30th Annual International IEEE EMBS Conference
Vancouver, British Columbia, Canada, August 20-24, 2008

Characterization of DNA Bio-bonds for Meso-Scale Self-assembly

A. Abbaci [μ], B. Daunay [μ], D.S. Haliyo [μ], S. Régnier [μ], R. Boyd* and A. Cuenat*
[μ] *Institut des Systèmes Intelligents et de Robotique (ISIR)*
Univ. Paris 6 - CNRS, BP 61, 92265 Fontenay aux Roses, email: ahlem.abbaci@isir.fr
** National Physical Laboratory (NPL), Hampton Road, Teddington, Middlesex, TW11 0LWNPL*

***Abstract*— In this paper, we have investigated the use of DNA hybridization as the basis for the production of new mesoscale components. AFM experimental results are studied and compared to two theoretical approaches: molecular and thermodynamic. We explain how and why DNA hybridization process can provide a good bond to self assemble components, and how molecular modelling methods allow further understanding of the physical mechanism of this process.**
Furthermore, the strength interaction of DNA complementary strands is measured and analyzed using statistical tools. These results are then compared to the theoretical approaches.

I. Introduction

While techniques for synthesizing nanostructures at the molecular level and manufacturing complex forms in micro-scale form bulk is progressing, assembling parts and components from different materials remains an important challenge in nano- and micro-technology. That's why self-assembly becomes an interesting approach for the assembly of micro- and nano-scale components.
Self-assembly is defined as "Spontaneously generating order in a system of components" [1]. It allows us to surpass some of the limitations of traditional techniques of assembly. These are the complexity of the assembly process and the manipulation of too small components. In the past ten years, many research laboratories started to explore this idea and today, self-assembly is becoming an increasing popular field of research. The main idea is to place several components with some specific characteristics (electrostatic, photonic, geometric, ...etc.), in a particular environment, using specific interactions between components to place them. In spite of the growing interest in this approach, the existing techniques in this domain lack maturity for inclusion into industrial development processes. Mainly, there are important constraints in material and geometry of the components which make the process more complicated than individual pick-and-place operations.
Capillary forces seems to be a popular solution to guide the self-assembly process [2]. Alternatively, different self-assembly procedures based on other forces are investigated: electric force field in [3] or magnetic forces in [4]. In general, works in the literature on self-assembly do not lean on any measurements or any mechanical and physical characterization of the process. It becomes obvious and necessary to orientate the development of new technologies in the self-assembly field.
The use of the biological assembly processes as they are found in the nature appears like an interesting solution. Mimicking or integrating existing biological processes in artificial self-assembly would be an efficient approach. The dimensions of the involved molecules being rather in nano-scale, the self-assembly approaches differ if the assembled components are in micro [5] or nano-scales [6][7][8].
This paper is focused on both experimental and theoretical approaches. We study the binding of two components; each of them has one functionalised surface with tethered oligos. The oligos tethered on a component's surface are complementary to the ones tethered on the opposite surface. The molecular, experimental and statistical aspects are exposed to explain the different difficulty levels. The first section is focused on the structural bio-bond aspects. The second section deals with the experimental aspects of the considered system, and exposes the statistical analysis performed on the obtained data. In the last section, different theoretical approaches are considered.

II. DNA as bio-bond for self-assembly

Today, the DNA hybridization is the one of the most promising biological process for self-assembly purposes. To characterize this process suitable for meso-scale self-assembly, we must consider the DNA structure. It can be regarded as written with four-letter alphabet of 4 letters A, T, C, G which can offer a great programming potential allowing geometrical conformations and specific recognitions between components to assemble.
The DNA molecule is composed of two strands: two nucleotidic chains which form the double helix. The nucleotides in each chain are always complementary. The diameter of the double helix is about 2 nm. The length between two bases on one chain is about 0.34 nm, and 10 bases approximately form one helix tour. The complementarity of the two chains is materialised by the hydrogen bonds which are more resistant than a Van der Waal interaction ($\sim$x10). Its length is about 2 Å. DNA hybridization assemble two complementary DNA strands under particular conditions. This process has been used for a long time in DNA micro-arrays technology [9]. It was shown that this biological process is controlled by some key parameters. Each of these parameters have been studied and defined in the literature: The environmental parameters such as **temperature** assessed from the chemical proprieties of the solution and the DNA sequence [10] using the Nearest Neighbor model [11], the **ionic composition** of the solution, especially salt concentration, and the intrinsic

DNA parameters such as the **DNA sequence composition** [10] and its **length** [12].

