



HAL
open science

The shape of DNA

Guillaume Devailly

► **To cite this version:**

| Guillaume Devailly. The shape of DNA. 2019. hal-02958372

HAL Id: hal-02958372


<https://hal.inrae.fr/hal-02958372>

Submitted on 5 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The shape of DNA

2019-07-23 ·  opinion (<https://gdevailly.netlify.com/categories/opinion/>), R
(<https://gdevailly.netlify.com/categories/r/>)



(<https://twitter.com/intent/tweet?text=The%20shape%20of%20DNA&url=https%3a%2f%2fgdevailly.netlify.com%2fpost%2fthe-shape-of-dna%2f>)



(<https://www.facebook.com/sharer.php?u=https%3a%2f%2fgdevailly.netlify.com%2fpost%2fthe-shape-of-dna%2f>)



(<https://www.linkedin.com/shareArticle?mini=true&url=https%3a%2f%2fgdevailly.netlify.com%2fpost%2fthe-shape-of-dna%2f&title=The%20shape%20of%20DNA>)



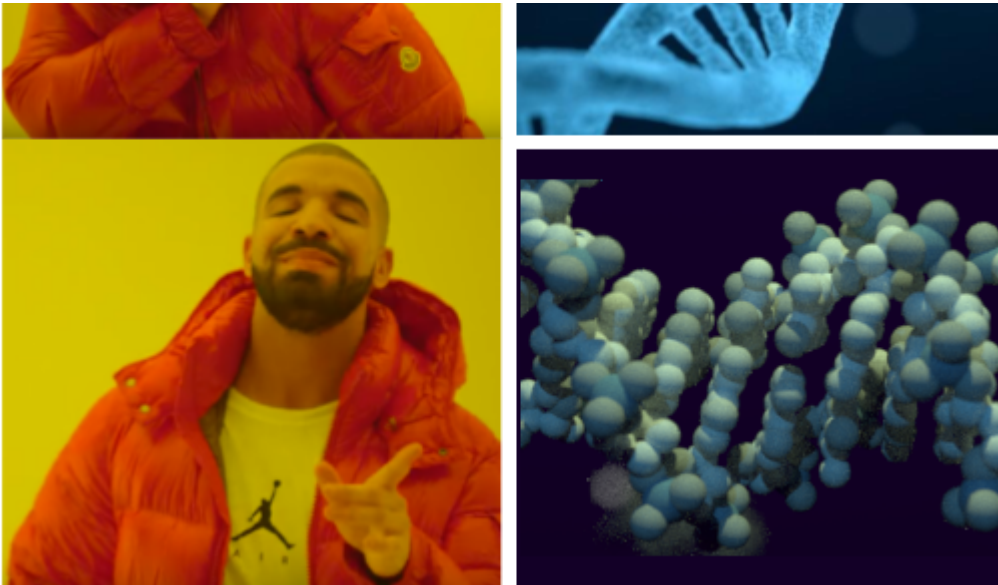
(<http://service.weibo.com/share/share.php?url=https%3a%2f%2fgdevailly.netlify.com%2fpost%2fthe-shape-of-dna%2f&title=The%20shape%20of%20DNA>)



(<mailto:?subject=The%20shape%20of%20DNA&body=https%3a%2f%2fgdevailly.netlify.com%2fpost%2fthe-shape-of-dna%2f>)

Graphical abstract:






Molecular biologists and biochemists are quite sensitive on the topic of DNA representation. Many of them will be triggered when the looping of the DNA is misrepresented as a left-handed helix instead of a right-handed one (what we mean by that is that when one looks at DNA from 'within' the helix, they turn clock-wise as they get more distant). I am not quite offended by this mistake, as I hardly remember if (B-)DNA is right-handed or left-handed, and what right-handed and left-handed means in the context of an helix.

I am paying more attention to the fact that the DNA double-helix has **two grooves of unequal size**, one big **major groove**, and a narrow **minor groove**. It matters because transcription factors can bind either to the major groove or to the minor groove, and the rules governing their sequence specificity won't be the same in each case.

I sometimes do complain about that:






Guillaume Devailly
@G_Devailly

Should someone tell them about the whole major/minor groove thing?

