Proc. Estonian Acad. Sci. Phys. Math., 1998, **47**, 4, 282–289
https://doi.org/10.3176/phys.math.1998.4.04

A PROPOSAL TO USE SOLID STATE SPECTROSCOPY FOR DNA SEQUENCING

Karl K. REBANE

Institute of Physics, University of Tartu, Riia 142, 15014 Tartu, Estonia; e-mail: rebanek@fi.tartu.ee

Received 9 February 1998

Abstract. Single impurity molecule detection and spectroscopy open new possibilities for sequencing of nucleotides in a DNA molecule by means of laser spectroscopy. In addition to the approach realized by W. P. Ambrose, P. M. Goodwin and other researchers (*Exp. Tech. Phys.*, 1995, **41**, 237–248), two further steps are proposed: 1. to fix the jet with the sequence of nucleotides on a proper solid carrier and to form the identification card, and 2. to cool the identification card to liquid helium temperatures and thus to increase the spectral selectivity and reliability of the study by several orders of magnitude.

Key words: DNA sequencing, single molecule spectroscopy, zero-phonon lines, identification card.

1. INTRODUCTION

The motivation of this paper comprises two points: 1. to propose the methods of solid state spectroscopy for the study of DNA sequencing and other DNA properties [1,2]; 2. to display an important potential application of single molecule detection (SMD) and, especially, single impurity molecule spectroscopy (SMS) [3]. It is shown that not only SMD [4–6] but also SMS, besides being a novel exciting academic topic in solid state physics and chemistry, opens prospects for essential applications in various fields, in particular, in molecular biology.

Optics provides a possibility of parallel data storage and processing and, consequently, a potentially very high speed and enhanced reliability (see, e.g. persistent spectral hole burning space-and-frequency-domain [7,8] and time-and-space-domain holographies [9–14]).

The main obstacle in the application of conventional optics to DNA sequencing is that the distances between adjacent nucleotides d in a DNA molecule are hundreds of times shorter than the wavelengths of visible light λ

($\lambda >> d$). Thus the conventional optical resolution in space is hundredfold less than required to investigate the positions of nucleotides in DNA.

In [4–6] it is shown that the nucleotides can be cut off one-by-one from the free end of an anchored DNA molecule and carried away with a stream of liquid. The sequence in the stream is supposed to be the same as in the DNA molecule, but the distances between adjacent nucleotides can be made much larger than the wavelength λ. Therefore, SMD can be applied to recognize the kind of a nucleotide (one or all four kinds of nucleotides are marked with molecules emitting characteristic fluorescence) and to find the sequence.

In this paper two further steps as compared to [4–6] are proposed. First, to transfer the nucleotides preserving their order in the sequence from the liquid stream onto a solid carrier to form an *identification card* (IC). The nucleotides on the IC can be identified and the sequence recognized by means of solid state SMD.

The second step is to cool the IC down to liquid helium temperatures, i.e. 4.2 K and lower, and thus to increase the spectral resolution [14–19]; SMD is to be replaced by SMS. In principle, a spectral resolution higher by 4–5 orders of magnitude can be achieved. This will make the results more reliable and provide possibilities to check and to some extent correct the errors which might have been collected in the procedure of cutting, transporting, and fixing the nucleotides. The prerequisite for applying SMS of very high spectral resolution is that markers are found which have narrow and intense *zero-phonon lines* (ZPL) in their absorption spectra [1,2,19–21].

2. IDEAL SCHEME

The scheme comprises the following steps. (Later we will discuss possible deviations from the ideal case and also the ways how to overcome them to some extent.)

The first four steps 2.1–2.4 are the same as in [4–6].

2.1. The DNA molecule, in which one or all four kinds of the nucleotides are marked with molecules emitting spectrally specific fluorescence, is anchored by one end to a base (a polymeric bead) and placed in a flow of liquid.

2.2. Using enzymes, the nucleotides are cut off one-by-one from the free end of DNA and carried away in the stream preserving the sequence as it was in DNA.

