Look-and-Say Biochemistry: Exponential RNA and Multistranded DNA

Óscar Martín

1 BACKGROUND AND INTRODUCTION.

Look-and-Say sequences (*LS-sequences*, for short) are built by beginning with an arbitrary string of numerical digits as a zeroth term, or *seed*, and then describing the present term to get the next. For example, take 1 as seed (to generate the *standard* LS-sequence). We can describe it as "one one," so the first term is 11. This is "two ones," so the next term is 21. Now we have "one two, one one," and so we get 1211. These are the first terms in the standard LS-sequence:

```
    1
   11
   21
  1211
 111221
 312211
13112221
1113213211
   ...
```

Technically, we must assume a base as large as needed for our numeral system. For instance, a string of n consecutive 1s is described as n1, and this must be considered a two-digit string, no matter how large n is.

John H. Conway was the first to study LS-sequences in depth. Virtually every result known today is included in his paper [1]. For the convenience of the reader, the rest of this section summarizes Conway's findings. Proofs are omitted.

As in [6], we denote the *description function* by α; that is, if the seed is s, the first term is $\alpha(s)$, and the kth is $\alpha^k(s)$. We set $\alpha^0(s) = s$. We also say that $\alpha^k(s)$ is a *descendant* of s.

The key observation is that some strings split into two or more substrings, in the sense that these substrings do not interfere with each other in later terms of the sequence. For example, take the string 1511. It splits as 15·11 because the descendants of the left half always end in 5, while the right half never starts with this digit (as we will shortly see). Thus, if s splits as $s = lr$, then $\alpha^k(s) = \alpha^k(l)\alpha^k(r)$ for every k. Indeed, $\alpha(\mathtt{1511}) = \mathtt{111521} = \alpha(\mathtt{15})\alpha(\mathtt{11})$, etc.

We signify the repetition of digits with superscripts in the usual way, so that, for example, $\alpha(\mathtt{1511}) = \alpha(\mathtt{151}^2) = \mathtt{111521} = \mathtt{1}^3\mathtt{521}$. (Note, however, that the superscript in $\alpha^k(s)$ stands for iteration, not repetition.) Superscripts are always maximal. For instance, when $\mathtt{x}^1\mathtt{y}^2$ is written, it is understood that $\mathtt{x} \neq \mathtt{y}$.

The following are Conway's first easy results. When $t = \alpha^k(s)$ for some string s, Conway says that t is k days old.

The One-Day Theorem. *No one-day-old string contains substrings of the following types:*

- $\mathtt{yxzx}$ *beginning in an odd position of the string*
- $\mathtt{x}^k$ *with* $k \geq 4$
- $\mathtt{x}^3\mathtt{y}^3$

About the first of these three types of substrings, note that if $\mathtt{yxzx}$ begins in an odd position, then it describes the string $\mathtt{x}^y\mathtt{x}^z$, which should instead be written as $\mathtt{x}^{y+z}$. The two other types of substrings, whether they begin in an odd or an even position, are special cases of the first type. Similar reasoning gives us information about two-day-old strings:

The Two-Day Theorem. *No digit* $\mathsf{4}$ *or larger can be born on the second day or later. Also, a two-day-old string cannot contain any substring of the form* $\mathtt{3x3}$ *(or, in particular,* $\mathsf{3}^3$*).*

The next result is only a little more difficult than the previous ones, but it does require a careful case study.

The Starting Theorem. *Let* s *be a string that is at least two days old. If* $s = \mathsf{22}$*, then* $\alpha^k(s) = \mathsf{22}$ *for every* k*. In every other case the initial digits of the descendants of* s *ultimately cycle in one of two ways:*

$$\begin{tikzcd}
& \mathsf{1}^3\ldots \arrow[dr] & \\
\mathsf{1}^1\mathtt{x}^1\ldots \arrow[ur] & & \mathsf{3}^1\mathtt{x}^{[\neq 3]}\ldots \arrow[ll]
\end{tikzcd}
\qquad
\begin{tikzcd}
& \mathsf{2}^2\mathsf{1}^3\ldots \arrow[dr] & \\
\mathsf{2}^2\mathsf{1}^1\mathtt{x}^1\ldots \arrow[ur] & & \mathsf{2}^2\mathsf{3}^1\mathtt{x}^{[\neq 3]}\ldots \arrow[ll]
\end{tikzcd}$$

Observe that the cycle to the right is the same as the one to the left with extra digits $\mathsf{22}$ prepended (that is, added to the left). We adopt the convention that, in order to represent any digit or superscript that has a special property, we enclose that property in brackets. For example, $[\geq\mathsf{4}]$ represents any digit greater than or equal to $\mathsf{4}$. Three ellipsis points signify that the string may or may not have more digits in that direction.

On the other hand, the final digit of a string does not change in subsequent terms; that is, the last digit of $\alpha^k(s)$ is the same for any k. With this in mind, we can state the conditions for a split to occur.

The Splitting Theorem. *A string* s *at least two days old splits if and only if it is one of the following ten types (the centered dots mark the splits):*

$$\ldots[\geq\mathsf{4}]\cdot[\leq\mathsf{3}]\ldots$$
$$\ldots\mathsf{2}\cdot\mathsf{1}^1\mathtt{x}^1\ldots$$
$$\ldots\mathsf{2}\cdot\mathsf{1}^3\ldots$$
$$\ldots\mathsf{2}\cdot\mathsf{3}^1\mathtt{x}^{[\neq 3]}\ldots$$
$$\ldots\mathsf{2}\cdot[\geq\mathsf{4}]^1\ldots$$
$$\ldots[\neq\mathsf{2}]\cdot\mathsf{2}^2$$
$$\ldots[\neq\mathsf{2}]\cdot\mathsf{2}^2\cdot\mathsf{1}^1\mathtt{x}^1\ldots$$
$$\ldots[\neq\mathsf{2}]\cdot\mathsf{2}^2\cdot\mathsf{1}^3\ldots$$
$$\ldots[\neq\mathsf{2}]\cdot\mathsf{2}^2\cdot\mathsf{3}^1\mathtt{x}^{[\neq 3]}\ldots$$
$$\ldots[\neq\mathsf{2}]\cdot\mathsf{2}^2\cdot[\geq\mathsf{4}]^1\ldots$$

Since we have stated the starting theorem, an "ending theorem" would seem to be in order. Here it is.

The Ending Theorem. *The endings of any LS-sequence ultimately cycle in one of three ways:*

$$\begin{tikzcd}
\ldots 2^2 \arrow[loop above] & \ldots 2\cdot 31221132221222112112322211[>1] \arrow[d, shift left] \\
& \ldots 2\cdot 1311222113321132211221121332211[>1] \arrow[u, shift left]
\end{tikzcd}$$

$$\begin{tikzcd}
\ldots 2\cdot 311322113212221 \arrow[r] & \ldots 2\cdot 13211322211312113211 \arrow[d] \\
\ldots 2\cdot 12322211331222113113112211 \arrow[u] & \ldots 2\cdot 1113122113322113111221131221 \arrow[l]
\end{tikzcd}$$

(Splits are marked, though they are unimportant here.)