Ewan Birney @ewanbirney

The @emblebi Comms team has made this key credit card reminder for DNA handedness - you know you are in genomics Comms when you've got one of these

7:55 PM · Dec 13, 2018 

 3  See Guillaume Devailly's other Tweets

As others do:



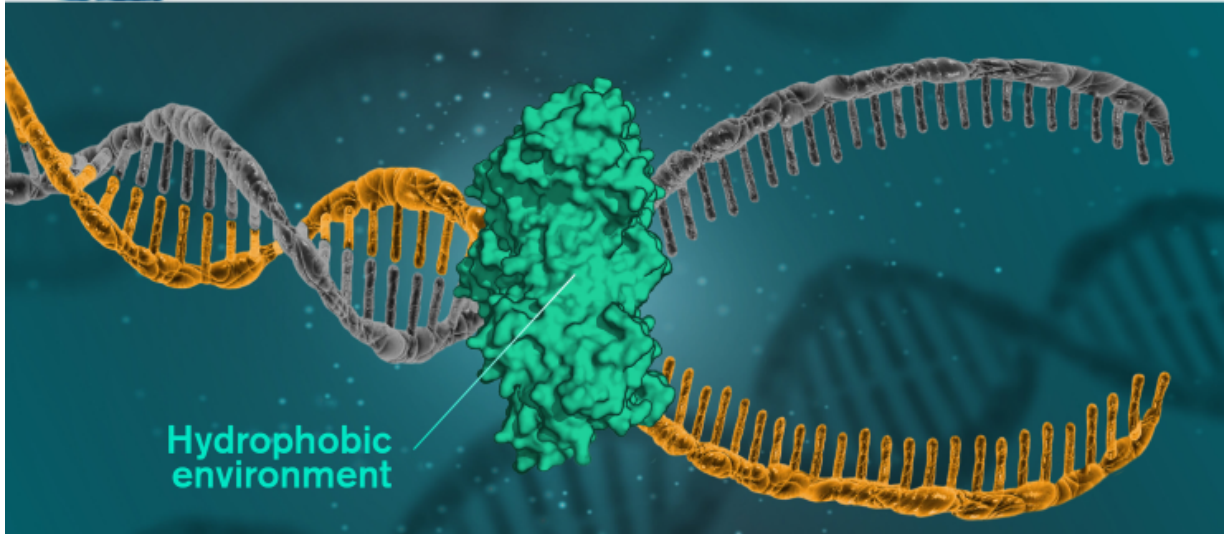
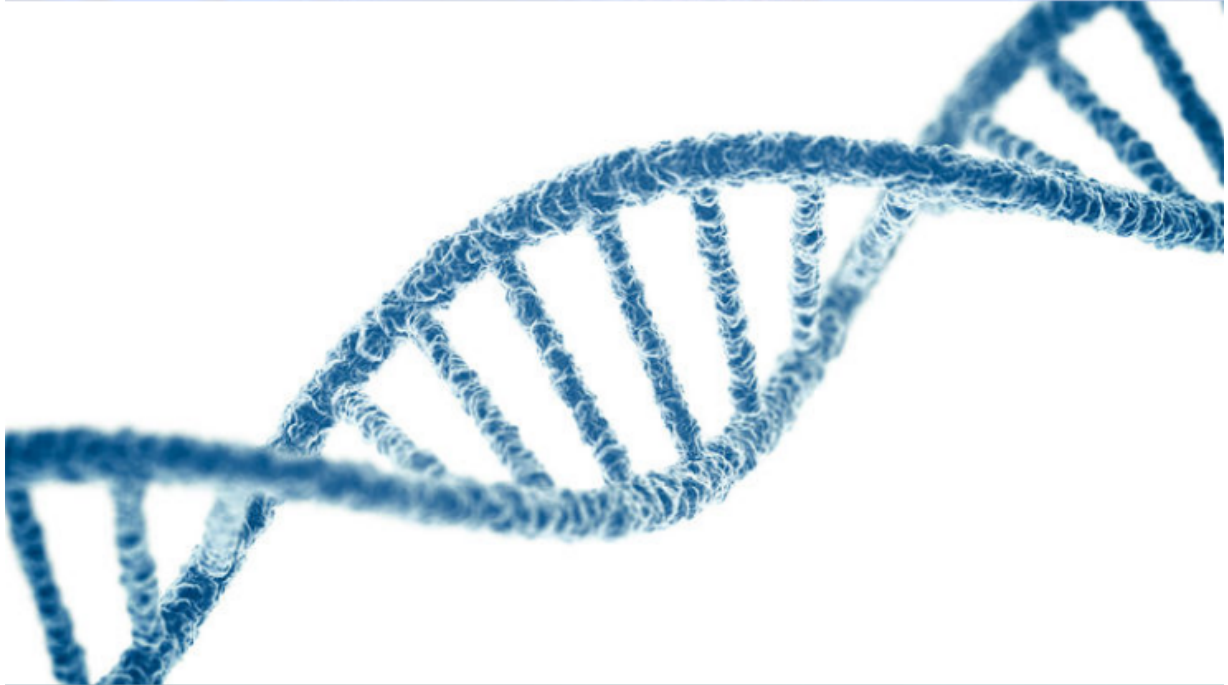
And guess what, complaining on twitter can sometime have an impact!



