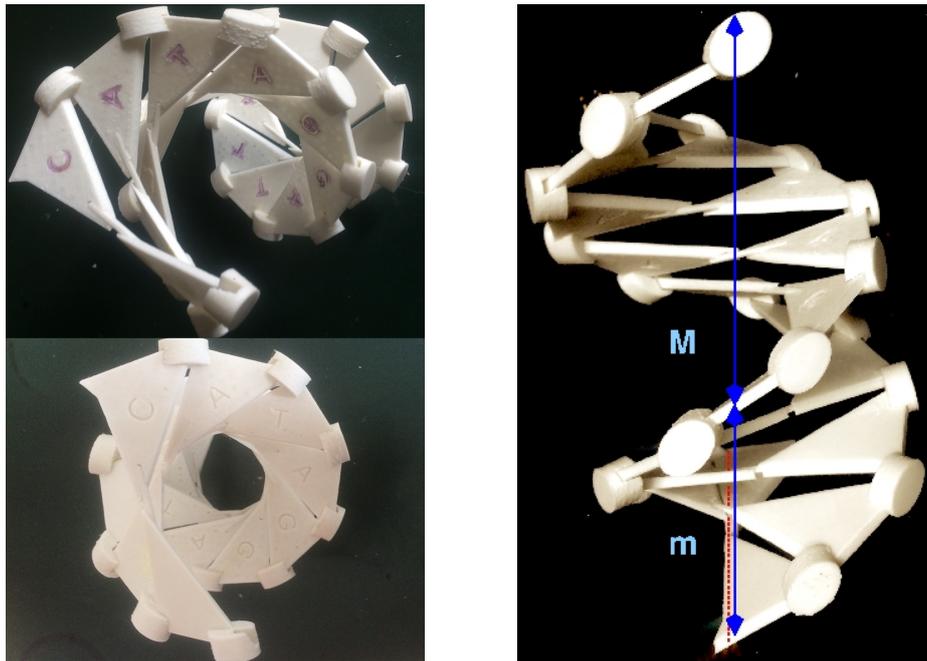


Figure 11: Double helix of the monomeric triangles of DNA CATAGGATTA polymer, with internal hole drawn by pairing bonds, and labelled grooves.

The double helix shape of DNA affords spatial compression and robustness to the double strand that is similar to a rope, wound again and again at further winding levels (nucleosomes, loops, fibers, chromosomes). In conclusion, DNA is a crucial passage through which proto-genomes reach the modern form of genomes and cells may increase the size of their biological memories by moving toward the more complex and efficient forms of *eukaryotes*, from which multicellular organisms will stem [9].

### 3. Double helix of monomeric triangles

The development of the 3D-printable model of a DNA double helix shown at the end of the previous section was carried out using the open-source 3D-modeling language OpenSCAD and related software [16]. This is a visual interpreter of a full-fledged functional programming language, whereby it proves straightforward to translate the mathematical structure of DNA into the modular programming constructs of OpenSCAD, that follow the Constructive solid geometry (CSG) paradigm [7]. However, while the abstract model considered in Figures 9 and 10 is composed of 2D-triangles arranged in 3D-space, its materialization shown in Figure 11 consists of solid components. These include those which represent monomeric triangles as well as those which form their binders, either for concatenation along the antiparallel helices or for complementary pairing. The design of these components for 3D-printing was derived from that of a visual model where component instances are suitably positioned in 3D-space. The design of this model is presented here, while suggestions for possible educational use of the present work are given in the next section, including practical advice and tips about assembling and manipulation of the resulting physical model. Implementation issues and related workarounds, together with further advice about 3D-printing the model are dealt with in Section 4.

#### 3.1. Design goal and requirements

The idea of developing an open-source 3D-printable model of a DNA double helix was mainly motivated by the following goal: to materialize a sample DNA model, fit for educational use, while only taking geometric and informational features of DNA structure into account.

Intended use scenarios include such basic tasks as: butterfly assembly, viz. pairing of complementary monomers (such a pair is named “butterfly” in the model source code, following its poetic view by [20]); single strand assembly, viz. concatenation of two or more monomers along a single helix; double helix assembly, viz. concatenation of two or more butterflies.

More complex tasks may be conceived as well, for example DNA template-driven duplication, or even those inspired by DNA recombination processes, see Section 6.

The aforementioned basic tasks actually drive the model design too, where a distinction is drawn between the *visual model*, with components positioned in 3D-space according to the required geometry, and the *printing model*, where each individual component, unconnected to any other component, is positioned at the center of the horizontal plane. Components in the printing model are derived from corresponding ones in the visual model mainly by simple roto-translations (but see also the second constraint in Section 3.2 below).

Two further requirements are set for the model design:

- *scalability*: a single scale factor is defined, to determine the size of the components, so that different model sizes result from different values of the scale factor, while respecting proportions and overall model geometry;
- *parametrization* with values close to given measures of the DNA B model [23]: parameters of the model geometry are defined, with values determined by the scale factor; in a few cases, small deviations from reference values are taken, for implementation convenience—for each case, motivation and details are given in Section 3.3 below.

### 3.2. Design constraints

While design requirements determine most of the 3D-model geometry, certain design aspects come from other sources. Details about the constraints that thereby arise and some preliminary practical advice are given in Section 4. Two different kinds of constraints are readily identified, as follows.

1. The available 3D-printer features a single extruder, which fact poses a severe constraint on the printing model, viz. it sets an upper bound (approximately  $50^\circ$  in our case) on the *overhang* angle, that is, the angle between the vertical line and any surface that delimits from below the volume of the object to be printed.
2. The biochemistry of DNA bindings tells that concatenation is established by phosphodiester bonds, and is thus stronger than pairing, which occurs by hydrogen bonds. This affects the relationship between components in the visualization model and their derived variants in the printing model, where empirically determined, different *tolerances* between contact surfaces of connected components make these slightly differ from their visual counterparts. It ought to be mentioned, too, that actual tolerances also depend on the choice of the printing materials, which need not be (and as a matter of fact are not) the same for all components.

Other constraints relate to the actual accuracy that may be achieved whenever volumes present very thin parts or a large height/width ratio.

### 3.3. Design decisions

An outline of the design decisions relating to the model geometry is presented next, while the relevant mathematical details are deferred to Section 3.4. Hereafter, the cylinder axis is assumed to lie on a vertical line—as in the visual model.

Reference measure approximations were the first design decisions to be made, according to various criteria among which the most relevant ones are: geometric simplicity, realizability under the aforementioned design constraints, ease of use of the printed model with respect to the given educational goal. Here is a list of the most significant deviations from the reference measures reported in the cited literature. However, most of our values agree with the so-called “textbook” version of the DNA B model.

