Biochimica et Biophysica Acta 1810 (2011) 603–611

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Design and experimental validation of a generic model for combinatorial assembly of DNA tiles into 1D-structures

Aude Laisne [a], Eric Lesniewska [b], Denis Pompon [a,*]

[a] Centre de Génétique Moléculaire, CNRS, Avenue de la Terrasse, F91190 Gif-sur-Yvette, France
[b] Institut Carnot Bourgogne, UMR CNRS 5209, Université de Bourgogne, F-21078 Dijon, France

ARTICLE INFO

Article history:
Received 4 December 2010
Received in revised form 1 March 2011
Accepted 2 March 2011
Available online 22 March 2011

Keywords:
Combinatorial assembly
Deoxyribonucleic acid tile
System biology
Modeling
Kinetic control
Gene synthesis

ABSTRACT

Background: Quantitative modeling of the self-assembly of DNA tiles leading either to defined end-products or distribution of biopolymers is of practical importance for biotechnology and synthetic biology.
Methods: The combinatorial process describing tile assembly was implemented into a generic algorithm allowing quantitative description of the population of significant species accumulating during the reaction course. Experimental formation and characterization by optical and electrophoresis approaches of copolymers resulting from the self-assembly of a limited number of half-complementary tiles were used to define and validate generic rules allowing definition of model parameters.
Results: Factors controlling the structure and the dynamic of the oligomer population were evidenced for assemblies leading or not to defined end-products. Primary parameters were experimentally determined using rapid mixing experiments. Adjustment of simulations to experimental profiles allowed definition of generic rules for calculation of secondary parameters that take into account macro- and microenvironment of individual hybridization steps. In the case of copolymers, accurate simulation of experimental profiles was achieved for formation of linear assemblies.
Conclusions: Overall length of species and structure of the DNA regions flanking the hybridization sites are critical parameters for which calculation rules were defined. The computational approach quantitatively predicted the parameters affecting time-course and distribution of accumulating products for different experimental designs.
General significance: The computational and parameter evaluation procedures designed for the assembly of DNA tiles into large 1D-structures are more generally applicable for the construction of non-DNA polymers by extremities-specific recognition of molecular blocks.

1. Introduction

Linear assembly of DNA tiles is mainly involved in the domain of gene synthesis [1,2] and combinatorial library constructions [3]. A plethora of synthesis options that relies on basic molecular biology techniques and that aims to guarantee successful synthesis in cost-effective time and conditions have been described [4–6]. However, no modeling approach was proposed to describe the combinatorial structure of mono-dimensional DNA assemblies at the level of the time dependence of the distribution of all possible sub-assemblies. Controlled tuning of the process is particularly critical to optimize directed evolution based protein engineering. Substitution of the current trial-and-error approaches by modeling would thus constitute a significant advantage [7]. Moreover, new emerging technologies, like the construction of networks of interacting peptidic domains [8,9] using self-assembly of synthetic nucleoprotein structures, are also demanding for modeling of linear assemblies [10,11]. When single step assembly of a preformed DNA–protein covalent structure is targeted, reversible conditions requiring high temperature must be avoided, owing to the limited thermal stabilities of proteins. Ligation or polymerase based approaches remain possible but are also problematic in these geometrically crowded systems. For these reasons, assembly processes kinetically controlled at low temperature appear interesting but require quantitative modeling to limit kinetic traps. Two experimental settings can be considered: one, typical for gene or library synthesis which involves a finite combinatorial allowing full description of all intermediate species, and the other, corresponds to unlimited block-copolymers for which an open combinatorial structure has to be managed. The two cases can be modeled but formation of repetitive copolymers involving a limited number of

Abbreviations: ss, single-stranded; ds, double-stranded

* Corresponding author at: CGM, CNRS, bât 23B, Avenue de la Terrasse, 91190 Gif-sur-Yvette, France. Tel.: +33 169823680; fax: +33 169823682.
E-mail address: pompon@cgm.cnrs-gif.fr (D. Pompon).

0304-4165/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbagen.2011.03.002

tiles constitutes a simpler model when investigation of general rules for parameter determination is targeted. However, such model raises the problem of the combinatorial explosion of possible reactions during the formation of unlimited copolymers. Evaluation of individual rate constants represented a challenge that was approached by the determination of a core of experimental rate constants and the definition of rules for the evaluation of the whole model parameters from the seeding set. The influence of parameters such as the length dependence of ss-DNA and the structure of flanking sequences on the hybridization rates were taken into account. Indeed, the length dependence of hybridization rates between ss-DNA was investigated [12–15] but few were reported in the case of species of variable lengths assembled through hybridization of ss-DNA segments of constant length. The present paper addresses the problem of modeling end-recognition driven self-assembly of tiles involving very large or even unlimited number of intermediate species. It specifically develops ways allowing determining critical model parameters for DNA assemblies using combination of experimental approaches and extrapolation rules. The core part of the modeling is however widely independent of the chemical nature of the building blocks and can equally apply to a large range of biopolymers. This approach aims to fill a gap between current modeling approaches either targeting to describe at a local level the interaction between tiles [16,17] or, in contrast, focusing on the global organization [18] of resulting supramolecular assemblies.

