Polish J. Chem., **78**, 1007–1017 (2004) **REVIEW ARTICLE**

Genomic Maps and Novel Approaches to Sequencing of Repetitive *versus* Non-repetitive DNA*

by W. Szybalski

McArdle Laboratory, University of Wisconsin, 1400 University Avenue, Madison, WI 53706-1599, USA

(Received April 28th, 2004)

As first demonstrated by Avery *et al.*, DNA is a macromolecule which governs most aspects of life [1]. Thus our role as chemists was to determine the structure of this macromolecule, synthesize it, and possibly relate its structure to the genetic function. The first task was to determine the subunit structure of DNA, namely the structure of bases and their organization in relation to the deoxyribose and phosphate backbone. This was done in decades around the 1950-s. Independently, and around the same time, the concept of genes and the gene maps emerged as to relate the linear structure of DNA to its function. Next came the visualization of DNA by electron microscopy (EM) and its physical mapping using the heteroduplexes between DNA strands of various mutants. This permitted a precise way of measuring the length of DNA, positioning various deletions or other rearrangements and relating these to the genetic and transcriptional maps. The final step was the precise sequencing of DNA, either by the now abandoned chemical based method or by the presently used enzymatic procedure, which led to progressively more genomes being sequenced. Taken together, all this important scientific milestones led to our present day understanding of the chemical structure and function of DNA in relation to the 'puzzle of life'!

However, it was soon realized that the precise entire sequence of DNA could not be determined for many genomes, especially the eukaryotic ones, because they contain numerous long stretches of highly repetitive sequences, which defy the present computerized overlap procedure required for aligning of fragments and determining the final sequence. Therefore, we had to develop novel strategies to accurately sequence repetitive elements of DNA, as outlined here. These comprise construction and use of special transposons and pBAC/oriV vectors, both equipped with very rare cutting sites. Transposons (Tn) allow determination of 500–1000 nucleotide (nt) sequences on both sides of their insertion, whereas the very rare cutting sites (like I-SceI, PI-SceI or our Achilles heel cleavage sites) allow precise mapping of the positions of the insertions, using pulsed field gel electrophoresis (PFGE) or other physical means, including electron microscopic (EM) mapping. Thus we had to return to some of our earlier methods of physical mapping, which together with transposon-associated priming would allow sequencing of large eukaryotic genomes to be completed. This would be the final triumph of the structural chemistry of the DNA macromolecules which are the essence of genomes and genomics.

Key words: genomics, DNA sequencing, repetitive DNA

---

*Dedicated to Prof. E. Borowski on the occasion of his 75th birthday.

1. My early days of chemistry: future relationship to genetics and DNA mapping

Our biological existence and biology in general are based on the macromolecular chemistry as instructed by DNA molecules. Thus as chemists, our role was to determine the structure of DNA and to confirm it by the synthesis of DNA. This was done during the past several decades. Very little was known, however, about the macromolecular structure and function of DNA at the time when I entered the field of chemistry, first as at high school level in 1937–39 at the VIII Gimnazium in Lwów, Poland, and subsequently as the chemistry undergraduate student in 1939–44 at the Politechnika Lwowska (PL), also in Lwów, Poland, followed by the doctoral studies combined with teaching in 1945–49 at the Politechnika Gdańska (PG). But even then, although this was a ghastly time of the second World War (WWII), accompanied by Soviet and Nazi atrocities and followed by the British and American mindless but shameful and nearsighted betrayal of Poland, I was fascinated by the fields of macromolecular chemistry and genetics. I tried to educate myself whenever finding any 'leisure' time between my studies, enemy terror, aerial bombardments and my partisan activities.

