Self-assembly programming of DNA polyominoes.

Hui San Ong[1], Mohd Syafiq-Rahim[2], Noor Hayaty Abu Kasim[3], Mohd Firdaus-Raih[2], and Effirul Ikhwan Ramlan[1,4,*]

[1] Natural Computing Laboratory, Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya, 50603, Kuala Lumpur, Malaysia
[2] School of Biosciences and Biotechnology, Faculty of Science and Technology and Institute of Systems Biology, Universiti Kebangsaan Malaysia, 43600, Bangi, Malaysia
[3] Department of Restorative Dentistry, Faculty of Dentistry, University of Malaya, Kuala Lumpur, Malaysia
[4] Centre of Research for Computational Sciences and Informatics for Biology, Bioindustry, Environment, Agriculture, and Healthcare (CRYSTAL), University of Malaya, 50603 Kuala Lumpur, Malaysia

* effirul@um.edu.my

ABSTRACT

Fabrication of functional DNA nanostructures operating at a cellular level has been accomplished through molecular programming techniques such as DNA origami and single-stranded tiles (SST). During implementation, restrictive and constraint dependent designs are enforced to ensure conformity is attainable. We propose a concept of DNA polyominoes that promotes flexibility in molecular programming. The fabrication of complex structures is achieved through self-assembly of distinct heterogeneous shapes (i.e., self-organised optimisation among competing DNA basic shapes) with total flexibility during the design and assembly phases. In this study, the plausibility of the approach is validated using the formation of multiple 3 x 4 DNA network fabricated from five basic DNA shapes with distinct configurations (monomino, tromino and tetrominoes). Computational tools to aid the design of compatible DNA shapes and the structure assembly assessment are presented. The formations of the desired structures were validated using Atomic Force Microscopy (AFM) imagery. Five 3 x 4 DNA networks were successfully constructed using combinatorics of these five distinct DNA heterogeneous shapes. Our findings revealed that the construction of DNA supra-structures could be achieved using a more natural-like orchestration as compared to the rigid and restrictive conventional approaches adopted previously.

Keywords: DNA polyominoes, molecular programming, self-assembly, DNA nanotechnology, DNA nanofabrication

1. Introduction

Self-assembly allows DNA molecules to naturally fuse together and form supra-structures (Mao et al., 2000; Seeman, 1982; Winfree, 1998). The spontaneous reaction via Watson-Crick base pairing allows the formation of discrete structures with high precision and efficiency. Common approaches in constructing DNA supra structures include DNA origami (Han et al., 2011; Kuzuya and Komiyama, 2010; Marchi et al., 2014; Rothemund, 2006), molecular tiles (Winfree, 1996), parallelograms from Holliday junctions (Mao et al., 1999) and single stranded modular motif (Wei et al., 2012; Yin et al., 2008). These conventional approaches have their limitations (Ke et al., 2012; Ong et al., 2015; Pinheiro et al., 2011; Wei et al., 2012; Yin et al., 2008). Crucially, the intricate and meticulous sequence design phase in which the nanostructures were fabricated by generating a definitive set of DNA sequences. This study attempts to address this issue by allowing the structures to be constructed autonomously using distinct interchangeable components.

This is achieved through the formation of desired conformations from a combination of distinct heterogeneous shapes. This increases the flexibility of constructing DNA nanostructures since the formation of the structures is achieved through the self-organisation of the competing DNA shapes without pre-fixed configuration. The core principle is to allow the most preferred shape and sequence combinations to take precedence (i.e., survival of the fittest). For instance, if n sets (where n is more than 1) of DNAs are initially designed to assemble into the desired conformations, in cases where a single set of the structure collapsed, the remaining n -1 sets would be capable to form the target structure. In fact, individual units inside the n -1 sets can replace the incompatible unit of the original set. This interchangeability is a key aspect of the approach. Every component in each set is modular,

whereby the failure of any particular unit would not affect the completeness of the set. The mechanism allows a specific substitution (i.e., to replace any incompatible shapes) or the replacement of the entire shape configurations to be executed. Therefore, total programmability (i.e., pre-fixed configuration of binding between shapes) is not promoted in this approach, the formation of the structures is dependent entirely on the self-organised characteristics of the molecule. This eliminates dependency on successful wet lab implementation of a particular set. The mechanism employed promotes molecular orchestration (Zauner, 2005), and in this instance, a mixture of multiple potential sets that self-organised themselves into the desired configurations (i.e., many to one relationship, where extraction of successful configurations could be made regardless of the sets).

The construction of DNA nanostructures (Amir et al., 2014; Benenson et al., 2004; Ding and Seeman, 2006; Douglas et al., 2012) begins with the sequence design steps. Various strategies such as strain minimization, sequence symmetry minimization and free energy minimization are employed by programs such as SEQUIN (Seeman, 1982), Tiamat (Williams et al., 2008), Uniquimer-3D (Zhu et al., 2009) and GIDEON (Birac et al., 2006) to aid in the sequence generation. In fact, the design of 3D DNA origami structures is now supplemented by software packages such as caDNAno (Douglas et al., 2009) that incorporate a graphical user interface. Recently, a program called Polygen has been developed to aid in the construction of complex atomistic covalently linked DNA nano-cages (Alves et al., 2016).

In this study, a computational tool leveraging on the aforementioned strategies such as sequence symmetry, is extended towards optimising and designing a set of less stringent sequences. Our model encourages the competition between DNAs to occur in an effort to promote sequence to structure flexibility. A tool is then created to provide mapping for the competitive shapes by delineating all probable paths taken by the DNA to form the structures using graph theory (Biggs et al., 1986). This is essential since molecular self-assembly is asynchronous with a multitude of errors (Rothemund et al.,

2004), and probable shapes (i.e., the "best" unit) must compete with the partially probable shapes (i.e., the "next best" unit) during the assembly process at all time. The mapping of the paths strategy exerted in this work could therefore provide insights into the fundamental basis of the structure construction for the end user.

2. Material and methods

2.1 Fundamental Concepts in DNA polyominoes

This section begins by presenting the fundamental concept in the proposed schema, DNA polyominoes. Polyominoes shapes (monomino, tromino and tetrominoes) were used as the representative in demonstrating the feasibility of using multiple elementary blocks in structural assembly. The hierarchical schema in DNA polyominoes starts with an elementary block, followed by shapes and then larger structural formations. As a basis, each block used two single-stranded DNAs to form a block. Then, multiple units of these blocks assembled into a shape. Different shapes would then assemble into a larger structure (Fig. 1).

Each of these shapes can comprise of one or more connector on the horizontal sides of the shape. Its function is to enable the shape to bind to another shape that had a matching connector and thus forming larger structures. In the context of DNA sequences, the matching connector was defined as DNA with complementary sticky ends. In total, eight distinct shapes were used and labelled with specific alphabets. All DNA shapes (Fig. 2) are comprised of four single stranded DNAs except for shape I (made from two single stranded DNAs).

2.2 DNA segmentation

A computational protocol was developed to map the interaction (intermolecular binding) between DNA nucleotides based on the principle of binding dependencies (Ramlan and Zauner, 2013) between nucleotides. We implemented an undirected graph representation, in which DNA segments are stored as nodes that are connected using edges. This allows the automated system to compute all probable paths taken by each DNA segment (node) in forming the structures. The protocol required each DNA strand to be separated into different segments based on perfect complementarity (i.e., where all the intended bases hybridize as specified in the design) between its pairs (Fig. 3).

2.3 Construction of the free energy and binding affinity matrices

In order to construct the energy matrix, each segment is represented as node, or vertex. Thermodynamics free energy between each node were calculated using the program DuplexFold (Reuter and Mathews, 2010). Default parameters (for the program) were used with the "DNA" parameter setting. The free energy profile with n number of nodes resulted in a matrix with n number of rows and columns as follows:

$$\begin{bmatrix} X_{1,1} & \dots & X_{1,n} \\ \vdots & \dots & \vdots \\ X_{n,1} & \dots & X_{n,n} \end{bmatrix}$$

$$X_{i,j} = \Delta G_{i,j} \; ; \; i, j = 1, 2 \; \ldots\ldots n \; ; \; n = \text{Total number of nodes}$$

The energy matrix is converted into a binding affinity matrix. The free energy at each position $\Delta G_{i,j}$ will then be divided by the lowest energy in each row $\left(min\left(\Delta G_{i,j_{1\ldots n}}\right)\right)$, resulting into the probable binding affinity value between every node. The value of 1.0 indicates the lowest free energy (strongest binding) between all available bindings. For instance, when the binding affinity between node 1 and node 2 is 1.0, it indicates that the binding strength of node 1 with node 2 is the strongest compared to other available binding with the remaining nodes (e.g. 3, 4, 5…etc). The formula for binding affinity is as follows:

134 $$\text{Binding affinity for } P_{i,j} = \frac{\Delta G_{i,j}}{min(\Delta G_{i,j_{1\ldots n}})}$$

135 Edges connecting nodes with 1.0 binding affinity values represent the most favourable binding
136 among the *n* number of nodes. However, in circumstances where no edges carry the most favourable
137 binding affinity values (1.0), the highest value will take precedence (in our implementation binding
138 affinity must be above the threshold value of 0.7).
139

140 **2.4 Binding affinity graph: computing the probable paths**

141 **2.4.1 Determining the start point**

142 In order to determine the start point, the melting temperature (T_m) for every DNA pair $X_{i,j}$ was
143 calculated using UNAFold-3.8 (Markham and Zuker, 2008). The DNA pairs with T_m value equal or
144 higher than quartile 3 were selected as the start point. For each pair, the strongest node (with lowest
145 free energy) were selected as the start point and the remaining of the nodes (within the pair) would act
146 as the sticky ends. These sticky ends would then operate as precursors in determining the node to be
147 selected next (Supplementary Fig. S1).
148

149 **2.4.2 Greedy search phase**

150 The graph would only proceed to the node with the binding affinity value of $P_{i,j} > 0.7$. The default
151 value is fixed at 0.7 to ensure that the graph is restricted to favour only strong estimation values (i.e.,
152 representative of the preferable binding interactions). Lower assignment of threshold generates
153 convoluted paths full of weak interactions, which only complicates the search process. For every new
154 node, two conditions will be considered; the emergence of one or more new sticky end(s) and the non-
155 availability of sticky ends. The initial value of every node starts at 1.0. The DNA uptake rate is set at
156 0.001 probability. Whenever a node is selected, the value of that node will be deducted by the DNA
157 uptake rate. The formula for node concentration calculation is as follows:
158 $[Node_{NewCurrent}] = [Node_{Current}] - [Node_{UptakeRate}]$.