2. Materials and methods

2.1. DNA-tile hybridization and gel analysis

Oligonucleotides were purified as described in the supplemental protocol 1 in Appendix A. Approximately 3.5 μl of the 4 polymerizing DNA-tiles (~1 μM final) in phosphate buffered saline (PBS) were separately deposited on the wall of a tube. The reaction was initiated by centrifugation and vortexing. Incubation was performed at 24 °C or 42 °C for variable times. For gel analysis, reactions were quenched by addition of a premix of the four terminating DNA-tiles (~1.5 μM final) and the reaction continued for 1 h before deposit on 3.5% Nusieve GTG agarose gel that was run in 0.5× TBE at 50 V and 4 °C for 2 h. Images of ethidium bromide stained gels were submitted to trapezoidal deformation to correct heterogeneity of DNA migrations. Data from different exposure times were concatenated to increase dynamic range of signal without saturation.

2.2. DNA hybridization kinetic analysis by rapid-mixing experiments

Half-complementary ss-DNA-tiles were loaded at the same concentration (~0.8 μM in PBS buffer) in the two stopped-flow syringes. After rapid mixing at 24 °C, 400–1600 spectra (40 to 1240 s) were recorded and hypochromic changes at 260 nm were fitted to exponential or hyperbolic laws to estimate initial rates and amplitudes of reactions. Second order rate constants were calculated from the ratio between the initial slope and the extrapolated total amplitude of the traces. For self-assembly of 4 ss-DNA tiles to form polymers, an equimolar mix of tiles (Fig. 3, model 2) was loaded into one syringe in PBS buffer, set to pH 11.8 with 20 mM NaOH. The second syringe was loaded with PBS supplemented by 20 mM HCl in order that the 1:1 (vol./vol.) mix in the observation chamber exhibits a pH of 7.4. Hybridization traces were recorded and rate constants calculated as previously. Pre-formed DNA-tile hybrids (Fig. 3, model 3) were obtained by incubation for 30 min at 24 °C of an equimolar mix of half-complementary tiles in PBS buffer pH 7.4. Two different types of hybrids were loaded into separated stopped-flow syringes, and after rapid mixing, hybridization traces were recorded and rate constants calculated as previously.

2.3. Experimental data processing and visualization of simulation output procedures

Expected electrophoresis motilities of species and their relative staining responses were computed based on experimental calibrations with ss- and ds-species of known structures. Molecular marker lines were scanned and peak shapes modeled using polynomial modified Gaussian distributions. The calibration curves (log plots) were fitted on polynomial laws. Experimental profiles were first corrected for inhomogeneous migrations between lanes using a non-linear local expansion/compression algorithm then were processed to improve band resolution and reduce noise: gel profiles were deconvoluted using peak shapes of molecular markers degraded about 2.1-fold by increasing their half-width. This was necessary to take into account the broader electrophoretic migration of segmented polymers. Finally, experimental and simulated data were reconstructed using the same virtual resolution typical of ds-DNA.

3. Simulation procedures

The flow chart for the simulation of the 1D self-assembly of DNA tiles is represented on Fig. 1. This computational process was specifically designed to simulate assembly of bifunctional tiles by base-pairing but can as well apply to the polymerization of any kind of linear block-polymers provided of limited adaptations. The algorithm comprises two main complementary sections: one, depicted on the left panel, aims to generate the collection of possible species which can be formed by random association of members and descendant of a seeding set of tiles, and the other (right panel) uses the generated data to automatically build up and solve differential equations describing the system. Depending on the structure of the seeding set, two types of situations can be encountered giving rise, respectively, to finite and infinite series. When combinatorial convergence is possible, all species fulfilling the association rules are generated. In contrast, in the case of non-converging sets (for example, formation of unlimited block-copolymer), only a part of possible species is generated. In such a case, a critical point is that the combinatorial procedure will always generate species which mainly contribute to the observed population in given experimental conditions.

3.1. Combinatorial generation of species and of formation paths

As described in Fig. 1, left panel, the initial seeding set is first transferred into an accumulator which stores both species description and their possible formation paths (step 1). The algorithm then enters an iterative process in which all possible couples of species present in the accumulator are tested for fulfilling criteria for a coupling process (step 2) involving either one, the other, or both of their extremities (cyclization). Allowed couplings are defined by heuristic rules, which, in the case of DNA tiles, rely on base-pairing stability but which could differ, for example, corresponding to chemical reactivity in case of some types of polymers. Species formed as the result of a successful coupling event (step 3) are compared to the content of accumulator. Novel species are created and stored. For already existing ones, the new formation path is added to the species descriptor. Non-converging combinatorial processes have to be stopped after a specific number of cycles, thus truncating species generation. During the last cycle, storage process differs: paths for existing species are registered in the accumulator while new generated species are stored in a sandbox. The concentration of species accumulated within the sandbox must remain neglectable as compared to total concentration of species at the final time simulated. Otherwise, calculation must be restarted using a larger number of combinatorial cycles or the simulation stopped at a shorter time to decrease the sandbox content.