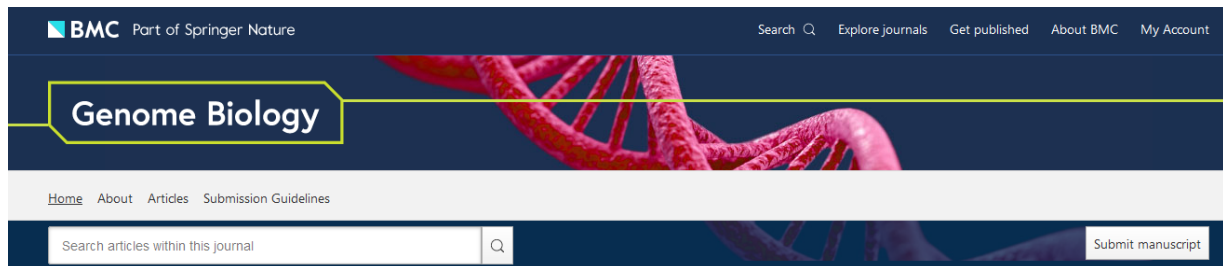
My other pet peeve is when artistic views of DNA show unnecessary “asperities” in the texture of DNA to “make it more real” **while still representing a very basic rope-ladder shape.**

For examples:





These visualisations are so popular that any efforts to name and shame will results in either naming the entire universe or hand picking some for biased reasons. I'll hand pick one nonetheless: the homepage of Genome Biology (<https://genomebiology.biomedcentral.com/>). ☹

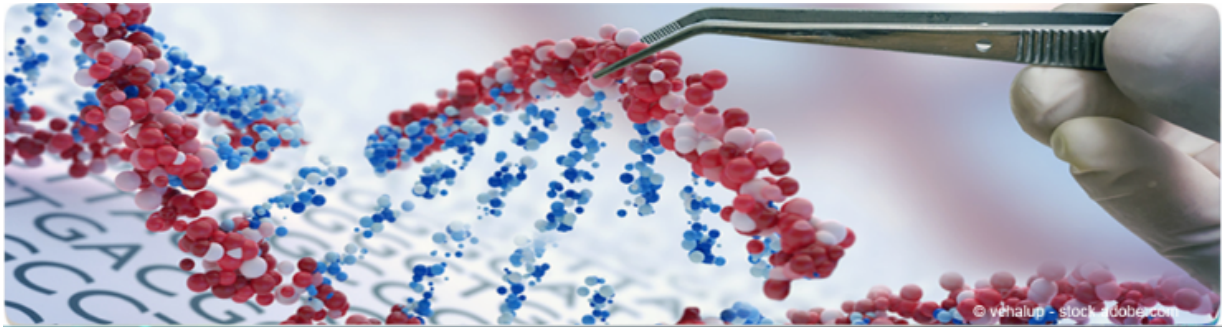


It seems that at some point, some artists began to realise that the DNA rope-ladder was actually made of atoms. They decided to represent atoms as spheres, which make sense. But instead of positioning the spheres on realistic positions, or by simply downloading available 3D models, they did that:





And the endless list of bad DNA visualisation goes on.