III. DNA Bio-bond AFM Experimentations

In order to evaluate the attachment between two meso-scale components based on the use of DNA hybrization, we have initiated an experimental approach using Atomic Force Microscopy (AFM). The objectives of this approach is to obtain numerical values for the force binding of two DNA strands complementary populations.

A. Materials and Methods

DNA strands were fixed on the AFM tip and their complementary on the substrate. By bringing them into contact (approach step), and then dissociate the formed double helixes (retreat step), the interaction force between the two strands' population was obtained.

The length of the used oligonucleotides is of 75 bases and is about 25.4 nm. The nucleotide sequence is the following (5'-3'):

S = CAA-ATA-CCG-TGG-GAC-GAC-ACG-CAC-CGG-CAG-TGC-GCA-GGC-AGC-GTCGGA-CAC-AAC-ACG-CTT-ACG-GCC-CTC-AAC-ACT

This sequence was chosen because of its reduced number of mismatches comparing to its high interaction energy. Oligonuleotides were purchased commercially from Eurogentec Company. The termination at 5' is chemically modified with Amine and the melting temperature is 76.6 C.

We could test and compare two different modified surfaces' substrates: the non-blocked and blocked substrates. The blocked ones differ from the non-blocked ones because of the addition of blocking agents, an alkylamine $C_2H_5NH_2$ which allows to eliminate the non-specific interactions between the tethered oligos on the tip and the substrate.

Both cantilevers and substrates are made of silicon. Different parameters are studied in these measurements. These are: AFM cantilever stiffness k: 0.1 N/m, 0.03 N/mn, cantilever speed v: 0.1 $\mu m/s$, 1 $\mu m/s$, 5 $\mu m/s$, substrate preparation : with (B) or without (NB) blocking agents, scan type: 256 iteration on the same coordinates of the substrate, and 1μm x 1 μm surface sweep on 1024 iterations (32 x 32)). Table I summarizes all experiments acheived in this study.

	v = 0.1 $\mu m/s$	v = 1 $\mu m/s$	v = 5 $\mu m/s$
k = 0.1 N/m	Exp2 : 256 i + B	Exp1: 256 i + B Exp5: 256 i + NB Exp6: 1024(s) i+ B	Exp3 : 256 i + B
k = 0.03 N/m		Exp4 : 256 i + B	

TABLE I
OVERVIEW OF MEASUREMENTS WITH DIFFERENT PARAMETERS, I: ITERATION, B: BLOCKED SAMPLES, NB: NON-BLOCKED SAMPLES, S: SCAN.

B. Data Analysis

Exp1, 2, 3, 4 and 5 include 256 measures that correspond to the iteration (repetition) of the AFM cantilever's approach/retreat process on one sample's point. Exp6 consists in 1024 measurements in a scaning mode. Figure 1 shows the approach/retreat curve for one iteration (10^{th}) in tne Exp2 data set. The breaking force value F corresponds to the difference between the minimal value of the curve and the intersection value of the approach and retreat curve.

The variance value V(X) represents the distance between the points' set of this iteration. It is given by the following expression: $V(X) = \sum_i (p_i(x_i - \bar{x})^2)$ where x_i represents the (i^{th}) force value on the approach retreat curve, $\bar{x}$ represents the mean and p_i represents the ratio of each x_i. Even if this variance value has no physical meaning, it allows to compare the curves' profile for the 256 repetitions of experience 1 to 5, and for the 1024 repetitions for experience 6. Figure 2 represents the variance curves. The curve's monotony allows us to deduce that a dependency between the measurements along time exists. Indeed, for v =1 $\mu m/s$ (Exp1, Exp4, and Exp5) the variance increases and vice versa for v = 0.1$\mu m/s$ (Exp2) and 5 $\mu m/s$ (Exp3). The represented curve is monotonous but decreasing in the scaning mode (Exp6). Therefore, there is a high variability in the data which does not only depend on the velocity but also on the cantilever stiffness.

Two points appear to be very important: the high variability between recorded data which implies the non-repeatability and the dependency between the successive measurements.

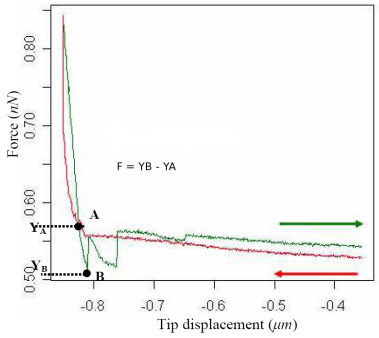

Fig. 1. Force (nN)/distance (μm) curve. F is the breaking force value. (red) appraoch step, (green) retreat step. Several discrete steps appear. We suppose that the most probable reason is the hydrogen bonds breaking and the reorganization of the molecule or ultimately DNA breaking.

