Challenges **2011**, *2*, 37-42; doi:10.3390/challe2010037

OPEN ACCESS

challenges

ISSN 2078-1547

www.mdpi.com/journal/challenges

Communication

How to Preserve Documents: A Short Meditation on Three Themes

Jan Petr [1,*], Václav Ranc [2], Vítězslav Maier [1], Pavlína Ginterová [2], Joanna Znaleziona [2], Radim Knob [2] and Juraj Ševčík [2]

[1] Regional Centre for Advanced Technologies and Materials, Department of Analytical Chemistry, Palacký University in Olomouc, 17. listopadu 12, CZ-77146 Olomouc, Czech Republic; E-Mail: vitezslav.maier@upol.cz

[2] Department of Analytical Chemistry, Palacký University in Olomouc, 17. listopadu 12, CZ-77146 Olomouc, Czech Republic; E-Mails: vasekranc@mac.com (V.R.); ginterovapavlina@seznam.cz (P.G.); jznaleziona@gmail.com (J.Z.); rknob@seznam.cz (R.K.); sevcik@prfnw.upol.cz (J.Š.)

* Author to whom correspondence should be addressed; E-Mail: secjpetr@gmail.com; Tel.: +420-585-63-4416; Fax: +420-585-63-4433.

Received: 4 January 2011; in revised form: 23 February 2011 / Accepted: 25 February 2011 / Published: 2 March 2011

Abstract: The capability to present electronic media that can preserve information is highly restricted to few decades (e.g., a lifetime of DVD media does not exceed 100 years), and therefore the question of how to preserve documents for more than thousands or millions of years presents a challenging task. In this article, we discuss three thinkable possibilities for long-term data storage: (i) self-assembly systems, (ii) chirality, and (iii) nucleic acids. These systems have, in our opinion, added-value regarding functionality and storing capability. Self-assembly systems form 3D structures, which could reflect any information more precisely than a 2D structure, and therefore they could be used as a training information package. Chirality provides the next added value in the possibility of using an interval of <−1; 1> for storing the data (fuzzy logic) and could be also interesting in increasing the storage capacity if using compounds with more chiral centers, such as polysaccharides. Finally, nucleic acids represent a method of storage in which the reading step is developed and probably will be still active if people inhabit the Earth, which will realize the whole process of writing/storing and reading easier.

Keywords: human document project; information; preservation; self-assembly systems; chirality; nucleic acids

1. Introduction

The capability of present electronic media to preserve information is highly restricted to few decades, as was described e.g., at the beginning of the first workshop of the Human Document Project [1]. Indeed, the question how to preserve documents for thousands or millions of years has been opened before, e.g., The Rosetta Project, Gene banks, *etc.* The task is not easy as noted by Andreas Manz in the introduction of this issue [2]. We should think about more than technologies how to store the data. There are more factors affecting the storage, such as life cycles, erosion, catastrophes, *etc.* Interestingly, there is one example of the power of nature in the area of Chernobyl. It is not more than 24 years [3] ago that a set of unfortunate events led to a well-known radioactivity catastrophe closing that area. Looking at today's pictures of the city of Pripyat near Chernobyl [4], the erosion effects together with the expansion of plants started to change that area. Indeed, the area is still radioactive, but the effect of the nature on the human non-treated area is clearly visible. Nowadays, we are looking for systems that should "survive" such events or we are looking for new systems that have some added-value regarding, for example, functionality or storing capability. In this article, we present our view on three possibilities that could be applied for preserving any information.

2. Chemical Reactions and Self-Assembly Systems

Chemical reactions that are not in the equilibrium state lead, under defined conditions, to certain products [5]. Similarly, if some components are presented together in the defined environment, they form higher-structures which used to be called self-assembly systems [6]. These examples represent an interesting possibility to preserve information, in our opinion. The reaction systems (chemical reactions, self-assembly reactions) are dependent only on the defined environment. Moreover, there are many reactions that could proceed in different environments, having the same products. This could be profitable for the information storage, too, because deviations in the environment where the reactions should take place could be expected in thousands of years. Maybe the reaction with atmospheric oxygen could be a good starting point. If the system of information is placed into a box with an inert atmosphere, opening this box will start the oxidation processes that could produce the information package. The advantage of this possibility over others is due to the fact that human beings need oxygen for breathing, so if there are still living human like species it could be supposed that the atmosphere will contain oxygen. However, as noted from our time, oxygen is also marked as a destroying agent and some historical artifacts are studied in an inert atmosphere where the starting reaction (oxidation) cannot proceed. When imaging future times, people probably would try to stop the reaction on the "ancient" object and hence stop the reaction that forms the information package. In this regard, we should inform readers that the reaction should take place or that we should have a reaction system that proceeds in a non-oxidative environment with an initiation reaction by oxygen.