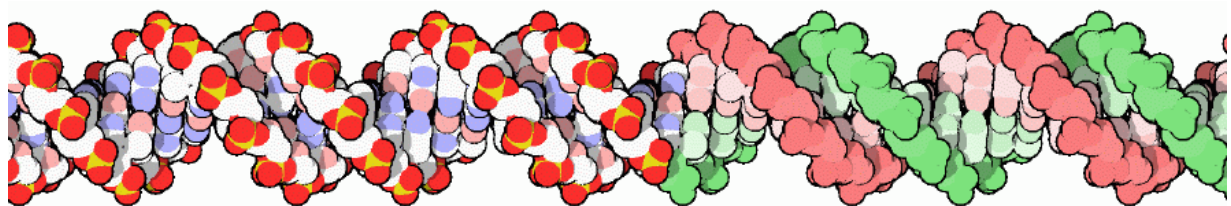


I find those view annoying, because we have known the exact structure of the B DNA helix for a long time, with many 3D models freely available! One could take those, apply a bumpy fluorescent blue texture in front of a black background and produce similarly impressive artistic views that would be **much closer to the reality**.

Let me try to be clearer:

- my rant is not against the *rope and ladder* view of DNA. It is a useful representation, but for diagram and such.
- it is not against artistic views, they are both useful and often spectacular when well done.
- it is about one **bad** type of artistic view that is representing DNA as a “rope ladder” while adding asperities to make it more *realistic*, ignoring that the atomic structures of DNA had been known since before I was born.

Thankfully, the brilliant PDB’s molecule of the month team (<http://pdb101.rcsb.org/motm/23>) offers us some hope. Let’s all admire the DNA molecule:



But enough complaining for today, let’s try to be a bit more constructive and build our own **scientifically accurate artistic view of DNA with R!**

First we download and parse the PDB structure 1BNA , which is a structure file of a 12 nucleotide long DNA double helix in B shape. Let’s use the CRAN package

(<https://cran.r-project.org/package=bio3d>) bio3d :

```
library(bio3d)

dna <- read.pdb("1BNA")
## Note: Accessing on-line PDB file

head(dna$atom)
## type eleno elety alt resid chain resno insert x y z o
## 1 ATOM 1 05' <NA> DC A 1 <NA> 18.935 34.195 25.617 1 64
## 2 ATOM 2 C5' <NA> DC A 1 <NA> 19.130 33.921 24.219 1 44
## 3 ATOM 3 C4' <NA> DC A 1 <NA> 19.961 32.668 24.100 1 31
## 4 ATOM 4 04' <NA> DC A 1 <NA> 19.360 31.583 24.852 1 37
## 5 ATOM 5 C3' <NA> DC A 1 <NA> 20.172 32.122 22.694 1 46
## 6 ATOM 6 03' <NA> DC A 1 <NA> 21.350 31.325 22.681 1 48
## segid elesy charge
## 1 <NA> 0 <NA>
## 2 <NA> C <NA>
## 3 <NA> C <NA>
## 4 <NA> 0 <NA>
## 5 <NA> C <NA>
## 6 <NA> 0 <NA>
```

We keep only the lines and columns that we will use, notably getting rid of a few water molecules.

```
library(dplyr)
##
## Attachement du package : 'dplyr'
## The following objects are masked from 'package:stats':
##
## filter, lag
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union

dnamod <- filter(dna$atom, type == "ATOM" ) %>%
  select(x, y, z, atom = "elesy")

head(dnamod)
## x y z atom
## 1 18.935 34.195 25.617 0
## 2 19.130 33.921 24.219 C
## 3 19.961 32.668 24.100 C
## 4 19.360 31.583 24.852 0
## 5 20.172 32.122 22.694 C
## 6 21.350 31.325 22.681 0
```

We use the `rayrender` package (<https://github.com/tylermorganwall/rayrender>) to hand-craft our scientifically accurate artistic view of a DNA molecule. It is a quite new ray-tracing tool, not particularly designed to render macro-molecules. This means that we have to do some dirty work ourselves.