2.3. The stream is hydrodynamically organized into a thin and well-oriented jet. The velocity of the jet is matched with the frequency of the cutting procedure in such a manner that the distances between the nucleotides in the jet considerably exceed the fluorescence excitation wavelengths and also the size of the spot illuminated by the exciting light to avoid simultaneous excitation of two or more nucleotides.

2.4. The jet with nucleotides is directed through the spot of laser illumination which excites the spectrally specific fluorescence. The kind of the nucleotide is recognized by SMD. Repeating the detection for all the marked nucleotides reveals the sequence in the jet which is supposed to be the same as it was in DNA before cutting.

Laser light has to be sharply focused in space, first, to provide high excitation intensity required for SMD and, second, to guarantee that only one nucleotide is under excitation at one time of measurement. No very high spectral sensitivity is required, because at room temperature the bands of absorption and fluorescence are supposed to be broad – hundreds and thousands of cm^{-1} of bandwidth.

To continue the scheme into the domain of solid state physics and spectroscopy, including the very high resolution ZPL spectroscopy, I propose to proceed as follows.

2.5. Let the jet hit the surface of a solid substance (band or plate of polymer, paper, gel, etc.) moving with a proper velocity v through the spot of hitting in such a manner that the nucleotides are caught into the solid material and fixed on its surface or in the bulk near to the surface. The sequence in the liquid jet is transformed into a sequence fixed and preserved in solid material – this is the identification card (IC).

Changing the speed v enables one to vary the distances between the nucleotides and select the proper ones. If the frequency of nucleotides cut-off v is 10^3 sec^{-1}, the velocity v of 1 cm per second places the nucleotides at the distance $v/c = 10^{-3}$ cm, i.e. at 20 wavelengths of the light of $\lambda = 5 \times 10^{-5}$ cm. The IC of DNA or a section of it is convenient to preserve and to study by optical methods. Nucleotides can be recognized in the first place more or less in the same way as in [4–6], i.e. by applying SMD of relatively low spectral resolution.

An obvious benefit is that the sequence fixed on the IC is not destroyed after the first measurement but can be preserved for a long time. The IC can be investigated repeatedly in various aspects, using various methods, in different laboratories. The reliability can be enhanced, more information from a cut into nucleotides DNA extracted.

Optical methods can provide large-scale parallelism of measurement and data processing. The ICs of two DNAs, considered to be identical, can be prepared and relatively easily compared with each other. The comparison of two supposedly identical ICs from the same source (patient), prepared from DNAs taken at different periods of her or his life, seems also exciting.

Note that it is possible to mark and additionally mark the nucleotides or change their marking on the IC. This can also be done at elevated or lowered temperatures.

Cooling of the IC, in which the conventional markers not having narrow and intense lines of absorption are used, can increase to some extent the spectral resolution and also enhance the coherence, useful for parallel processing. A further possible advantage is that the yield of fluorescence of some molecules

may increase by orders of magnitude with decreasing temperature and thus new classes of markers can be found.

Actually, an IC comprises more information than can be read out by means of SMD. The spectral selectivity of investigating the IC can be enhanced by 4–5 orders of magnitude if we proceed from SMD to SMS. An important prerequisite for high-resolution SMS, as already mentioned above, is to find the markers which bring into the absorption spectra very narrow and intense ZPLs – the optical analogues of the Mössbauer γ-resonance line (see, e.g., [15,17]). The prospects of finding that kind of markers are satisfactory: thousands of impurity centres, formed by ions, atoms, lattice defects, and, especially, organic and inorganic molecules, introduced in thousands of various glassy, single crystal, polycrystalline, polymer matrices, display fine ZPLs in their low-temperature (i.e. usually below 10 K) spectra of absorption (and also fluorescence). Some guidelines from chemistry, where organic molecules could favourably display ZPLs, are given in [22,23]. The main optical criterion is the smallness of the Stokes' shift between the conjugated absorption and fluorescence bands [15,17].

Once the proper markers are found, the IC has to be cooled to liquid helium temperatures and studied at the level of the very high accuracy provided by SMS which actually is a chapter of ZPL spectroscopy.