- Number of butterflies spanning a  $360^\circ$  double helix rotation around the cylinder axis: in the visual model this is parameter  $nm = 10$ , while the literature sets it to 10.5. Besides geometric simplicity, ease of use is a motivation for this, too, for it affords a simple visual check of correct concatenation of just five butterflies by checking the parallelism of the horizontal projections of the start and end sides of each helix.
- The geometric simplicity of the previous choice yields a slightly larger, yet somewhat elegant value of the *twist angle*  $\varphi = 36^\circ$  between the horizontal projections of the concatenated sides of two adjacent monomers in either helix. This is the vertex angle of the golden triangle, which is an isosceles triangle, thus with each base angle twice wide as the vertex angle. This ratio greatly simplifies several calculations in Section 3.4. The triangle owes its name to the golden ratio it features between the length of the duplicated side and that of the base side.
- The *axial rise* of each butterfly, with reference values around  $3.4 \text{ \AA}$ , is approximated to  $3.57 \text{ \AA}$  in the visual model, resulting from a  $35.7 \text{ \AA}$  measure of the *helix pitch*, that is the cylinder height for a  $360^\circ$  turn of the double helix, in order to get an *axial rise angle*  $\chi = 30^\circ$ . The motivation for this choice is the convenience of a  $60^\circ$  relative inclination between a butterfly wing pair, for it enables their efficient pairing by a connector with hexagonal cross-section and with two edges respectively coming out of (vertically) opposite faces of the paired wings. However, please note that  $35.7 \text{ \AA}$  is the reference measure of the helix pitch in the cited literature, thus the deviations of the twist angle and axial rise values in the visual model from their reference values may be seen as just consequences of the design decision about the  $nm$  parameter (see above).

- The *minor/major groove ratio*, that is set to  $12/22$  in the DNA B model, is implicitly approximated to  $3.5/6.5$  in the visual model, for geometric simplicity according to the following rationale. The reference value of this ratio comes from a measure of the grooves vertical span in a complete turn of the helix. The same ratio must obviously partition the cylinder horizontal cross-section, where the border between the two grooves is established by the projection of the wing pairing axis on it (please note that the pairing axis does neither intersect the cylinder axis nor belong to any horizontal plane, see next approximation).

More precisely, the intersections of the horizontal projection of the pairing axis with the circular border of the cylinder cross-section partitions the  $2\pi$  span of the latter according to the minor/major groove ratio. In terms of the  $\varphi$ -sectors determined by the wing horizontal projections, our choice means assigning  $6.5 \varphi$ -sectors to the major groove and  $3.5 \varphi$ -sectors to the minor groove. This corresponds to phase angle  $\omega = 126^\circ$ , which is the angular span of the minor groove. The fact that most of the monomers' biological material is placed in the major groove [3] is represented in the visual model by letting  $3 \varphi$ -sectors separate paired wings in the minor groove, with the pairing axis accounting for a further  $\varphi/4$  of the horizontal angular span of each wing assigned to the minor groove, whereas  $5 \varphi$ -sectors separate paired wings in the major groove, with the pairing axis accounting for a further  $3\varphi/4$  of the horizontal angular span of each wing assigned to the major groove. As we are going to mention next, this induces a further, very small deviation from reference measures.

- The cited literature for the DNA B model reports a *base pair tilt* angle of about  $-6^\circ$  between a line drawn through the paired bases and a line perpendicular to the helix axis. In the visual model this is defined as the angle between the pairing axis and its horizontal projection, see Figure 13, where the pairing axis is the red line. It seems convenient to assign the same vertical positioning to paired wings, viz. the same vertical coordinate to the midpoints of their concatenation sides. Under this choice for the visual model, one gets a *pairing tilt angle*  $\psi = -5.8^\circ$  (the sign reflects the relative orientation with respect to the directed vertical axis for each helix). The calculation of this value is in Section 3.4.1.

Other design decisions relate to the shape and relative fitting of components. About the latter, it proves convenient to connect wings by distinct connector components for concatenation and for pairing, since these are smaller pieces that require less printing time and material. Moreover, double helix assembly and reassembly operations prove easier if connectors are separate components.

The wing horizontal projection basically is a triangle whose shape is determined by the cylinder radius (10 Å), the aforementioned twist angle  $\varphi$ , and the minor/major groove ratio, as the latter occurs in the calculation of the pairing vertex position. The triangle base side is also taken as a side of an adjoined thin stripe that extends the wing outwards, to represent the DNA backbone; the modified triangle is actually a pentagon. Linear extrusion of this 2D-shape renders a solid wing, which undergoes two volume subtractions as carvings on one of the (almost) triangular faces:

- the *pairing slot* where the pairing connector is designed to slide in;
- the monomer base identifier, viz. one of the four 3D-letters A, G, C, T.

The length of the pairing slot uniquely corresponds to the monomer base, because of the following design decision: to represent complementarity by exploiting a geometric feature of DNA base complementarity, viz. the different radial lengths of the four bases, together with their different molecular masses, whereby pairing is subject to the constraint that the sum of the radial lengths of paired bases should fit in the available space while also filling it, and that the molecular mass of different base pairs be (almost) the same. In the visual model design, pairing slots have different, base-dependent lengths while pairing connectors are all identical components.

Both pairing slot and base identifier carvings are made on the same face of any wing, that is the upper face in the printing model because of the implementation constraint that sets the overhang upper bound. However, in the visual model the carved wing faces show opposite vertical orientation, according to double helix antiparallelism, with relative inclination of  $2\chi = 60^\circ$ . Another implementation constraint led to the decision of positioning the wing slot carvings a little away from the theoretical pairing axis, which runs through the pairing vertices of a butterfly wings; without this decision, the slot hole would leave too thin margins at the wing tips, which would thus prove either very fragile when wings are printed out of rigid material or too slack when flexible material is made use of. The relative inclination of paired wings by  $60^\circ$ , together with their geometry enable one to calculate the distance of a pairing connector, with hexagonal cross-section, from the theoretical wing pairing vertex, such that just one of the six edges of the connector's cross-section comes out of the carved surface of either wing. It is easy to realize that the connector's outcoming edges in the two wing faces are different, and that they are neither contiguous nor opposite.

Finally, the design of the concatenation connector aimed at robustness and strong grip, hence the connector is carved in this case, since it may well be designed substantially thicker than the inserted wings. No calculation is needed in this case for the carvings, since these may be obtained by wing subtraction in the visual model, with the outcome then turned into a printing model by a suitable roto-translation.

### 3.4. A few mathematical underpinnings

A reasonable selection of mathematical details are presented in this section, that is aimed at explaining how the previously introduced design decisions are effected in the model geometry. The influence that certain 3D-printing implementation constraints had on its actual form is worked out in Section 5.