159 The value of every node is evaluated during each cycle. The search will continue until the values of
160 any node became nil (Table 1).

161

162 **2.5 DNA annealing**

163 Oligonucleotides were purchased from Integrated DNA Technologies Pte. Ltd. (USA). The
164 complexes were formed by mixing stoichiometric quantities of DNA in an annealing buffer (40 mM
165 Tris base, 2.5 mM EDTA, and 13 mM $MgCl_2$) and annealing process from 90°C to 40°C for three
166 hours using Eppendorf Mastercycler Pro S thermocycler (Eppendorf, Hamburg, Germany). To form
167 the individual DNA shapes, four different oligonucleotides were mixed stoichiometrically in an
168 annealing buffer and the final concentration was set to 0.5 μM.

169

170 **2.6 Gel electrophoresis**

171 The results of the annealing reactions were analyzed using non-denaturing gel electrophoresis
172 containing 4% and 5% polyacrylamide gel (29:1 acrylamide:bisacrylamide), 0.75 mm thick and run at
173 approximately 12V/cm-1 for 2 hours at 4°C. The running buffer contained 10 mM $MgCl_2$ and 1X TBE
174 (89 mM Tris base, 89 mM Boric acid and 2 mM EDTA pH8.3) and the loading buffer contained 0.25%
175 Bromophenol blue tracking dye and 30% glycerol. GelRedTM Nucleic Acid gel stain (Biotium, US)
176 was used to stain the gel.

177

178 **2.7 Sample preparation and atomic force microscopy (AFM) imaging**

179 **2.7.1 Preparation of mica surface**

180 A 0.1% APTES ((3-aminopropyl) triethoxysilane) solution was prepared in ultrapure water. Then
181 a drop (2 μL) of 0.1% APTES solution was deposited onto the freshly cleaved mica surface and the
182 surface was rinsed with ultrapure water (20 μL) after 5 minutes incubation at room temperature.

183

184 **2.7.2 Sample preparation for AFM imaging**

185 The samples were diluted to 0.2 ng/μL with a buffer (40 mM Tris-HCl (pH 7.6), 13 mM $MgCl_2$,
186 2.5 mM EDTA). 2 μL of the sample solution was placed onto the APTES-treated mica surface for 5
187 minutes and the surface was later rinsed with the buffer (20 μL) to remove unbound molecules.

188

189 **2.7.3 Atomic Force Microscopy (AFM) imaging**

190 The AFM images were collected using high-speed AFM (Nano Live Vision, Research Institute
191 of Biomolecules Metrology Co., Tsukuba, Japan). The images were collected in tapping mode.

192

193 **3. Results**

194 **3.1 Formation of** 3 x 4 **DNA network using DNA polyominoes**

195 The size of the DNA network is fixed at 3 x 4 (i.e., 3 horizontal rows, and 4 vertical columns).
196 Different configurations of heterogeneous shapes (monomino, tromino and tetrominoes) were
197 generated to conform to the layout as illustrated in Fig. 4.

198

199 DNA sequences representing the respective shapes are generated using the autonomous protocol
200 developed in our previous work (Ong et al., 2015). The program focuses on the stacking and merging
201 of blocks to form DNA shapes. The program (Ong et al., 2015) relies on dependency information of all
202 nucleotides positions with inter-binding linkage between different DNA strands (i.e., DNA-DNA
203 binding). Details of the dependencies are available in the Supplementary Table S1-S5. The
204 intermolecular bindings between various DNA shapes are "loosely" programmed using complementary
205 sticky ends. The sticky ends are positioned at the intersection point, where different shapes are adjacent
206 to each other. The default lengths of the sticky ends (for all DNA shapes) are set to 10 nucleotides. In
207 order to further exploit the self-organisation ability of the molecule, the placement of matching sticky
208 ends should be randomly placed. This will create an environment where optimisation between

competing shapes (i.e., survival of the most stable assembly) will help the stability of the desired structures as well as allowing total modularity to be enforced. However, in this study, the complementary sticky ends were predefined to ensure that different configurations of the 3 x 4 DNA network are attainable during wet-lab validation.

Molecular representation of our 3 x 4 DNA network showed that Set 1, 2, 3 and 4 have the same DNA shape compositions (i.e., the four heterogeneous DNA with different orientations). The size of set 1 is smaller as compared to set 2, 3, and 4. This is because set 1 only requires 25 nucleotides in each basic unit; the remaining sets require 40 nucleotides for their basic units. Set 5 has a different DNA shapes configuration. Compared to the existing techniques of DNA nanofabrication, our proposed approach increases the degree of freedom in designing the desired structure two-folds. Existing techniques focuses only on the sequence diversity of the design phase (i.e., sequences that conform to the scaffolds), while our approach introduces the combinatorics of the polyominoes shape into the equation thus allowing diversity not only in sequence, but also in the heterogeneous shapes composition (i.e., many sequences to many shapes configurations that conform to the desired structure).

3.2 Gel electrophoresis and atomic force microscopy (AFM) imaging

DNA sequences for each shape were added sequentially during the gel electrophoresis procedure (Fig. 5). AFM images of the structure were captured. Comparison with AFM images was conducted and the findings are encouraging. Successful clearly defined formations of DNAs that resemble the designed structures can be observed (Fig. 6).

All polyominoes shapes, with the exception of monomino, are single crossover DNA tiles or Holliday junctions. In contrast to the double-crossover (DX) motif which is structurally rigid (Li et al., 1996), the structure of the Holliday junction motif is inherently floppy (Rothemund, 2005). This is

because the four-way junction of the motif alternates between one of two different "stacked-X" conformations (Duckett et al., 1988; Murchie et al., 1989), thus forming an approximately 60^{o} angle (Mao et al., 1999) between the two DNA helices. Given this natural profile, the final structure captured using the AFM is floppy as the self-assembly of multiple Holliday junctions has an approximately 60° native between the two DNA helices as observed in the figure.

4. Discussion

To address the complexity of determining the "many sequences to many shapes configurations" allowances introduced in our approach, we have annotated the base pairing probability using the concept of undirected graphs (i.e., each node/vertex can be visited more than once; with no emphasis on the order of the path taken. In our implementation, "nodes" represent DNA segments while the "edges" represent binding affinity between nodes. The decision of traversing any of these nodes are dependent on the free sticky ends resulted from prior binding (edge) (Fig. 7). As long as the new sticky ends have a probability value of more than the defined threshold value (0.7), it is predicted to be able to bind to the existing parent DNA (node). This process will be repeated iteratively for each node (similar to a greedy search where all paths are traversed).

Therefore, the number of graphs is equivalent to the number of potential structures that can be generated from a set of DNA strands. This number includes both the desired and misfolded structures. For example, the shapes configuration of set 5 produces 31 different graphs with only 21 graphs indicating the formation of the desired structure. Thus, there are 10 misleading paths that are biased towards unfavourable folding leading to the formation of mismatch structures. The number of occurrences for binding affinity close to 1.0 indicates the level of competition between the unintended nodes (i.e., not design to form base pair). Thus, the number of competitions is linear to the number of graphs that will be generated. Our search revealed that set 4 has the highest number of graphs

generated, followed by set 3, 2, 1 and 5 respectively (Table 2). This contributed to the higher number of binding affinities with values near to 1.0.

The value $P_{i,j}$ represents the relative binding affinity between each DNA segment estimated using the thermodynamics free energy from the program Duplexfold (Reuter and Mathews, 2010). $P_{i,j}$ has the value of 1.0, if the intended binding between nodes is a perfect complementary pair. In our calculation, partially complement ($P_{i,j} < 1.0$) of DNA segments is still considered. However, these partially complements segments have the tendency to create false routes (causing the emergence of sticky ends) and eventually resulted in the formation of false structures or miscellaneous aggregates. The correct graphs are representations of all the nodes visited exactly once and the edges taken by each node are correctly linked as designed, regardless of the starting points. The order of the completed routes will provide a blueprint for the DNA sequences to form the desired structures.

Acknowledgements

We gratefully acknowledge Professor Hiroshi Sugiyama from the Department of Chemistry, Graduate School of Science, Kyoto University and Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University and Dr. Yuki Suzuki from the Department of Chemistry, Graduate School of Science, Kyoto University for their assistance and expertise in AFM imaging. Acknowledgement is extended to both Dr. Zamri Radzi and Ms. Nabila Farhana from the Department of Paediatric Dentistry & Orthodontics, Faculty of Dentistry, University of Malaya in providing AFM imaging for the study. This research is supported by the High Impact Research Grant UM.C/625/1/HIR/MoE/FCSIT/002 (H-22001-00-B0002) from the Ministry of Higher Education, Malaysia and University of Malaya.

Author contributions statement

Conceived and designed the experiments: HSO MSR MFR EIR. Performed the experiments: HSO MSR. Analyzed the data: HSO MSR MFR EIR. Contributed reagents/materials/analysis tools: HSO

MSR NHAK MFR EIR. Wrote the paper: HSO MSR NHAK MFR EIR.

Additional information.

Competing financial interests. The authors declare no competing financial interests.

References

Alves, C., Iacovelli, F., Falconi, M., Cardamone, F., Morozzo Della Rocca, B., de Oliveira, C.L., Desideri, A., (2016) A Simple and Fast Semiautomatic Procedure for the Atomistic Modeling of Complex DNA Polyhedra. J Chem Inf Model. 56, 941-949.

Amir, Y., Ben-Ishay, E., Levner, D., Ittah, S., Abu-Horowitz, A., Bachelet, I., (2014) Universal computing by DNA origami robots in a living animal. Nature Nanotechnology 9, 353–357.

Benenson, Y., Gil, B., Ben-Dor, U., Adar, R., Shapiro, E., (2004) An autonomous molecular computer for logical control of gene expression. Nature 429, 423.

Biggs, N.L., Lloyd, E.K., Wilson, R.J., (1986) Graph Theory 1736-1936. Oxford University Press, New York.

