Soft Matter

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/softmatter

New concept for molecular engineering of artificial enzymes: A multiscale simulation†

Pavel V. Komarov,[a,b] Pavel G. Khalatur,[c,*] Alexei R. Khokhlov[a,c,d]

[a]Institute of Organoelement Compounds, Russian Academy of Sciences, Moscow, 119991 Russia
[b]Department of Theoretical Physics, Tver State University, Tver, 170100 Russia
[c]Department of Advanced Energy Related Nanomaterials, Ulm University, Ulm D-89069, Germany
[d]Physics Department, Moscow State University, Moscow, 119992 Russia

AUTHOR INFORMATION

*Corresponding Author
E-mail: khalatur@polly.phys.msu.ru | pavel.khalatur@uni-ulm.de

ABSTRACT: We propose a new concept for designing of artificial enzymes from synthetic protein-like copolymers and non-natural functional monomers which in terms of their affinity to water can be divided into two categories: hydrophobic and hydrophilic. The hydrophilic monomers comprise catalytically active groups similar to those in the corresponding amino acid residues. A key ingredient of our approach is that the target globular conformation of protein-like, core-shell morphology with multiple catalytic groups appears spontaneously in the course of controlled radical polymerization in a selective solvent. As a proof of the concept, we construct a fully synthetic analog of serine hydrolase, e.g. α-chymotrypsin, using the conformation-dependent sequence design approach and multiscale simulation that combines the methods of "mesoscale chemistry" and atomistic molecular dynamics (MD). A 100-ns GPU-accelerated MD simulation of the designed polymer-supported catalyst in aqueous environment provides valuable information on the structural organization of this system that has been synthesized in our Lab.

KEYWORDS: polymer-supported catalyst, computationally designed enzymes, biomimetic approach, protein-like copolymers, multiscale simulation

†Electronic supplementary information (ESI) available: **Movie 1**. Temporal evolution of a copolymer globule having a protein-like core-shell microstructure at mesoscale. The copolymer is built up from hydrophobic (H) and hydrophilic (P) monomers. H segments are given in gray, P segments are in green, catalytic triads are depicted as red triangles with spheres at their center. **Movie 2**. Temporal evolution of a single globule of the designed copolymer in water environment during a 10 ns molecular dynamics simulation. Like enzymes, this water-soluble "synzyme" - a functional analog of chymotrypsin - is tailored in such a way that it has a core-shell microstructure, with hydrophobic core and hydrophilic envelop. Catalytic triads are formed randomly from synthetic monomers imitating protein amino acid residues. The hydrocarbon backbone is represented by a two-tone "hose", in which chain segments connected with hydrophobic units are given in grey, while segments connected with hydrophilic units are colored green. The side chains of the monomer units are shown as sticks. Catalytic triads are depicted as red triangles. Water is not shown.

1. Introduction

Many important synthetic and industrial processes are catalyzed by organic and inorganic catalysts. Catalytic methods for transformation of complex molecules is of special importance in the area of fine chemicals production and development of new materials with desired properties. The grand challenges in catalysis is the design and synthesis of catalytic structures with controlled reactivity and high durability. Experimental characterization of catalytic sites is however challenging, and it is rarely possible to perform their systematic adjustments to probe the influence of structural variations on performance. For a long time, the search for new and improved catalysts was characterized mostly by the use of empirical, trial-and-error methods rather than the result of rational concepts.

One of the most efficient and highly evolved catalysts are enzymes.1 They catalyze a large variety of biologically relevant chemical transformations, often with a speed and specificity unrivalled by conventional catalysts. Various enzymes are currently widely used in many contemporary industries such as food, feed, pharmaceuticals, etc. It is therefore natural to adopt enzymes as a model for the development of their synthetic analogues – biomimetic copolymers which are sometimes referred to as *synzymes* or *theozymes* (short for theoretical enzymes).

The rational design of synzymes from amino acid residues is a promising area of research with the great potential. Understanding the relationship between the primary sequence of a naturally selected protein and its native three-dimensional conformation is one of the key problems here. Typically, proteins fold to organize a very specific globular conformation, known as the protein's native state, which is in general stable in water under physiological conditions. Predicting a functional tertiary structure from a polypeptide sequence has presented a long-standing challenge to the scientific community. Given that the polypeptide sequence has such a strong effect on protein folding, efforts have focused on ways to design the targeted patterns of amino acid residues to induce folding. However, due to the large complexity of proteins and their catalytic sites, the molecular design of functionally and structurally equivalent copolymers, either by applying the complex synthetic strategies known in protein chemistry or with the aid of computational methods, is met with great difficulties. There are a number of books and excellent review articles that cover this subject (see, e.g., Refs. 2-12). Despite some progress in this rich and rapidly growing field, one can state that the current success rate is rather modest. Besides, it should be kept in mind that the use of many naturally occurring enzymes as biocatalysts for industrial applications has limitations by reason of their low stability, especially at elevated temperatures and in organic solvents.[13] In this context, it is reasonable to consider an alternative way to designing biomimetic catalysts.

There are two main paradigms in the sequence design problem. In protein science, it is assumed that the order of amino acids in a polypeptide chain should be the same for a particular protein so that all sequences of a given type are structurally identical copies: it is impossible to distinguish one individual sequence from another.[14] The synthesis of a polypeptide of significant length (e.g. ~10^2 residues) can in principle be performed using the Merrifield synthesis (peptide synthesis machine) that involves attaching the C-terminus of the peptide chain to a polymeric solid and the step-by-step addition of amino acids to the growing chain.[15,16] It is clear that due to the synthetic challenges inherent in generating precise sequence-specific polymers in quantity, this direction may at best serve to laboratory purposes.[16] On the other hand, if we deal with fully synthetic copolymers obtained, for example, by radical polymerization, the occurrence of a certain degree of sequence disorder is inevitable. Accordingly, when we are speaking about the properties of these copolymers, we mean, explicitly or implicitly, that averaging over many different sequences has been carried out.[14] In many cases, the synthetic design paradigm based on the rather simple methods of polymer chemistry can be more promising. In what follows, we rely on this approach.

For reproducing enzyme-like behavior with a synthetic polymer, one needs the corresponding monomers, in which the required functionality is built into their chemical structure, and an efficient and robust synthetic route for producing the polymer with appropriate monomer sequence distribution along the chain. What is of paramount importance is that the primary copolymer sequence should ensure spontaneous folding of the chain into compact water-soluble globules that serve as a matrix with a high-local density of catalytically active groups. It is this feature that provides the main novelty of our methodology.

Biomimetic chemistry, the term introduced by Breslow,[17-19] is currently widely used to synthesize various tailor-made catalytic systems, ranging from relatively small organic compounds, which imitate working features of these naturally occurring compounds, to linear and branched macromolecules with different functional groups, molecularly imprinted polymers, etc.[20-22] Early attempts came from the work of Overberger and Salamone on poly-4(5)-vinilimidazole[23] and Klotz on polyethyleneimine (PEI).[20,24-26] In particular, Klotz and co-workers have reported the synthesis of a PEI-based polymer containing dodecyl chains for binding sites and imidazole moieties for functional catalytic groups.[24] This highly branched polymer with synthetically incorporated catalytic groups catalyzed the hydrolysis of phenolic sulfate esters by a mechanism resembling that of the natural enzyme called IIA sulfatase. To mimic the catalytic action of the serine protease class of enzymes (e.g., α-chymotrypsin), Karmalkar at al. have proposed a molecular imprinting technique that positions hydroxyl, carboxyl and imidazole

groups within a hydrogel matrix.[27] The template molecule, 2-([(isobutyrylamino) caproyl]-l-phenylalanyl)2-aminopyridine, was removed, and the catalytic activity of the molecularly imprinted material was checked for the hydrolysis of nitrophenyl esters. The imprinted hydrogel catalyzed the hydrolysis at a rate comparable to that of chymotrypsin, whereas the non-imprinted hydrogel exhibited only modest activity. Over the last years, a wide range of structurally diverse molecular systems to mimic the functions of natural enzymes have sprung up.[28-35] Although some of these and other fully synthetic systems have demonstrated a high catalytic activity, they can be hardly qualified as true enzyme models. Indeed, most of protein catalysts have a compact globular conformation with more or less sphere-like shape and maintain structural integrity and functionality in aqueous media under physiological conditions, while organic polymers designed as enzyme mimics are not shape-persistent objects and do not operate in water because they are typically usefully soluble only in organic solvents.[21]

In this communication, we report an inexpensive and reliable approach to designing bio-inspired polymer-supported catalysts from non-natural monomers with polymerizable double bonds. As a proof of the concept, the computer-aided molecular design of a synthetic analog of α-chymotrypsin will be performed and the salient results will be discussed. This enzyme is chosen as a representative prototype for artificial catalyst. Although each design scheme is unique in detail for a particular enzyme, all would rely on the same principles. Using multiscale simulation and so-called conformation-depended sequence design,[14] we will show that it is possible to map the most essential properties of the natural target system onto a relatively simple synthetic copolymer capable of folding spontaneously in aqueous media under physiological conditions into a compact and well-packed three-dimensional structure with multiple catalytic centers.

The rest of the paper is organized as follows. First, the main ideas behind our design concept are considered in detail. In the next section, we discuss how this approach can be applied to constructing α-chymotrypsin-inspired copolymer. We then give a brief outline of our multiscale simulation procedure. The next section provides a detailed description of the models and simulation methods used in this study. We then present and discuss the main results. Finally, we conclude the paper by highlighting the key observations and specifying future research directions.