A preliminary remark is in order. As already mentioned, the geometric model is highly parametric, viz. a scale factor allows one to 3D-print it in any viable size by simply setting the scale parameter to the desired value. Moreover, different values for the previously mentioned parameters of the DNA double helix could be chosen, with three exceptions:

- *concatenation twist angle*  $\varphi = 36^\circ$ . This means that the number of base pairs per turn is fixed to 10. The only reason for this rigidity is the convenience of the vast amount of properties enjoyed by the golden angle, that greatly simplifies several calculations in the present case.
- *axial rise angle*  $\chi = 30^\circ$ . A practical reason comes into play in this case, viz. the convenience of materializing the pairing connector with a nearly hexagonal cross section (see Figure 17), to facilitate the wing pairing assembly with correct relative inclination.
- *minor/major groove ratio*  $= 3.5/6.5$ . This is not even a designated parameter, as this value is implicit in the calculation of the pairing axis position using Euclid's angle bisector theorem (*Elements*, Proposition VI.3), as we are going to see right away.

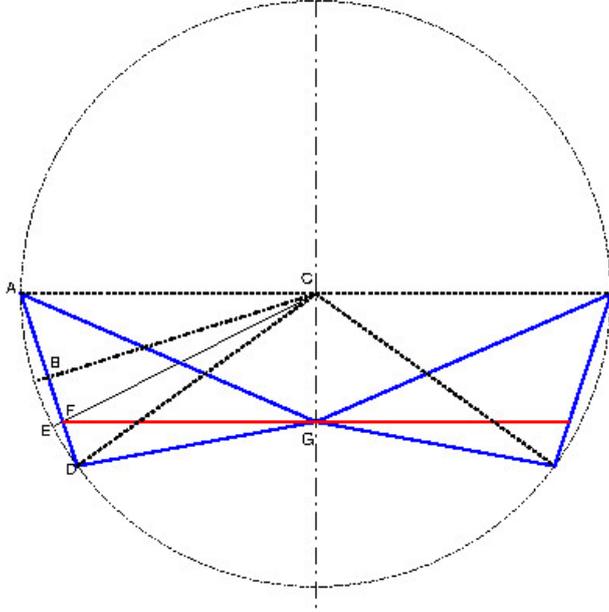


Figure 12: Calculation of pairing measures

### 3.4.1. Pairing measures

A schematic drawing of the horizontal projection of wing pairing is displayed in Figure 12, with wing triangles drawn in blue, pairing vertex G, and additional elements drawn on its left side that are meant as a visual support to the calculation of the exact position of vertex G, and of the length of the pairing axis projection (the red line) through it. Here is a list of picture elements, with their geometric characterization and either defined or computed measures whenever applicable:

- $ACD = \varphi = \pi/5$ , sector assigned to the wing triangle on the left side;
- $|AC| = |EC| = |DC| = r$ , cylinder radius; point E is the left border between the grooves on the circle, where the right border is its mirror point behind the paired wing;
- BC: median of triangle ACD, thus  $ACB = BCD = \varphi/2$  and  $|BC| = r \cdot \cos(\varphi/2)$ ;
- CF: angle bisector of BCD, thus  $BCF = FCD = \varphi/4$ , with point F being the pairing axis left end; let  $b = |EF|$ , thus  $|FC| = r - b$ ;
- $CFG = ACF = 3\varphi/4$ ;
- $|BC| = (r - b) \cdot \cos(\varphi/4) = r \cdot \cos(\varphi/2)$ ;
- $|FC| = r - b = r \cdot \cos(\varphi/2) / \cos(\varphi/4)$ ;
- $|GC| = |FC| \cdot \sin(3\varphi/4) = r \cdot \cos(\varphi/2) \cdot \sin(3\varphi/4) / \cos(\varphi/4)$ ;
- $|FG| = |FC| \cdot \cos(3\varphi/4) = r \cdot \cos(\varphi/2) \cdot \cos(3\varphi/4) / \cos(\varphi/4)$ .

While the value of  $\varphi$  and the position of vertex G fully determine the triangle shape, it is useful to find the ratio  $f = |AF|/|AD|$ , to ease the positioning of the pairing axis in the wing design. By Euclid's angle bisector theorem one gets:  $|BF|/|FD| = |BC|/|DC| = \cos(\varphi/2)$ .

Let  $h=|AB|=|BD|$ ,  $k=|BF|$ , then  $|FD|=h-k$ , so  $k/(h-k)=\cos(\varphi/2)$ , hence  $k=h \cdot \cos(\varphi/2)/(1 + \cos(\varphi/2))$ , whence

$$f = \frac{h+k}{2h} = \frac{1}{2} \cdot \left(1 + \frac{\cos(\varphi/2)}{1 + \cos(\varphi/2)}\right) = 1 - \frac{1}{2(1 + \cos(\varphi/2))}$$

It is left as an exercise to the reader to prove that  $|GC| = f r \sin \varphi$  and that  $|FG| = r(1 - f(1 - \cos \varphi))$ . The latter proves useful to get a relation between the ratio  $f$  and the three main angle parameters  $\varphi, \chi, \psi$ , independent of  $r$ , as follows.

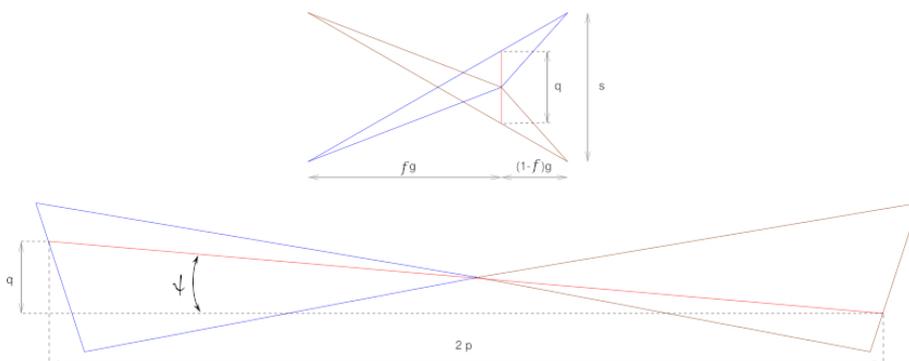


Figure 13: Calculation of pairing tilt

Let  $p = |FG|$  denote the half-length of the horizontal projection of the pairing axis, while  $q$  denotes the length of its vertical projection. The latter is a fraction of the butterfly axial rise  $s$ , by a ratio that relates to  $f$  by pairing specularity according to the following rationale. Consider the paired triangles in 3D space, with  $z$ -axis the cylinder axis and  $y$ -axis the symmetry axis between their horizontal projections as in Figure 12. By assigning them the same vertical span, their projections on the  $yz$ -plane are as shown in the upper drawing of Figure 13, where  $g = r \cdot \sin(\varphi)$  is the length of their  $y$ -projection. Then the ratio  $q/s$  is easily found to be  $q/s = f - (1 - f) = 2f - 1$ .