Birac, J.J., Sherman, W.B., Kopatsch, J., Constantinou, P.E., Seeman, N.C., (2006) Architecture with GIDEON, A Program for Design in Structural DNA Nanotechnology. J Mol Graph Model 25, 470–480.

Ding, B., Seeman, N.C., (2006) Operation of a DNA robot arm inserted into a 2D DNA crystalline substrate. Science 314, 1583.

Douglas, S.M., Bachelet, I., Church, G.M., (2012) A logic-gated nanorobot for targeted transport of molecular payloads. Science 335, 831.

Douglas, S.M., Marblestone, A.H., Teerapittayanon, S., Vazquez, A., Church, G.M., Shih, W.M., (2009) Rapid prototyping of 3D DNA-origami shapes with caDNAno. Nucleic Acids Res 37, 5001-5006.

Duckett, D.R., Murchie, A.I.H., Diekmann, S., Kitzing, E.V., Kemper, B., Lilley, D.M.J., (1988) The structure of the Holliday junction, and its resolution. Cell 55, 79–89.

Han, D., Pal, S., Nangreave, J., Deng, Z., Liu, Y., Yan, H., (2011) DNA origami with complex curvatures in three-dimensional space. Science 332, 342–346.

Ke, Y., Ong, L.L., Shih, W.M., Yin, P., (2012) Three-Dimensional Structures Self-Assembled from DNA Bricks. Science 338, 1177-1183

Kuzuya, A., Komiyama, M., (2010) DNA origami: Fold, stick, and beyond. Nanoscale. Review. 2, 310-322.

Li, X., Yang, X., Qi, J., Seeman, N.C., (1996) Antiparallel DNA Double Crossover Molecules As Components for Nanoconstruction. J. Am. Chem. Soc. 118, 6131–6140.

Mao, C., LaBean, T.H., Reif, J.H., Seeman, N.C., (2000) Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. Nature 407, 493–496.

Mao, C., Sun, W., Seeman, N.C., (1999) Designed Two-Dimensional DNA Holliday Junction Arrays Visualized by Atomic Force Microscopy. American Chemical Society 121, 5437–5443.

Marchi, A.N., Saaem, I., Vogen, B.N., Brown, S., LaBean, T.H., (2014) Toward Larger DNA Origami. Nano Letter 14, 5740–5747.

Markham, N.R., Zuker, M., (2008) UNAFold: software for nucleic acid folding and hybridization. Methods Molecular Biology 453, 3-31.

Murchie, A.I.H., Clegg, R.M., Kitzing, E.V., Duckett, D.R., Diekmann, S., Lilley, D.M.J., (1989) Fluorescence energy transfer shows that the four-way DNA junction is a right-handed cross of antiparallel molecules. Nature 341, 763–766.

Ong, H.S., Rahim, M.S., Firdaus-Raih, M., Ramlan, E.I., (2015) DNA Tetrominoes: The Construction of DNA Nanostructures Using Self-Organised Heterogeneous Deoxyribonucleic Acids Shapes. PLoS ONE 10, e0134520.

Pinheiro, A.V., Han, D., Shih, W.M., Yan, H., (2011) Challenges and opportunities for structural DNA nanotechnology. Nature Nanotechnology 6, 763–772.

Ramlan, E.I., Zauner, K.-P., (2013) In-silico design of computational nucleic acids for molecular information processing. Journal of Cheminformatics 5, 22.

Reuter, J.S., Mathews, D.H., (2010) RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics 11, 129.

Rothemund, P.W.K., (2005) DNA self-assembly with floppy motifs – single crossover lattices. Foundations of Nanoscience, Self-Assembled Architectures and Devices, Proceedings of FNANO'05. J.H. Reif eds, pp. 185–186.
Rothemund, P.W.K., (2006) Folding DNA to create nanoscale shapes and patterns. Nature 440, 297–302.
Rothemund, P.W.K., Papadakis, N., Winfree, E., (2004) Algorithmic self-assembly of DNA Sierpinski triangles. PLoS Biol. 2, e424.
Seeman, N.C., (1982) Nucleic-acid junctions and lattices. J Theor Biol 99, 237–247.
Wei, B., Dai, M., Yin, P., (2012) Complex shapes self-assembled from single-stranded DNA tiles. Nature 485, 623-626.
Williams, S., Lund, K., Lin, C., Wonka, P., Lindsay, S., Yan, H., (2008) Tiamat: a three-dimensional editing tool for complex DNA structures. In: Goel, A., Simmel, F.C., Sosík, P. (Eds.), The 14th International Meeting on DNA Computing Proceedings, Czech Republic: Silesian University in Opava, pp. 112–121.
Winfree, E., (1996) On the computational power of DNA annealing and ligation. In: Lipton, R.J., Baum, E.B. (Eds.), DNA-based computers. American Mathematical Society, Providence, Rhode Island, pp. 199–221.
Winfree, E., (1998) Algorithmic self-assembly of DNA. California Institute of Technology.
Yin, P., Hariadi, R.F., Sahu, S., Choi, H.M.T., Park, S.H., LaBean, T.H., Reif, J.H., (2008) Programming DNA Tube Circumferences. Science 321, 824-826.
Zauner KP. (2005). From Prescriptive Programming of Solid-State Devices to Orchestrated Self-organisation of Informed Matter. Unconventional Programming Paradigms. 2005;3566:47-55.
Zhu, J., Wei, B., Yuan, Y., Mi, Y., (2009) UNIQUIMER 3D, a software system for structural DNA nanotechnology design, analysis and evaluation. Nucleic Acids Research 37, 2164-2175.

Figure Legend

Fig. 1. Self-organisation of DNA polyominoes. (A) The formation of DNA polyominoes shapes using a single or multiple basic blocks. Each block may or may not have connector(s) to form inter-assembly between multiple polyominoes shapes. (B) The conceptual illustration of the assembly for a desired DNA configuration. The polyominoes shapes are assembled using complementary connectors (case 1). The assembly of polyominoes shapes would not occur without the presence of the connector motifs (case 2) or when non-complementary connector exists (case 3). DNA strands are used to assemble each individual polyominoes shape. Different DNA strands are labelled as DNA 1, DNA 2, DNA 3 and DNA 4. Whenever there is a presence of a connecter, its corresponding region (in another DNA sequence) will have sticky end to enable two polyominoes shapes to bind together. The assembly of four DNA strands used to form polyominoes shapes will twist the double-stacked DNA strands at an approximately 60° angle (Mao et al., 1999), which results in the DNA polyominoes shape to be floppy.

Fig. 2. Conceptual representation of the formation of DNA polyominoes shapes. (A-H) represents the formation of T-shape, W-shape, F-shape, E-shape, V-shape, L-shape, B-shape and I-shape. The resulted four-way junction in the DNA polyominoes shapes (except for I-shape) are structurally floppy. Basic blocks were used to form four long continuous single-stranded DNAs (ssDNAs). DNA strands were represented as DNA 1, DNA 2, DNA 3 and DNA 4. The arrows in the DNA strands indicated the 5' to 3' direction.

Fig. 3. An example segmentation for DNA heterogeneous shapes (A) DNA duplex (I-Shape) and (B) Holliday Junctions (B, E, W, I, T, F, L, V-shape).

Fig. 4. The nucleotides arrangements of the 3 x 4 DNA network for (A) Set 1 (B) Set 2 (C) Set 3 (D) Set 4 and (E) Set 5. The arrows represent 5' to 3' terminal and the dotted lines represent complementary binding. The symbol (*) on the 3 x 4 DNA network (right) represents the location of the sticky ends used for intermolecular binding between different DNA shapes.

Fig. 5. Gel electrophoresis result for (A) Set 1 on 8% non-denaturing gel (B) Set 2 on 5% non-denaturing gel (C) Set 3 on 5% non-denaturing gel (D) Set 4 on 4% non-denaturing gel and (E) Set 5 on 5% non-denaturing gel.

Fig. 6. AFM images showed the image size of (A) Set 1 (B) Set 2 (C) Set 3 (D) Set 4 and (E) Set 5 at 100 nm x 75 nm. Each region of the image is labelled with orange color and the AFM images are compared with the predicted representation (Design of 3 x 4 DNA network). The final structures captured in the AFM images are structurally floppy due to the single crossover lattices in each DNA polyominoes shape (except I-shape) that has a native angle of approximately 60°.

Fig. 7. The connectivity map for each node in (A) Set 1 (B) Set 2 (C) Set 3 (D) Set 4 and (E) Set 5. Black lines indicate the binding affinity between the respective nodes, which is equals to 1.0. Blue dashed lines indicate nodes that are derived from the same DNA strands, which are then used to decide on the emergence of potential sticky ends binding region. Orange lines reveal the nodes with the

binding affinity value of $0.7 < P_{i,j} < 1.0$. The colour legends represent the type of DNA shapes involved in the configuration of the network.

Tables

Table 1. Algorithm for computing all probable paths.

```
1:  Split DNA into different segments (Node), N
2:  Define bounded node (form base pairing)=N_b
3:  Define unbound node (free sticky ends)=N_u,
4:  Initialise all initial node concentration, [N] = 1.0
5:      For each N_u do
6:          Check probability matrix
7:              If P_e > ThresholdValue, 0.7
8:                  Record new node, N_TempoNew bind to N_b
9:                      For each N_TempoNew do
10:                         Check all nodes concentration, [N] in the solution
11:                         If  [N_all] > 0% then
12:                             N_TempoNew is bind to N_u
13:                             Compute new Sticky Ends, N_u
14:                             Record N_u
15:                             Update latest total solution concentration
16:                             [N_Latest] =[N_Current]-[N_UptakeRate]
17:                         else
18:                             No binding, [N_NewCurrent]= [N_NewCurrent]
19:                     end for
20:             end if
21:     end for
```

Table 2: Summary of the numbers of graphs generated through the searches.