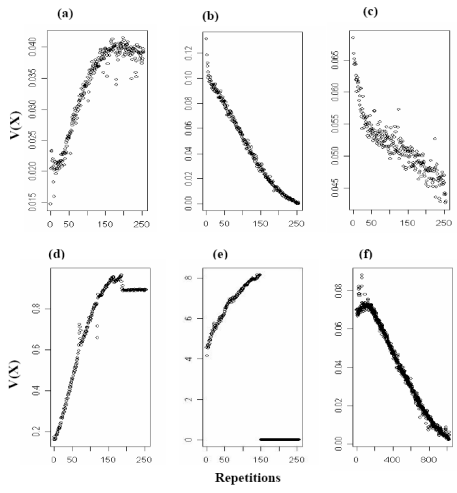

Fig. 2. Variance curves for the entire data during 256 repetitions for **(a)** Exp1 , **(b)** Exp2 , **(c)** Exp3 , **(d)** Exp4 , **(e)** Exp5 and **(f)** Exp6.

In the following, "the cut data" corresponds to the retreat part of the approach/retreat curve, and "the entire data" to

the whole curve (figure 1). Here, the cut data are compared to select the significant data which will be used to extract the pertinent information for the force intercation F. There are three phenomena, which are responsible for the monotony of variance curves: the samples wear (break, pulling up, etc...), the strand/strand entrainment to the approach/retreat processes and the whole experimental system entrainment. The hypothesis consisting in considering that the data are exploitable if the monotony of the cut data variance curve disappears is done. In fact, it will imply the independence of the measures for the considered experience. Therefore, two data groups could be distinguished: the exploitable and the non-exploitable data.

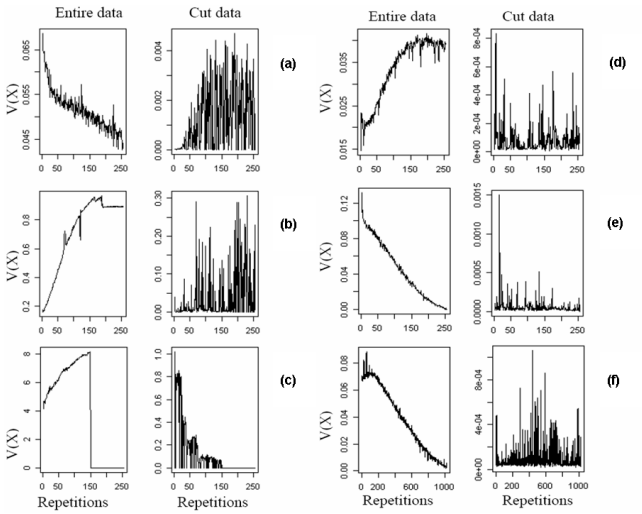

Fig. 3. Variance function of repetition curve for each experience: left (entire data) and right (cut data). **(a)** Exp3, **(b)** Exp4, **(c)** Exp5, **(d)** Exp1, **(e)** Exp2, **(f)** Exp6.

In the case of Exp3, Exp4 and Exp5, the monotony doesn't disapears. That implies that these data are not exploitable. Indeed, the loss of sensitivity and the specificity of the recorded signal are due respectively to the induced noise by cantilevers type (figure 3 (b)), the used non-blocked substrate(figure 3 (c)) and the high velocity (figure 3 (a)). In case of Exp1, Exp2 and Exp6, the monotony disapears which implies that the data are exploitable. Using this variance analysis methode, the exploitable data are distinguished from the non-exploitable data. Furthermore, we determined the conditions to recognize the significant measures. The velocity must be lower than 0.1 $\mu m/s$, the used cantilever must not be too flexible ($k > 0.03N/m$), and substrates have to be blocked.

C. Force interaction strength

The extracted force interaction F is represented in figure 4. On one hand, the mean value of the interaction force is 0.05 nN and is the same for all the exploitable data. The variance remains very low (green line). On the other hand, in the the non-exploitable data, there is high variability (high variance) and F is higher then it is in the exploitable data and varies from an experience to the other.