On the other hand, the axial rise angle  $\chi$  is related to  $s$  by  $\tan(\chi) = s/(2r \sin(\varphi/2))$ , whereby the pairing tilt  $\psi$  satisfies the following equation (see the lower drawing in Figure 13, that shows the  $xz$ -projection of the paired triangles):

$$\tan \psi = q/2p = (2f - 1)s/2p = (s/2r) \cdot \frac{2f - 1}{1 - f(1 - \cos \varphi)} = \frac{(2f - 1) \sin(\varphi/2)}{1 - f(1 - \cos(\varphi))} \cdot \tan \chi.$$

### 3.4.2. Wing shape geometry

As mentioned in Section 3.3, the triangular shape considered above for monomers is actually modified by adjoining a thin stripe, displayed in Figure 14, to represent the backbone built by their concatenation.

Figure 15 illustrates the butterfly wing design, where the following elements are found:

- an inner circle, which delimits the backbone radial extent;
- the  $\varphi$ -sectors assigned to the butterfly, delimited by radial broken lines;
- the pairing axis and pairing vertex.

The previously considered triangles, indicated by dotted lines, are only displayed for illustrative purpose, viz. to show how the original triangle is actually turned into a pentagon. The concatenation direction is as follows: the backbone end that lies in the minor groove is the concatenation tail (the lower end in the figure), the one in the major groove (upper end in the figure) is the concatenation head.

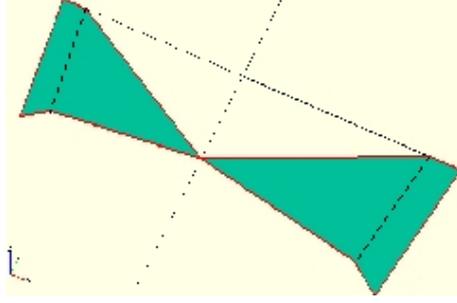


Figure 14: 2D projection of butterfly wings with trapezoidal backbone

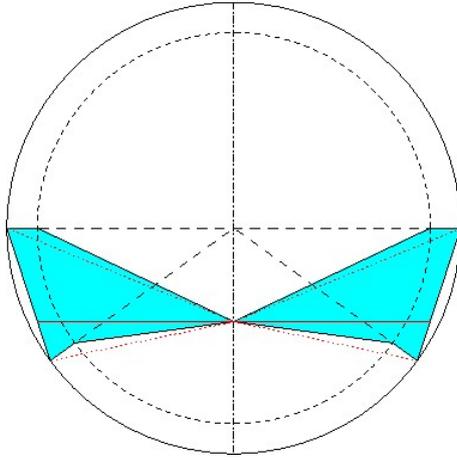


Figure 15: Butterfly 2D-shape design with pairing axis

The 3D wing design starts with that of its 2D projection on the horizontal plane. The pentagon on the right side of Figure 15 is formally represented as a 5-vertex sequence, in clockwise order starting from the pairing vertex, with each vertex specified by its coordinates. Let  $\beta r$  ( $\beta < 1$ ) be the radius of the inner circle in Figure 15; the 5-vertex sequence is:

$$((0, -fr \sin \varphi), (\beta r, 0), (r, 0), (r \cos \varphi, -r \sin \varphi), (\beta r \cos \varphi, -\beta r \sin \varphi))$$

The design of the 3D shape of butterfly wings requires consideration of how is it affected by the required axial rise and base pair tilt angles. Moreover, a twofold 3D carving of the 3D wing volume has to be designed, viz. to cut off a monomer base identifier and a corresponding pairing slot to host the pairing connector. The latter should enforce the required  $2\chi$  relative inclination of butterfly wings. This is consistent with the process of DNA template driven duplication, where monomers may bind to complementary ones in the template strand in arbitrary order, while concatenation in the new strand is performed afterwards; if pairing bonds would allow arbitrary relative inclination of the paired bases, then concatenated build-up of the new strand would be jeopardized.

A detailed analysis of the pairing connection design is carried out in Section 3.4.3, while here are a few details about the lifting of the pentagonal wing 2D-projection to a suitable polyhedron, according to the given values of  $\chi$  and  $\psi$ , together with its base identifier carving.

Basic CSG operations that help one to solve the 3D lifting problem are: linear extrusion, translation, rotation, intersection. The 2D pentagon on the left in Figure 15 is first translated, to get pairing along the  $x$  axis, with the intersection of pairing axis and backbone side at the origin. Figure 16 displays the intersection of a linear extrusion of the pentagon with an aptly placed disk. The disk inclination is visible in the left side picture, while the right side picture is a top view that shows the base identifier carving, which is made on the upper disk surface by subtraction of a linear extrusion of the 2D letter text.

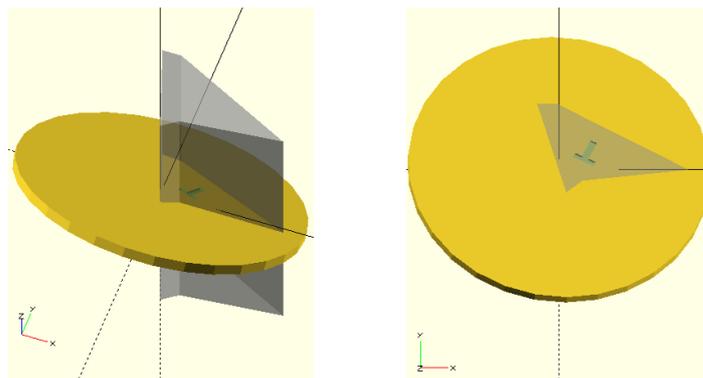


Figure 16: 3D wing shape design

Some care has to be taken with the composition order of the rotations applied to the disk cylinder. Let  $R_d(\rho)$  denote a rotation by angle  $\rho$  around axis  $d$ , the rotation applied to the disk with  $z$  symmetry axis is  $R_y(\psi) \circ R_x(\chi) \circ R_z(\varphi/2)$ . The rotation around the  $z$  axis is only needed because of the letter carving, in order to get it aligned with the (longest) side of the pentagon.

### 3.4.3. 3D wing pairing and antiparallelism

The 3D shape of a butterfly wing is independent of which side of the butterfly it is placed on: up to complementarity, either wing coincides with its paired one after an  $\omega$  rotation around the cylinder  $z$ -axis and a  $180^\circ$  rotation around the pairing axis, see Figure 18. Thus, the design problem for the pairing connection consists in devising a suitable cross-section of the pairing connection that would allow only one relative inclination between the paired wings, viz. by  $2\chi = 60^\circ$ .

A solution to this problem is illustrated in Figures 17 and 18.

Recall that the pairing connection is parallel to the theoretical pairing axis but translated a little away from it, for constructive reasons (explained in Section 4, with calculation details in Section 5). The connection cross-section, which is displayed on the left of Figure 17, is first designed as the extension of a regular hexagon with two right triangles that have their hypotenuses respectively coincident with two sides of the hexagon that are neither adjacent nor opposite to each other. This yields an irregular hexagon that has a single symmetry axis in between the triangular extensions.