Combinations	Number of correct graphs	Number of graphs	Binding affinity $0.7 < P_{i,j} < 1.0$
Set 1	17	200	0.74, 0.76
Set 2	16	469	0.71, 0.72, 0.75, 0.77
Set 3	16	605	0.72, 0.72, 0.72, 0.73, 0.75
Set 4	12	757	0.72, 0.72, 0.72, 0.73, 0.77, 0.79, 0.84
Set 5	21	31	0.73

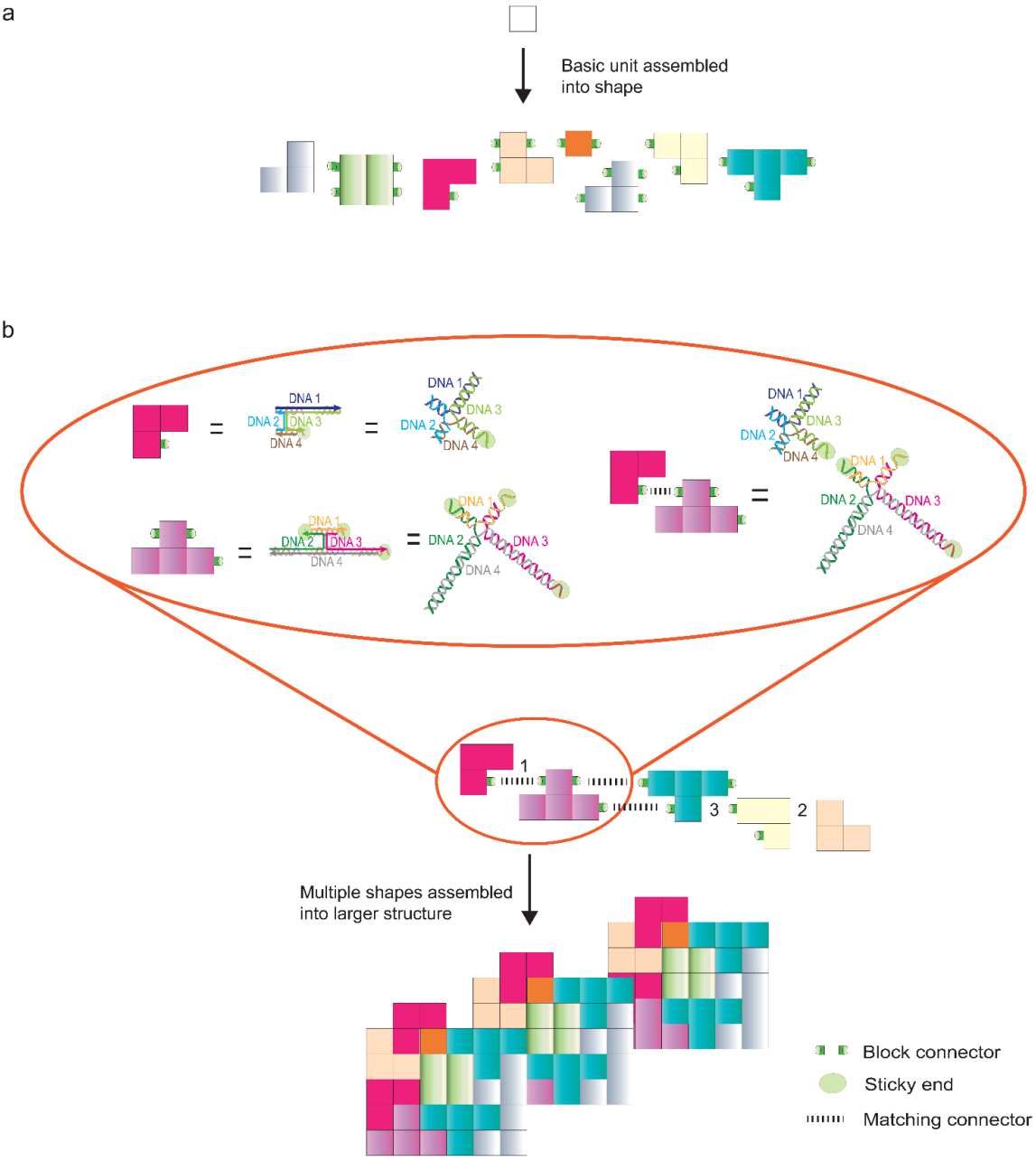

a
Basic unit assembled
into shape
b
DNA 1
DNA 2
DNA 3
DNA 4
1
2
3
Multiple shapes assembled
into larger structure
Block connector
Sticky end
Matching connector

Figure 1

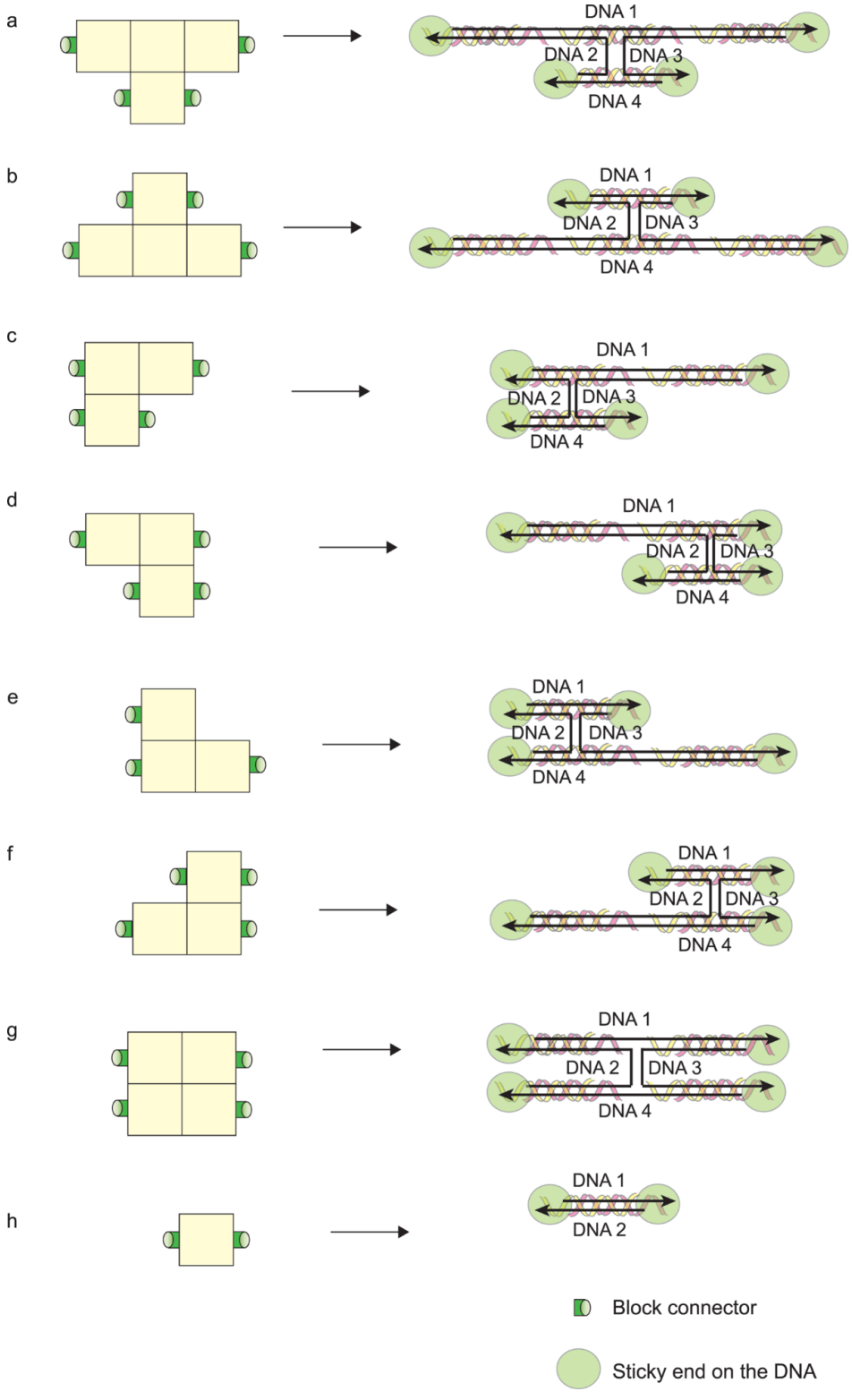

a
DNA 1
DNA 2
DNA 3
DNA 4
b
DNA 1
DNA 2
DNA 3
DNA 4
c
DNA 1
DNA 2
DNA 3
DNA 4
d
DNA 1
DNA 2
DNA 3
DNA 4
e
DNA 1
DNA 2
DNA 3
DNA 4
f
DNA 1
DNA 2
DNA 3
DNA 4
g
DNA 1
DNA 2
DNA 3
DNA 4
h
DNA 1
DNA 2
Block connector
Sticky end on the DNA