A string, or substring, that cannot be split is called an *atom* or an *element*. Ninety-two of these elements are very special in that they appear in any LS-sequence (I will be more precise about this point shortly). In analogy with the ninety-two naturally-occurring chemical elements, Conway called these *common elements* and gave them the corresponding atomic numbers, symbols, and names, from hydrogen (H_1) to uranium (U_{92}). (As a matter of fact, a few of the first ninety-two real-world chemical elements do not occur in nature.) We often accompany a chemical symbol with its corresponding atomic number as subscript. This is nonstandard notation, but is useful in our context. For instance, H_1=22 and Mn_{25}=111311222112. The "periodic table" containing the relevant information about the common elements can be found in appendix A. The reader will most likely want to consult it as he or she reads on.

There are two other important elements in the periodic table:

- neptunium: Np_{93}=1311222113321132211221121332211[$\geq$4]
- plutonium: Pu_{94}=31221132221222112112322211[$\geq$4]

These have an infinity of isotopes—one for each digit that can appear in the last place. These are called *transuranic elements*. Note that digits greater than 3 can appear only at the end of transuranic elements, and not at all in common elements. An element that is neither common nor transuranic is called *exotic*. For example, 1, 2, and 23 are exotic, as is 3333 or any other element with four or more consecutive equal digits.

Now, according to the information in the periodic table, the description of a common element is a common element or a compound (that is, a concatenation) of common elements. For instance,

- $\alpha(\mathrm{U}_{92}) = \alpha(3) = 13 = \mathrm{Pa}_{91}$
- $\alpha(\mathrm{Rn}_{86}) = \alpha(311311222113) = 1321132\cdot 1322113 = \mathrm{Ho}_{67}\mathrm{At}_{85}$

This observation leads us to the first exciting result.

The Chemical Theorem.

1. *When a seed is a common element, all its descendants are compounds of common elements.*

2. *When a seed is a common element other than the stable* H_1=22*, any sufficiently late descendant is a compound of all the common elements simultaneously.*

3. *For any seed other than* H_1 *any sufficiently late descendant includes all the common elements simultaneously (and possibly other elements).*

4. *If a term of an LS-sequence includes a transuranic element, its description (that is, the next term of the sequence) includes the same isotope of the other transuranic element.*

Now we come to what, according to [1], is "the finest achievement" of this theory.

The Cosmological Theorem. *For any string s and for k satisfying $k \geq 24$, $\alpha^k(s)$ is a compound of only common and transuranic elements.*

Or, as Conway puts it, all exotic elements have disappeared as of day twenty-four after the Big Bang.

The proof of this result seems very difficult. Indeed, the first purported, complete published proof has appeared only recently in [6] and [5]. Mike Guy found an exotic element that needs the twenty-four days stated in the theorem, thus showing that this bound is best possible. He called the element Methuselum. Its string of digits is

$$\texttt{2233322211}[\geq 4],$$

with an isotope for each choice of the last digit.

Finally, linear algebra is needed to prove the following:

The Arithmetical Theorem.

1. *The asymptotic rate of growth of the terms of an LS-sequence is independent of the seed. It is a constant λ ("Conway's constant") whose value is approximately* 1.303577. *Formally, if $|s|$ is the length (number of digits) of s, then*

$$\frac{|\alpha^{k+1}(s)|}{|\alpha^k(s)|} \to \lambda$$

as $k \to \infty$. The constant λ cannot be expressed by radicals. It is the largest root of a polynomial of degree seventy-one that is irreducible over the rationals.

2. *All the common elements occur in each LS-sequence and do so in definite asymptotic proportions that are independent of the seed.*

This completes the exposition of Conway's "weird and wonderful chemistry." My aim in this article is to extend it by investigating self-descriptive and cyclically descriptive strings, that is, strings s such that $\alpha(s) = s$ or $\alpha^k(s) = s$ with $k \geq 1$, respectively.

The sad truth is that the only self-descriptive string is 22, the hydrogen atom, and there are no sets of cyclically descriptive strings. In order to check this, try to build such strings starting with any two initial digits. For instance, a self-descriptive string whose first two digits were 14 would have to start with "one 4." Thus, it is impossible. Neither can a string beginning with 14 be in a cyclically descriptive set. If it were, then another string of the set would begin with 4x, and yet another with xxxx, which is not a valid description.

We can find some partially self-descriptive strings like $2\underline{2211322}$ and $\underline{223332}22$. The first describes its underlined substring: $\alpha(2211322) = 22211322$; the second is described by its underlined substring: $\alpha(22333222) = 223332$. None of them is especially interesting.

This is slightly disappointing. Nevertheless, observe that in the previous discussion we have tacitly assumed that strings were finite. If we allow for infinite strings, interesting things happen that lead us to a biochemistry based on Conway's chemistry. The objective of this paper is to explore the biomolecules that arise.

2 RNA.

The same argument, by case study, that shows the impossibility of building finite self-descriptive strings (except H_1) also proves that no semi-infinite string, either to the right or to the left, can be self-descriptive. In order to find something interesting, we need to look for partially self-descriptive strings. First, we state the last negative result of this paper.

The Negative Theorem. *No semi-infinite string describes a substring of itself. That is, there are no strings s and t such that s is finite, t is semi-infinite, and $\alpha(t) = st$ or $\alpha(t) = ts$.*

Proof. Suppose, for instance, that there are strings s and t such that $\alpha(t) = st$ (i.e., st describes t). Then s is all we need to know, for we can write down s and then, aware that it is a description of something, can append to its right what it describes—the beginning of t. As shown in Figure 1, we imagine the job being done by two people: the reader 👁, who reads the description (st), and the writer ✎, who writes whatever 👁 is describing (namely, t). Both 👁 and ✎ are shown in their initial positions. When s has been read 👁 continues to read whatever ✎ has already written.

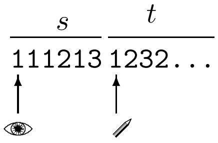


Figure 1: Reader and writer in their initial positions.

We want to compare the speeds of movement of 👁 and ✎. We assume that both work in steps. In each step 👁 reads two digits (11, for example) and ✎ writes a *run* of equal digits described by 👁 (1, in this example). First, note that because st is supposed to be a description of something, the one-day theorem applies to it. In particular, there can be no run of four or more equal digits in st. Accordingly, no digit greater than 3 will appear in it. Also, runs of three equal digits cannot begin in even positions. We use the terms 1-*run*, 2-*run*, and 3-*run* for runs of the specified lengths.

We conclude that 👁 always advances two positions, while ✎ advances one, two, or three positions. It follows that at each step 👁 can gain or lose at most one position. Thus, in the event that 👁 happens to catch up with ✎ at some point, there must be a step in which both are working on the same position. Whatever remained to the right would be a self-descriptive string, which we know cannot exist. This means that the only way string st can exist is if 👁 never catches ✎. It seems unlikely that this could actually happen, because a description usually moves λ times faster than what it describes (the arithmetical theorem). Nonetheless, we prove it.

One way in which 👁 and ✎ would move at the same pace is if ✎ also advanced two positions in every step; that is, if t were composed entirely of 2-runs (from some position on). But the description of such a string, which is st itself, cannot comprise only 2-runs. This possibility can be excluded, so there must be an infinity of 1-runs or 3-runs in t.