2. Enzyme-inspired copolymers or multiple catalytic sites vs. single site

Our concept for designing artificial enzymes is predicated on the following facts that govern the functional properties of these heteropolymers. (i) Enzymes are composed of amino acid residues which can be divided in terms of their affinity to water into two main categories: hydrophobic (H) and hydrophilic or polar (P) residues.[36] (ii) The HP sequence pattern is a key

factor that determines both secondary and tertiary protein structure: the position of H and P residues along the polypeptide backbone is very important for structure formation and folding. It is known, for example, that for the same HP pattern, the particular side chain is not as important as the hydrophobic or polar character of the amino acid.[37] (iii) The catalytic activity of the majority of enzymes is realized in a globular state when most of the H residues are in the dense core of the globule while the P residues form the envelope of the core, thereby preventing the globules from aggregation and enzymatic degradation in aqueous environment. (iv) This core-shell microstructure is associated with a specific sequence distribution of the monomers comprising the copolymer chain, which is capable of folding spontaneously to a stable native state. (v) The catalytic function of enzymes depends upon the integrity of protein conformation and is directly linked with the spatial organization of active-site amino acids, which are usually disposed close to enzyme surface. It is this well-organized three-dimensional conformation that determines the macroscopic properties and function of enzymes. (vi) Complex secondary/tertiary protein structures are not a prerequisite for catalysis and natural proteins do not have a monopoly on biological catalysis.4

As noted in the Introduction, the main idea pursued in this study is to construct bio-inspired polymer catalysts directly synthesized via the radical copolymerization of synthetic H and P monomers which mimic natural amino acids. Such copolymers should form non-aggregating globules in a solution and serve as a support for catalytically active groups. Since natural enzymes contain typically only one catalytically active group (with the exception of multidomain enzymes, where different enzyme activities are located in multiple catalytic centers on different parts of the protein chain), a promising way to control and facilitate the catalytic activity of the designed copolymer is to assemble many catalytically active groups at the surface of a specifically constructed water-soluble globule. These systems can actually be regarded as surface nano-reactors capable of concentrating reactants in limited volume where spatial confinements of reactants can boost reactivity significantly. Indeed, it is known that the speeding-up of the reaction rate, for example, in micelle or dendrimer solutions can reach hundreds of times as a result of such a redistribution of reactants.[20] It is therefore believed that the presence of many surface catalytically-active groups, each one having the structure similar to that in its natural counterpart, along with the ability to concentrate reactants in the limited volume of the designed copolymer globules can control and accelerate target reactions. It is natural to call this class of putative biomimetic catalysts "enzyme-inspired nanoreactors".

Of course, one of the main drawbacks anticipated for these systems lies in the fact that catalytic sites will exhibit a highly dynamical behavior, in which their structure will undergo

continuous changes both in the geometric arrangements of catalytic groups and their spatial orientations. Natural enzymes are fairly rigid objects with small atomic fluctuations and their active site is typically contained in a structurally-rigid hydrophobic environment. Although some of these features are not expected for the synthetic analogous, their absence would be compensated by the presence of multiple catalytically active groups on the surface.

3. Chymotrypsin-inspired copolymer

As an example, we consider α-chymotrypsin that belongs to a class of serine hydrolases. Its active site is built from residues Ser195, His57 and Asp102, which represents a well known "catalytic triad" (CT) of hydroxyl, carboxyl and imidazole groups participating in the catalytic cleavage of ester or amide bonds (cf. Figure 1a). Serine hydrolases with the Ser-His-Asp triad are among the most studied and representative group of hydrolases. Apart from the serine hydrolases, the Ser-His-Asp catalytic triad also occurs in several lipases. Although amino acids would be the most obvious choice of material from which to construct a synthetic enzyme, other types of building blocks can be selected for this purpose.

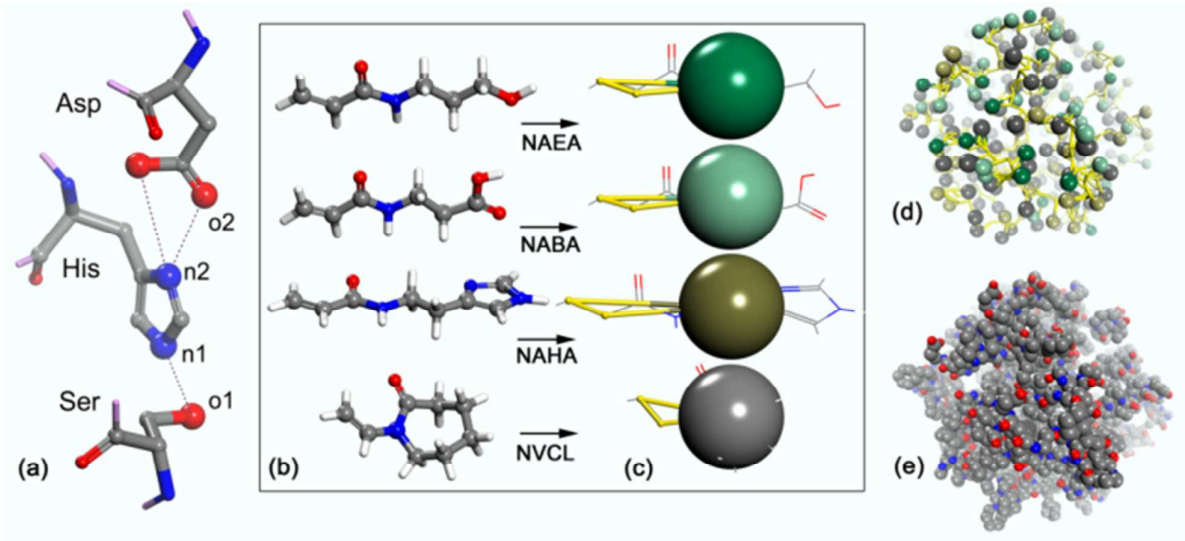


Figure 1. (a) Ball and stick representation of the catalytic Ser-Asp-His triad of α-chymotrypsin. Active site combines a binding pocket, nucleopholic hydroxyl, carboxyl and imidazole group in an organized array of hydrogen-bonded amino acids. The amino acid residues are in three-letter code. Atoms composing the triad are also specified. (b) Hydrophilic (NAEA, NABA, NAHA) and hydrophobic (NVCL) monomers mimicking amino acid residues (for the conformity, see Table 1). (c) In mesoscale simulation, atomistic structures of the monomers with polymerizable double bonds are approximated by coarse-grained (CG) "composite" particles consisting of non-reactive beads (large spheres) and reactive linkers (small yellow balls). Beads are normal CG particles; they participate in all site–site interactions. On the other hand, linkers are "semivirtual" sites - they participate in intramolecular interactions only (namely, in bond stretching and angle bending interactions) but can form covalent bonds; that is, they can react. (d) The CG monomers react with each other in water environment to form a polymer chain that gradually collapses into a compact structure. (e) After a long equilibration of the CG system, a reverse mapping procedure is applied to give atomistic structure of the designed copolymer.

The synthetic analog of α-chymotrypsin can be constructed from (at least) four types of copolymerizing monomers. One of them should be hydrophobic in order to provide the formation of a compact globular core, while three other monomers should be hydrophilic and include catalytically active groups: hydroxyl (-OH), carboxyl (-COOH), and imidazole (IM). It is assumed that the copolymerization of these monomers at varying stoichiometries can appropriately mimic both the protein hydrophobic matrix and the enzyme catalytic site. The main problem to be resolved by computer-aided design is that overall polymer-supported catalytic structure should naturally be created in a self-organized and concerted manner to produce shape-persistent core-shell and, at the same time, rather flexible configuration that allows reactants to be delivered to the catalytic site. To achieve this, both catalytically active groups and specific building blocks responsible for the formation of soluble (non-aggregating) globules have to be incorporated into the designed copolymer sequence. To that end, we will use the conformation-dependent sequence design (CDSD), the design scheme that takes into account a strong coupling between the conformation and primary structure of copolymers during their synthesis.[14,38,39] By controlling the polymerization conditions it is possible within the framework of this approach to create synthetic polymers with chemical functionalities and fold them into desired conformations.

The possibility to obtain the copolymers capable of forming core-shell microstructures via the CDSD-based redox-initiated free-radical copolymerization of the monomers differing in hydrophilicity/hydrophobicity has been for the first time demonstrated in the experimental work performed in our Lab.[40,41] It was shown that such polymerization process can automatically lead to copolymers with a specific sequence and composition that allows the macromolecules to self-assemble into non-aggregating globules in aqueous solution. Very recently, this methodology was extended in our Lab[42] in order to implement the design scheme described above. The synthesis was accomplished using copolymerization of four monomers, N-acryloyl-2-ethanolamine (NAEA), N-acryloyl-β-alanine (NABA), N-acryloyl-histamine (NAHA) and N-vinylcaprolactam (NVCL), differing in hydrophilic/hydrophobic characteristics (cf. Table 1 and Figure 1b).

Table 1. Synthetic monomers and their correspondence to amino acid residues.