Additionally, chemical reactions (self-assembly processes) lead to the 3D structures that represent another possibility of information storage. As known, for example, from the cinema, the 3D structure reflects the situation more precisely than any 2D structure. The use of reactions could form the information into a more "plastic" view. Maybe this would be useful as a training set of information (considering the use of more information packages, e.g., training, basic, middle, and advanced).

3. Chirality

Chirality represents an interesting feature of all living and non-living species. It is well known that enantiomers are different in their spatial orientation and they can be distinguished only in the chiral environment. Enantiomers usually act differently in the human body, e.g., one enantiomer has teratogenic effects while the second can be used as a sedative [7,8].

Mixing enantiomers could provide an interesting feature useful for data storage. Enantiomers (*R*) and (*S*) of a compound with one chiral center have the same value of optical rotation but with different sign. When mixing (*R*)- and (*S*)- enantiomers in different ratios, the value of optical rotation is proportionally changed. Theoretically, this represents an interval $<-1; 1>$ formed from *R*:*S* mixtures. As all intervals, this one also contains an infinite number of values of Real numbers (and theoretically an infinite number of *R*:*S* mixtures). From mathematics, this principle is similar to the fuzzy logic which is nowadays widely used in many branches in everyday life [9]. In this view, the principle of the fuzzy logic could be also interesting for storing data.

Let us think about the *R*:*S* mixtures in details. Supposing one million molecules of e.g., lactic acid, which represents 1.5×10^{-16} g (the relative molecular weight of lactic acid is 90.08), then there are one million possibilities to store the data, e.g., by ratio 1:999,999, 500,000:500,000, *etc.* This ratio could open a possibility to store e.g., a letter from all the languages needed for the storage (English, Chinese, Spanish, Hindi, *etc.*). The next task relates to coding and decoding (recording and reading) the information. In the reading point, measuring the optical rotation could be useful since it is fast and relatively easy for operation. Also the possibility to evaluate the *R*:*S* ratio is given by different analytical approaches like circular dichroism spectroscopy, liquid chromatography or capillary electrophoresis [10,11]. The question about recording the information is much more complicated. First, the synthesis of such ratio could be done maybe by micromanipulation synthetic techniques. However, this method represents a very labor-intensive task and probably will be not considered as the best way to store the data. However, there is a next hypothetical possibility. Let us imagine a chiral compound that changes the optical rotation based on energy added to the system, e.g., by a laser beam having definite wavelength. Then the synthesis is not problematic (in the view of necessity of having e.g., 1:999,999 ratio) and could be done by classical chemical reactions resulting in some *R*:*S* ratio. Then the information (the ratio) could be changed by the energy transfer from the light. However, it is very important to design this chiral compound interacting with the light to have only a short wavelength window, and if possible, from an artificial source. This would cause that the ratio to be stable for a long time. Indeed, there are more marking requirements for such a compound like stability within temperature, humidity, irradiation, *etc.* However, the possibility of the multiple information storage given by the chirality seems to be very promising. Moreover, the compound used may have more than one chiral center, as polysaccharides have. There is the next important fact connected with polysaccharides: their basic units (single carbohydrates) form sequences (the next dimension of the information storage), and therefore polysaccharides could be useful as multidimensional information carry systems. Indeed, this way to preserve the data is more advanced and probably should be incorporated in advanced information packages (if using more information packages).

4. Nucleic Acids

The last idea deals with nucleic acids (NAs). It is well-known that NAs are the basic molecules of all the living species, coding their primary as well as advanced characteristics, *i.e.* structure of body, sex, disposition to illness, *etc.* They are replicated in each cell during their division and they are passed from generation to generation [12].

There are two major advantages connected with nucleic acids as a storing medium: (i) NAs are transferred from generation to generation, which gives them the possibility to keep the information in living (both in natural or artificial) organisms, (ii) we are able to decode the information from NAs by sequencing. The first advantage is connected with the idea of where to store the data. Maybe we could store some information in the human genome in the future (this is exactly done now but in a different way of thinking, as described in disadvantages) or more practically in small organisms which are not resistant to changes in the environment or catastrophes, such as *Tardigrades*, who is able to survive a nuclear explosion. Interestingly, *E. coli* seems to be tested for the data storage (in this case for the storage of The Declaration of Independence) as recently appeared on the news web [13].

