# Supplementary Materials

## Equivalence between bond breakage and dilution

Here we demonstrate that the set of equations (1-2 ) also describes the system in which intra-polymer bonds are unbreakable, while concentrations of all molecules are diluted at rate $\beta$ and fresh monomers of type $i$ are supplied at rate $\phi_i = \beta \cdot c_i$. Indeed, the dilution adds terms $-\beta \cdot d_{ij}$ to the r.h.s. of equations 1. The dynamics of individual monomer concentrations (both bound and unbound), $c_i$, are given by equations $\dot{c}_i(t) = \phi_i - \beta c_i(t)$. After a brief transient regime, all monomer concentration $c_i(t)$ reach a steady state value $\phi_i/\beta = c_i$ so that Eqs. 2 become automatically satisfied. In this way the problem with dilution is exactly mapped onto the problem where bond breakages are allowed.

## Topological analysis of the undirected graph

Here we discuss the undirected graph representation of the pool of heteropolymers. Since our system always contains pairs of mutually complementary 2-mers $ij$ and $j^*i^*$ (with exception of self-complementary 2-mers $ii^*$), one can represent this pair with a single undirected edge connecting $i$ to $j^*$. These edges form an undirected graph shown in Fig. 3b. Note that due to these rules an edge connecting vertices $i$ and $k$ in this graph represents a pair of 2-mers $ik^*$ and $ki^*$ shown as two edges in Fig. 3a or two matrix elements in Fig. 2c. For simplicity in Fig. 3b we did not assign weights to these symmetric edges. Each edge of this graph corresponds to exactly one equation in the set of Eqs. 5. Hence, the number of undirected edges is equal to $(N + N_{sc})/2$ and according to Eq. 6 it cannot exceed $Z$. On the other hand, based on network topology, this number can be expressed as $Z - N_{comp} + N_{cycles}$. Here $N_{comp}$ is the number of connected components of the graph, while $N_{cycles}$ - the number of independent cycles defined as the minimal number of edges one needs to cut to remove all cycles. The inequality (6) means that the number of independent cycles cannot be larger than the number of components. Furthermore, this inequality must be also satisfied for each of the individual connected components of the graph because the number of equations (edges) cannot exceed the number of independent variables (the number of vertices in this components). In other words, *each of the components may contain no more than one cycle*. Graphs with this property are known as ”pseudoforests” [27]. For unicyclic components (i.e. those that include exactly one cycle), the numbers of edges and vertices are equal to each other, while for each of the tree (cycle-free) components, the difference between these two numbers is equal to $1$. This gives a topological interpretation to the number of surviving 2-mers $N$ in terms of the number of tree-like components of the undirected graph, $N_{trees}$:

$$N = 2Z - N_{sc} - 2N_{trees} \quad , \tag{7}$$

which automatically leads to inequality (6).

One can also demonstrate that only the cycles of odd lengths (1,3,5, etc.) are allowed in our system. Indeed, for a hypothetical even-length cycle $i_1 - i_2^* - i_3 - i_4^* - ...i_{n-1} - i_n^* - i_1$, one can construct a combination of Eqs. 5 of the following form:

$$\frac{\Lambda_{i_1 i_2}\Lambda_{i_3 i_4}...\Lambda_{i_{n-1} i_n}}{\Lambda_{i_2 i_3}...\Lambda_{i_{n-2} i_{n-1}}\Lambda_{i_n i_1}} = 1 \tag{8}$$

Here $\Lambda_{ij} = 1 - \Delta_{ij} = \lambda_{ij} \cdot \lambda_{j*i*} \cdot l_i \cdot r_j \cdot r_{i*} \cdot l_{j*}$ All the variables $l_i$ and $r_i$ at the left-hand-side of this equation cancel, making it an invariant that depends only on ligation rates. Therefore this equation cannot be satisfied for a generic matrix $\lambda_{ij}$, which rules out the existence of even cycles in most of the cases.