We next demonstrate that there needs to be a 1-run between any two 3-runs and a 3-run between any two 1-runs and that this is an impossible requirement. Indeed, we have seen that 3-runs have to begin in odd positions, so there are an odd number of 1-runs between any two 3-runs. On the other hand, for every 1-run 👁 gains one position on ✎, whence there must be at least one 3-run to compensate for it so that ✎ does not get caught. Because 👁 begins to the left of ✎, we can allow a bounded number of additional 1-runs. Finally, it is not difficult to see that the description of a string with this structure (1-runs and 3-runs interleaved) does not have this same structure. The $\alpha(t) = ts$ case is handled in a similar manner. □

Semi-infinite, partially self-descriptive strings do exist the other way, that is, strings that are described by substrings of themselves. Before getting to them, we remember a very few concepts of real-world biochemistry.

As the reader probably remembers from school, in real-world biochemistry the most important types of molecules, the ones that convey genetic information, are nucleic acids: RNA and DNA. They are complex and very long molecules composed of chains of smaller molecules called nucleotides. The shape of RNA molecules is that of a helix that biochemists call an α-helix. The shape of DNA molecules is that of a double helix: two helical strands interleaved. The primary structure of a nucleid acid is its mere sequence of nucleotides; the secondary structure refers to its helical shape. In each strand of a DNA molecule, each nucleotide has a "base" that pairs a particular base in the other strand, thus keeping the molecule in one piece. Now, back to the Look-and-Say world, where we find structures similar to nucleic acids.

The RNA Existence Theorem.

1. *Some semi-infinite strings are described by substrings of themselves. That is, there exist strings s and t, called respectively the "seed" and the "tail", such that s is finite, t is semi-infinite, and $\alpha(st) = t$ or $\alpha(ts) = t$. Such strings st or ts are called "RNA molecules."*

2. *For strings infinite to the right, the only seeds that do not generate RNA molecules are* `1`*,* `22`*,* `333`*,* `4444`*,* `55555`*,... (i.e., any seed of the form $\mathtt{x}^x$).*

3. *For strings infinite to the left, the only seeds that do not generate RNA molecules are formed when a digit is repeated an arbitrary number of times (i.e., the seeds of the form $\mathtt{x}^y$).*

In order to build an RNA molecule, we first need to write an appropriate seed. Then we can begin describing it left to right or right to left. This involves the same idea used earlier, but now ✎ writes a description corresponding to the text 👁 is reading (Figure 2).

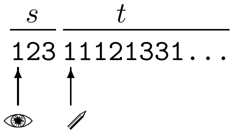


Figure 2: Reader and writer again in their initial positions.

An argument very similar to the proof of the negative theorem shows that 👁 will never catch ✎, so there will always be something to describe, and the process will not end. In fact, after the RNA composition theorem is stated we will be able to show that ✎ indeed moves λ times faster than 👁. For the seeds mentioned in parts 2 and 3 of the RNA existence theorem, it is impossible to perform the process coherently. We consider some examples.

Let the seed be 233 in order to build an RNA molecule to the left. First, we must prepend the description of the ending 33. The run whose description is to be written in the next step is underlined. Thus, at the start, we have $\mathtt{2\underline{33}}$. Here is the sequence of steps:

$$\begin{array}{r}
\mathtt{2\underline{33}} \\
\mathtt{23\underline{2}33} \\
\mathtt{122\underline{3}233} \\
\mathtt{131\underline{22}3233} \\
\mathtt{2213\underline{1}223233} \\
\mathtt{11221\underline{3}1223233} \\
\ldots
\end{array}$$

If we use the seed 3 instead, the only possible first step is to $\mathtt{\underline{2}33}$, and no further step is possible. If we prepend 12 we get 12233, with two 2s together, making the description incorrect. Prepending 22 does not work either, nor does anything else. Something similar to this always happens at a very early stage to the seeds mentioned in statement 3 of the RNA existence theorem.

Now, beginning with the seed 233 we build to the right:

$$\begin{array}{l}
\mathtt{\underline{2}33} \\
\mathtt{2\underline{33}12} \\
\mathtt{233\underline{1}223} \\
\mathtt{2331\underline{22}311} \\
\mathtt{233122\underline{3}1122} \\
\mathtt{2331223\underline{11}2213} \\
\ldots
\end{array}$$

Seeds such as 333, or the others mentioned in part 2 of the RNA existence theorem, do not even allow for a first step. Just try!

In some special cases an RNA molecule is the juxtaposition of the terms of the LS-sequence generated by the same seed. For example, if we take vanadium (V_{23}=13211312) as seed and append to it its description, titanium (Ti_{22}=11131221131112) we arrive at

$$V_{23}Ti_{22} = \mathtt{13211312{\cdot}11131221131112}.$$

A split condition appears between the two elements and is marked with the dot. Given that at the end of titanium we are describing the end of vanadium, a new split occurs there. Now we have to append the description of Ti_{22}, which is Sc_{21}, and continue in this way to get the primary structure of an RNA molecule, that is, the sequence of its components (nucleotides in real-world biochemistry, but atoms in our case):

$$V_{23}Ti_{22}Sc_{21}Ho_{67}Pa_{91}H_{1}Ca_{20}Co_{27}Dy_{66}Th_{90}H_{1}K_{19}Fe_{26}Tb_{65}Ac_{89}H_{1}\ldots$$

Not every RNA molecule is composed solely of common elements, even when the seed is. For example, if we take as seed an atom of tin (Sn_{50}=13211) and append to it its description, indium (In_{49}=11131221), the atoms fuse, producing the exotic element 1321111131221.

A split condition always appears at some point of any RNA molecule (indeed, at infinitely many points). In the following I write only about RNA molecules semi-infinite to the right, but the results are valid for both directions.

The RNA Splitting Theorem. *In an RNA molecule there occur infinitely many split conditions.*

Proof. As already noted we need only be sure that at least one split condition occurs, because another will occur when the first is being described, and so on. We have found no way to prove this other than with the aid of a computer program. The program is available as files `RNASplit.*` at the web address `http://www.uam.es/oscar.martin/downloads.html`. It is written in CWEB, so the reader should be able to master it. (CWEB is a literate programming tool, that is, a programming and documenting tool that stresses the importance of explaining first to humans and, secondarily, to the computer, what a program is intended to do.) We will briefly explain the workings of the program.

As in [1] we use the word "chunk" to describe a finite substring of a term of an LS-sequence or an RNA molecule. For the program to be useful it is a prerequisite that the chunk still to be described continues to get longer as we advance in the construction of the molecule. More precisely, given any integer N, there is a point on the molecule such that ✎ will always be at least N digits ahead of 👁 from that point on. If this were not the case, 👁 and ✎ would eventually have to move at the same pace, which an argument such as the one in the proof of the negative theorem rules out.

Because the future of the molecule depends only on the chunk still to be described, we can consider this chunk to be an *alternative seed*. We build as much of the molecule as needed so that the alternative seed is large enough (namely, twenty-five digits long, see below). We are interested in the description of this alternative seed, the description of its description, and so forth, and in determining whether a split condition happens in any of them.