Just before and also during WWII I read a few books written by famous theoretical physicists on the effect of radiation on the living cells ("target theory"), which referred to the 'mystery of life'. These authors were searching for a new physical 'principle' governing life. These books were fascinating, but rather misleading and therefore not too useful. However, for me the turning point was when I read the paper by Avery *et al*., which convinced me that it must be DNA, a chemical macromolecule, which is the 'life principle' [1]. I read this publication with fascination, although I found it totally by chance, when the first shipment of post-WWII American journals arrived in Gdańsk, Poland, as a gift from the United Nations Relief and Rehabilitation Agency (UNRRA). I believed then (maybe naively) that a chemical engineer like me would be sufficiently trained to tackle the chemistry of DNA and also the essence of genetics that later was followed by genomics. I also decided to study the more practical but related field now known as biotechnology, to which I was first exposed during the lectures by Professor Adolf Joszt, in 1943, during Nazi occupation, at the Politechnika Lwowska in Nazi-occupied Lwów, in eastern Poland. My rather recent essays summarize this period of my scientific experience [2–4].

In this manner my life became closely associated with microbial and molecular genetics, especially when I joined Prof. Winge's group at Carlsberg Laboratory in Copenhagen [3] and later became the staff member at the famous Cold Spring Harbor Laboratories in USA. There, my colleague Al Hershey and his technician and my dear friend Martha Chase used a radiochemical approach to provide additional evidence for the pivotal genetic role of phage DNA ([2], while Barbara McClintock was studying programmed variation in maize, now known to be caused by transposons (which we used also in the present work, and also studied earlier as IS (Insertion Sequence) elements, designated so by Fiandt *et al*. [5].

At that time, I studied resistance to antibiotics and other drugs and identified mutations in several bacterial species [6]. After moving to the Rutgers Institute of Microbiology (where I met again Edward Borowski who carried out an important research there), and since 1960 to University of Wisconsin, I extended my studies also to human cells, and investigated the nature of their DNA by combining physicochemical techniques with genetic transformation methods, to further understand the relationship between the chemistry of DNA and its biological/genetic function. That led to our discoveries of (i) DNA transformation of human cells (gene therapy), (ii) *in vitro* synthesis of the biologically active transforming DNA (normal and chemically modified by halogenation, glucosylation *etc*.), (iii) the preparative separation of DNA strands, and (iv) their application in studying development of phage λ, (v) its heteroduplex mapping and (vi) transcriptional mapping [7].

2. Physical maps of DNA

Physical mapping was our first approach to precisely characterize the large DNA molecules at the time when efficient nucleotide sequencing techniques were not yet developed. Mapping permitted localization of genetic rearrangements and thus also to measure the size and position of genes while defining these in the terms of base pairs.

(i) Electron microscopy(EM) of heteroduplexes

As soon as Kleinschmidt's methods of EM visualization of DNA molecules were perfected, they were introduced to Madison Campus by Hans Ris and his talented student, Barbara Chandler. It occurred then to us that by annealing separated DNA strands from various deletion or substitution mutants of phage λ (which we were preparing routinely at that time) it would be possible to create heteroduplex molecules in which the positions of deletions could be precisely measured in the relation to the molecule ends. Deletions would be visualized as single-stranded loops originating in one point of double-stranded DNA, and substitutions as regions of single-stranded non-homology. I contacted Hans about possible collaboration, but found out that he had just 'fired' Barbara, who in the meantime got married and changed her name to Barbara Westmoreland. I easily convinced Hans to 'un-fire' Barbara, and our series of exciting experiments has began. In the first experiment, I annealed at 65°C the DNA strands of wild-type λ with λ mutant *b*2*b*5. After attempting to visualize the heteroduplexes by EM, Barbara declared that it was a failure (actually she used a much stronger word), since there was to much DNA, all of it was broken and/or tangled and the background was very grainy. She suggested to use ten times less DNA, an ammonium carbonate buffer which would totally evaporate under EM vacuum, and a milder method of annealing. I followed her suggestions by reducing tenfold the DNA concentration, and decided to use 50% formamide to reduce the annealing temperature to 37°C. The resulting experiment was an unbelievable success! The *b*2 element proved to be a beautiful single-stranded deletion loop and the *b*5 element a substitution, as visualized by single-stranded bubble of non-homology. The preparations were picture perfect [8]. What a joy! This work was followed for several years in our

laboratory by a series of talented investigators, including Mike Fiandt, Zdenka Hradecna, Grażyna Konopa, Liz Szybalski and the visitor, Ena Mosharafa.