Note that really large organic molecules, e.g. proteins or DNA, can have very high density of low-frequency excitations (compared to that of small- or medium-size molecules which have molecular weight up to a few hundreds). These are thermally excited even at 1 K. There is a possibility that in case of such really large molecules as impurities in solids good ZPLs show up only when the IC is cooled down to deci-, centi- or millikelvins. It is also useful to apply high hydrostatic pressure to suppress spectral diffusion (caused by dynamic inhomogeneity). Thus, good hopes to have nice ZPLs even in proteins do exist, provided the Stokes' shift in spectra is not too large.

In the first place, very narrow ZPLs provide very high spectral resolution. Further, as is well known, the homogeneous linewidth Γ_{hom} is bound to the phase memory time τ_2 or the coherence time of the excited electronic state, $\tau_2 \sim \Gamma_{hom}^{-1}$. More precisely, the ZPL's homogeneous linewidth is [18,19]:

$$\Gamma_{hom}(T) = 1/\pi\tau_2(T) = 1/\pi\tau_2^*(T) + 1/2\pi\tau_1(T),$$

where T is temperature, τ_2 is the full dephasing time, τ_2^* is the pure dephasing time, and τ_1 is the energy relaxation time.

For quite a number of molecular (and other) impurities at 2 K the characteristic homogeneous ZPL's linewidth is about Γ_{hom} ($T = 2$ K) ~ $10^{-3} - 10^{-4}$ cm^{-1}, which gives the estimate for $\tau_2(T)$ about a nanosecond. This is quite a sufficient time interval to initiate and measure coherent optical phenomena, e.g. photon echoes, time-and-space-domain holography. Note that, besides a narrow homogeneous linewidth, these phenomena require a broad inhomogeneous one.

Because of the inhomogeneous structure of the IC's basic material, into which the nucleotides are introduced, the inhomogeneous broadening should be large enough (see, e.g., [12,17,18,20]). Thus, besides the very high spectral resolution, the narrow ZPLs provide also a good possibility of studying the IC via coherent phenomena.

2.6. As mentioned above, the read-out of the IC at a very high spectral resolution level can help overcome the flaws collected in the process of making the IC.

2.6.1. The frequency positions of ZPLs are, because of their narrowness, very sensitive to the state of the solid matrix, and not only in the vicinity. Single impurity molecule spectroscopy experiments show that a replacement of a neighbouring impurity molecule at a distance of up to tens of nanometres away by another slightly different molecule can shift the position of ZPL by more than its linewidth. Thus the spectral resolution provides a powerful tool not only to identify the nucleotide but also quite in detail tells about the environment of it.

If not a single nucleotide but a larger section of DNA, comprising 2, 3 or more nucleotides, is cut off, the high spectral resolution of ZPLs allows us to well understand the situation and identify not only the nucleotides but also their positions in the section. In principle, it is even possible to tell at which site the marked nucleotide is actually positioned in quite a big piece of DNA. If the non-controllable inhomogeneity was small, that kind of identification should be possible for a DNA section comprising a hundred and even more nucleotides [12].

2.6.2. If some nucleotides happen to be positioned nearer to each other than the wavelength λ, the spectral resolution helps again correctly understand the situation and recognize the nucleotides [2].

2.6.3. The nucleotides are in fact implanted not along a line narrower than the wavelength but with some random side deviations, and actually the distribution of nucleotides on the IC covers a broader band than the diameter of the sharply (close to the diffraction limit, i.e. excitation wavelength λ) focused laser beam used in SMD.

Single impurity molecule *microscopy* provides a considerably better possibility [24]. In this version of SMS the spectrally highly selective laser beam illuminates an area (with the diameter D) much broader than λ ($D >> \lambda$) of the IC's surface, e.g. $D^2 = 0.1 \times 0.1$ cm^2. Many nucleotides are simultaneously under illumination, but only that fraction of them whose ZPLs are in resonance with the laser frequency is actually excited and emits specific fluorescence. By means of highly selective spectral scanning of the laser frequency the fluorescence via ZPL excitation spectra of all of the selected nucleotides can be measured, the kind of each nucleotide recognized, and its position in the illuminated space-frequency field found. Actually, in this approach a detailed frequency-and-space-domain picture of the distribution of the nucleotides is measured. The nucleotides which have deviated from the main line of implantation farther than

D can be found in one or in a few additional spatial scans of relatively low accuracy (each accompanied by scanning of the laser frequency). The approach via SMS microscopy seems especially attractive when applied in concert with comparison of two ICs made from two supposedly identical DNA molecules.