Monomer	Functional group	Amino acid	Polarity, H/P
N-acryloyl-2-ethanolamine (NAEA)	Hydroxyl	Ser	P
N-acryloyl-β-alanine (NABA)	Carboxyl	Asp	P
N-acryloyl-histamine (NABA)	Imidazole	His	P
N-vinylcaprolactam (NVCL)			H

The H monomer (NVCL) was chosen to be sufficiently hydrophobic to drive globule formation, but not too hydrophobic such that the resultant copolymer would become insoluble in

water. The monomers NAMEA, NABAL and NAHA were chosen to be polar and similar (with respect of their functional groups) to Ser, Asp and His amino acid residues in α-chymotrypsin, respectively. Their structures and relative fractions in the copolymer chain were such as to provide enough hydrophilicity to maintain solubility in aqueous environment. The copolymer was synthesized by copolymerization in water at a temperature close to the phase separation point of the NVCL-based homopolymer solution, according to the procedure previously detailed.[40,41] This CDSD regime corresponds to the so-called copolymerization with simultaneous globule formation:[43] when the growing macroradical chain is long enough it forms a globule. In this way, the protein-like structure with predominantly hydrophobic core and hydrophilic outer envelope emerges. The catalytic behavior of the product was tested by monitoring the hydrolysis of *p*-nitrophenyl-propionate (*p*NPP). It turned out that when the prepared copolymer existed in a globular state, it catalyzed the reaction with remarkable efficiency, while strongly swollen polymer coils were practically inactive. This indicates that catalysis is driven primarily by specific, localized enzyme-like effects differing dramatically from the trivial nonspecific effects exhibited by the same monomers with the same catalytically active groups but distributed in solution more or less randomly. It is supposed that the acceleration of the hydrolysis by the designed copolymer is most likely due not only to the formation of H/P interfaces but also to local accumulation of catalytically active groups in its outer regions exposed to water.[41,42] Importantly, the catalytic activity of these copolymers is observed at much lower concentrations than that for their counterparts with alternating or random HP sequences.

These experimental results are encouraging but overall highlight considerable room for improvement in the field. In particular, it is not clear which of the synthesized structures are most likely to be catalytic, which copolymer composition and sequence will provide the maximum number of catalytic sites and how the sites are distributed over the globule. Another important question is whether the catalytic arrangement of a synthetic enzyme would be maintained in a dynamic and aqueous environment. Computer simulations can provide detailed information that often cannot be obtained experimentally.

4. Multiscale simulation strategy

In general, our multiscale simulation strategy includes the following four stages: (i) Atomistic structures of the polymerizing monomers (Figure 1b) are coarse grained into beads and springs (Figure 1c), and their atomic potentials are replaced by effective site-site potentials between beads. (ii) The coarse-grained monomers react with each other in water environment via radical polymerization with simultaneous globule formation and are incorporated into a copolymer chain

until a desired chain length is reached (Figure 1d). (iii) A reverse mapping procedure is used to recover and refine the atomistic structure of the system (Figure 1e). (iv) After equilibration, large-scale fully atomistic molecular dynamics (FA-MD) is performed to determine microstructural properties of the obtained copolymer. If necessary, the system can again be brought back to the CG representation (Figure 1d), the simulation can be continued starting from stage (iii) to modify or relax the system, and then go back to the atomistic level (Figure 1e) and study the changes in the properties.

5. Models and simulation techniques

Mesoscale simulation was performed using the method of dissipative particle dynamics (DPD).[44] Needless to say, this approach concerns no atomic details, but it is designed to reveal the most significant features of a complex polymer system. One of the most important parts of any mesoscale simulation is the adequate coarse-graining of the molecular system. In our DPD simulations, atomistic structures of the polymerizing monomers (Figure 1b) were approximated by coarse-grained (CG) *composite particles* (Figure 1c) consisting of non-reactive beads (large spheres) and reactive linkers (small yellow balls).[45] Beads are normal DPD particles; they participate in all interactions that include a conservative force, a dissipative force and a stochastic force. On the other hand, linkers are "semivirtual" sites: they participate in intramolecular interactions (namely, in bond stretching and angle bending interactions) but nonbonded conservative forces do not act on them. Another their important property is that they can form covalent bonds. This approach, which allows one to probe longer length- and timescales while making direct connection to more rigorous atomistic models, has demonstrated its utility in studying a variety of complex slow relaxing systems, such as highly cross-linked polymer networks and nanocomposites.[45-48]

In the CG approach used herein, the beads were placed at the center-of-mass of each of the monomer (Figure 1c), while the linker positions coincided with the positions of the corresponding carbon atoms which can bond together in the course of polymerization to form the polymer hydrocarbon backbone. Thus, each monomer molecule in our model is a single bifunctional DPD particle with two reactive sites. The connectivity of the DPD sites within a molecule was implemented by means of a simple harmonic bond potential. Copolymer chains were modeled as strings of H and P beads connected by linkers. It was assumed that the water bead (W) is represented by a single sphere containing eight water molecules at a density of 1 g/cm^3, so that both the water bead and the monomer beads are approximately of the same size.

The short-range (repulsive) interactions between different types of beads and polymer–solvent interactions were described by Flory–Huggins (FH) parameters χ. The hydrophilic

monomers NAEA, NABA and NAHA, being well soluble in water,[42] have close molar volumes and solubility parameters and the corresponding homopolymers demonstrate practically the same characteristic ratios, $C_\infty \approx 10$.[49] Therefore, for simplicity we set for them $\chi_{W,\alpha} = 0$ and $\chi_{\alpha,\beta} = 0$ (α,β = NAEA, NABA, NAHA). On the other hand, NVCL is a sparingly soluble amphiphilic compound that can provide the precipitation of a NVCL–containing copolymer from aqueous solution at certain chain length and temperature.[41,42] Using an extended FH model[50] and molecular simulation to calculate the compatibility of the binary NVCL/H_2O mixture, we obtained an estimate of 4.2 for $\chi_{W,NVCL}$. Finally, we set $\chi_{NVCL,\alpha} = 0$ (α = NAEA, NABA, NAHA).

Our simulation box with periodic boundary conditions was cubic and the global bead density was $\rho = 3\sigma^{-3}$, where σ denotes the characteristic bead size. The volume fractions of monomers and water, φ_m and φ_W, were 6.4 vf% and 93.6 vf%, respectively; that is, we simulated a diluted solution. The volume fractions of P monomers in the reaction mixture were $\varphi_P = \varphi_{NAEA} = \varphi_{NABA} = \varphi_{NAHA}$, in accordance with the experiment.[42] The fraction of P monomers in solution was controlled by the parameter $\varphi_P = 3\varphi_P/\Sigma_\alpha\varphi_\alpha$. It was varied from 0.3 to 0.7. By varying the φ_P value, one can control the hydrophilic/hydrophobic ratio f_P of the resultant copolymer and its morphologies in line with experiment.

To simulate chemical reactions, we used the concept of "mesoscale chemistry" in which reactions are probabilistic. The simulation scheme was similar to that described elsewhere.[43,45] The results were obtained by starting in a disordered configuration with randomly distributed DPD particles. It is assumed that there is a single "seed" radical in the reaction volume. This radical is transferred to a randomly selected H monomer and the polymerization process starts. One of the linkers at the chain end acts as a growth "active" site. An "active" site changes to an "inactive" one after linkage with next monomer linker that in turn becomes "active". The mechanism of chain propagation is governed by the local concentration of reactive monomers near the "active" end of the growing macroradical and some reaction probability $p_R = 2\times10^{-5}$, which is defined by reaction rate constants and adjusted to keep quasi-equilibrium conditions.[43,45] It is evidently that the reaction rate constants are very close to each other for the system studied here. A depolymerization reaction is not allowed and chain transfer reactions that increases the total number of chains are absent. Note that under conditions considered here, the chain growth is a reaction controlled process at initial and intermediate stages of polymerization. Much later, when the concentration of unreacted monomers becomes low enough and the diffusing species take a long time to diffuse to the growing macroradical end, the diffusion controlled regime is established. After the copolymer chain reaches a required length of N monomeric units (N = 500 or 1000), the polymerization process stops, the unreacted monomers are replaced by water beads,

the copolymer solution is relaxed and a long productive run (~10^6 time steps) is then performed. Since DPD provides high rates of molecular relaxation, it gives us an opportunity to investigate the polymerization on a large timescale, up to high monomer conversions.

The mesoscale simulations reported in this study were carried out using our highly scalable DPD code based on the domain decomposition parallelism and developed for performing computations on massively parallel supercomputers.[51]

Reverse mapping. After equilibration at mesoscale, a reverse mapping procedure (RMP) was used to recover the atomistic structure of the system (Figure 1e). The RMP includes the least-squares best superposition of the DPD coordinates and the atomistic templates via the quaternion method and another specific procedure that we call "chemical cleaning" (CC). The CC is a cascade of bond redistributions, checks for ring catenation and "spearing" effects, recalculation of the partial charges, setting the correct force field types for atoms depending on their surroundings, adding/removing hydrogen atoms, etc (for more detail, see Ref. 45). A refinement procedure, which combines energy minimization followed by MD-NPT simulation, was also applied as a convenient way to allow for further structure improvement. As a result, we ended up with fully relaxed atomistic structures of the copolymer in aqueous environment.