This is almost like calculating the terms of an LS-sequence, but with an important difference—the chunks we consider now are not isolated strings, but parts of something bigger. We want to be sure that the chunk eventually produces a split, whatever its *environment* be. For example, we want to describe `...123...`. This could occur as the (underlined) inner part of $\texttt{...2}\underline{\texttt{123}}\texttt{1...}$, which is described by $\texttt{...1}\underline{\texttt{112}}\texttt{13...}$, or of $\texttt{...21}\underline{\texttt{123}}\texttt{331...}$, which is described by $\texttt{...2}\underline{\texttt{112}}\texttt{33...}$, etc. We know for sure only that the string `112` (shown underlined) is a part of the description.

Accordingly, at each step we compute the standard description, from which we then remove all unsure digits: the last two digits (the description of the last run) and the first one (the count for the first run). The following shows the evolution of the chunk `3211213`:

`...3211213...` → `...312211211...` → `...311222112...` → `...32·13221...`

A split occurs in three steps, as shown (maybe the reader would like to check this split in the list of split conditions in the splitting theorem).

Here is an example in which the string vanishes:

`...1311122...` → `...11331...` → `...123...` → `...112...` → `...1...`

If we study all the possible chunks of length twenty-five we find that sixteen steps, tops, are required to reach a splittable string in any case. Some chunks of length twenty-four vanish, showing that twenty-five is the minimum lenght required.

□

The next result provides key information about the structure of RNA molecules.

The RNA Composition Theorem.

1. *An RNA molecule is from some digit on composed of only common and transuranic elements.*
2. *An RNA molecule includes all the common elements, and they occur in the same proportions as in LS-sequences.*
3. *No RNA molecule is eventually periodic with finite period.*

Proof.

1. Once the first split condition occurs, others will appear when describing the first, so the string gets divided into independent chunks that reproduce the terms of an LS-sequence. For instance, with the seed 33 the associated RNA molecule is divided into chunks as shown by the dots

   332·312·131112·11133112·312·3. . . ,

   which can be rewritten 332$\mathrm{Zn}_{30}\mathrm{Cu}_{29}\mathrm{Ni}_{28}\mathrm{Zn}_{30}$. . . Part 1 of the theorem follows from Conway's cosmological theorem for LS-sequences. By the way, a seed like 131 does not seem, at first glance, to produce any transuranic elements, but it does, because a digit 4 appears soon in the RNA molecule it generates: 131111341. . . .
2. This is a direct consequence of part 1 and Conway's results.
3. If an RNA molecule had period p, its description (i.e., its own tail) would have to have period $p\lambda$, which is not an integer.

□

A small digression is in order here to explain the word "finite" in statement 3 of the theorem. We have hitherto restricted the discussion to nonfinite strings of length ω, but there is no reason we cannot consider strings whose lengths are ordinals greater than ω. Many of the results are easily extended to longer strings, while some of them, like aperiodicity, just hold for length ω. To see that infinite periods are possible we will get a little ahead of ourselves. In section 3, where we study DNA, we will see that there exist cyclically descriptive sets of semi-infinite strings. For instance, there exist strings s_1, s_2, and s_3 semi-infinite to the right such that $\alpha(s_1) = s_2$, $\alpha(s_2) = s_3$, and $\alpha(s_3) = s_1$. Consider the string s of length ω^2 that we get by concatenating s_1 with s_2 to its right, then with s_3, then s_1, s_2, and so on: $s = s_1s_2s_3s_1s_2s_3s_1 \ldots = (s_1s_2s_3)^\omega$. This s can be regarded as an RNA molecule with seed s_1 and with period $\omega \cdot 3$.

The same idea provides counterexamples to other results. For example, the string s can be seen as describing a part of itself, $\alpha(s_1s_2s_3s_1 \ldots) = s_2s_3s_1s_2 \ldots$, which is counter to the conclusion of the negative theorem. This ends the digression. Henceforth we confine ourselves, as far as nonfinite strings go, to those of lenght ω.

Experimental methods to observe the secondary structure (that is, the spatial conformation) of these objects are not available, so we have to hazard a guess:

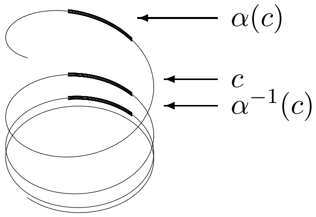


Figure 3: RNA secondary structure.

The RNA Secondary-Structure Conjecture. *We propose that RNA molecules assume the shape of an exponential helix with parametric equations*

$$x = \cos 2\pi t, \quad y = \sin 2\pi t, \quad z = \lambda^t$$

for real t. As this helix is related to the description function α, it seems natural to agree with biochemists and call it an "α-helix."

Figure 3 shows the structure of such an α-helix (for the present the reader can ignore the labels and the bold arcs).

To justify this conjecture we have to make several nontrivial assumptions about the attractive interactions within RNA:

- Apart from the strong interactions that keep a strand in one piece, there are weak attractive interactions between a chunk of a molecule and the chunk it describes.
- The interactions have unlimited radius of action.
- The interaction's strength obeys an inverse square law.

Because the helix is stretched upward, the chunk of RNA delimited by $t = t_1$ and $t = t_2$ has its description *just above it* (i.e., it is delimited by $t = t_1 + 1$ and $t = t_2 + 1$).

Moreover, we can prove that this structure is stable. To see this take an arbitrary chunk c, the middle chunk of the three bold ones in Figure 3. Its description $\alpha(c)$ is the upper bold chunk, and what c describes, $\alpha^{-1}(c)$, is the lower bold chunk. Now $\alpha(c)$ is λ^2 times longer than $\alpha^{-1}(c)$, but lies at a distance λ times farther from c. We conclude that the resulting interactions balance, so the structure is stable.

Other weak interactions will appear between unrelated regions of an RNA molecule, one of which happens to describe the other. These account for its undoubtedly complex tertiary structure (that is, its three-dimensional overall shape, caused by bends and foldings of the helical threads).

3 DNA.

We now shift the focus from isolated strings to sets of strings. We call a set of cyclically descriptive strings a *DNA molecule* and each of its strings a *strand*. That is, if a DNA molecule has n strands $S_1, \ldots, S_n$, we have $\alpha(S_1) = S_2$, $\alpha(S_2) = S_3$, $\ldots$, $\alpha(S_n) = S_1$. We use the terms *duplex DNA*, *triplex DNA*, and so on to indicate the number of strands, just as biochemists do.

The dependencies among the proofs in this section require that we make statements about DNA composition even before we know that DNA exists. That is why the following theorem begins with a proviso:

The DNA Composition Theorem. *Suppose that DNA molecules exist. Then the following assertions hold:*

1. *Infinitely many split conditions occur in any strand of a DNA molecule.*
2. *DNA molecules are formed exclusively by common and transuranic elements.*
3. *Each strand of a DNA molecule involves all the common elements, and they occur in the same proportions as in LS-sequences and RNA molecules.*
4. *DNA molecules do not necessarily include transuranic elements. Nevertheless, when they occur they do so in each strand of the molecule, but not both elements can appear in the same strand. Only one transuranic isotope can occur in a given DNA molecule (thus, the number of strands has to be even in this case). Ultimately, if transuranic elements occur, they do so infinitely often, but with asymptotically null frequency.*
5. *No strand of a DNA molecule is eventually periodic.*

Proof.