(ii) Transcriptional mapping

As soon as physical maps become available using our heteroduplex mapping, we decided to apply them for mapping the position of phage λ transcripts, by quantitative hybridization of mRNA with various λ deletion mutants and plotting the amount of mRNA hybridized along the physical map of λ [9]. These were the first experiments on the precise locating the transcripts on the physical map, something, which at present consumes the effort of thousands of molecular biologists. However, this is now aided by the knowledge of the exact nucleotide sequence, something we did not have at our early times.

(iii) Restriction mapping

(a) Using polyacrylamide gel electrophoresis (PAGE)

The advent of restriction enzymes brought in the era of gene mapping by restriction analysis. The sizes of restriction fragments were measured by their migration on PAGE. Moreover, strategies became available for mapping the positions of restriction fragments along the physical map. Use of restriction fragments as probes also simplified transcriptional mapping, since it was often difficult to secure all the convenient deletion mutants.

(b) Using pulsed field gel electrophoresis (PFGE)

The original PAGE method was well adapted to measure the size of relatively small DNA molecules up to about 10 kb. To separate and size DNA molecules with sizes of many kb and up to several MB, it became necessary to use DNA migration through very long gels. Since very long PAGE were impractical, David Schwartz, still as a student of Bruno Zimm at UCSD, tried to guide DNA while tumbling them along a very long and winding road to separate large DNA fragments on a reasonably short gel [10]. This is the principle of the PFGE. In the past, we have successfully used this approach to separate all the yeast chromosomes and to prove that we could cut only one of them, when using our Achilles heel cleavage (AC) technique, as described by Koob and Szybalski [11,12] and reviewed by Szybalski [13]. (Two scientists from PG participated in this work: Prof. Józef Kur and Adam Burkiewicz [12]).

(c) Using optical mapping (OM)

It was again David Schwartz, who developed a new way of restriction mapping of individual DNA molecules. He and his colleagues used restriction endonucleases to cut DNA that was gently bound to glass surfaces [14]. Each cut was visualized as a small gap in the continuity of the stained DNA using fluorescence microscopy. Thus the lengths of individual fragments were measured and their end-to-end alignment directly observed.

3. Sequencing

Of many DNA sequencing methods, including a briefly popular purely chemical method, only the enzymatic method of Sanger, twice a Nobel laureate, is presently in use [15]. Its principle is to use short [about 20 nucleotide (nt) long] primers that match a known part of DNA sequence under investigation, and extend them with DNA polymerase into the unknown region for 500–1000 nt, while using a mixture of all four nucleoside triphosphates and one randomly terminating dideoxynucleoside triphosphate. Truncated products of the reaction are electrophoresed and the sequence is read by some of the constantly improving automated techniques.

At present there are two basic approaches: (i) sequencing only the ends of the totally random population of DNA fragments and then assembling the entire sequence (establishing the contigs and filling-in the gaps), or (ii) use of the so called 'primer walking', when after each determination of 500–1000 nt, the new primer is designed based on the end of the just determined sequence. Dr. T. Kaczorowski of Gdańsk has pioneered this second approach where the new primers were produced by ligating the appropriate hexamers from our pre-existing hexamer library [16]. This method, however, although quite suitable for automation, was not generally adopted for genomic sequencing.

The currently adopted method uses only two carefully designed primers, while a random population of DNA fragments is cloned in special vectors in which the primer-complementary sequences are located at both sides of the cloning site. The most sophisticated vectors, one kind of which we have designed and patented (single-copy for stability and multi-copy for providing more DNA for the actual sequencing [17,18]) comprise two different primer sites placed at each end of the cloned fragment.

As an alternative to the use of random cloning for determination of the end sequences, one could use also transposons that carry two divergent primer-complementary sites and which could insert at random, *in vitro* or *in vivo*. As described in section **3ii**, below, this approach is helpful for sequencing repetitive DNA.

The current strategy is to randomly sequence high enough numbers of 500–1000-nt segments of DNA; so as to cover several fold the entire DNA that is to be sequenced. The next step after sequencing, presently the most demanding task, is alignment of all the 500–1000-nt segments of the sequence as to obtain the contigs and assemble them to obtain the correct sequence of the entire DNA under investigation.