Atomistic simulation. For the atomistic molecular dynamics (MD) simulations, we employed the class II consistent force field PCFF (polymer consistent force field)[52] and the LAMMPS software package[53] compiled with GPU accelerated functions. The PCFF is an *ab initio* based force field in which the total potential energy is represented as a sum of intramolecular terms (the energy contributions for bond, bond angle, torsion and out-of-plane angle coordinates as well as the energy contributions for diagonal and off-diagonal cross-coupling terms between internal coordinates) and intra- and intermolecular non-bonded interaction terms (a Lennard–Jones "9-6" potential for the van der Waals interactions and a Coulombic term for electrostatic interactions). The PCFF force field was parameterized against a wide range of synthetic polymers. Since our focus was on a synthetic polymer in the present study, we used just this force field, not biologically-oriented force fields like AMBER, CHARMM, GROMOS, etc. MD runs were performed under NPT conditions at T = 300 K and P = 1 atm, using a Nosé−Hoover thermostat/barostat with relaxation constants of 1 and 2 ps for the thermostat and barostat, respectively. Integration time step was 1 fs. Pair non-bonded interactions – long-range Coulomb forces handled via the PPPM (particle-particle particle mesh) technique and van der Waals forces – were calculated on 32 NVIDIA Fermi GPUs at a cut-off distance of 18 Å, while the calculation of intramolecular bonded interactions in molecules and a spatial

hashing of the atom data into an array of atom bins was carried out on 320 CPUs of a hybrid high performance computing platform at MSU.

Our simulations were organized as follows. First, a relatively small 70,580–atom system obtained from the DPD runs for a 500-unit copolymer chain was simulated in a cubic periodic box, which extends to at least 20 Å between the copolymer and the edge of the box. Next, the box was replicated twice along the *x* and *y* directions, producing the initial box volume 2,859,680 $Å^3$, and the resultant 282,320–atom solution containing four chains was simulated during 100 ns after equilibration. Post-processing of the trajectories yielded data in the form of pairwise distance distributions, pair correlation functions, solvent accessibility, hydrogen bond formation, the number of catalytic triads, etc.

6. Results and discussion

6.1. Coarse-grained model

Protein-like core-shell structure. One of the main objectives of the mesoscale simulation was to reveal some general trends and regularities characteristic of our design scheme. First, we analyze the effect of copolymer composition on the chain conformation.

As with experiment, the key is to increase the hydrophilic fraction of the copolymer f_P in order to provide a core-shell morphology of polymer globules. In the simulation, the formation of the core-shell globules was observed only when $f_P > 0.3$. This result is in qualitative agreement with the ^{1}H NMR experimental data[42] showing that the aqueous solution of a copolymer with molecular weight ranging from 15 to 40 kDA and the NVCL:NAEA:NABA:NAHA composition of 0.630:0.102:0.150:0.118 (mole/mole) is stable, while the phase separation of the solution with the formation of polymer rich droplets occurs at $f_P < 0.1$.

The H and P density profiles, $\rho_H(r)$ and $\rho_P(r)$, for four independently generated 1000-unit copolymers are presented in Figure 2a, where the center of mass of the copolymer is taken as the origin ($r = 0$). The equilibrium chain conformation corresponds to a globule having a well-defined dense core filled with H beads distributed more or less uniformly and an outer shell where the density of P beads dominates (Figure 2b). Although there is rather strong penetration of P beads into the globular core, we do not observe a homogeneous structure of the globule with chain groups homogeneously distributed inside the globule. One may say that an intramolecular microphase separation of H and P chain groups takes place inside a protein-like core-shell macromolecule. The proximity of the results obtained from different runs demonstrates that the design scheme proposed is quite robust.

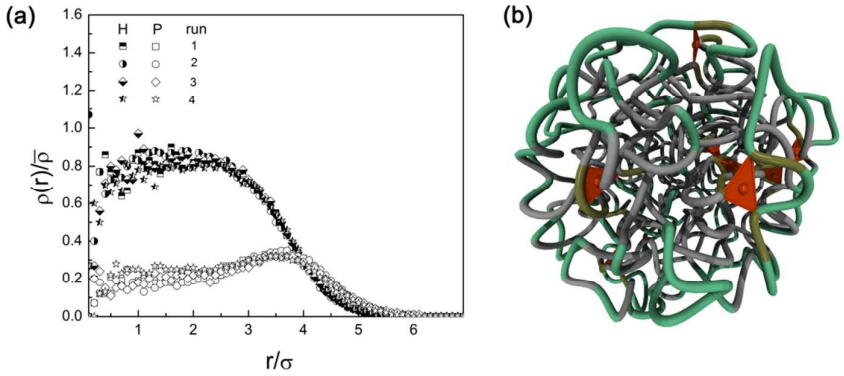


Figure 2. (a) Density profiles $\rho(r)/\overline{\rho}$ showing the distribution of H and P units around the center of mass of copolymers generated in four independent runs at $f_P = 0.4$ ($\bar{\rho}$ is the average particle density inside globule). (b) Typical snapshot of a mesoscale globule (H units are given in gray, P units are in green, catalytic triads are depicted as red triangles).

The reasons leading to the formation of core-shell morphologies can be explained as follows (for more detail, see e.g. Refs. 14 and 54). As long as the current chain length is small, the growing macroradical remains in a coil-like conformation and monomers of different solubility (hydrophobicity) are statistically distributed along the chain. However, when the chain length becomes sufficiently large, the copolymer tends to collapse due to the loss in translational entropy of free H species (NVCL in our case) after their incorporation into the growing chain. For flexible-chain polymers, a globule is formed already for short enough chains, $N \sim 10$. As a result, a hydrophobic–hydrophilic interface develops. The presence of the H/P interface dramatically changes the reaction conditions because the local reaction volume becomes strongly inhomogeneous and the microenvironment of the radical differs from that of the overall solution. Furthermore, the growing macroradical affects its own microenvironment. This means that the polymerization proceeds in a self-consistent manner: the chain conformation, on the one hand, and the type of attached monomers and the probabilities of their addition, on the other hand, are mutually dependent. This CDSD regime is closely related to the so-called bootstrap effect.[55] The hydrophobic core of emerging globule selectively absorbs H species from the solution, thereby increasing the probability of their attachment to the "active" chain end when it is located inside the core. When the growing chain end emerges at the globule surface and, thus, enters a more hydrophilic microenvironment, it becomes enriched in P units. This facilitates the microsegregation of H and P segments: the H segments are accumulated predominantly within the core, whereas the P segments are arranged at the periphery and stabilize the core. Since the chain conformation and the monomer sequence distribution in copolymer is determined by the structure of the macroradical as a whole, the formation of copolymer sequences with specific long-range correlations of a gradient type is observed.[56] Note that this process has a long-term memory, implying that a polymer of any length never "forgets" that it forms a globule and, as a

result, the gradient-like primary structure is always preserved. Certainly, speaking of gradient copolymers, we have to consider their ensemble generated by the same synthetic process, not a single macromolecule which can, in principle, have an arbitrary statistics.[56]

Formation of catalytic triads. Simulation at the mesoscale allows us to explore qualitatively the triad formation mechanism. To that end, we should first define what is a potentially active catalytic triad. We assumed that a triad is a group consisting of three chain beads NAEA, NAHA and NABA and the distances separating them do not exceed the bead size σ. As noted in the Introduction, the enzyme active site is typically contained in a hydrophobic pocket formed by second-shell residues. This pocket along with overall shape-resistant polypeptide matrix provide the highly sophisticated and complex "traffic control" by which reactants are delivered to the catalytic site. Apparently, such a model cannot be developed in detail with the mesoscopic representation. Therefore, we restricted ourselves to a much simpler analysis by establishing an additional criterion that must be met for a triad to exist. This criterion is defined as the number of hydrophobic NVCL beads, n_H, in the second coordination sphere around the triad center-of-mass. The average number of triads $\overline{n}_{TR}$ that can be found in the globule using both the distance and hydrophobic criteria is averaged over the DPD trajectory and all independent runs. Figure 3 shows $\overline{n}_{TR}$ as a function of the copolymer composition f_P and n_H.

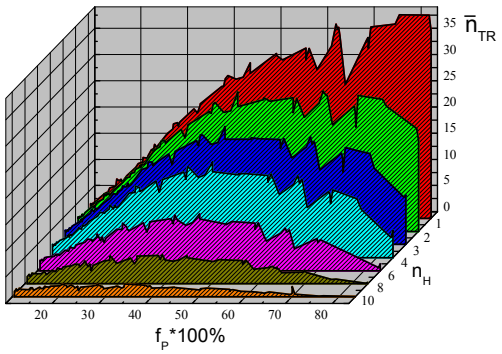


Figure 3. Average number of catalytic triads in 1000-unit copolymer as a function of the copolymer composition f_P and the number of hydrophobic beads n_H in the second coordination sphere around the triad center-of-mass. The n_{TR} value is averaged over both the DPD trajectory (10^6 time steps) and four independent runs.

It is seen that at very small n_H, the average number of triads increases monotonically with the degree of copolymer hydrophilicity f_P, as expected. When n_H is large ($n_H > 10$), the triads cannot actually be detected. In the intermediate range of n_H, the $\overline{n}_{TR}$ value exhibits non-monotonic behavior as a function of f_P and passes through the maximum close to $f_P = 0.4$-0.6. The essential fact is that the number of triads observed for this copolymer composition is much larger than unity, i.e. the value characteristic of natural enzymes.

It is natural to expect that with the two criteria defined above, the catalytic triads should mainly occur at the H/P interface. As seen from Figure 4, this is indeed the case if the density of hydrophobic pocket is not too large. However, we find that most of the triads becomes concentrated in the hydrophobic core when $n_H > 6$. Such triads will be inactive in reality, since the reagent access to them is restricted.