1. The computer-aided proof of the RNA splitting theorem establishes this part of the present theorem as well.
2. For each strand S in an n-stranded DNA molecule, $S = \alpha^{kn}(S)$ whenever $kn \geq 24$. In view of statement 1, the cosmological theorem gives us the result.
3. This is an immediate consequence of statement 2 and Conway's results.
4. This follows from the previous parts of the theorem and the discussion that ensues after the DNA existence theorem is presented.
5. If a strand of an n-stranded DNA molecule had period p, its nth order description (i.e., itself) would have period $p\lambda^n$, which is not an integer (remember that λ cannot be expressed by radicals).

   It is important to note here, as in the RNA composition theorem (see the digression following it), that our strings are of length ω and only finite periods are allowed.

□

Suppose that we have a duplex DNA molecule that is semi-infinite to the right, as the one depicted on the left in Figure 4. Choose one of its strands, say S, and any initial substring s (as a string of atoms). In particular, s can probably (see the next theorem) be taken to be just the first atom of the strand S. The other strand T has to begin with $\alpha(s)$, so S must begin with $\alpha^2(s)$, and this has to have the form $\alpha^2(s) = sv$ for some nonempty sring v. Continuing in this manner, we can construct the entire molecule, which is *generated* by the

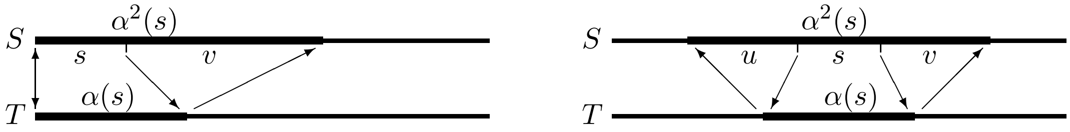


Figure 4: Generating semi-infinite and bi-infinite DNA.

seed s. Conversely, any atom A with the property that $\alpha^n(\mathrm{A}) = \mathrm{A}v$ can be used in this way to generate an n-stranded molecule.

This observation leads us to the proof of the existence of DNA. Since we know that DNA is completely composed of common and transuranic elements, we can agree to express its strands as strings of atoms instead of strings of digits. This is convenient, because atoms are always separated by splits, ensuring that their descriptions do not fuse or become mixed up.

The DNA Existence Theorem.

1. *If a semi-infinite n-stranded DNA molecule exists, then it can be generated by a seed of one of the following two types:*
   - *an atom* A *such that* $\alpha^n(\mathrm{A}) = u\mathrm{A}$ *or* $\alpha^n(\mathrm{A}) = \mathrm{A}v$ *for nonempty strings* u *or* v*, respectively*
   - *two atoms* A *and* H*, where* H *is the hydrogen atom, such that* $\alpha^n(\mathrm{AH}) = u\mathrm{AH}$ *or* $\alpha^n(\mathrm{HA}) = \mathrm{HA}v$ *for nonempty strings* u *or* v*, respectively*
2. *If a bi-infinite n-stranded DNA molecule exists, then it can be generated by a seed of one of the following three types:*
   - *two atoms* A *and* A′ *such that* $\alpha^n(\mathrm{AA'}) = u\mathrm{AA'}v$*, where neither* u *nor* v *is empty, in which case the molecule is the juxtaposition of two semi-infinite DNA molecules of opposite directions generated by* A *and* A′*, respectively*
   - *three atoms* A*,* A′*, and* H*, where* H *is the hydrogen atom, such that* $\alpha^n(\mathrm{AHA'}) = u\mathrm{AHA'}v$*, and neither* u *nor* v *is empty, in which case the molecule is the juxtaposition of two semi-infinite DNA molecules of opposite directions generated by* A *and* A′*, respectively, with* H *in between them*
   - *a single atom* A *such that* $\alpha^n(\mathrm{A}) = u\mathrm{A}v$*, where neither* u *nor* v *is empty*
3. *DNA molecules of each kind mentioned in parts 1 and 2 exist: bi-infinite, semi-infinite to the right, and semi-infinite to the left. There are only a finite number of DNA molecules, modulo identification of molecules containing different isotopes of the same transuranic element.*

Proof.

1. The argument given prior to the statement of the theorem proves this. Observe that $\alpha^n(\mathrm{AH}) = u\mathrm{AH}$ is equivalent to $\alpha^n(\mathrm{A}) = u\mathrm{A}$.
2. We want to show that for every bi-infinite DNA molecule we can find a string of atoms s that generates the molecule (i.e., $\alpha(s) = usv$, where neither u nor v is empty). We choose a strand S and consider an arbitrary chunk of atoms s_1 in S. Now $\alpha^n(s_1)$ also

lies in the strand S. Say, for example, that it lies to the left of s_1. We can repeat the argument of 👁 and ✎ to demonstrate that by moving to the right we will reach a chunk of atoms, call it s_2, such that $\alpha^n(s_2)$ lies to the right of s_2. The chunk s from s_1 to s_2, both included, is the desired string (see the right-hand side of Figure 4).

Now look at the first (or the last) atom of s. If it is not included in its nth order description, then it can be trimmed from s, and what remains will still generate the entire molecule. Continuing to trim s in this way, we arrive at one of the three possibilities given in the statement.

3. This assertion follows from the existence of atoms that satisfy the conditions of parts 1 and 2. Examples and a complete discussion follow this proof.

□

We now identify all the distinct DNA molecules. In his paper [3], Hilgemeier drew a digraph representing the "evolution" of common elements. We have included it as appendix B. The digraph has a node for each common element and an edge from one element to another if the latter is in the description of the former. According to the DNA existence theorem, in order to build DNA molecules we must find elements that are included in their own descriptions of some order; in other words, we must find cycles in Hilgemeier's graph. As most of the arrows go from elements of higher atomic numbers to elements of lower ones, we need to look for arrows going in the opposite direction. There are no cycles of length one apart from the stable H_1, so there can be no monostranded DNA.

There are only two possibilities for DNA semi-infinite to the right. They are given by the following cycles:

- $\mathrm{Ho}_{67} \to \mathrm{Dy}_{66} \to \mathrm{Tb}_{65} \to \mathrm{Ho}_{67}\mathrm{Gd}_{64}$
- $\mathrm{Zn}_{30} \to \mathrm{Cu}_{29} \to \mathrm{Ni}_{28} \to \mathrm{Zn}_{30}\mathrm{Co}_{27}$

With each of these we can form triplex DNA. Here are the first steps of its construction for the former cycle:

start	*second step*	*third step*	*fourth step*
Ho	Gd	Pm Ca Zn Ar Mn	Ce Cl Zn Co P Ti Mg
Dy	Eu Ca Co	Nd K Cu Cl Cr Si	La H Ca Co S Cu S...
Tb	Sm K Fe	Pr Ar Ni S V Al	Ba H K Fe P Ni Mn...

There are two additional DNA molecules that are semi-infinite to the right. They are obtained by attaching a hydrogen atom (22) to the left of each strand of the two given molecules.