(i) Non-repetitive DNA: assembly of fragments into contigs and beyond

Such aligning is simply done by a computerized search for the sequence overlaps between the great multitude of the 500–1000-nt segments of sequence, as shown schematically in Fig. 1A. At present, this procedure is routinely used for sequencing the increasing numbers of various prokaryotic and eukaryotic genomes.

However, this approach fails for the long, over 500-nt stretches of repetitive sequence, because repetitions lead to spurious alignments. At present, the sequences of many eukaryotic genomes remain incomplete, since although many of them are

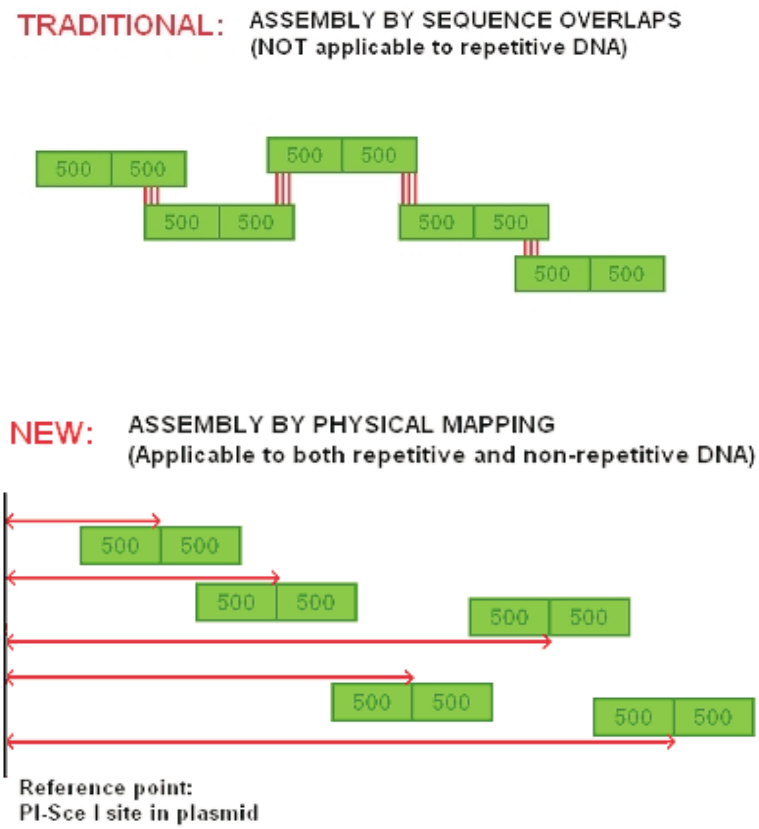

Figure 1. Assembly of a series of two adjoining 500-nt sequences determined and collected from sequencing the library of the random Tn*5* transposon insertions.
1A TRADITIONAL (upper figure): Assembly of the 500-nt sequences is performed by the traditional computerized search for the sequence overlaps that is well adapted for the non-repetitive DNA [Section **3*E***(I)] but would fail for the repetitive DNA [Section **3*E***(II)]. Short vertical lines represent the sequence overlaps.
1B NEW (lower figure): Assembly of the 500-nt sequences is performed by the establishment of the precise map of all the Tn*5* insertions. Vertical line at the left margin represents the reference PI-SceI site in the BAC plasmid core. Horizontal lines represent the length of the SceI-SceI fragments determined by PFGE (Sections **3ii*E*a-c**) or by any of the alternative approaches (Section **3ii*E*d**).