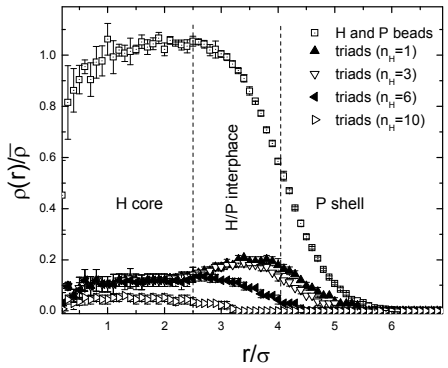


Figure 4. Density profiles $\rho(r)/\overline{\rho}$ showing the distribution of all beads and triads around the copolymer center of mass at $f_P = 0.4$ ($\overline{\rho}$ is the average particle density inside globule). The triads are determined for different values of n_H, as indicated in brackets. Error bars are the standard deviation of the mean, and averages are over four independent runs.

6.2. Atomistic model

For atomistic simulations, we reconstructed one of the generated mesoscale copolymers with the NVCL:NAEA:NABA:NAHA composition of 0.444:0.178:0.194:0.184 (mole/mole) and the degree of hydrophilicity $f_P = 0.556$, which is close to the composition that provided the maximum of $\overline{n}_{TR}$ in the mesoscale simulation (cf. Figure 3). Four identical globules were placed in a periodic box filled with water, as described above.

Molten globule state. In a quick look, we note in Figure 5 that the designed polymer globules swell in water. They resemble a so-called molten protein globule or compact intermediate that conserves a globular conformation and maintains structural integrity in solution but has increased solvent-exposed hydrophobic surface area and a dynamic tertiary structure without the tightly packed, glass-like interior characteristic of native proteins.[57] The specific feature of the molten globule state with native-like secondary structure is also the presence of large outer loops similar to those well seen in Figure 5, although the α-chymotrypsin-inspired copolymer has no regular elements of secondary structure due to lack of polypeptide backbone. These flexible hydrophilic loops, fluctuating around rigid core, stabilize partially folded conformations by increasing the overall system entropy. The solvent environment necessarily plays a critical role in these processes. The hydrophobic segments enriched with NVCL monomers perform double functioning: first, they stabilize the globular structure as a whole and,

second, they can provide hydrophobic areas near the catalytically-active triads (a rough analogy of hydrophobic pocket). Although structural disorder in proteins is generally considered to reduce catalytic efficiency, this widely held assumption has been challenged by the finding that enzymes can combine high catalytic activity with the highly dynamic properties of a molten globule (see e.g. Refs. 58 and 59).

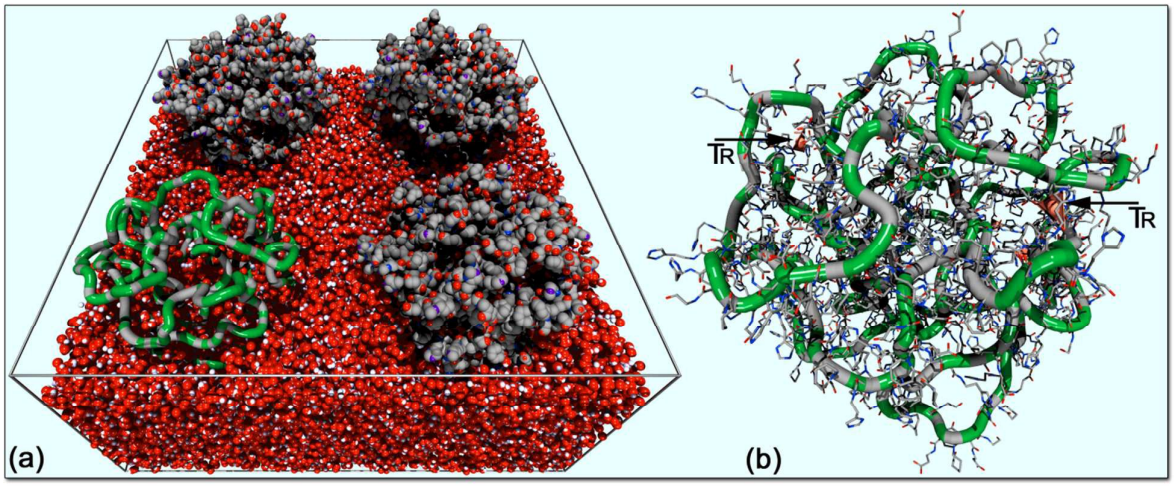


Figure 5. (a) Representative snapshot of the atomistic system simulated in a box with periodic boundary conditions. Shown are four copolymer globules surrounded by water. The formation of a core-shell protein-like morphology is evident. For one of the globules, only hydrocarbon backbone chain is depicted in the form of a flexible "hose" whose geometry is determined by a three-dimensional B-spline. Chain segments connected with hydrophobic H units are given in dark grey, while segments connected with hydrophilic P units are colored green. The top water layer is invisible for better visual clarity. At 27^0C and P = 1 atm, each water molecule has a volume of 30.1 Å^3, so that the volume occupied by four molecules of the 10,580-atom copolymer ($C_{3539}H_{5574}N_{684}O_{783}$) is 451,560 Å^3 (2,859,680 - 80,004×30.1) and the polymer volume fraction in solution is ~0.16. (b) Typical core-shell conformation of the designed copolymer. The hydrocarbon backbone is represented by a two-tone "hose", while the side chains of the monomer units are shown as sticks. Catalytic triads (TR) are depicted as pink triangles with spheres at their center. Arrows indicate two of six triads present in this conformation. Hydrogen atoms of the globules in (a) and (b) are invisible for clarity.

Aggregative stability. The solubility and aggregative stability in water is necessary for enzymes to be catalytically active. It is therefore important to examine the solution behavior of the designed copolymers and their stability towards aggregation. Aggregation can be considered as an equilibrium process leading to a size distribution of the intermolecular aggregates. To describe this process, we calculated the average aggregation number m, which can be estimated from the center-of-mass pair correlation function, $g_{COM}(r)$. The average number of polymers $n(r)$ around a given polymer within a distance r is defined as $n(r) = 4\pi\rho_c \int_0^r dx g_{COM}(x) x^2$, where $\rho_c = n/V$ is the number density of the macromolecules in the volume V. Two polymers belong to an aggregate if their centers-of-mass are located closer to each other than some characteristic "aggregation distance" r_a. In other words, $1 + n(r_a)$ chains form a connected cluster, so that $m = 1 + n(r_a)$. Certainly, the choice of r_a is to some extent arbitrary. For simplicity, we set the

parameter r_a to be the characteristic distance between non-interacting uniformly distributed particles. For our system, this gives r_a = 35.5 Å. At this distance, we found that the average aggregation number does not exceed 1.05. Hence, we conclude that the globules do not aggregate on the time scale of the simulation. This is confirmed by the analysis of snapshots (Figure 6) which also show that a degree of exposure of non-polar units to solvent is much less than that of polar units. Although the existence of separated single-chain globules is a prevalent structural motif, nevertheless, we detected the occasional formation of intermolecular aggregates: the globules ("unimers") can stick together, forming pairs and (very seldom) trimers, but they maintain their morphological integrity in these finite-size aggregates, so that for the system studied here, no large-scale aggregation is observed. In this regard, the designed copolymer behaves like a typical protein with a very small tendency to aggregate. Water-soluble non-precipitating fractions of the copolymer synthesized from the same co-monomers as in our simulation have also been found in the experimental work.[42]

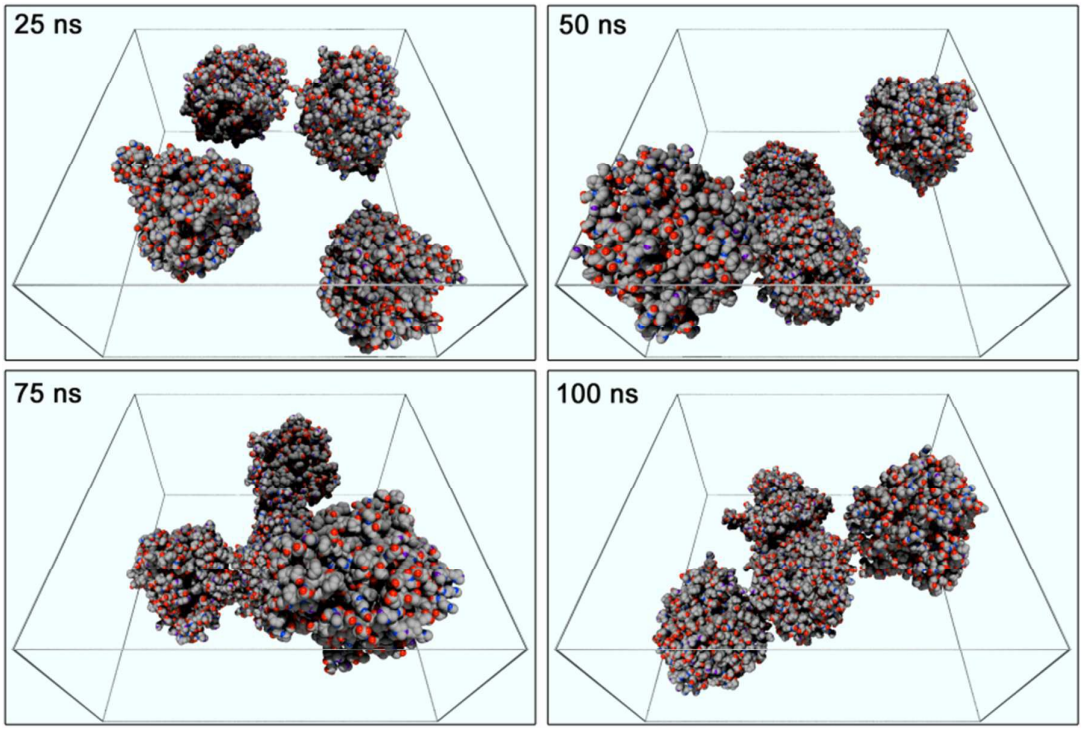


Figure 6. Snapshots from stretches of the MD trajectory at different time. Carbon atoms are shown in grey, oxygen atoms in red and nitrogen atoms in blue. For clarity, solvent molecules and hydrogen atoms are not shown.