The following cycles produce DNA molecules that are semi-infinite to the left, three of each are duplex DNA and the other quadruplex DNA:

- $\mathrm{Li}_3 \to \mathrm{He}_2 \to \mathrm{Hf}_{72}\mathrm{Pa}_{91}\mathrm{H}_1\mathrm{Ca}_{20}\mathrm{Li}_3$
- $\mathrm{W}_{74} \to \mathrm{Ta}_{73} \to \mathrm{Hf}_{72}\mathrm{Pa}_{91}\mathrm{H}_1\mathrm{Ca}_{20}\mathrm{W}_{74}$
- $\mathrm{Tc}_{43} \to \mathrm{Mo}_{42} \to \mathrm{Nb}_{41} \to \mathrm{Er}_{68}\mathrm{Zr}_{40} \to \mathrm{Ho}_{67}\mathrm{Pm}_{61}\mathrm{Y}_{39}\mathrm{H}_1\mathrm{Ca}_{20}\mathrm{Tc}_{43}$
- $\mathrm{Pu}_{94} \to \mathrm{Np}_{93} \to \mathrm{Hf}_{72}\mathrm{Pa}_{91}\mathrm{H}_1\mathrm{Ca}_{20}\mathrm{Pu}_{94}$

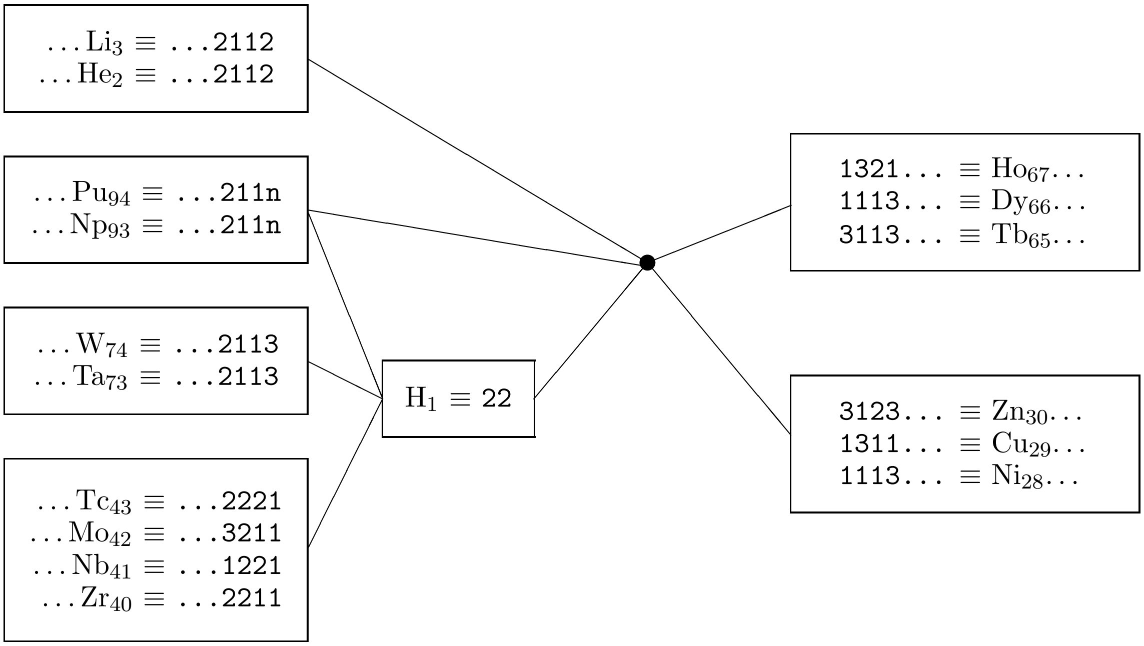


Figure 5: Juxtaposition of semi-infinite DNA produces bi-infinite DNA.

The last cycle actually gives rise to a family of DNA molecules, one corresponding to each isotope of Pu_{94} and Np_{93}. They do not show up in Hilgemeier's graph because only common elements appear there. Hydrogen atoms can also be attached to all the DNA molecules associated with these four cycles except the first, because both Li_3 and He_2 end with the digit 2, so a fusion would occur with H_1 (22).

There are no other semi-infinite DNA molecules. We next look for bi-infinite DNA. First, the previous molecules can be juxtaposed to form DNA with six or twelve strands. We have to duplicate, triplicate, or quadruplicate each strand in the molecule in order to do this. Figure 5 exhibits all the bi-infinite DNA molecules obtained in this way. The big dot means nothing: it is just a way to clarify that any of the strings to its left can be juxtaposed with any to its right. Note that some molecules require hydrogen bonds, while others do not admit them. Nevertheless, the DNA molecules in which transuranic elements appear may or may not include hydrogen bonds. For instance, here is a part of a six-stranded DNA molecule:

... H Ar Hf Pa H Ca W	H	Ho Gd Pm Ca Zn Ar Mn...
... H Cl Lu Th H K Ta	H	Dy Eu Ca Co Nd K Cu...
... H Ar Hf Pa H Ca W	H	Tb Sm K Fe Pr Ar Ni...
... H Cl Lu Th H K Ta	H	Ho Gd Pm Ca Zn Ar Mn...
... H Ar Hf Pa H Ca W	H	Dy Eu Ca Co Nd K Cu...
... H Cl Lu Th H K Ta	H	Tb Sm K Fe Pr Ar Ni...

All these molecules are generated by compound strings: Li_3Ho_{67}, $W_{74}H_1Zn_{30}$, and so on. Also, these are the only cases of DNA that cannot be generated by single atoms.

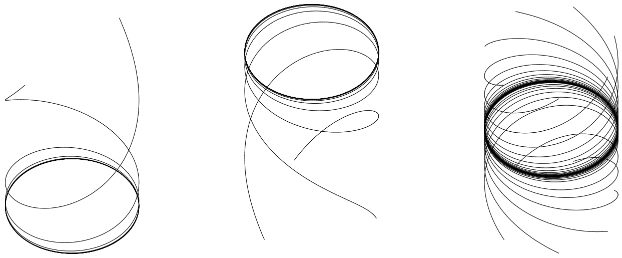


Figure 6: DNA structures.

Other bi-infinite DNA molecules correspond to the following twenty-one arrows in Hilgemeier's graph:

$\mathrm{Ta}_{73} \to \mathrm{Pa}_{91}$
$\mathrm{I}_{53} \to \mathrm{Ho}_{67}$
$\mathrm{Te}_{52} \to \mathrm{Eu}_{63}$
$\mathrm{Sb}_{51} \to \mathrm{Pm}_{61}$
$\mathrm{Rh}_{45} \to \mathrm{Ho}_{67}$
$\mathrm{Ru}_{44} \to \mathrm{Eu}_{63}$
$\mathrm{Nb}_{41} \to \mathrm{Er}_{68}$

$\mathrm{Y}_{39} \to \mathrm{U}_{92}$
$\mathrm{Ge}_{32} \to \mathrm{Ho}_{67}$
$\mathrm{Ga}_{31} \to \mathrm{Eu}_{63}$
$\mathrm{Ga}_{31} \to \mathrm{Ac}_{89}$
$\mathrm{Sc}_{21} \to \mathrm{Ho}_{67}$
$\mathrm{Sc}_{21} \to \mathrm{Pa}_{91}$
$\mathrm{Sc}_{21} \to \mathrm{Co}_{27}$

$\mathrm{P}_{15} \to \mathrm{Ho}_{67}$
$\mathrm{Mg}_{12} \to \mathrm{Pm}_{61}$
$\mathrm{Be}_{4} \to \mathrm{Ge}_{32}$
$\mathrm{Be}_{4} \to \mathrm{Ca}_{20}$
$\mathrm{He}_{2} \to \mathrm{Hf}_{72}$
$\mathrm{He}_{2} \to \mathrm{Pa}_{91}$
$\mathrm{He}_{2} \to \mathrm{Ca}_{20}$

Except for H_1, every element occurs in some cycle. Also, as we know, every element occurs in every DNA molecule. The lengths of these cycles (i.e., the number of strands in the molecules) is diverse—only twelve, which coincides with one of the molecules obtained earlier by juxtaposing semi-infinite DNA, nineteen, and seventy-one repeat. The shortest cycle has length seven, and the longest cycle has length ninety. The element U_{92} is in only one cycle; all other elements, except H_1, occur in more than one. In addition, atoms with atomic numbers between 53 and 62, both included, can generate up to sixteen different DNA molecules.