First we centre the molecule coordinates, to be able to point the virtual camera to the centre of the DNA molecule by simply targeting the XYZ point: $c(\theta, \theta, \theta)$.

```
summarise_if(dnamod, is.numeric, min)
##           x           y           z
## 1  4.025  8.032 -11.401

summarise_if(dnamod, is.numeric, max)
##           x           y           z
## 1 26.506 34.195 31.084

dnamod <- mutate(dnamod, x = x - mean(x), y = y - mean(y), z = z - mean(z))

summarise_if(dnamod, is.numeric, min)
##           x           y           z
## 1 -10.69357 -12.94741 -20.2247

summarise_if(dnamod, is.numeric, max)
##           x           y           z
## 1 11.78743 13.21559 22.2603
```

We create a look-up table containing some details about each atom in the molecule. We use some standard, non-artistic, colours first:

```
unique(dnamod$atom)
## [1] "O" "C" "N" "P"

elref <- tribble(
  ~atom, ~atomic_mass, ~color,
  "C", 12, "lightgrey",
  "N", 14, "darkblue",
  "O", 16, "darkred",
  "P", 31, "darkorange"
)
```

Now let's build our model. `Rayrender`'s function `sphere()` is not entirely vectorised. The `x`, `y`, `z` parameters seem vectorised: you can render several spheres with one function call by passing vectors in `x`, `y`, `z` parameters. But as far as I understand, they have to share the non vectorised `material` parameter. Therefore we call

`sphere()` 4 times, one for the carbon atoms, one for the oxygen, one for the nitrogen, and one for the phosphorus, using `purrr::map_dfr()` to be fancy:

```
library(rayrender)
library(purrr)

mol <- map_dfr(
  unique(dnamod$atom), # one iteration for each tom type, C, N, O, P
  function(ato) {
    melref = filter(elref, atom == ato)
    dnamo = filter(dnamod, atom == ato)
    sphere(
      x = dnamo$x, y = dnamo$y, z = dnamo$z,
      radius = 0.4 * melref$atomic_mass^(1/3),
      # radius is proportionnal to the cubic root of the atomic mass.
      # 0.4 is a magic number leading to a visualisation that doesn't look like
      material = diffuse(color = melref$color)
    )
  }
)
```

We now render our first scene. Rayrender is basically ray-tracing on CPU, which means it is quite slow. So it's best to render small scenes first to find the appropriate parameters, before rendering a beautiful one once you are happy with the scene.

A scene is composed of objects (here described in the `mol` tibble), and a camera that has a location, and point to something. We centred our object, so we tell the camera to point at `c(0, 0, 0)`. It's probably best if the camera is somewhere outside of the object. I played a bit with the `lookfrom` parameter, and found the values `c(60, 0, -10)` reasonably satisfying. So behold the following rendering:

```
render_scene(
  mol,
  lookfrom = c(60, 0, -10),
  lookat = c(0, 0, 0),
  parallel = TRUE, width = 800, height = 500, samples = 100,
  backgroundhigh = "white", backgroundlow = "white"
)
```