Density profiles. Since the designed structure swells in water, it is instructive to understand where water is located within the globule. Figure 7 shows the radial density profiles of the individual species, i.e., histogrammed occurrences of the copolymer and water atoms around the globule center of mass, averaged over the MD trajectory. As seen, water is mainly located close to the hydrophilic periphery, as expected, but also can penetrate rather deep into the

core although the polymer structure does not open up enough in 100 ns. Interestingly, the density distribution across the globule is not uniform. This evidences a structuring (albeit somewhat weakly pronounced) of the water molecules with different segments of the macromolecule. As a result, the total material density inside the globular copolymer can slightly exceed 1 g/cm^3 (dashed line in Figure 7).

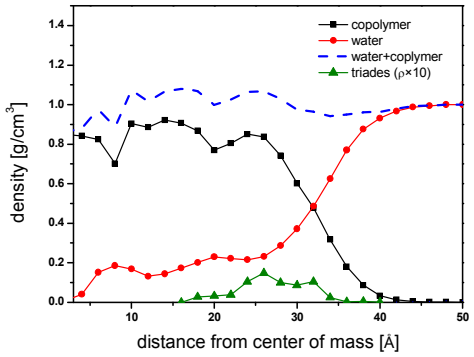


Figure 7. Time-averaged radial density profiles of copolymer, water and catalytic sites. The center of mass of the copolymer is taken as the origin (r = 0).

Radial distribution functions. In order to examine the local water environment forming in the vicinity of polymer, we calculated radial distribution functions (RDFs) $g_{A\text{-}B}(r)$ which define the probability of finding an atom A within a spherical shell of volume $4\pi r^2 dr$ at a distance r from an atom B, relative to that expected from their random distribution in the system at the same density and temperature. Polymer atoms were classified as hydrophilic (all oxygen and nitrogen atoms of the pendant functional groups of NAEA, NABA and NAHA; hydrogen atoms connected to oxygen or nitrogen atoms of these groups) and hydrophobic (all heavy atoms of the polymer backbone and NVCL). In Figure 8a-d we present the distribution functions $g_{W\text{-}B}(r)$ characterizing the correlations between the water oxygen atoms (W) and the terminal groups of the side chains (B = NAEA, NABA, NAHA and NVCL). Since water can penetrate into the globule, all copolymer atoms feel the effect of these water molecules. The unsurprising result is that the hydrophilic side chains of NAEA, NABA and NAHA are most hydrated (Figure 8a-c). From these RDFs, there is a preference for the water molecules to be associated near the functional hydroxyl and carboxyl groups rather than the imidazole group. The first sharp peaks with a maximum located at around 2 Å are due to the contacts between water and hydrogen atoms, while the next peaks and shoulders reflect the water structuring in the vicinity of the heavy charged atoms of these terminal groups. This suggests that water is bound to the polar sites on the hydrophilic monomer units, which are found to be exposed to the solvent, as well seen in Figure 5b. Of course, water molecules near exposed polar areas are not conserved, undergoing fast exchange with bulk solvent, and not all "surface" side chains are totally exposed to solvent. On the other hand, for the hydrophobic NVCL side chains and the aliphatic hydrocarbon backbone,

the first peak of RDF shifts toward longer distances (Figure 8d and e). As a result, the hydrophobic areas of the designed copolymer are strongly shielded from water, thus resembling the behavior of some hydrophobic residues in proteins.

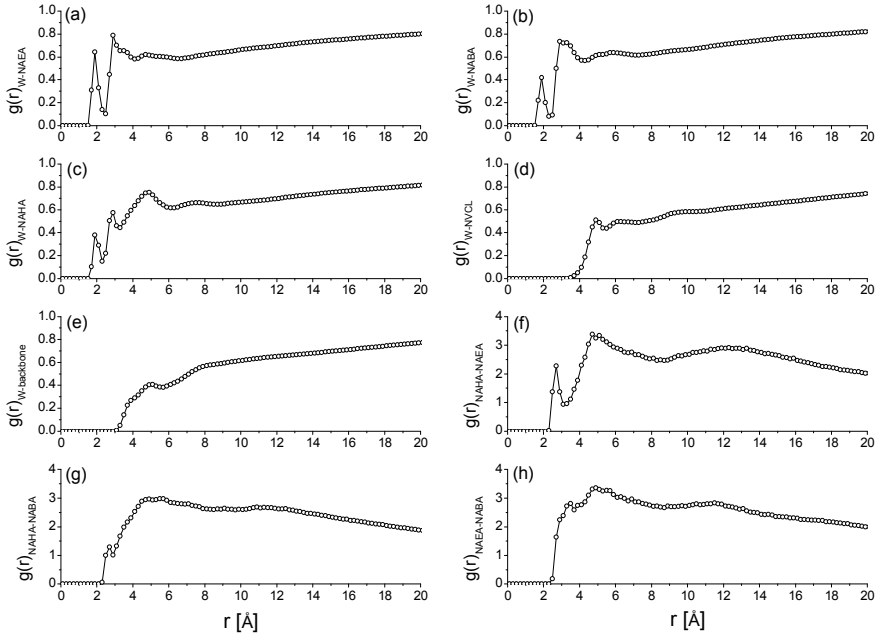


Figure 8. Radial distribution functions $g_{A\text{-}B}(r)$ calculated for different atoms A and B. (a)-(d) The solvent-solute RDFs which are associated with the water molecules and the atoms of the terminal groups of NAEA, NABA, NAHA and NVCL. (e) The water-hydrocarbon backbone RDF. (f)-(h) The RDFs characterizing the interaction between the end groups of the hydrophilic side chains. The RDFs approach unity at long distance.

Polymer swelling. By simulating the designed copolymer in water environment and in vacuum it is possible to address how the solvent affects the conformation and conformational mobility. Figure 9a presents the time evolution of the radius of gyration, *R*, calculated separately for the H and P chain segments of the hydrocarbon backbone. It is seen that the polymer remains stable and retains its globular conformation in solution till the end of the 100 ns MD run. The radii of gyration found for the copolymer surrounded by water are higher than those in vacuum, confirming solvent-induced polymer swelling. The solution structure might thus significantly differ from that in vacuum. The equilibrium swelling coefficients α_H and α_P, which reflect the solvent effect, are very close to each other (1.185 and 1.187 for the H and P segments, respectively), suggesting that both H and P domains of the copolymer undergo a more or less uniform swelling. It should noted that the polymer swelling in water is primarily due to the extension of hydrophilic surface side chains upon solvation (cf. the snapshot shown in Figure 5b). Because of surface tension acting on the hydrated polymer, simulation in water produces a structure more spherical than the vacuum structure.

For the solvated copolymer macromolecules, it is also found that the presence of water appears to increase the conformational mobility, leading to low-frequency, relatively large-

amplitude dynamic fluctuations throughout the entire MD simulation, while less relevant small-amplitude fluctuations are observed for both the systems (Figure 9a). It is well known that the conformational dynamics of natural proteins is often a critical element for many aspects of their function.[60] In particular, a certain degree of internal protein mobility is essential for biocatalysis because atomic fluctuations are required in most chemical reactions catalyzed by enzymes.[60] It appears, however, that the conformational mobility of the α-chymotrypsin-like synthetic copolymer is generally higher as compared to that typical for natural proteins. This is because the major structural component of proteins is the polypeptide backbone that has fewer degrees of freedom than the more flexible hydrocarbon backbone of the designed copolymer. As a result, it does not have the tightly packed conformation of the native protein and water can penetrate not only into the outer hydrophilic shell but also into the copolymer core (see Figure 7).

Hydrogen bonds. One of the main factors stabilizing core-shell globular conformation is hydrogen bonding, which reduces the solvent exposure of hydrophobic moieties. In order to count the number of hydrogen bonds, n_{HB}, formed throughout the MD trajectory, we used the HBonds plugin entering the VMD program[61] with the "standard" set up for hydrogen bond: maximum donor-acceptor D–A distance = 3.0 Å, minimum D–H–A angle = 20°. The number of hydrogen bonds per monomer unit vs. simulation time is plotted in Figure 9b-d for different pairs of the hydrophilic side chains. Most of the hydrogen bonds, which are quite stable during ~0.5 ns (remaining at 3 Å), is formed between the hydroxyl oxygen and nitrogen atoms of the polar monomers NAEA and NAHA (cf. Figure 9b). Some "strong" hydrogen bonds between the NAHA and NAEA donor and acceptor groups appear at early times (after ~0.1 ns) and preserve thereafter for a relatively long time. Average residence times of such donor/acceptor group having one H-bond at any given time is approximately 0.2 ns. The n_{HB} value averaged over time indicates that more than 50% of the –OH and >NH groups permanently forms intramolecular H-bonds. It is noteworthy that >NH groups are also considered to be capable of accepting a hydrogen in an H-bond.[62] The >NH and $-COO$ groups are much less likely to form hydrogen bonds (Figure 9c). Because water can make strong hydrogen bonds to both sp^2-hybridized carbonyl/carboxyl and sp^3-hybridized hydroxyl oxygen atoms, which are the primary acceptors of water hydrogen atoms and capable of accepting up to two H-bonds from water, one may expect the water molecules to have some effect on intramolecular hydrogen-bond stability, taking into account the possibility of water penetration into the globule (Figure 7). This effect is probably the main reason behind less pronounced participation of the $-COO$ groups in the formation of H-bonds (Figure 9c and d).