This exhausts the list of possibilities. We remark that a cycle of length n can be used to generate mn-stranded molecules for any m by repeating every strand m times. These are degenerate constructs, and we do not consider them proper DNA. Thus, the following result is immediate:

The DNA Self-Replicating Property. *Any (nondegenerate) DNA molecule can be completely recovered from any of its isolated strands.*

We conclude with a hypothesis about the spatial conformation of DNA.

The DNA Secondary-Structure Conjecture. *We propose that DNA molecules assume the shape of a multiple exponential helix. For n-stranded DNA the parametric equations are*

$$x = \cos 2\pi(t + i/n), \quad y = \sin 2\pi(t + i/n), \quad z = \lambda^t \quad (i = 0, \ldots, n-1)$$

for real t.

Figure 6 illustrates DNA molecules with two, three, and seven strands.

The rational for this proposal is the same as that for the RNA secondary structure. Now the helixes are interleaved and evenly spaced (which we achieve by adding i/n to t). Once again, the description is just "above" or "below" what it describes.

4 APPENDIX A: THE PERIODIC TABLE.

This table, adapted from [1], contains all the relevant information about the ninety-two common elements and the two transuranic ones. The first column on the left contains the atomic number and the second the chemical symbol. The third column gives the string of digits that defines the element. The fourth column is the description of the element (i.e., the function α applied to it) written as a string of atoms.

The ordering of the common elements is such that the element of atomic number n appears in the description of the element of atomic number $n+1$. Also, note that all the descriptions of common elements are compounds of common elements, as stated by the chemical theorem.

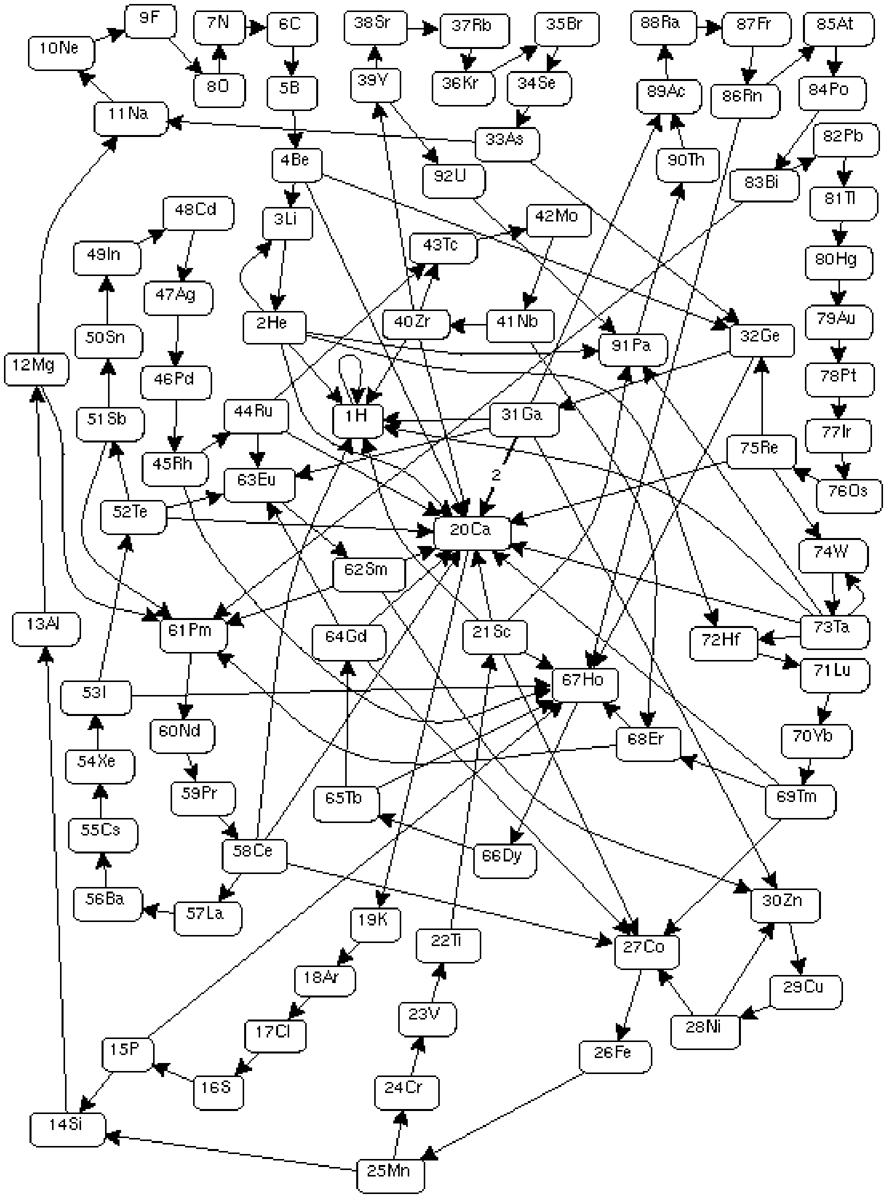
Element		**String**	**Description**
94	Pu	312211322212221121123222 11[≥4]	Np_{93}
93	Np	13112221133211322112211213322 11[≥4]	$\mathrm{Hf}_{72}\mathrm{Pa}_{91}\mathrm{H}_{1}\mathrm{Ca}_{20}\mathrm{Pu}_{94}$
92	U	3	Pa_{91}
91	Pa	13	Th_{90}
90	Th	1113	Ac_{89}
89	Ac	3113	Ra_{88}
88	Ra	132113	Fr_{87}
87	Fr	1113122113	Rn_{86}
86	Rn	311311222113	$\mathrm{Ho}_{67}\mathrm{At}_{85}$
85	At	1322113	Po_{84}
84	Po	1113222113	Bi_{83}
83	Bi	3113322113	$\mathrm{Pm}_{61}\mathrm{Pb}_{82}$
82	Pb	123222113	Tl_{81}
81	Tl	111213322113	Hg_{80}
80	Hg	31121123222113	Au_{79}
79	Au	132112211213322113	Pt_{78}
78	Pt	111312212221121123222113	Ir_{77}
77	Ir	3113112211322112211213322113	Os_{76}
76	Os	1321132122211322212221121123222113	Re_{75}
75	Re	111312211312113221133211322112211213322113	$\mathrm{Ge}_{32}\mathrm{Ca}_{20}\mathrm{W}_{74}$
74	W	312211322212221121123222113	Ta_{73}
73	Ta	13112221133211322112211213322113	$\mathrm{Hf}_{72}\mathrm{Pa}_{91}\mathrm{H}_{1}\mathrm{Ca}_{20}\mathrm{W}_{74}$
72	Hf	11132	Lu_{71}
71	Lu	311312	Yb_{70}
70	Yb	1321131112	Tm_{69}
69	Tm	11131221133112	$\mathrm{Er}_{68}\mathrm{Ca}_{20}\mathrm{Co}_{27}$
68	Er	311311222	$\mathrm{Ho}_{67}\mathrm{Pm}_{61}$
67	Ho	1321132	Dy_{66}
66	Dy	111312211312	Tb_{65}
65	Tb	3113112221131112	$\mathrm{Ho}_{67}\mathrm{Gd}_{64}$
64	Gd	13221133112	$\mathrm{Eu}_{63}\mathrm{Ca}_{20}\mathrm{Co}_{27}$
63	Eu	1113222	Sm_{62}
62	Sm	311332	$\mathrm{Pm}_{61}\mathrm{Ca}_{20}\mathrm{Zn}_{30}$
61	Pm	132	Nd_{60}