Finally, reasons of ergonomics suggested to smooth either extension's corner by a small cut, parallel to the hypotenuse, since those corners correspond to connector's edges to be pushed by finger pressure when assembling a butterfly. This decision makes the hexagon extensions get trapezoidal shapes. The pairing connector fits into the wings to be paired after a  $90^\circ$  rotation of its cross-section, as displayed on the right side of Figure 17, where the wing tails are found at the right end.

The upper picture in Figure 18 displays a sample wing pairing, where the monomer letters A, T appear carved on vertically opposite faces of the paired wings, consistently with antiparallelism of their pairing. Please [take it into account](#) that the direction of the  $x$ -axis appears reversed in the picture since the view is in front of the major groove side of the paired wings. The connector part of the picture is magnified in the lower picture of Figure 18, where it may be appreciated how the connector's trapezoidal corner which

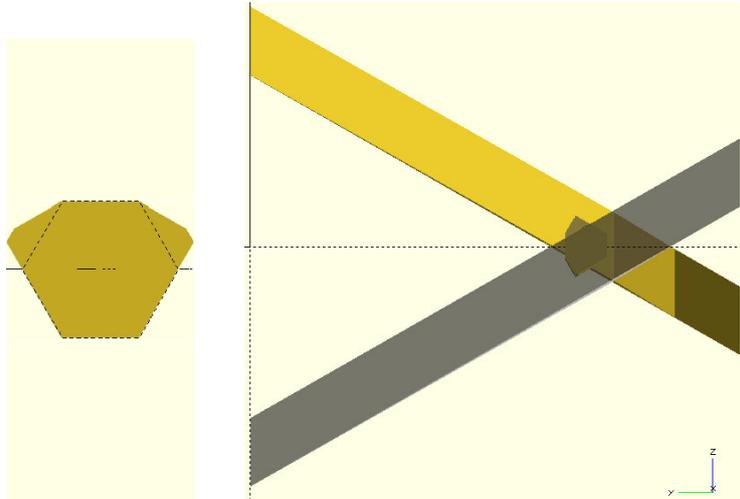


Figure 17: Wing pairing by unique fitting to connector. In either wing, one connector's trapezoidal corner comes *outside* of a wing face whilst the other one fits *inside* the wing, and the one which comes outside the face of either wing is the same trapezoidal corner that fits inside the other wing.

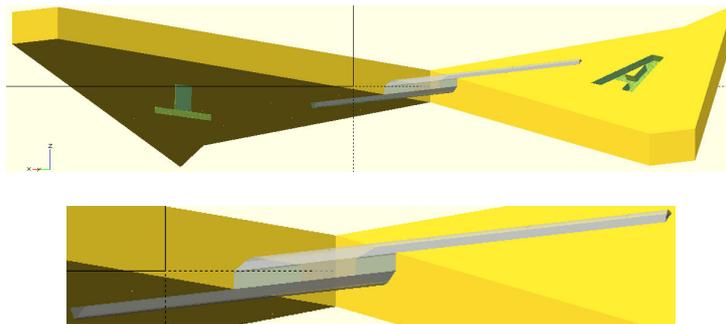


Figure 18: Wing pairing connection

comes outside of a face of either wing is the same which fits inside the other wing. This implies that, once the connector is pushed inside a wing, its fitting into the other wing is unique.

Please [note](#) that the view direction in Figure 17 is that of the pairing axis, which has a  $\psi$  inclination over the horizontal plane (the horizontal line in Figure 18); the corresponding rotation of the pairing connector leads the tail of either wing to the same vertical level as the head of the other wing, with the front wing (the transparent one in Figure 17, wing A in Figure 18) heading downwards and the back wing heading upwards. This fact materializes the antiparallelism of the DNA double helix in the visual model.

Figures 17 and 18 show that the axis of the pairing connector does not coincide with the geometrical pairing axis. This is also, perhaps more clearly apparent from the top view of paired wings with their connector shown in Figure 19, where a small translation of the pairing axis may be appreciated, along the  $y$ -direction in the major groove. This was required by a constructive implementation constraint that was briefly mentioned at the end of Section 3.2, viz. the convenience of avoiding very thin parts, which in this case would arise if the pairing slot should break the wing tip; see Section 4 for further motivation. Details about the calculation of the translation length are given in Section 5.

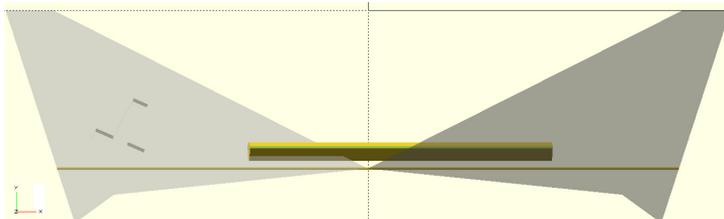


Figure 19: Parallel translation of pairing connector away from pairing axis

#### 3.4.4. Geometrical rendering of complementarity

Complementarity of DNA monomers relies on both geometric and chemical complementarity of their embedded base pairs (A, T) and (G, C) [27]. The chemical features of complementarity are not reflected in the visual model (with one exception, see below), nor so is the relative spatial positioning of the hydrogen bonds which tie the AT or CG base pairs. The only geometric feature that is chosen in the visual model to represent complementarity is the difference in radial length of purines (A, G) and pyrimidines (T, C), combined with the difference in molecular mass. This yields the length order  $l_G > l_A > l_T > l_C$ . Each monomer's base length is represented by the length of the slot, carved on its surface, that hosts the pairing connector, while pairing connectors are all of the same length. Therefore, letting the  $l$  function represent the slot length, complementarity of paired wings is represented by the condition  $l_A + l_T = l_G + l_C$ , together with the aforementioned order condition.

There are several ways of implementing these conditions. The solution in the visual model is formalized independently of the connector's length by assigning a smallest fraction  $\frac{m}{n}$  of its total in-slot length (see below) to the shortest base, and the complementary fraction  $\frac{n-m}{n}$  to the longest base. The choice for the positive integers  $m$  and  $n$  is guided by practical considerations, as follows. Given that the unit corresponds to a fraction of the total in-slot length, it proves convenient to let the unit correspond to the difference  $l_A - l_T$  as well as to  $l_G - l_A = l_T - l_C$ . Since  $n$  corresponds to  $l_A + l_T$ , it must be the sum of two consecutive integers, hence odd, and greater than 3 for it corresponds to the sum of the slot lengths of complementary bases. Setting  $n = 5$  proved inconvenient since the resulting smallest slot length  $l_C$  was too short to ensure a stable grip on the pairing connector, therefore an optimal value was sought for  $n = 2k + 5, k > 0$  to get a best compromise between stability of the pairing connector inside the shorter slot and sufficient discrimination between different slots. It took a little trial-and-error to fix  $k = 3$  as an optimal solution. Therefore, by posing  $m = 4, n = 11$ , and representing the base length order by the integers in  $[0, 3]$  in decreasing order, base types are mapped onto their length indices by function  $i_b$ , and the fraction of the total in-slot length assigned to base of type  $b$  is  $\frac{7 - i_b}{11}$ . In other words, if  $l_\Sigma = l_A + l_T = l_G + l_C$  denotes the total in-slot length, then by taking  $l_\Sigma/11$  as slot length unit, then the four bases in decreasing length order respectively get slots long 7, 6, 5, 4 units.