The second advantage is related to the fact, that it is possible to sequenate NAs. It could be implied that storing the information will be useful only in the case of persistent life. And if we are at the point of having NA sequences of various organisms [14,15], we could believe that the next generations will also have appropriate knowledge and ability to sequenate NAs. From this point, there is only a need to say to our future generation "Hey, here is the information, just read the NA sequence." And this is, in our opinion, easier than, for example, the need of saying more in the previous task of storing the data in chiral compounds: "Hey, here is the information; you need a laser with a definite wavelength, such construction, *etc.*"

However, there is a big disadvantage related to the use of NAs for the data storage. It is connected with the replication process and defects. For example, the DNA sequence changes over years (DNA sequences of *Homo erectus* and *Homo sapiens* are different) and it is also subjected to immediate change, e.g., by irradiation (the correction mechanisms of the metabolism are not able to react on such stimulants). These changes in the genome form new populations (organisms are adapting to a new environment), which is also some kind of information. In this view, humans also carry information in the genome, mainly about the environmental factors and stimulators. However, these changes could be very problematic if we need to store the data without any alternative. Maybe, the selection of an appropriate organism could limit such effects, but this needs to be covered in the development process.

5. Conclusions

In this article, we discussed three challenging possibilities for long-term data storage: (i) self-assembly systems, (ii) chirality, and (iii) nucleic acids. These systems are not the only which could be used for the preservation of the information but they have, in our opinion, added-value regarding functionality and storing capability. Self-assembly systems form 3D structures, which could reflect any information more precisely than 2D structures (if necessary or welcome). They could be used as a training information package. Chirality gives the next added value in the possibility of using an interval <−1; 1> for storing the data (fuzzy logic). Moreover, the use of compounds with more chiral centers could be also interesting in increasing the storage capacity. And finally, nucleic acids represent a way of storage in which the reading step is developed and probably will be still active if people inhabit the Earth.

Acknowledgements

The financial support of the research by the Operational Program Research and Development for Innovations—European Social Fund (project CZ.1.05/2.1.00/03.0058) is gratefully acknowledged. Authors thank Pavel Neužil (KIST Europe, Saarbrücken, Germany) for his stimulation to attend the first HuDoP meeting.

References and Notes

1. The Human Document Project. Available online: http:// www.humandocument.org (accessed on 11 December 2010).
2. Manz, A. The human document project and challenges. *Challenges* **2010**, *1*, 3-4.
3. Wikipedia. Chernobyl disaster. Available online: http://en.wikipedia.org/wiki/Chernobyl_disaster (accessed on 11 December 2010).
4. Wikipedia. Prypiat city. Available online: http://en.wikipedia.org/wiki/Prypiat_(city) (accessed on 11 December 2010).
5. Earnshaw, A.; Greenwood, N. *Chemistry of the Elements*, 2nd ed.; Butterworth-Heinemann: London, UK, 1997.
6. Mann, S. Self-assembly and transformation of hybrid nano-objects and nanostructures under equilibrium and non-equilibrium conditions. *Nature Mat.* **2009**, *8*, 781-792.
7. Bentley, R. The nose as a stereochemist. Enantiomers and odor. *Chem. Rev.* **2006**, *106*, 4099-4112.
8. Eriksson, T.; Björkman, S.; Höglund, P. Clinical pharmacology of thalidomide. *Eur. J. Clin. Pharmacol.* **2001**, *57*, 365-376.
9. Mukaidono, M. *Fuzzy Logic for Beginners*, 1st ed.; World Scientific Publishing Company: London, UK, 2001.
10. Busch, K.W.; Busch, M.A. *Chiral Analysis*, 1st ed.; Elsevier Science: New York, NY, USA, 2006.
11. Chankvetadze, B. *Capillary Electrophoresis in Chiral Analysis*, 1st ed.; Wiley: Chichester, UK, 1997.
12. Berg, J.M.; Tymoczko, J.L.; Stryer, L. *Biochemistry*, 6th ed.; W. H. Freeman and Company: New York, NY, USA, 2007.
13. Flatley, J.L. Scientists developing *E. coli* bacteria that stores, encrypts data. Available online: http://www.engadget.com/2010/12/21/scientists-developing-e-coli-bacteria-that-stores-encrypts-data (accessed on 25 December 2010).

14. Kidd, J.M.; Cooper, G.M.; Donahue, W.F.; Hayden, H.S.; Sampas, N.; Graves, T.; Hansen, N.; Teague, B.; Alkan, C.; Antonacci, F.; *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **2008**, *453*, 56-64.
15. Harris, T.D.; Buzby, P.R.; Babcock, H.; Beer, E.; Bowers, J.; Braslavsky, I.; Causey, M.; Colonell, J.; Dimeo, J.; Efcavitch, J.W.; *et al.* Single-molecule DNA sequencing of a viral genome. *Science* **2008**, *320*, 106-109.