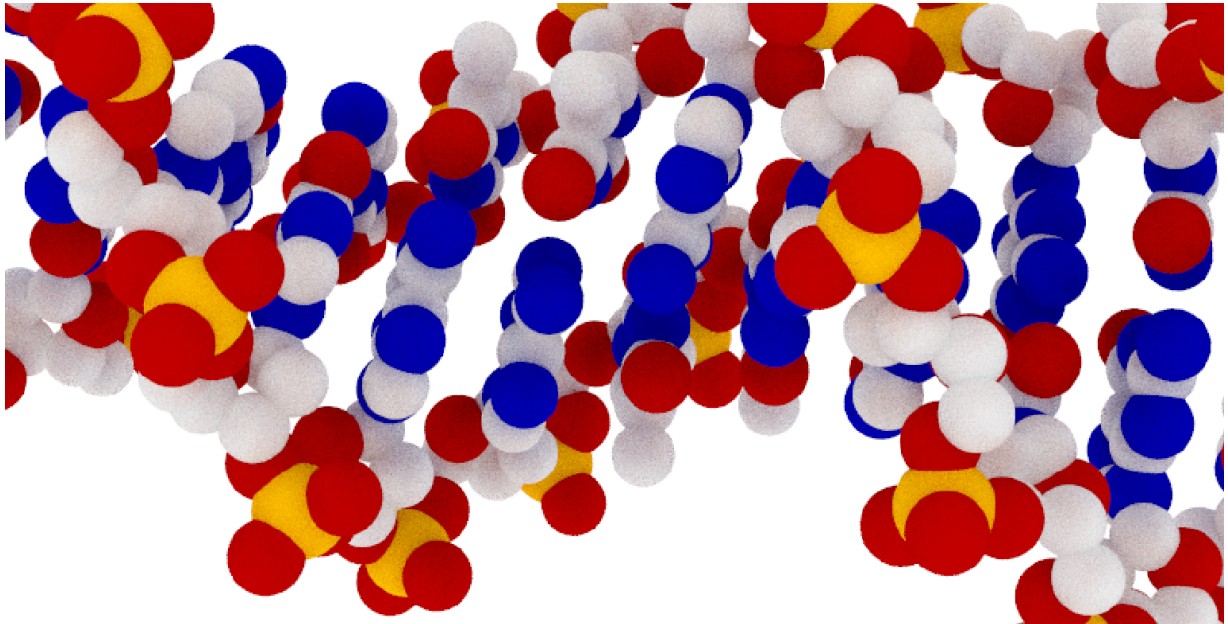


Figure 1: Scientifically accurate, non artistic view of DNA.

Now let's try to render this non-artistic, a bit boring, view of DNA into an artistic one. I'm no artist, but let's put a darker background, two light sources of different colours, and an invisible metallic mirror in the background to generate some dirty reflections to make the scene less empty. Let's also pick fairly neutral colours for the atoms, and fiddle with the camera aperture and focal distance .

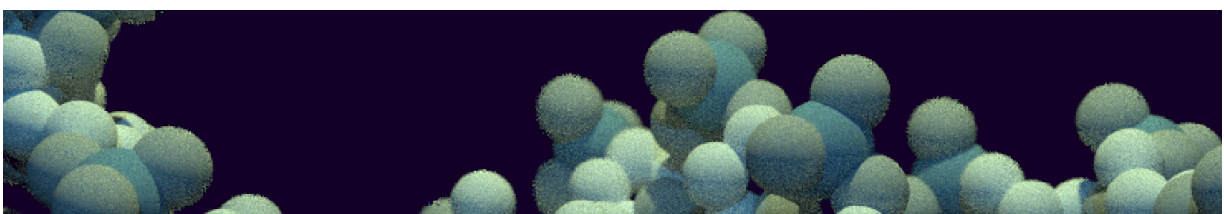

```

elref <- tribble(
  ~atom, ~atomic_mass, ~color,
  "C", 12, "azure1",
  "N", 14, "azure3",
  "O", 16, "azure4",
  "P", 31, "steelblue4"
)

mol <- map_dfr(
  unique(dnamod$atom),
  function(ato) {
    melref = filter(elref, atom == ato)
    dnamo = filter(dnamod, atom == ato)
    sphere(
      x = dnamo$x, y = dnamo$y, z = dnamo$z,
      radius = 0.4 * (melref$atomic_mass)^(1/3),
      material = diffuse(color = melref$color)
    )
  }
) %>%
  # first light, magic numbers
  add_object(sphere(
    x = 120, y = 50, z = -10, material = light("dodgerblue", intensity =
  )) %>%
  # second light, magic numbers
  add_object(sphere(
    x = 20, y = 70, z = -10, material = light("darkolivegreen1", intensit
  )) %>%
  # a big spheric mirror somewhere on the back
  add_object(sphere(
    x = -450, y = -100, z = 160, radius = 400, material = metal(fuzz = 0.
  ))

render_scene(
  mol,
  lookfrom = c(60, 0, -10),
  lookat = c(0, 0, 0),
  aperture=0.8, focal_distance = 55, clamp_value = 5,
  ambient_light = TRUE, backgroundhigh = "#010005", backgroundlow = "black"
  parallel = TRUE, width = 800, height = 500, samples = 600
)