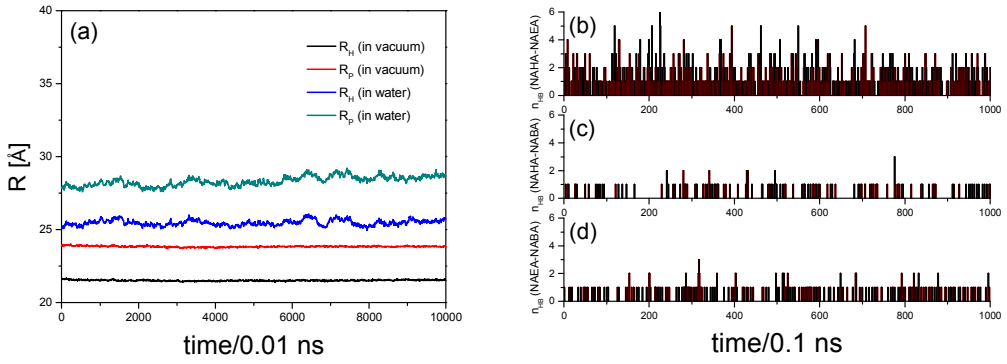


Figure 9. (a) Gyration radii as a function time for hydrophobic (R_H) and hydrophilic (R_P) backbone segments of the copolymers *in vacuo* and from the solution simulation. The gyration radii remain stable till the end of the runs. They are averaged over 10 ps to reduce fluctuations. (b)-(d) The number of intramolecular hydrogen bonds vs. time for the copolymer solution (a sampling interval of 100 ps is used).

Catalytic sites. First, we analyze the radial distribution functions that illustrate the spatial distribution of hydrophilic side-chain termini capable of forming catalytic centers (Figure 8f-h). The first peaks at ≈2.7-3.0 Å arise from the hydrogen bonding partners in the first shell of the side chain. The pronounced second (main) peaks at 4.9-5.1 Å are associated with the first coordination shell formed by one of the side-chain termini around the second one, giving an average first coordination shell size of approximately 0.4 contacting polar end group. Inspection of randomly chosen structures indicated that it is not uncommon to find clusters from two neighboring polar groups that are clearly interacting with one another directly. When comparing the RDF curves corresponding to different end groups, we observe that the NAHA-NABA and NAEA-NABA main peaks are broader, smaller in magnitude and slightly shifted to higher distances as compared to the NAHA-NAEA main peak.

To identify putative catalytic sites, simple geometric criteria based on experimental data and the spatial distribution of side-chain termini (Figure 8f-h) were used. Structure of the catalytic triad (Ser195-His57-Asp102) of serine proteases is shown in Figure 1a. For a typical catalytic triad arrangement, the C_α atom distances of Asp to His, His to Ser and Asp to Ser averaged over 16 experimentally determined crystal structures from a protozoan, a fungus, a plant, etc. are 6.4 ± 0.01, 8.4 ± 0.03 and 9.8 ± 0.02 Å, respectively.[63] These data were the basis for constructing a catalytic triad template used for the triad identification. The nitrogen-attached hydrogen atom of an imidazole ring can shuttle from one nitrogen to another and can therefore be present in any of the two nitrogen atoms. Taking this into account, we assumed that both the imidazole nitrogens (n1 and n2, see Figure 1a) are equivalent. The same is true for the two oxygen atoms of any carboxyl group. The following distance constraints between atoms were set for the hydroxyl-imidazole-carboxyl structural motif: o1-n1 (n2) $\le r_{TR}^{(1)}$ and n2 (n1)-o2 $\le r_{TR}^{(2)}$, where $r_{TR}^{(1)}$ and $r_{TR}^{(2)}$ are considered as parameters. The three-dimensional structures that fit with

these criteria were incorporated into the final dataset, while the hydrogen bond angle constraints often imposed on the $SerO^{\gamma}$–$SerH^{\gamma}\cdots HisN^{\varepsilon 2}$ and $HisN^{\delta 1}$–$HisH^{\delta 1}\cdots AspO^{\delta 2}$ angles for amino acid residues to differentiate between active and inactive proteins were ignored. For brevity, we restrict ourselves below only the case when $r_{TR}^{(1)} = r_{TR}^{(2)} = r_{\mathrm{TR}}$. It should be emphasized that in the present study, the process of triad formation is considered only from a geometrical viewpoint rather than a dynamical one, i.e., without including certain time-dependent correlations, which is out of the scope of this work.

The average number of triads, $\overline{n}_{\mathrm{TR}}$, is presented in Figure 10a as a function of r_{TR}. The distribution function over the number of triads calculated at r_{TR} = 5 Å is given in Figure 11b. With the use of the geometry-based triad identification, one sees that $\overline{n}_{\mathrm{TR}}$ increases rapidly with r_{TR} increasing, as it should be. What is most important is the observation that the value of $\overline{n}_{\mathrm{TR}}$ is significantly greater than unity for reasonable cutoff distances $r_{\mathrm{TR}} \sim 5$ Å, thereby confirming our concept of multiple catalytic sites. In this case, the number of the randomly occurring, potentially active catalytic centers can reach 30 and more per molecule (see Figures 5b and 10b). It is natural to expect that in the context of chemical transformations, this will facilitate catalysis. The inspection of Figure 7 suggests that triads are mainly localized at the hydrophobic/hydrophilic interface where there is relatively free access both to water and to reactants. The same conclusion can be drawn from the mesoscale simulation (Figures 2 and 4).

It has to be born in mind that in reality, not all but just a few spontaneously occurring triads will be active due to their "structural failure" or short lifetime, so that the simulation gives an upper bound on the number of active catalytic centers. Besides, due to the highly dynamic nature of the bio-inspired enzyme, some of the triads can exhibit *on-off* changing states, a behavior similar in a certain respect to that found for e.g. enzymes of unicellular organisms which demonstrate both steady states and oscillatory reactive patterns.[64] It should further be mentioned that there are multidomain enzymes, in which the transference of a substrate from one catalytic site to the next often occurs in order to prevents diffusion of the substrate or unstable intermediates into the bulk solvent without their chemical transformation.[65] It may well be that the spatial proximity of the multiple functional groups in the designed structure will contribute to such a process. Finally, we note that binding of a ligand on a protein can change the flexibility and dynamics of certain parts of the protein and this local perturbation propagates through the structure over long distances.[66] As result, the conformational lability of a protein-ligand complex and individual catalytic centers can be significantly reduced. These issues, which are closely related to dynamic aspects of enzyme-inspired synthetic copolymers, will be discussed in a separate publication.

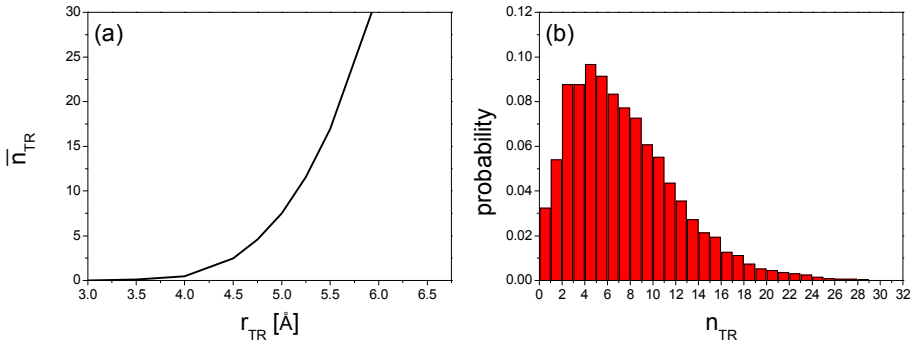


Figure 10. (a) Average number of catalytic triads as a function of the cutoff distance r_{TR}. (b) Distribution function over the number of triads at r_{TR} = 5 Å.

Next, we ask the question: What is the relationship between the number of triads and the polymer size? At first glance, the answer seems obvious: the denser structure, the greater the number of triads. It turns out, however, that this is not true. In Figures 11a and 11b, we plot the values of n_{TR} for different conformations (time frames) with the corresponding values of R_H and R_P; the histogram distributions of R_H and R_P are shown in Figures 11c and 11d, respectively. From these data, we find out that the maximum number of triads (up to 30) occurs for the structures whose size is close to the mean size, not for compact or expanded conformations. This may hint for a correlation between catalytic activity and solvent quality, pH, ionic strength, temperature, etc.