well sequenced, they contain a considerable number of what we call 'black holes' corresponding to the not yet sequenced repetitive DNA regions. The reason for these 'black holes' is that using present aligning approaches based on the sequence overlap it is impossible to align the newly acquired 500–1000-nt repetitive sequences, since often sequences of each of these 500–1000-nt stretches are identical or nearly identical. Typical examples of such highly and medium-repetitive sequences include the centromeric and other heterochromatic regions in higher eukaryotic genomes. The sequence segments from these regions of genome are impossible or nearly impossible

to assemble, and for this reason have not been resolved in the "Whole Genome Sequencing Projects" for many different organisms. Therefore, to enable completion of sequencing of many genomes, it became necessary to develop some innovative methods to deal with these repetitive sequences. To this end, we have just develop a novel transposon-initiated/primed sequencing systems [based on the Epicentre (Madison. WI, USA) EZ::TN™<oriV/KAN-2> transposon system], that would allow sequencing the repetitive DNA. The next segment of this review represent our new, hitherto unpublished research as conceived in 2002-3 and being executed in my lab in 2003-4 in co-operation with a Madrilène group specified in Acknowledgments and in Mendez Lago *et al*. [19].

(ii) Repetitive DNA: Combined mapping and sequencing of 500–1000-nt fragments permit their assembly into the final sequence

It was soon realized that the sequencing task could not be totally completed because many genomes, especially the eukaryotic ones, contain numerous long stretches of repetitive sequences, which defy the present aligning methods required for sequencing. Therefore, we had to develop novel strategies, which are for the first time outlined here. These comprise the use of specially constructed transposons and pBAC/oriV vectors, which we have equipped with very rare cutting sites [17,18]. Transposons allow to determination of 500–1000-nt sequences on both sides of their insertion site, whereas the very rare cutting sites [like I-SceI or PI-SceI, or our AC sites] allow to map precisely the positions of the insertions, using the PFGE or other physical means, including heteroduplex EM mapping or OM. Thus we had to return to some of our earlier methods of physical mapping, which together with transposon-associated priming would allow completion of sequencing of large eukaryotic genomes. That would amount to the final triumph of the structural chemistry of the DNA macromolecules and thus genomes.

It might be amusing to recall here that it was once believed that the DNA sequence is a repeated tetramer of all four bases. This was proposed in 1920 by Phoebus A.T. Levene, and was used as argument that DNA must be a structural polysaccharide component and thus of no genetic importance.

Below, I will describe step-by-step the essence of our sequencing procedure for repetitive DNA on the actual examples (see also [19]):

(A) Cloning of DNA fragments to be sequenced in their entirety

A 100-200 kb DNA fragment, either non-repetitive (control) or repetitive (180 kb from *Drosophila* heterochromatic centromeric region), was cloned into a bacterial artificial chromosome (BAC) vector or into our pBAC/oriV vector, the latter permitting the amplification of DNA prior to any next step. For our purpose these vectors must contain in their backbone a very rare restriction site, *e.g.*, PI-SceI or I-SceI.

(B) Construction of the Tn5 transposon with PI-SceI and/or I-SceI very rare sites

Dr. Jadwiga Wild has modified the Epicentre EZ::TN™<oriV/KAN-2> system by adding the very rare restriction sites PI-SceI or I-SceI. These transposons contained also a selection marker (kanamycin resistance or KAN or Km^R), the *oriV* origin of replication (requiring the TrfA replication-initiating protein), and two divergent priming sites.

One could also use other very rare restriction sites of intronic enzymes or our AC or RecA-AC sites [13].

(C) Transposon insertion library

Using an *in vitro* transposition procedure, Maria Mendez Lago has decorated the BAC clones with our modified transposons, which were to be inserted on average at every 400 bp. Such plasmids do contain now two rare restriction sites, one on the BAC backbone and the other next to the priming sites in Tn*5*. That allowed us a precise measurements of the distance between these two sites (corresponding to the priming sites and the reference point on the BAC backbone), by simply measuring the length of the SceI-SceI restriction fragments, as detailed below. The BAC clones after acquiring the Tn*5*/*oriV* transposons became amplifiable when grown in the *trfA*-carrying hosts described by Wild *et al*. [18].

(D) Sequencing with primers complementary to Tn5 priming sites

Using two kinds of divergent primers, two 500-nt sequences were determined for all the clones, in a routine manner.

(E) Assembly of the 500–1000-nt sequences

(I) For the **non-repetitive** (control) DNA clones

Two procedures could be used for assembly of non-repetitive DNA:

(a) Aligning is simply done by a computerized search for the sequence overlaps between the great multitude of the 500–1000-nt segments of sequence, as shown schematically in Fig. 1A and described in Section **3a**;

(b) Instead of using sequence overlaps, the position of all the 500-nt fragments is determined from the distance between each variable Tn*5* priming site and the reference point on the BAC backbone, as described in section **(II)**, below.