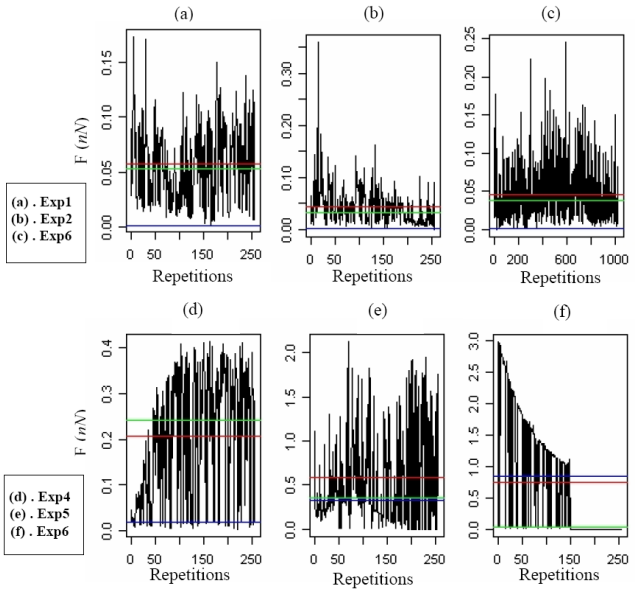

Fig. 4. F curves according to repetitions for the series: Exp1, Exp2, Exp3, Exp4, Exp5, and Exp6. In the Exp1, 2 and 6 the variance remains very low (blue line) while in Exp3, 4 and 5 it is too high. The mean value (red line) and the median value (green line) show the significance of the mean value in the exploitable data (Exp1, 2 and 6) and the non-exploitable data (Exp3, 4 and 5) because of the small or high distance between the two lines. The mean force value F_{exp} on the exploitable data remains at 50 pN.

We estimate the mean force value $F_{exp} = 50$ pN per contact area. This contact area is seen as a half sphere (the tip boundary) whose radius is lower than the strand length and estimated at $0.0125\mu m^2$. Therefore, the force between two components with a functionalized surface of $1\mu m^2$ is estimated at 4 nN for this DNA used sequence.

IV. Modelling of the DNA bio-bond

In order to evaluate the attachment between two components basing on the use of DNA hybrization, we have investigated also theoretical approaches. While the thermodynamical approach allows to predict the interaction energy involved in the DNA hybridization process, the molecuar approach allows to closely understand the DNA simple molecule behavior.

A. Molecular Approach

Using MOE (Molecular Operating Environment) software, the potential energy is computed according to a force field that must be chosen before (as CHARMM, etc). This energy is divided onto two terms: the bonded energy (bond stretch, bond angle bend, stretch-bend, out-of-plane, torsion and the non-bonded energy (Van der Waals, electrostatic, solvation, and restraint energies).

This energetic value can be sufficient to characterize our system. But the main difficulty is to make a relation between this energy and the experimental force data obtained using an AFM. . In order to achieve this it need to be converterd into a force. The external forces applied on a particle, corresponding to a given displacement, is equal to derivative of the energy according to the displacement. One

can consider that this detailed expression can be obtained directly from the force field. Even if this analytical energy can be obtained, this expression remains dependent of the force field used. To be force field independent, the main idea is to make an interpolation of the energetic values to approach the interaction energy profile between two molecules with a constructed analytical function. In [13] authors proposes a formulation for this interpolated energy and demonstrate its validity. This algorithm is applied on to two complementary DNA strands, using different lengths to obtain a first result which allows us to compare the molecular modelling approach to the experimental results. Using the Daunay's model. For an arbitrary sequence containing 75 base pairs, the interaction force is about F_{mol} = 8 nN.

B. Thermodynamic Approach

To predict the melting temperature Tm and the DNA molecules' stability in DNA microarray applications, many models were developed such as the Nearest Neighbour model (NN) [11]. The total difference in the free energy ΔG of the DNA folded and unfolded states can be approximated using such models. The NN model for nucleic acids assumes that the stability of a given base pair depends on the identity and the orientation of neighbouring base pairs.

$$\Delta G = \sum_i (n_i \Delta G(i) + \Delta G_{init} + \Delta G_{sym})$$

Where $\Delta G(i)$ are the standard free-energy changes for the 10 possible WatsonCrick NNs [11], n_i is the occurrence number of each nearest neighbour i, and ΔG_{sym} and ΔG_{init} are numerical parameters given in [11].
Different hybridized conformations are possible. Each of them corresponds to the gap of one strand along its complimentary. This gap induces new base paire formations, and thus the total energy changes. These different conformations correspond to mis-matchs. The unique perfect hybridized conformation corresponds to the perfect-match. For the DNA sequence used in the experience, 119 stable hybridized conformations are possible.For the experimental given sequence S, Δ G = -164.42 kcal/mol for the perfect matching corresponding to the maximum adhesion force estimated at 1141.27 pN. The involved mean force F_{thermo} in the strands separation is therfore estimated about 38.78 pN.