The in-slot length definition needs some care, since slot carvings end up on a slanted wing border, viz. the wing side on the major groove, see Figure 19. Conventionally, the slot ends at the intercept between wing side and pairing connector's axis. Section 5.2 gives details on a refined geometrical rendering of complementarity while taking the pairing connector translation into account.

## 4. Physical implementation issues

As it was mentioned in Section 3.2, the available 3D-printer and the choice of printing materials constrain the feasible design space. The available 3D-printing technology is known as Fused Filament Fabrication (FFF) as well as under the trademarked term Fused Deposition Modeling (FDM). In the printing process, one or more plastic filaments are heated and applied in layers via tiny nozzles, with a cooling fan lowering the plastic temperature soon after deposition, until an object emerges.

Only one filament may feed and be processed in a single-extruder, single-nozzle machine, like the one available for the present work. This feature is the source of the aforementioned upper bound on the overhang angle; as a matter of fact, when a second nozzle is available, then one may load the machine with a second *water-soluble* plastic filament, such as polyvinyl alcohol (PVA). Deposition of the latter builds the needed *support* structures that underlie those surfaces which exceed the overhang upper bound. The printed outcome is then immersed in water to get rid of the support.

The actual accuracy that may be achieved depends both on the machine and on the model local geometry. More accuracy means lower layer height, but then also longer printing time. In the present case, the smallest layer height available was 0.1 mm, but this was only used to print the wing pairing and concatenation connectors, whereas a 0.2 mm layer height proved sufficient to get acceptable wings printed in reasonable time.

The model local geometry determined the design decision to place wing pairing connectors a little away from the geometrical pairing axis, as explained in Section 3.3 and further detailed in Section 5.1 below. It is noteworthy, though, that the choice of the printing material also affects a high-level model parameter that relates to the connector-socket interface both for the pairing connections and for the concatenation connections. If a rigid material is made use of, then a small *positive* tolerance need be allowed at that interface, for actual holes tend to get a little smaller in the print than specified by the model. On the contrary, a small *negative* tolerance is adequate to ensure sufficient grip when flexible materials are employed. In the present case, an absolute value of 0.2 mm for the tolerance proves adequate in either case, with material-dependent sign, but it had to be set to -0.6 mm for the concatenation connectors when printed with flexible material. The tolerance is to be added to the appropriate dimension of the volume that is subtracted to make the corresponding hole. Please note that the tolerance is unrelated to the scale factor.

The prototype shown in Figure 11 was printed out of rigid polylactic acid (PLA), but the wing concatenation connections were not satisfactory (it was necessary to glue them). A better grip was obtained by printing the concatenation connectors with rigid PLA and the rest with a flexible polymer. The current version of the prototype is entirely printed out of flexible filaments, yet it is customizable for printing with rigid ones as well. A relevant parameter setting of the slicer (the software that generates the low-level code for the 3D-printer) is the *filling density* for the object's volumes: this was set to 40% for the (low volume) connectors, while to just 10% for the wings, which tend to bend the helix under their weight.

## 5. Mathematical details

Further mathematical details are presented in this Section, which is aimed at explaining how some of the previously introduced design decisions are actually effected in the 3D-printed model geometry, subject to constraints that arise from the aforementioned implementation issues.

### 5.1. Spatial position of wing pairing connector

As it was mentioned at the end of Section 3.4.3, constructive implementation constraints required a small translation of the pairing connector from the geometrical pairing axis position. The desired position is obtained by translation in the major groove along the y-direction, see Figure 19.

The calculation of the translation length is made in two steps, respectively indicated by the red and green arrows in Figure 20. The picture is drawn on the plane that is orthogonal to the pairing axis, with the pairing vertex G at the source point of the red arrow. The blue vertical line is the meeting line of the paired wing tips, whose tails are (not shown) on the right side. Wing thickness is relevant to the subject problem; the wing projections on the picture plane are the two regions, each delimited by a pair of parallel lines, whose intersection is the CPQR rhombus. To help relating the picture in Figure 20 to the previous pictures, consider that line CP represents the upper surface of downward heading wing A, on the right in Figure 18 and on the left in Figure 19, while line PQ represents the lower surface of its pairing mate wing, the upward heading one.

The yellow hexagon represents the section of the pairing connector, were it centered along the pairing axis (the trapezoidal extensions of the hexagon mentioned in Section 3.4.3 are not considered in this picture, since they are not relevant to the subject calculation).

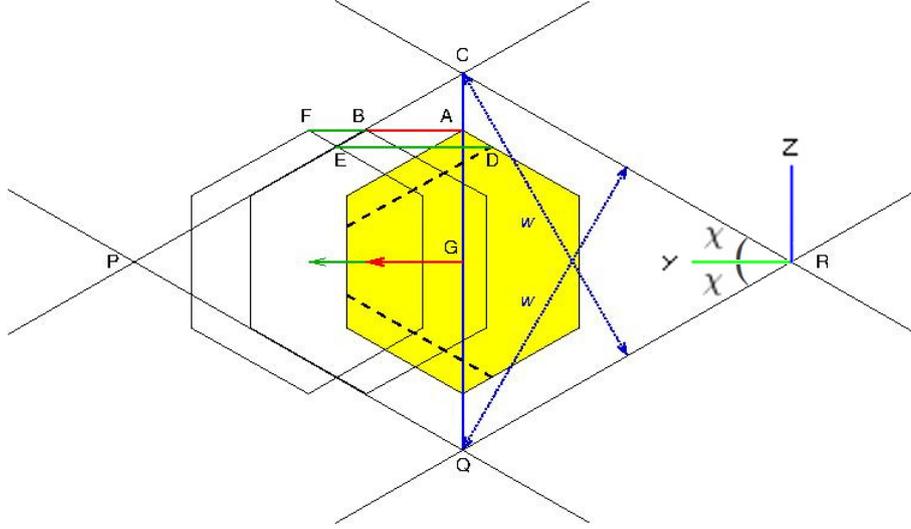


Figure 20: Length of the parallel translation of pairing connector

Now, as long as the translation keeps the hexagon inside the rhombus, then the pairing connector remains completely inside the paired wings, which fact is undesirable (according to the implementation constraint that sets the overhang upper bound, as explained in Section 4). The first calculation step is aimed at ruling out this possibility, by establishing a lower bound on the translation length. The calculated minimal translation makes two sides of the hexagon reach the rhombus border, meaning that two sides of the connector reach opposite faces of the paired wings, see the hexagon with top vertex B in Figure 20. However, as it may be seen in the picture, with the given ratio between the connector's thickness and the wing thickness, the minimal translation proves insufficient to let the pairing vertex fall outside the translated hexagon area. This means that the pairing slot to be carved on either wing would include the wing tip, which fact is undesirable, too (as argued in Section 4). An additional translation along the same direction may solve this issue.