Figure 2

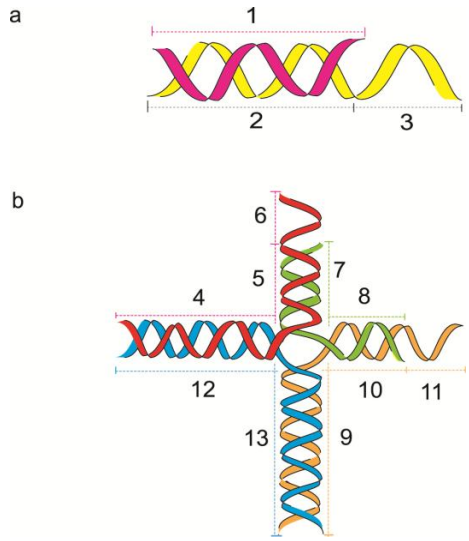


Figure 3

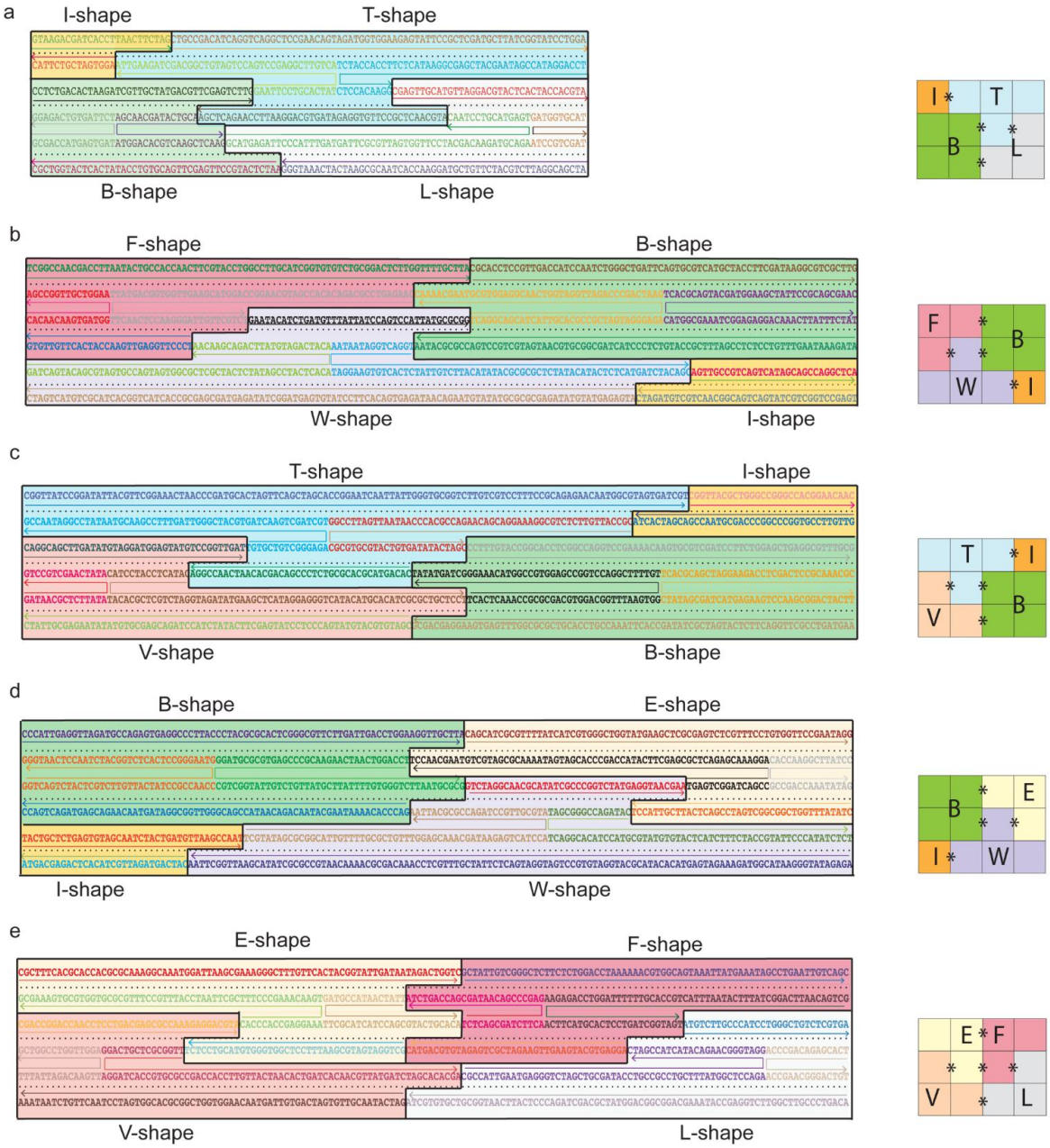
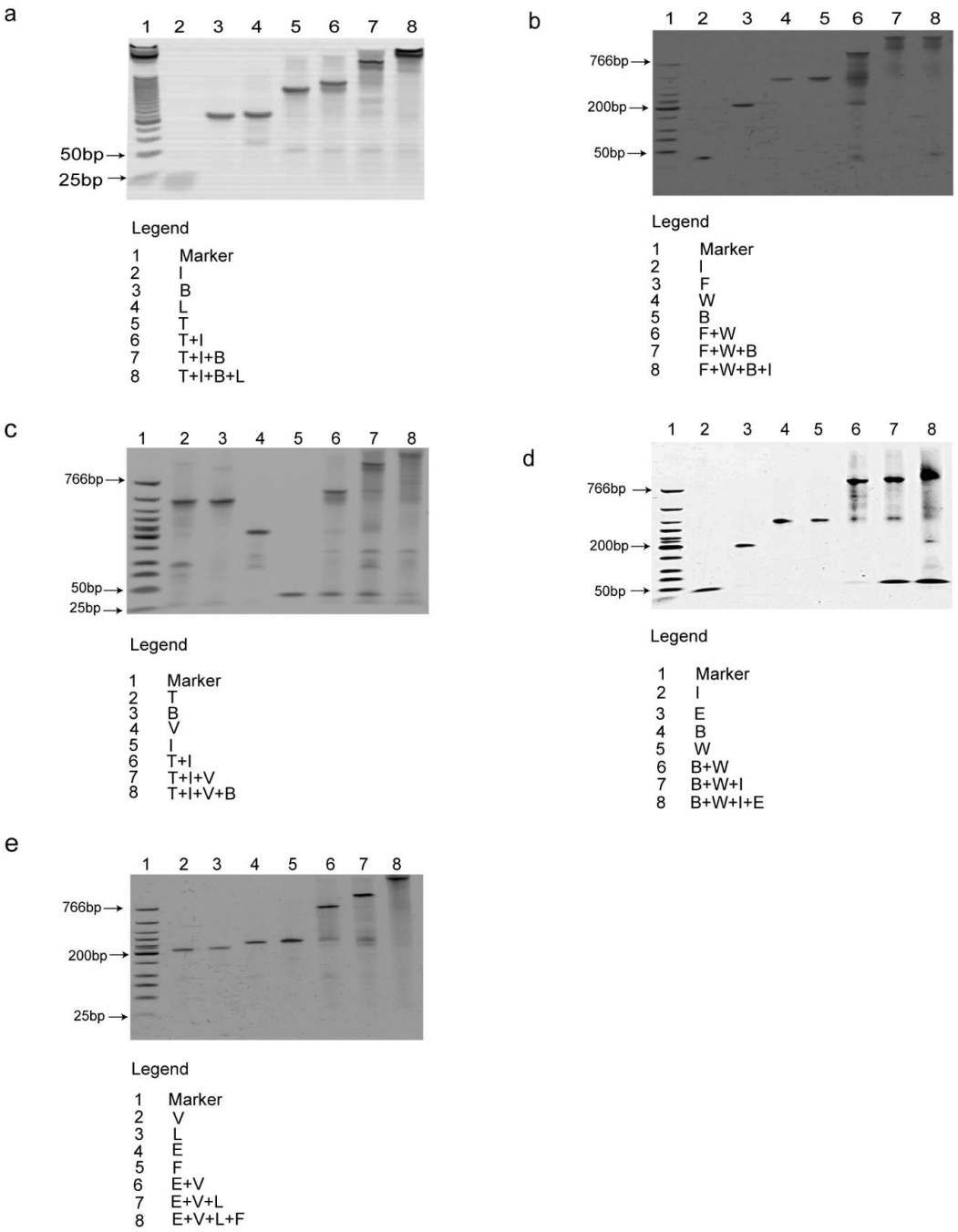

a
I-shape
T-shape
B-shape
L-shape
b
F-shape
B-shape
W-shape
I-shape
c
T-shape
I-shape
V-shape
B-shape
d
B-shape
E-shape
I-shape
W-shape
e
E-shape
F-shape
V-shape
L-shape