2.7. Four remarks.

2.7.1. The identification card can be preserved for a long time; if at low temperatures, then for a very long time, exceeding well the DNA source's (patient's) lifetime. As already mentioned above, the IC can be repeatedly marked, additionally marked, the marking changed.

2.7.2. Probably a proper way to transfer the nucleotides to the IC and fix them on/in it could also be to perform the cutting inside the IC's cover substance, which at the free end of DNA is kept liquid or gel-like. After the cut-off nucleotide is carried a little away, the substance is frozen solid. The other option is to shoot the nucleotide directly to the IC's surface passing at close distance to DNA.

2.7.3. The line, along which the nucleotides are implanted, can be of arbitrary configuration. A reasonable way seems to organize it as a spiral, analogically to the path of planting information on a compact audio disc (CD). The IC will be an analogue of the CD. The technique to control the movement of such a disc has been well worked out for the computer and audio CDs. Note that the information comprised in a DNA sequence (a few billion bits) is of the same order of magnitude as that of the music on an audio CD.

2.7.4. There can be a specific potential advantage for the future: if the study and manipulation of the natural original DNAs and the sequence of nucleotides in them is restricted or banned, good hopes do exist that to make and study ICs will stay free.

REFERENCES

1. Rebane, K. K. Can solid state high resolution spectroscopy contribute to DNA sequencing studies? *Exp. Tech. Phys.*, 1995, **41**, 2, 295–302.
2. Rebane, K. K. The possibilities of ultrahigh-resolution solid-state spectroscopy in the DNA nucleotide sequencing. *Phys. Solid State (USA)*, 1996, **38**, 11, 1902–1903.
3. Kador, L. Recent results of single-molecule spectroscopy in solids. *Phys. Stat. Sol. (b)*, 1995, **189,** 11–36.
4. Ambrose, W. P., Goodwin, P. M., Martin, J. C., and Keller, R. A. Single molecule detection and photochemistry on a surface using near-field optical excitation. *Phys. Rev. Lett.*, 1994, **72,** 1, 160–163.
5. Ambrose, W. P., Goodwin, P. M., Martin, J. C., and Keller, R. A. Alterations of single molecule fluorescence lifetimes in near-field optical microscopy. *Science,* 1994, **265**, 364–367.
6. Goodwin, P. M., Affleck, R. L., Ambrose, W. P., et al. Imaging biological molecule sensitivity using near-field scanning optical microscopy. *Exp. Tech. Phys.*, 1995, **41**, 237–248.