The second design step consists in calculating the length of the additional translation, which is represented by the green arrow in Figure 20, such that 1) the pairing vertex falls outside the hexagon area, and 2) the connector's grip on either wing is not lost. The latter requirement is made precise by [noting](#) that the grip would be completely lost if the translated hexagon would only have three sides (or less) inside either wing.

In practice, a substantially major fraction of the translated hexagon area should fall inside either wing. To this purpose it is estimated that the grip would be sufficient if at least  $3/4$  of either side of the hexagon partially outside the wings crossing section (the rhombus in the picture) would be kept inside it.

In Figure 20, each of the two broken lines partitions the hexagon area into the part kept inside a corresponding wing and the part outside of it, after translation. From the previous estimate an upper bound on the additional translation length is calculated, whereby it turns out that it suffices to solve the aforementioned issue. Indeed, the pairing vertex G falls outside the area of the hexagon with top vertex F in Figure 20.

Here are the calculations introduced so far. Let  $w$  denote the wing thickness (dotted blue arrows in Figure 20) and  $s$  denote the length of the hexagon side, then with reference to Figure 20 the following hold:

- $|GC| = \frac{w}{2 \cos \chi}$
- $|AG| = s$
- $|AB| = \frac{|AC|}{\tan \chi} = \frac{w}{2 \sin \chi} - \frac{s}{\tan \chi}$



Figure 21 schematizes the horizontal projection of this arrangement, with U and Y the pairing slot ends for a purine and a pyrimidine wing, respectively. The following relations formalize the connector’s length partition just introduced.

Let  $t_y = |\text{PG}| = w - \frac{3\sqrt{3}}{4}s$  be the y-translation of the pairing connector away from the pairing vertex, as calculated in Section 5.1,  $|\text{OG}| = fr \sin \varphi$  the distance of the pairing vertex from the double helix axis as established in Section 3.4.1,  $\beta r$  the radius of the inner circle in Figure 15. In the visual model, the x-projection of the connector’s total in-slot length,  $c_{ix}$ , is related to that radius by posing  $c_{ix} = \frac{3}{4}\beta r$ . This ratio is empirically found to yield a good compromise between the conflicting goals of having slots long enough to ensure stability of the pairing connection while short enough to afford fairly easy execution of the manual (re-)assembly operations which are proposed in Section 6 below. Then the x-projection of the connector’s off-slot length,  $c_{ox} = |\text{UY}|$ , is calculated by triangle similarity as  $c_{ox} = 2|\text{PG}|\frac{\beta r}{|\text{OG}|} = \frac{2\beta r t_y}{f r \sin \varphi} = \frac{2\beta t_y}{f \sin \varphi}$ . By taking the tilt angle  $\psi$  into account, see Figure 13, the connector’s length is thus  $c = \frac{c_{ix} + c_{ox}}{\cos \psi}$ . The slot carving on the wing model of type  $b$  is made by subtracting the connector’s model, centered at P and with  $\psi$  inclination, from the uncarved wing model, after x-translation of the former by  $t_x = (\frac{7-i_b}{11} - \frac{1}{2}) \cdot c_{ix}$ , and z-translation  $t_z = t_x \tan(\psi)$ , taking the (negative) tilt angle into account.

## 6. Educational use of the present work

The approach to model design adopted in this work followed a motivation that springs out of an envisaged educational use of the intended artifact. Since the very beginning, the model was conceived as a composable structure, consisting of monomer and connector components, whereby monomers could be arranged to form any DNA double helix at will. This feature distinguishes the present artifact from other simple models such as the Origami DNA paper model [5]. The latter is indeed attractive, for it is easy to make, does not require any particular machinery, and is well-suited for educational activities with kids. However, the lack of composability limits its usability for educational experiments relating to major DNA functions such as template-driven duplication, DNA recombination and the like (see sample activities considered below).

At a high level of abstraction, these functions comprise what may be viewed as a sort of “mathematical machine of essential life”, and it does not seem impossible to present at least some of them even to kids, especially with the aid of concrete objects they could see, touch and play with. In this perspective, the machine is in their hands and they may freely connect the DNA building blocks at will, but the geometry of the building blocks forces pairing and concatenation connections to follow a specific pattern, not unlike a mathematical machine forces a point, a segment or a figure to follow a specific trajectory, according to a mathematically determined law [4]. The building blocks “as a system of auxiliary, external stimuli, raises the effectiveness of the child’s activity considerably.” [26].

At higher education levels, the mathematical model, together with its OpenSCAD source code may deserve consideration for classroom and/or lab activities. For example, Sections 3 and 5 may be used as a case study in learning geometry and trigonometry; see also a couple of statements in Section 3.4.1 whose proof is left to the reader as an exercise (solvable by using the trigonometry addition and duplication formulae). Furthermore, when a 3D-printing lab is available, those mathematical details are the basis for a thorough understanding of the source code, which may be 1) inspected with the present documentation at hand, 2) employed to produce the required artifact components and, even more exciting, 3) modified to explore different modeling geometries and solutions. Given the current rising interest in coding activities at high school level, it would not be bad if this work could promote math-based coding as an alternative to the still dominant imperative programming paradigm.

Educational use of the artifact in its physical materialization deserves attention as well. The current prototype is not fully satisfactory in all respects, whereby further experiments may be pursued to overcome its weaknesses. Section 4 provides 3D printing labs with some preliminary advice, out of this case study, for physical model construction and adaptation to specific features of the available 3D printing resources—printer and filament in the first place.

Finally, what use can be made of the physical artifact itself, say a bunch of monomer and connector components? Here are a few suggestions, in increasing order of complexity and difficulty—the latter only due to the aforementioned weaknesses of the current prototype, as we are going to point out. Intended use scenarios include such basic tasks as:

- *butterfly assembly*: pairing of complementary monomers, where some difficulty may arise because of possible inaccuracies of the connector and of the pairing slots;
- *single strand assembly*: concatenation of two or more monomers along a single helix, while paying attention to get the required relative angles; here a difficulty is likely to occur, due to insufficient grip of the concatenation connectors on the wings; with the current prototype, the best result is obtained by printing connectors with a rigid material while wings with a flexible one, and by setting a small negative tolerance for their coupling, yet wings tend to slide out of connectors, especially when assembled in fairly long strands, so, there is room for further improvement and experimentation;
- *double helix assembly*: concatenation of two or more butterflies, while paying attention to the required geometry and antiparallelism of the resulting double helix construction; this task presents the same difficulty as the previous one, but may prove easier than the next one;
- *DNA template-driven duplication*: here monomers get first paired to complementary ones in a template single strand, possibly but not necessarily in sequential order; the concatenations of the newly paired monomers are to be made, as long as the first task is completed with all pairings in place: keeping the pairings stable while performing this second part of the task is not necessarily an easy job, because of the aforementioned weaknesses of the current prototype;
- *DNA recombination*: splitting two distinct double helices where an identical restriction site occurs, with restriction leaving “sticky ends” (see [19] for plenty of examples), followed by cross-assembly of the resulting four pieces into a different pair of double helices. A twofold challenge gives this exercise the top difficulty rank: first, to remove the pairing connectors at the restriction sites, together with the concatenation connectors at the site cleavage ends without breaking other connections in any of the four specified pieces; second, to connect the sticky ends of these pieces in two pairs of them as specified, while keeping stability of their existing connections.