Figure 4

474

475



476

477 **Figure 5**

478

479

480

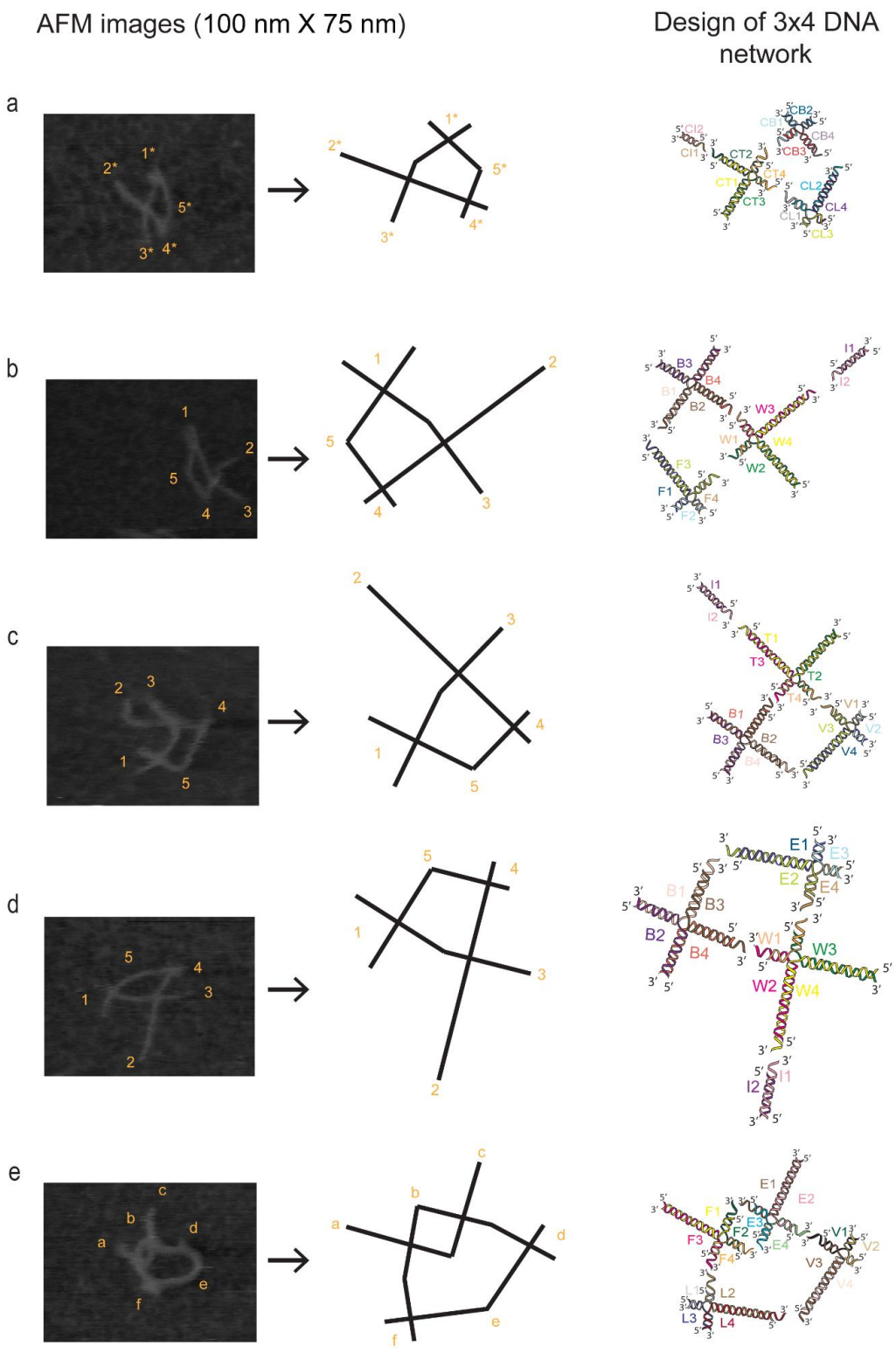


Figure 6

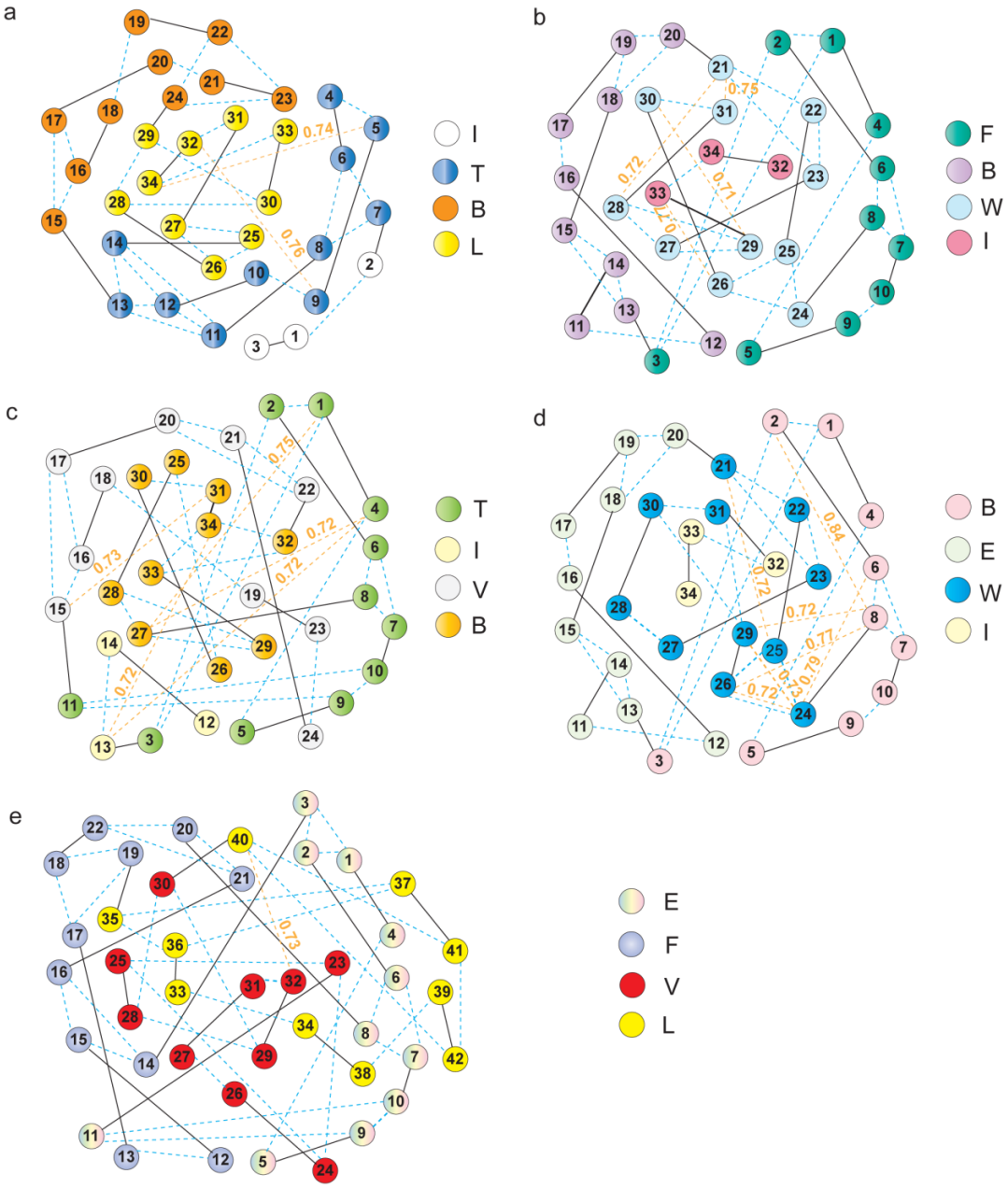


Figure 7

Figure 1 R
Click here to download high resolution image

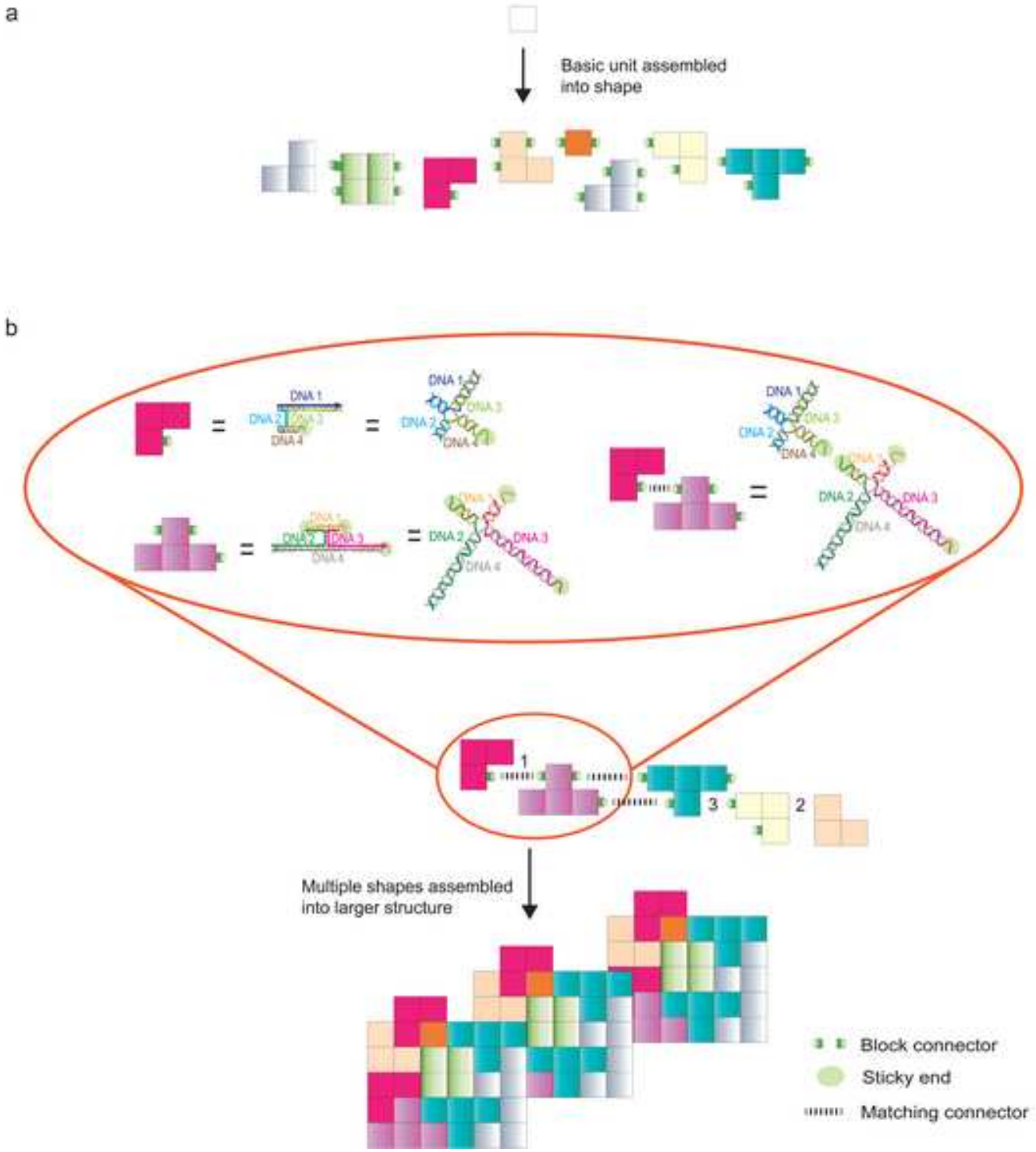


Figure 2 R

Click here to download high resolution image