```



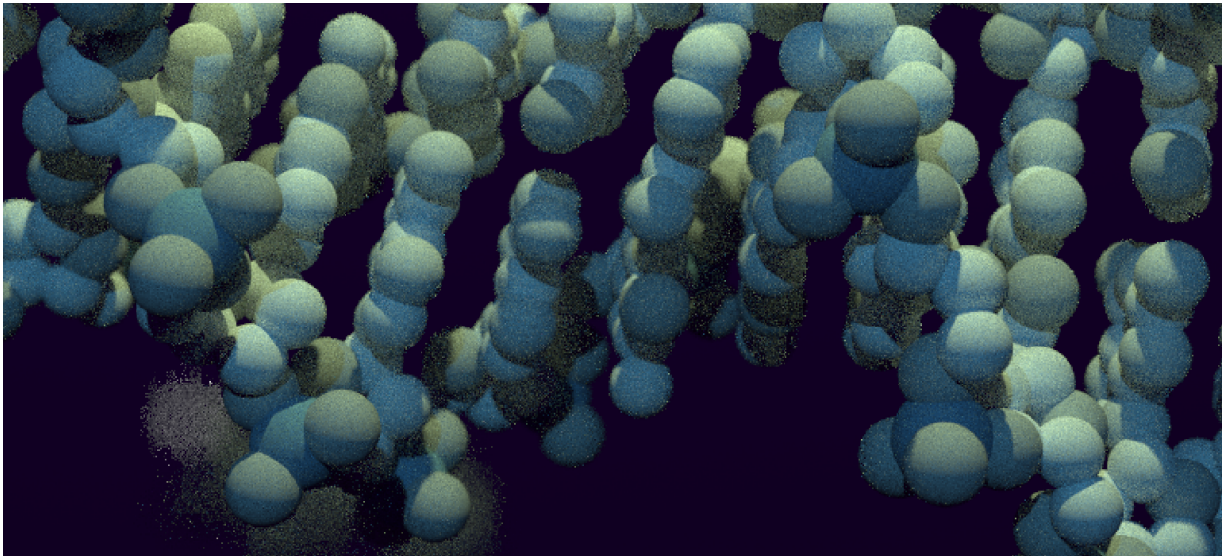


Figure 2: Scientifically accurate, totally artistic view of DNA.

Not bad! My main regret is that the PDB DNA structure 1BNA is only 12 nucleotide long. A longer DNA model might be more suitable for artistic representations, but I was too lazy to search for one or build one myself (too bad for you, 16-nucleotide long structure 3BSE (<https://www.rcsb.org/structure/3BSE>)).

To conclude, I'll offer a stay-at-home challenge: try to render your own scientifically accurate artistic view of the DNA molecule, for example by changing a few parameters from this post's script. Then post it on twitter mentioning me (G_Devailly (https://twitter.com/G_Devailly)), the winner will be retweeted! 🎉

Thanks to schneu for kindly providing some proofreading.

R session details:

```
sessionInfo()
## R version 3.6.3 (2020-02-29)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 18362)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=French_France.1252 LC_CTYPE=French_France.1252
## [3] LC_MONETARY=French_France.1252 LC_NUMERIC=C
## [5] LC_TIME=French_France.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] purrr_0.3.3      rayrender_0.9.0 dplyr_0.8.5      bio3d_2.4-1
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.4      highr_0.8        compiler_3.6.3   pillar_1.4.3
## [5] tools_3.6.3     digest_0.6.25    lattice_0.20-38  lubridate_1.7.8
## [9] evaluate_0.14   tibble_3.0.0     lifecycle_0.2.0  pkgconfig_2.0.3
## [13] rlang_0.4.5.9000 cli_2.0.2        rstudioapi_0.11  yaml_2.2.1
## [17] parallel_3.6.3  blogdown_0.18    xfun_0.12        emo_0.0.0.9000
## [21] stringr_1.4.0   knitr_1.28       raster_3.0-12    generics_0.0.2
## [25] vctrs_0.2.99.9010 grid_3.6.3       tidyselect_1.0.0 glue_1.4.0
## [29] R6_2.4.1        fansi_0.4.1      rmarkdown_2.1    bookdown_0.18
## [33] sp_1.4-1        magrittr_1.5     codetools_0.2-16 htmltools_0.4.6
## [37] ellipsis_0.3.0  assertthat_0.2.1 rayimage_0.2.5   stringi_1.4.6
## [41] crayon_1.3.4
```

DNA (<https://gdevailly.netlify.com/tags/dna/>)

rayrender (<https://gdevailly.netlify.com/tags/rayrender/>)

rant (<https://gdevailly.netlify.com/tags/rant/>)

R (<https://gdevailly.netlify.com/tags/r/>)

Related

- [How to shuffle ties after a sort? \(/post/how-to-shuffle-ties-after-a-sort/\)](/post/how-to-shuffle-ties-after-a-sort/)

© 2018 · Powered by the Academic theme (<https://sourcethemes.com/academic/>) for Hugo (<https://gohugo.io>), used through Blogdown (<https://bookdown.org/yihui/blogdown/>), hosted by Github (<https://github.com/gdevailly/siteperso>) and Netlify (<https://www.netlify.com/>).