(II) For the **repetitive** DNA clones

In this case, a computerized search for the sequence overlaps, as mentioned in section (***E*** I a) above, cannot be used; therefore, we have developed and further refined the procedure as outlined below.

(a) Length measurements by PFGE

Based on our earlier experience with PFGE [11], we used this method for measuring the length of the SceI fragments. The BioRad Gene Mapper permits optimization of electrophoresis conditions for the specific length of the SceI fragments, which when using appropriate assembly marker sizes allows measurement of the length of SceI-SceI fragments with about 1% precision, *i.e.*, 1 kb for the 100-kb fragments.

This should allow assembly of a map including all the 500–1000-nt sequences along the BAC clone. Upon SceI-mediated cutting the BAC clone, now with two SceI sites (one on BAC backbone and second on the inserted Tn*5*), two DNA bands are visible on the PFGE gels, and the size of these two has to be precisely determined. The two fragments are identified by Southern blotting using BAC-complementary probes.

(b) Tn*5* orientation

Since Tn*5* does insert with any of two orientations, it is necessary to determine the orientation of each transposon in every BAC clone. To identify these two orientations, Southern blots with Tn*5*-complementary probes had to be performed, using the same PFGE gels and the same techniques as in section (***E*** I), above.

These (***E*** I+II) Southern blots define the exact structure of each Tn*5*-decorated clone and both 500-nt sequences obtained with each of the two priming sites.

(c) Map of Tn*5* priming sites

All these measurements and blotting permit establishment of the exact map of all transposon inserts with their Tn*5* priming sites and thus in turn allow alignment of all of the 500-nt sequences along the BAC clone. Thus this map is the key to our procedure and is used instead of the overlap method, when the latter is impossible, as for the highly repetitive DNA. Our Tn*5* mapping procedure, as outlined in Fig 1B, permits the entire sequence of any DNA to be determined irrespective of whether it contains repetitive sequences or not.

(d) Alternative methods for the length measurements and establishment of a precise map of Tn*5* insertions

In addition to the PFGE, other methods are available for physical measurements of DNA length. The use of EM gives measurements considerably more precise than PFGE, but the method is cumbersome, unless some automation be developed.

Automation seems to be available for OM, but the precision of direct measuring might be too low because the size of fragment is very dependent on hydrodynamic forces. However, when as a part of the OM procedure one cuts the DNA fragments to be measured with restriction enzyme(s), and then automatically aligns all the fragments and gaps, the measurements will be necessary only for the terminal fragment, nearest to the Tn*5*, whereas all the others, as aligned by OM, will be in common. Thus the precision will be greatly increased because the actual measured fragments will be small. Instead of cutting with restriction enzyme(s) and aligning he gaps, one could mark DNA with one or more sequence-specific agents like methyl transferase or the oligo-RecA complexes [13], which could be made highly and individually visible by proper illumination and magnification. Such methods should be amenable to automation [20].

Another approach will be to establish first an OM map of the repetitive clone using several alternative restriction enzymes, and select the one which gives the most suitable restriction pattern. This would be followed by measuring the size of the SceI-SceI fragments using the following series of steps: (i) labeling the Tn*5*-proximal ends by filling-in the SceI site, (ii) partial digestion with the selected enzyme, and

(iii) PAGE and Southern blot of the products according to the principle of Smith and Birnstiel [21]. The next step would be the alignment of all these partial-digest PAGE-fractionated fragments and comparison of these SceI-SceI fragments, which should establish the map and thus the lengths of all the fragments. Consequently, this would permit to determine the entire sequence.

There is a considerable room for improvement of the methods for measuring of DNA length.

Epilogue

Research is not always moving along straight lines. As shown here, it is sometimes useful to remember the past kinds of experimentation, return to them, and use them to solve new problems, by combining them with new approaches. Thus combination of our experience with transposon elements together with precise mapping has permitted us to develop a novel method for sequencing the highly repetitive regions, for which the more modern sequencing strategies have failed.