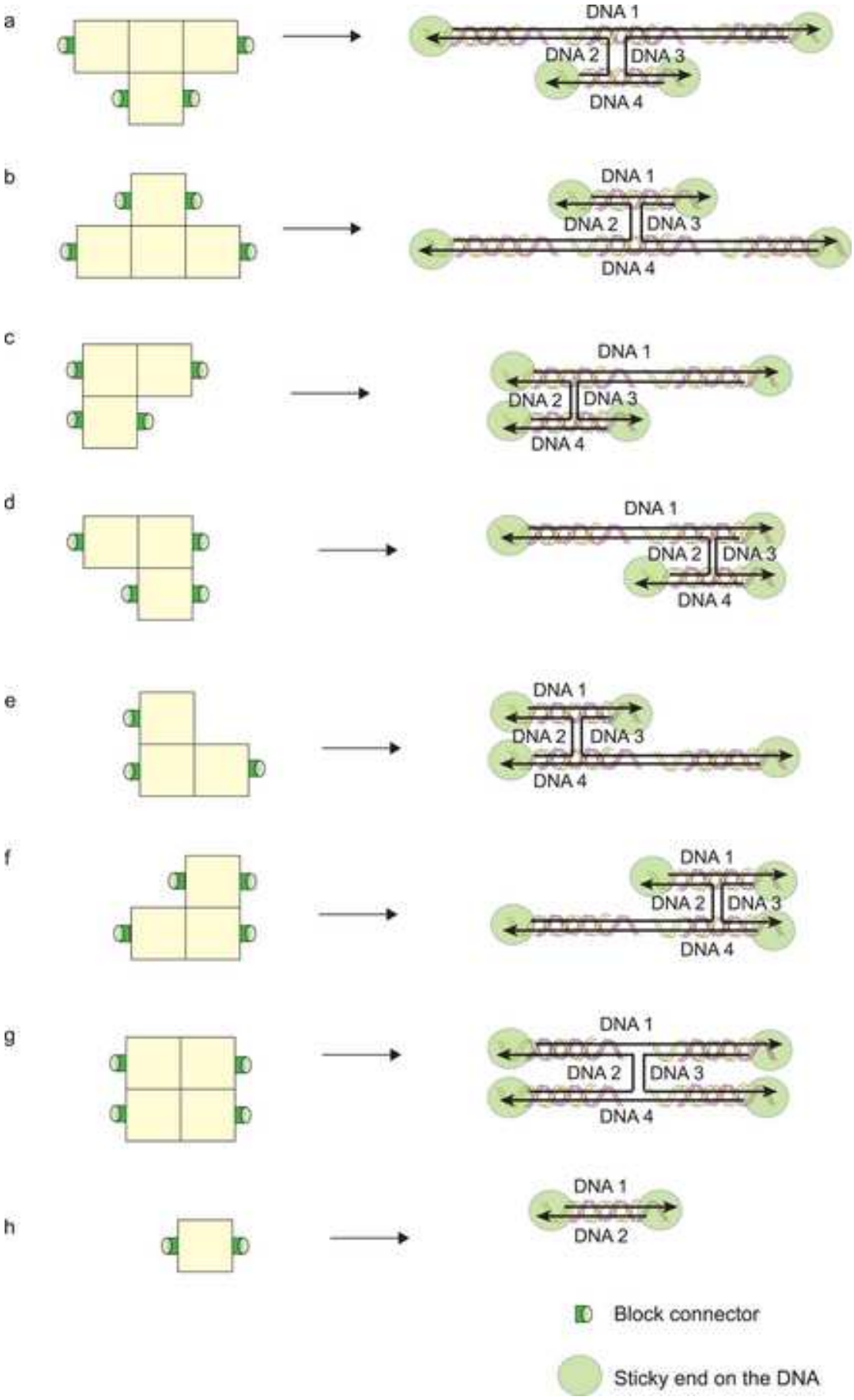


Figure 3 R
Click here to download high resolution image

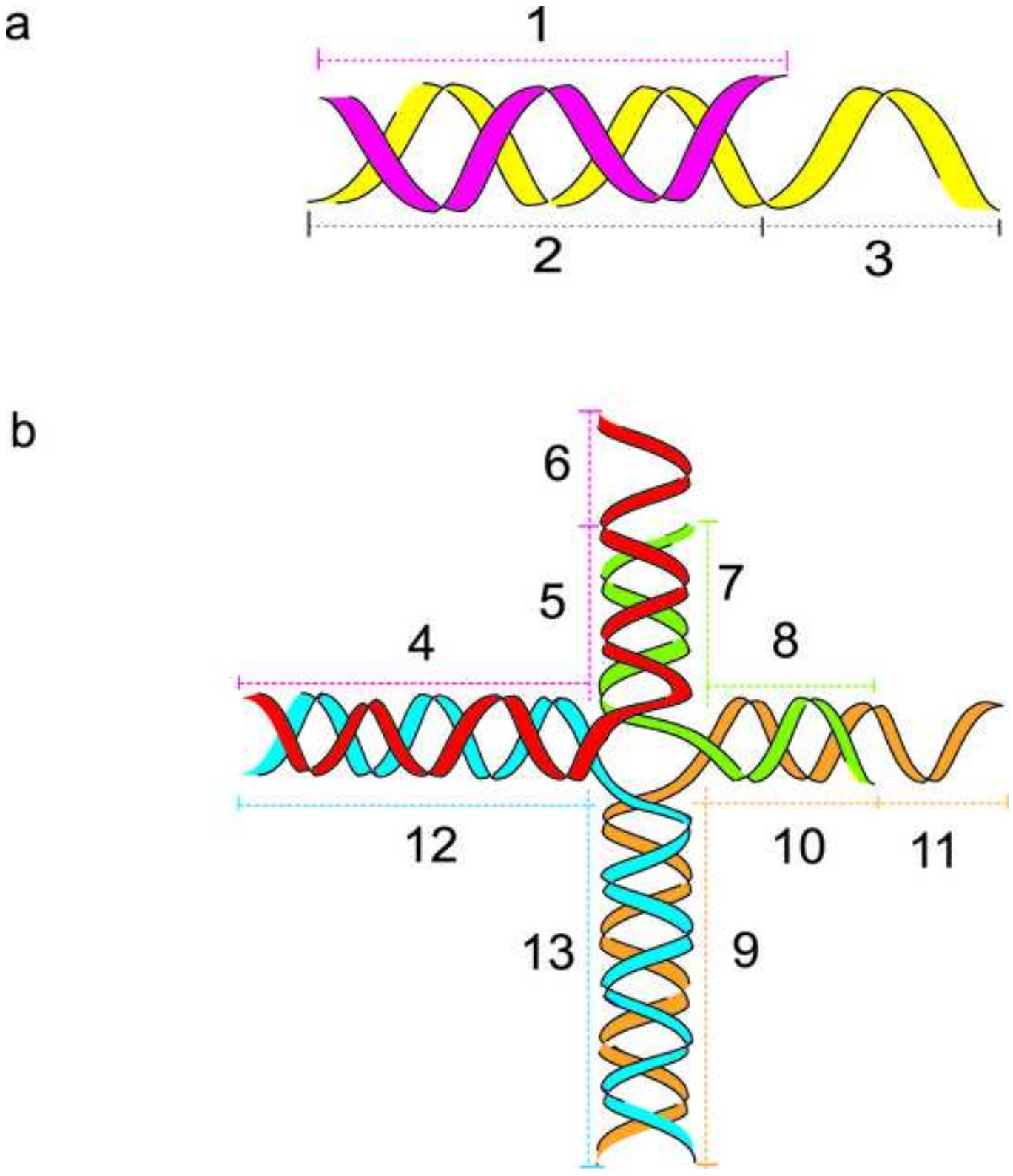


Figure 4 R

Click here to download high resolution image

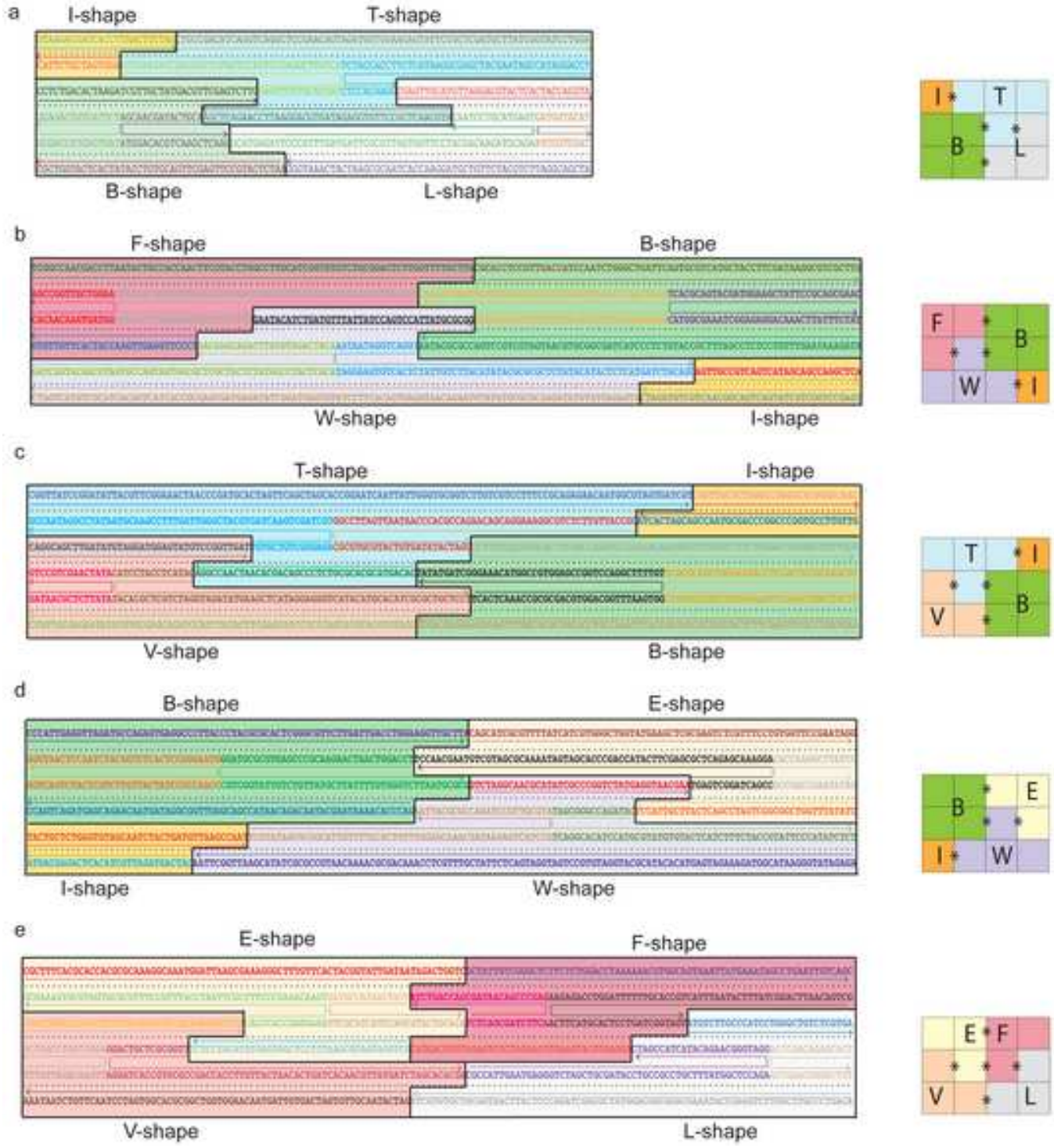
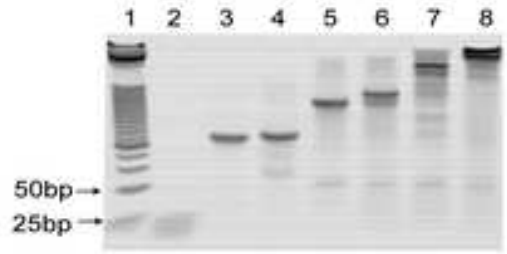