C. Discussion

Our force estimation, based on the molecular approach is based on considering the complementary strands to be rigid and the application point of the calculated interaction force is assumed to be the centre of mass. Thus, a segmentation of each strand onto small supposed rigid segments can be an accessible way to solve this problem. The molecular approach imporved with a segmentation can allow describing accurately the DNA bio-bond. From our experimental work, we knew that it is not possible to obtain a density value of strands tethered on the surface components. Furthermore, the non-repeatability of the hybridization force measurement experience, asks the question if our process can be described accurately by a deterministic approach.

V. Conclusion

It appears that the thermodynamical approach allows to be closer to the experimental reality because it considers both the intrinsic propereties of the DNA strands (elasticity, van der waal, hydrogen bonds , etc...) and the environment aspects such as the temperature and the ion concentration.
The AFM experiences allowed appreciating the non repeatability of the strands hybridization/separation phenomenon. The statistical analysis of the obtained data allowed us to determine the environmental conditions to obtain significant data, and to extract the force mean value. In addition, we obtained the mean value of 50 pN per 0.0125 μm^2 for the interaction force. We showed also, that the DNAs sequence length has to be smaller because of a great number of high possible conformations. Consequently, other experimental approaches have to be tried.

VI. Acknowledgments

This work was supported in part by the European GOLEM project (http://www.golem-project.eu/index.htm). The GOLEM project is supported by the Nanotechnology program (NMP) of the European Commission under the sixth framework program (FP6).

References

[1] M. Boncheva and G. M. Whitesides, "Making things by self-assembly," *MRS Bulletin*, vol. 30, pp. 736–742, 2005.
[2] *Integration of Silicon Circuits Components Onto Plastic Substrates Using Fluidic Self-Assembly*. International Conference on MEMS, NANO and Smart Systems, 2005.
[3] A. O'Riordan, P. Delaney, and G. Redmond, "Field configured assembly: Programmed manipulation and self-assembly at the mesoscale," *Nano Letters*, vol. 4, no. 5, pp. 761–765, 2004.
[4] H. Huang, J. N. Anker, K. Wang, and R. Kopelman, "Magnetically assisted and accelerated self-assembly of strawberry-like nano/microparticles," *J. Phys. Chem. B*, vol. 110, no. 40, pp. 19 929–19 934, 2006.
[5] H. McNally, M. Pingle, S. W. Lee, D. Guo, D. E. Bergstrom, and R. Bashir, "Self-assembly of micro- and nano-scla particles using bio-inspired events," *Applied Surface Science*, vol. 214, pp. 109–119, 2003.
[6] N. C. Seeman, "The design and engineering of nucleic acid nanoscla assembly," *Structural Biology*, vol. 6, pp. 519–526, 1996.
[7] W. M. Shih, J. D. Quispe, and G. F. Joyce, "1.7-kilobase single stranded dna that folds into a nanoscale octahedron," *Nature*, vol. 427, pp. 618–621, 2004.
[8] P. W. K. Rothermund, "Folding dna to create nanoscale shapes and patterns," *Nature*, vol. 440, pp. 297–302, 2006.
[9] Y. A. Chen, C.-C. Chou, X. Lu, E. H. Slate, K. Peck, W. Xu, E. O. Voit, and J. S. Almeida, "A multivariate prediction model for microarray cross-hybridization." *BMC Bioinformatics*, vol. 7, p. 101, 2006. [Online]. Available: http://dblp.uni-trier.de/db/journals/bmcbi/bmcbi7.html#ChenCLSPXVA06
[10] Y. Gao, L. K. Wolf, and R. M. Georgiadis, "Secondary structure effects on dna hybridization kinetics: a solution versus surface comparison," *Nucleic Acids Research*, vol. 34, no. 11, pp. 3370–3377, 2006.
[11] J. John SantaLucia, "A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics," *PNAS*, vol. 95, pp. 1460–1465, 1998.
[12] E. Carlon and T. Heim, "Thermodynamics of rna/dna hybridization in high-density oligonucleotide microarrays," *Elevier - Physica A*, vol. 362, pp. 433–449, 2006.
[13] B. Daunay, A. Micaelli, and S. Régnier, "Energy-field reconstruction for haptic-based molecular docking using energy minimization processes," RSJ International Conference on Intelligent Robots and Systems. IEEE, 2007.