Dedication

This paper is dedicated to honor the 75-th birthday of my very dear friend, Professor Dr. Edward (Edzio) Borowski [D.hc.], whom I first met in Gdansk as my student at the PG just after the WWII, then in 1950-s at the Institute of Microbiology at Rutgers University, and also many times at the PG and in other parts of the world. We both shared interest in biotechnology and antibiotics, and were honestly devoted to the applications of the true chemistry to the betterment of human health and life in general.

Edzio was also a graceful promoter of my D.hc. at the PG in 2001. I sincerely hope that our fruitful contacts will continue.

Acknowledgments

I am very thankful to Maria Mendez Lago and Dr. Alfredo Villasante, the members of the Madrilène group, and to Dr. Jadwiga Wild for the permission to cite our unpublished results and use the Figures prepared by Maria M.L. Dr. Peter Kay has helped greatly with editing this manuscript.

REFERENCES

1. Avery O.T., MacLeod C.M. and McCarty M., *J. Exp. Med.*, **79**, 137, (1944).
2. Szybalski W., in Stahl F.W., (Ed) *We Can Sleep Later*, *Alfred D. Hershey and the Origins of Molecular Biology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2000.
3. Szybalski W., *Genetics*, **158**, 1, (2001) http://www.genetics.org/cgi/content/full/158/1/1.
4. Szybalski W., *Acta Biochem Polonica*, **50** (No.2), pp. XVII–XXI, (2003).
5. Fiandt M., Szybalski W. and Malamy M.H., *Mol. Gen. Genet*., **119**, 223 (1972).
6. Bryson V. and Szybalski W., *Adv. Genet*., **7**, 1 (1955).
7. Blattner F.R., Burgess R.R., Dove W.F., Filutowicz M., Gourse R.L., Hradecná Z., Szybalski-Sandor B.A. and Wild J., *Gene*, **223**, 395 (1998); see also http://mcardle.oncology.wisc.edu/faculty staff//WSPubl.html.
8. Westmoreland B.C., Szybalski W. and Ris H., *Science*, **163**, 1343 (1969).
9. Hayes S. and Szybalski W., *Mol. Gen. Genet*., **126**, 275 (1973).
10. Schwartz D.C. and Cantor C.R., *Cell*, **37**, 67 (1984). See also Joppa B., Li S., Cole S. and Gallagher S., http://www.nal.usda.gov/pgdic/Probe/v2n3/puls.html.
11. Koob M. and Szybalski W., *Science*, **250**, 271 (1990).
12. Koob M., Burkiewicz A., Kur J. and Szybalski W., *Nucleic Acid Research*, **20**, 5831 (1992).

13. Szybalski W., *Curr. Opin. Biotechnology*, **8**, 75 (1997).
14. Aston C., Hiort C. and Schwartz D.C., *Methods Enzymol.*, **303**, 55 (1999).
15. Sanger F., Nicklen and Coulson A.R., *Proc. Natl. Acad. Sci. USA*, **74**, 5463 (1977).
16. Kaczorowski T. and Szybalski W., *Gene*, **223**, 83 (1998).
17. Szybalski W., Wild J. and Hradecna Z., US Patent No. 6, 472, 177.
18. Wild J., Hradecna Z. and Szybalski W., *Genome Research*, **12**, 1434 (2002).
19. Mendez Lago M., Wild J., Abad J.P., Martin-Gallardo A., Villasante A. and Szybalski W., *45th Annual Drosophila Research Conference*, March 24–28, 2004, *Abstract 981C*, p. 30, 2004. http://www.drosophiconf.org/.
20. Jing J., Reed J., Huang J., Hu X., Clarke V., Edington J., Housman D., Anantharaman T.S., Huff E.J, Mishra B., Porter B., Shenker A., Wolfson E., Hiort C., Kantor R., Aston C. and Schwartz D.C., *Proc. Natl. Acad. Sci. USA*, **95**, 8046 (1998).
21. Smith H.O. and Birnstiel M.L., *Nucleic Acids Res.*, **3**, 2387 (1976).