Figure 5 R

Click here to download high resolution image

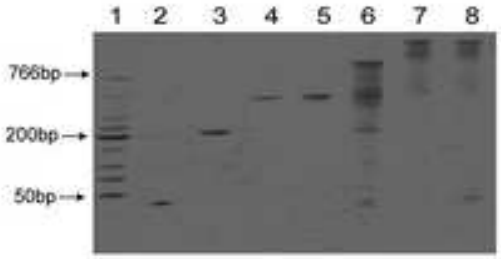
a



Legend

1	Marker
2	I
3	B
4	L
5	T
6	T+I
7	T+I+B
8	T+I+B+L

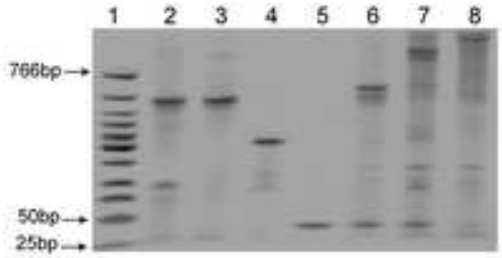
b



Legend

1	Marker
2	I
3	F
4	W
5	B
6	F+W
7	F+W+B
8	F+W+B+I

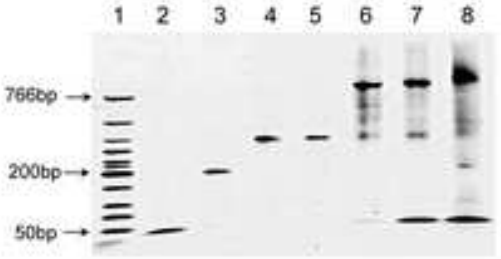
c



Legend

1	Marker
2	T
3	B
4	V
5	I
6	T+I
7	T+I+V
8	T+I+V+B

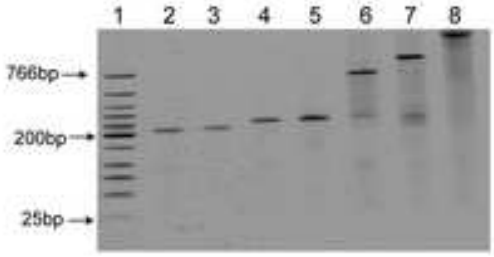
d



Legend

1	Marker
2	I
3	E
4	B
5	W
6	B+W
7	B+W+I
8	B+W+I+E

e



Legend

1	Marker
2	V
3	L
4	E
5	F
6	E+V
7	E+V+L
8	E+V+L+F

Figure 6 R

Click here to download high resolution image

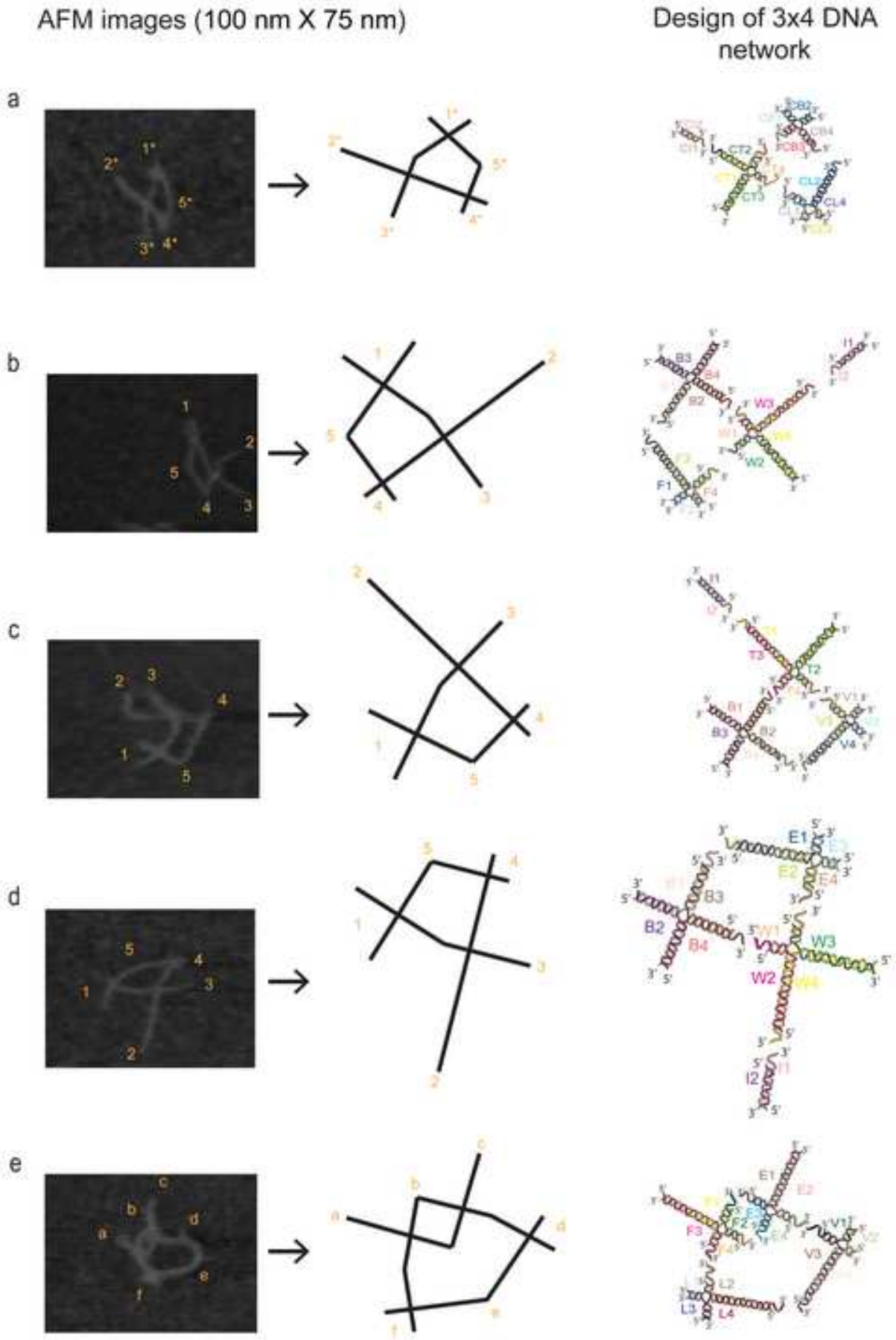
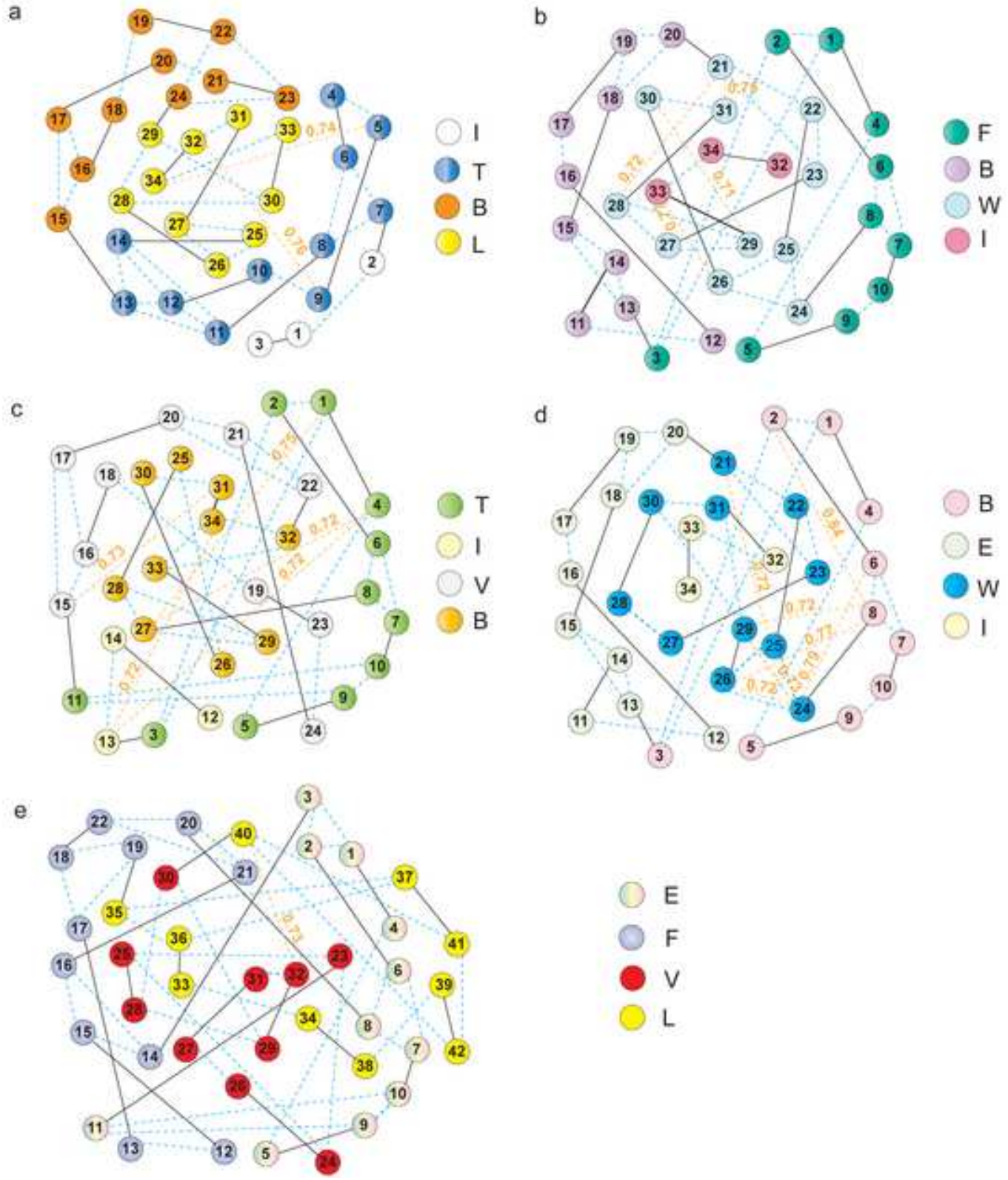


Figure 7 R
Click here to download high resolution image



Supplementary File R

Click here to download Supplementary File: Supporting Information.doc