7. Meixner, J., Renn, A., and Wild, U. P. Spectral hole-burning and holography. I. Transmission and holographic detection of spectral holes. *J. Chem. Phys.*, 1989, **91**, 6728–6736.
8. Renn, A., Meixner, J., and Wild, U. P. Spectral hole-burning and holography. II. Diffraction properties of two spectrally adjacent holograms. *J. Chem. Phys.*, 1990, **92**, 2748–2755.
9. Rebane, A., Bernet, S., Renn, A., and Wild, U. P. Holography in frequency selective media: Hologram phase and causality. *Opt. Commun.*, 1991, **86**, 7–13.
10. W. Mossberg. *Opt. Lett.*, 1982, **7**, 77.
11. Rebane, A., Kaarli, R., Saari, P., Anijalg, A., and Timpmann, K. Photochemical time-domain holography of weak picosecond pulses. *Opt. Commun.*, 1983, **47**, 173–176.
12. Saari, P. M., Kaarli, R. K., and Rebane, A. K. Picosecond time-and-space-domain holography by photochemical hole burning. *J. Opt. Soc. Amer. B*, 1986, **3**, 527–533.
13. Ollikainen, O., Rebane, A., and Rebane, K. Error-corrective optical neural network modelled by persistent spectral hole-burning. *Opt. Quantum Electron.*, 1993, **25**, 569–585.
14. Sild, O. and Haller, K. (eds.). *Zero-Phonon Lines and Spectral Hole Burning in Spectroscopy and Photochemistry*. Springer-Verlag, Berlin, 1988.
15. Rebane, K. *Impurity Spectra of Solids*. Plenum Press, New York, 1970.
16. Moerner, W. E. (ed.). *Persistent Spectral Hole Burning: Science and Applications*. Springer-Verlag, Berlin, 1988.
17. Rebane, K. Zero-phonon lines in the spectroscopy and photochemistry of impurity-doped solid matter. In *Zero-Phonon Lines and Spectral Hole Burning in Spectroscopy and Photochemistry* (Sild, O. and Haller, K., eds.). Springer-Verlag, Berlin, 1988, 1–19.
18. Rebane, K. K. and Rebane, L. A. Basic principles and methods of persistent spectral hole-burning. In *Persistent Spectral Hole Burning: Science and Applications* (Moerner, W. E., ed.). Springer-Verlag, Berlin, 1988, 17–77.
19. Rebane, A. and Rebane, K. K. Spectral hole burning and optical information processing. In *Current Trends in Optics* (Dainty, J. C., ed.). Academic Press, London, 1994, 177–194.
20. Rebane, K. K. Zero-phonon line as the foundation stone of high-resolution matrix spectroscopy, persistent spectral hole burning, single impurity molecule spectroscopy. *Chem. Phys.*, 1994, **189**, 139–148.
21. Rebane, K. K. and Rebane, A. Persistent spectral hole burning: Time- and space-domain holography. In *Molecular Electronics* (Mahler, G., May, V., and Schreiber, M., eds.). Marcell-Dekker Inc., New York, 1996, 257–302.
22. Renge, I. Correlation of the zero-phonon electronic transition probabilities (Debye–Waller factors) for molecular impurity centres in amorphous hosts with spectral matrix shifts. *Proc. Estonian Acad. Sci. Phys. Math.*, 1991, **40**, 3, 189–197.
23. Renge, I., Wolleb, H., Spahni, H., and Wild, U. Phtalonaphthaloeyanines: New far-red dyes for spectral hole burning. *J. Phys. Chem. A*, 1997, **101**, 35, 6202–6213.
24. Güttler, F., Irngairtinger, T., Plakhotnik, T., Renn, A., and Wild, U. P. Fluorescence microscopy of single molecules. *Chem. Phys. Lett.*, 1994, **217**, 4, 393–397.

ETTEPANEK KASUTADA TAHKISESPEKTROSKOOPIAT NUKLEOTIIDIDE JÄRJESTUSE MÄÄRAMISEKS DNA MOLEKULIS

Karl K. REBANE

Ühe lisandimolekuli spektroskoopia (ÜMS) ja detekteerimine (ÜMD) avavad uusi võimalusi lahendada molekulaarbioloogia keskset probleemi – nukleotiidide järjestuse määramist DNA molekulis.

Töödes [4–6] on realiseeritud nukleotiidide ükshaaval järjestikku äralõikamine üksiku DNA molekuli ühest otsast ja nende ärakandmine peene vedelikujoa poolt, milles nukleotiidide vahekaugus on tublisti suurem detekteeriva valguse lainepikkusest. Juga suunatakse laserkiire fookuspunkti, kus teostub ÜMD – spektraalselt selektiivselt ergastatakse igale markerile iseloomulikku luminestsentsi, mis identifitseeribki nukleotiidi liigi.

Käesolevas töös on pandud ette minna võrreldes töödega [4–6] kaks sammu kaugemale. Esiteks, kanda nukleotiidid vedelikujoaga tahkele kandjale säilitades nende järjestuse joas ja moodustada *tunnuskaart*. Teiseks, kui on kasutatud markereid, mille spektris on head foononvabad jooned, jahutada tunnuskaart vedela heeliumi temperatuurideni (s.o. alla 4,2 K), kus foononvabad jooned on väga kitsad ja kõrge tipuintensiivsusega, tõstes seega luminestsentsi ergastamise intensiivsust ja detekteerimise täpsust 4–5 suurusjärku, ühtlasi saades võimaluse rakendada ka koherentseid optilisi meetodeid.