## 7. Conclusions

The first learning experience out of this work is undoubtedly to be found in what did the authors learn about relationships between function and structure of DNA. This process went on not only by speculation and discussions about the motivating question addressed in this paper and the structural principles thereafter presented in Section 2, but also through a very fruitful back-and-forth interaction between speculation and model construction. The latter was instrumental to clarify vague ideas, to uncover further principles, and to give logical rigour to intuitions about relationships between those principles. The “learning by doing” paradigm is confirmed by this experience as an actual *understanding by doing*, in full agreement with a well-known Confucius’ aphorism.

The current version of the model source code is freely available on the Thingiverse platform [22], together with further documentation aimed at facilitating both its customization for different model sizes, 3D-printers or printing materials and the execution of the tasks proposed in Section 6. It ought to be mentioned that progress was made during the writing of the present paper, to overcome the aforementioned weakness of concatenation connections, by endowing them with the insertion of spikes that prevent their loosening during (re)assembly manipulations. The curious reader is welcome to explore further details about this solution in the model website. By no means should this model version be considered as the final version, since its adequacy to the educational purposes outlined in Section 6 badly needs extensive testing. Valuable feedback from experience of using it may lead to revisions and improvements. Suggestions are [thus](#) warmly welcome.

The link between structure and function of DNA turned out to be expressed by logical and geometrical reasons, but reconstructing these reasons by means of an essential model that abstracts from biochemical details enables one to grasp the underlying principles in a more evident way, hardly recognizable in the complexity of biochemical reality. The power of abstraction is just in forgetting aspects that are irrelevant to the analysis of investigated phenomena. In physics, laws and principles emerge just from (experimentally validated) mathematical abstraction. The present work showcases the advantage of a similar attitude in biology. It is reasonable to expect that analogous approaches could be fruitfully developed in many biological contexts where complexity needs simplification in order to uncover the whole logic that is inherent to wide classes of phenomena.

## Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

- [1] Alberts, B. et al.: *Essential Cell Biology*, Garland Science (2014)
- [2] Annunziato, A.: DNA Packaging: Nucleosomes and Chromatin. *Nature Education* 1(1):26 (2008)
- [3] Barocci, S.: La struttura degli acidi nucleici e loro varianti (2014), retrieved 10 October 2020. [http://silviare.altervista.org/chimica/ORGANICA\\_files/2%20La\\_struttura\\_degli\\_acidi\\_nucleici\\_e\\_le\\_loro\\_varianti.pdf](http://silviare.altervista.org/chimica/ORGANICA_files/2%20La_struttura_degli_acidi_nucleici_e_le_loro_varianti.pdf).
- [4] Bartolini Bussi, M.G., Maschietto, M.: *Macchine matematiche: dalla storia alla scuola*, Springer-Verlag (2006).
- [5] Bateman, A.G.: Origami DNA, based on Thoki Yenn's design, <https://www.yourgenome.org/activities/origami-dna>.
- [6] Bonnici, V., Manca, V.: Informational laws of genome structures. *Scientific Reports*, 6, 28840 (2016)
- [7] [https://en.wikipedia.org/wiki/Constructive\\_solid\\_geometry](https://en.wikipedia.org/wiki/Constructive_solid_geometry)
- [8] Gilbert, W.: The RNA World, *Nature*, 319 (6055), 618 (1986)
- [9] Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., Bartel, D. P.: RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science*, 292 (5520), 1319–25 (2001)
- [10] Manca, V.: On the logic and geometry of bilinear forms, *Fundamenta Informaticae*, 64, 1-4, 261–73 (2005)
- [11] Manca, V.: *Infobiotics: information in biotic systems*, Springer (2013)
- [12] Manca, V.: The Principles of Informational Genomics, *Theoretical Computer Science* (2017)
- [13] Manca, V.: A marvelous accident. The birth of life, *Journal of Proteomics & Bioinformatics*, 11(7) 135–7 (2018)
- [14] Manca, V.: From biopolymer duplication to membrane duplication and beyond. *J. Membr. Comput.* 1, 292–303 (2019) Information theory in genome analysis. In Rozenberg G., Salomaa A., Sempere J., Zandron C. (eds) *Membrane Computing, CMC16*, 1–16. LNCS 9504, Springer (2015)
- [15] Manca, V., Franco, G.: Computing by polymerase chain reaction. *Mathematical Biosciences*, 211, 282–98 (2008)
- [16] <https://www.openscad.org>
- [17] Orgel, L.: Origin of life. A simpler nucleic acid, *Science*, 290 (5495), 1306–7 (2000)
- [18] Petitjean, M.: Chirality in metric spaces, *Optim. Lett.* 14 329–38 (2020). <https://doi.org/10.1007/s11590-017-1189-7>
- [19] Pingoud, A. and Jeltsch, A.: Structure and function of type II restriction endonucleases, *Nucleic Acids Research*, 29 (18), 3705–27 (2001), <https://doi.org/10.1093/nar/29.18.3705>.
- [20] Salibra, E.: *sulla via di Genoard*, p. 64, Manni (2007).
- [21] Schrödinger, E., *What Is Life? the Physical Aspect of the Living Cell and Mind*, Cambridge University Press (1944)
- [22] Scollo, G.: *DNA3D - A 3D-printable model of DNA*, <https://www.thingiverse.com/thing:4732625> (2021).
- [23] Sinden, R.R.: *DNA Structure and Function*, Academic Press (1994).
- [24] Stewart, F.J. and Raleigh, E.A.: Dependence of McrBC cleavage on distance between recognition elements, *Biol. Chem.*, 379, 611–6 (1998).
- [25] Thomson, W.: Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light, Appendix H, Sect. 22, Footnote p. 619. Cambridge University Press Warehouse, London (1904)
- [26] Vygotsky, L.S.: *Mind in Society, The Development of Higher Psychological Processes*, Harvard University Press (1978).
- [27] [https://en.wikipedia.org/wiki/Complementarity\\_\(molecular\\_biology\)](https://en.wikipedia.org/wiki/Complementarity_(molecular_biology))