Element		String	Description
60	Nd	111312	Pr_{59}
59	Pr	31131112	Ce_{58}
58	Ce	1321133112	$\mathrm{La}_{57}\mathrm{H}_{1}\mathrm{Ca}_{20}\mathrm{Co}_{27}$
57	La	11131	Ba_{56}
56	Ba	311311	Cs_{55}
55	Cs	13211321	Xe_{54}
54	Xe	11131221131211	I_{53}
53	I	311311222113111221	$\mathrm{Ho}_{67}\mathrm{Te}_{52}$
52	Te	1322113312211	$\mathrm{Eu}_{63}\mathrm{Ca}_{20}\mathrm{Sb}_{51}$
51	Sb	3112221	$\mathrm{Pm}_{61}\mathrm{Sn}_{50}$
50	Sn	13211	In_{49}
49	In	11131221	Cd_{48}
48	Cd	3113112211	Ag_{47}
47	Ag	132113212221	Pd_{46}
46	Pd	111312211312113211	Rh_{45}
45	Rh	311311222113111221131221	$\mathrm{Ho}_{67}\mathrm{Ru}_{44}$
44	Ru	132211331222113112211	$\mathrm{Eu}_{63}\mathrm{Ca}_{20}\mathrm{Tc}_{43}$
43	Tc	311322113212221	Mo_{42}
42	Mo	13211322211312113211	Nb_{41}
41	Nb	1113122113322113111221131221	$\mathrm{Er}_{68}\mathrm{Zr}_{40}$
40	Zr	12322211331222113112211	$\mathrm{Y}_{39}\mathrm{H}_{1}\mathrm{Ca}_{20}\mathrm{Tc}_{43}$
39	Y	1112133	$\mathrm{Sr}_{38}\mathrm{U}_{92}$
38	Sr	3112112	Rb_{37}
37	Rb	1321122112	Kr_{36}
36	Kr	11131221222112	Br_{35}
35	Br	3113112211322112	Se_{34}
34	Se	13211321222113222112	As_{33}
33	As	11131221131211322113322112	$\mathrm{Ge}_{32}\mathrm{Na}_{11}$
32	Ge	31131122211311122113222	$\mathrm{Ho}_{67}\mathrm{Ga}_{31}$
31	Ga	13221133122211332	$\mathrm{Eu}_{63}\mathrm{Ca}_{20}\mathrm{Ac}_{89}\mathrm{H}_{1}\mathrm{Ca}_{20}\mathrm{Zn}_{30}$
30	Zn	312	Cu_{29}
29	Cu	131112	Ni_{28}
28	Ni	11133112	$\mathrm{Zn}_{30}\mathrm{Co}_{27}$
27	Co	32112	Fe_{26}
26	Fe	13122112	Mn_{25}
25	Mn	111311222112	$\mathrm{Cr}_{24}\mathrm{Si}_{14}$
24	Cr	31132	V_{23}
23	V	13211312	Ti_{22}
22	Ti	11131221131112	Sc_{21}
21	Sc	3113112221133112	$\mathrm{Ho}_{67}\mathrm{Pa}_{91}\mathrm{H}_{1}\mathrm{Ca}_{20}\mathrm{Co}_{27}$
20	Ca	12	K_{19}
19	K	1112	Ar_{18}
18	Ar	3112	Cl_{17}
17	Cl	132112	S_{16}
16	S	1113122112	P_{15}
15	P	311311222112	$\mathrm{Ho}_{67}\mathrm{Si}_{14}$

Element		String	Description
14	Si	1322112	Al_{13}
13	Al	1113222112	Mg_{12}
12	Mg	3113322112	$Pm_{61}Na_{11}$
11	Na	123222112	Ne_{10}
10	Ne	111213322112	F_9
9	F	31121123222112	O_8
8	O	132112211213322112	N_7
7	N	111312212221121123222112	C_6
6	C	3113112211322112211213322112	B_5
5	B	1321132122211322212221121123222112	Be_4
4	Be	111312211312113221133211322112211213322112	$Ge_{32}Ca_{20}Li_3$
3	Li	312211322212221121123222112	He_2
2	He	13112221133211322112211213322112	$Hf_{72}Pa_{91}H_1Ca_{20}Li_3$
1	H	22	H_1

5 APPENDIX B: HILGEMEIER'S GRAPH.

This is the digraph generated by the description function on common elements. It is reproduced here from [3, p. 142] with the kind permission of Mario Hilgemeier. There is a node for each common element (the atomic number precedes the symbol), and there is also an edge from an element to each element that appears in its description. Cycles in this graph allow us to build DNA molecules.



ACKNOWLEDGMENT. This paper is much better thanks to the very careful job of its anonymous referee.

References

[1] J. H. Conway, The weird and wonderful chemistry of audioactive decay, *Eureka* **46** (1985) 5–16; reprinted in *Open Problems in Communication and Computation*, T. M. Cover and B. Gopinath, eds., Springer-Verlag, New York, 1987, pp. 173–188.

[2] D. E. Gilbert and J. Feigon, Multistranded DNA structures, *Current Opinion in Structural Biology* **9** (1999) 305–314.

[3] M. Hilgemeier, One metaphor fits all: A fractal voyage with Conway's audioactive decay, in *Fractal Horizons: The Future Use of Fractals*, C. A. Pickover, ed., St. Martin's Press, New York, 1996, pp. 137–161; also available at `http://www.btinternet.com/~se16/mhi`.

[4] D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, 3rd ed., Worth, New York, 2000.

[5] R. A. Litherland, The audioactive package, available at `http://www.math.lsu.edu/~lither/jhc/doc.pdf`, 2003.

[6] ———, Conway's cosmological theorem, available at `http://www.math.lsu.edu/~lither/jhc/cct.pdf`, 2003.

ÓSCAR MARTÍN received his bachelor's degree from the Universidad Complutense de Madrid, in Spain, in 1989, specializing in computer science. He is at the moment trying to move his mathematical education a step further in the Universidad Autónoma de Madrid, where he is learning the theory of o-minimal structures. His interests also include recreational mathematics. He works for a financial company as a software developer.
Av. El Ferrol, 11. 28029 Madrid, Spain
oscar.martin@uam.es