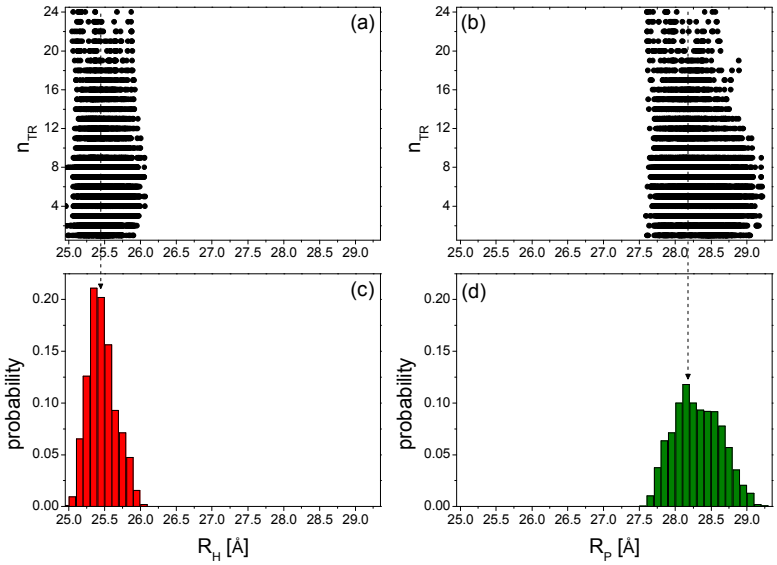


Figure 11. (a) and (b) The number of triads occurring in the conformations having the corresponding values of R_H and R_P (shown are the data obtained for each of 10,000 time frames of 100 ns MD trajectory) (c) and (d) The distribution functions over R_H and R_P ($\overline{R}_H$ = 25.5±0.2 Å and $\overline{R}_P$ = 28.3±0.3 Å).

Our simulation also indirectly corroborates the experimental observation, which suggests that *p*NPP hydrolysis acceleration by the designed copolymer in coil conformation is negligible as compared to that found for the globular state of the same copolymer.[42] Indeed, as stems from

the simulation performed in vacuum at elevated temperature (500 K), much above the globule-to-coil transition temperature, the number of triads is close to zero for the coil-like state. Thus, the proposed enzyme design protocol can lead to copolymers with a well pronounced conformation-dependent catalytic activity.

7. Concluding remarks

Based on the principles of the conformation-dependent design of copolymer sequences[14] and the methods of controlled radical copolymerization,[55] we have proposed a novel concept for rational designing of bio-inspired polymer-supported organocatalysts from polymerizing synthetic (non-natural) monomers mimicking the amino acid residues responsible for catalysis or/and binding to target molecules in natural enzymes. The focus was on molecular-level simulations to implement the proposed approach, to gain fundamental insight in the structural properties of the designed enzyme-like copolymers and to find the robust ways of their synthesis. A key ingredient of our approach is that the target globular conformation of protein-like, core-shell morphology can appear spontaneously in the course of controlled radical polymerization in a selective solvent. By combining four types of monomers with different hydrophobicity/hydrophilicity and polymerizable double bonds in *in silico* design, we have predicted the formation of stable polymer globules with a dense hydrophobic core rich in the hydrophobic monomer units and a hydrophilic shell containing various blocks with incorporated catalytically active groups. Compared to the enormous combinatorial diversity of four-letter copolymer sequences, the number of specific sequences leading to such a core-shell morphology is relatively small. One may say that our design methodology acts as a "sieve" that removes unsuitable copolymer sequences and leaves mainly desirable sequences.

As a case study, the design of a functional analog of chymotrypsin with self-assembly ability in a selective solvent was performed, using mesoscale and fully atomistic simulations. The functional groups present in the designed copolymer, viz., the hydroxyl group in N-acryloyl-2-ethanolamine (NAEA), the carboxyl group in N-acryloyl-β-alanine (NABA) and imidazole in N-acryloyl-histamine (NABA), also constitute the catalytic center of chymotrypsin, viz., serine, aspartic and histidine amino acid residues, respectively. Due to a specific monomer sequence distribution, the hydrophilic segments, bearing functional hydroxyl, carboxyl and imidazole groups, are mainly concentrated close to the globule surface, thus providing an increased probability for catalytic triad formation, whereas the hydrophobic segments, containing a varying number of N-vinylcaprolactam (NVCL) monomer units, are mainly located in a globular core. By carrying out mesoscale simulations on copolymers synthesized at various ratios of hydrophobic

(H) and hydrophilic or polar (P) monomers, we have shown that the copolymer hydrophilic fraction plays a crucial role in determining the final core-shell morphology. This kind of spatial arrangement of H and P segments was found to prevent polymer aggregation.

Although the enzyme-inspired synthetic copolymer was found from the atomistic simulation to be a highly dynamic system, whose structural lability is higher than that of semi-rigid proteins, and catalytic centers emerge more or less randomly, these drawbacks can be compensated by the fact that the number of potentially active catalytic triads can be approximately one or even two orders of magnitude higher than what is typical for natural counterparts. In this regard, each designed polymer globule with multiple catalytic sites behaves as a surface nanoreactor, maintaining its structural integrity in solution. Certainly, the conformational flexibility should reduce a "catalytic quality" of each individual chymotrypsin-like triad, but their quantity is capable of compensating this. In this context, we may recall the widely known military aphorism that "Quantity has a quality all its own".

As an extension of our approach, it would be interesting to incorporate covalent cross-links into the copolymer to bias its three-dimensional globular structure toward more rigid geometry. This can be accomplished by polymerizing multifunctional monomers. We anticipate that in this way it will be possible to obtain catalytically-active globular nanoparticles ("plastic enzymes") with enhanced activity and a reduced propensity to coalesce.

Acknowledgments

The authors gratefully acknowledge Prof. V. I. Lozinsky for valuable discussions and the Supercomputing Center at MSU for providing computing resources on the hybrid massively parallel machine LOMONOSOV-2. This work was supported by the Russian Science Foundation (grant no. 14-13-00544).

References

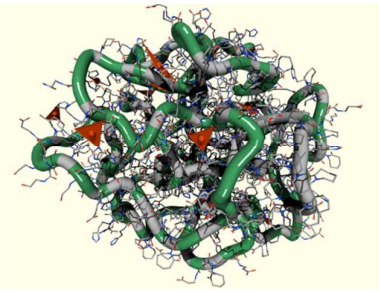
1. A. Fersht, *Enzyme Structure and Mechanism*, 2nd ed., W. H. Freeman, San Francisco, **1985**.
2. L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, D. Hilvert, K. N. Houk, B. L. Stoddard and D. Baker, *Science* **2008**, *319*, 1387−1391.
3. J. Damborsky and J. Brezovsky, *Curr. Opin. Chem. Biol*. **2009**, *13*, 26–34.
4. V. Nanda and R.l. Koder, *Nature Chem*. **2010**, *2*, 15−24.
5. J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St.Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael and D. Baker, *Science* **2010**, *329*, 309−313.
6. M. L. Zastrow, A.F.A. Peacock, J.A. Stuckey and V.L. Pecoraro, *Nature Chem*. **2012**, *4*, 118-123.
7. M. K. Tiwari, R. Singh, R. K. Singh, I.-W. Kim and J.-K. Leea, *Comput. Struct. Biotechnol. J*. **2012**, *2*, e201209002.
8. Y. Yin, Z. Dong, Q. Luo and J. Liu, *Progr. Polym. Sci.* **2012**, *37*, 1476−1509.
9. G. Kiss, V.S. Pande and K.N. Houk, *Methods in Enzymology* **2013**, *523*, 145–170.
10. F.-A. Chao, A. Morelli, J. C. Haugner, L. Churchfield, L. N Hagmann, L. Shi, L. R. Masterson, R. Sarangi, G. Veglia and B. Seelig, *Nature Chem. Biol*. **2013**, *9*, 81–83.
11. C. E. Tinberg, S. D. Khare, J. Dou, L. Doyle, J. W. Nelson, A. Schena, W. Jankowski, C. G. Kalodimos, K. Johnsson, B. L. Stoddard et al., *Nature* **2013**, *501*, 212-216.
12. J. Damborsky, J. Brezovsky, *Curr. Opin. Chem. Biol*. **2014**, *19*, 8–16.
13. H. Yang, J. Li, H.-D. Shin, G. Du, L. Liu and J. Chen, *Appl. Microbiol. Biotechnol.* **2014**, *98*, 23-29.
14. P. G. Khalatur and A. R. Khokhlov, *Adv. Polym. Sci*. **2006**, *195,* 1−100.
15. R. B. Merrifield, *J. Am. Chem. Soc*. **1963**, *85*, 2149–2154.
16. H. K. Murnen, A. R. Khokhlov, P. G. Khalatur, R. A. Segalman and R. N. Zuckermann, *Macromolecules* **2012**, *45*, 5229−5236.
17. R. Breslow, *Chem. Soc. Rev*. **1972**, *1*, 553.
18. R. Breslow, *Science* **1982**, *218*, 532−537.
19. R. Breslow, *Acc. Chem. Res*. **1995**, *28*, 146−153.
20. *Artificial Enzymes*, Ed. by R. Breslow, Wiley-VCH Verlag, Weinheim, **2005**.
21. *From Enzyme Models to Model Enzymes*, Ed. by A. J. Kirby and F. Hollfelder, The Royal Society of Chemistry, Cambridge, **2009**.

22. *Molecularly Imprinted Polymers: Man-made Mimics of Antibodies and their Applications in Analytical Chemistry*, Ed. by B. Sellergren, Elsevier, Amsterdam, **2001**.
23. C. G. Overberger and J. C. Salamone, *Acc. Chem. Res*. **1969**, *2*, 217-224.
24. H. C. Kiefer, W. I. Congdon, I. S. Scarpa and I. M. Klotz, *Proc. Nat. Acad. Sci. USA* **1972**, *69*, 2155−2159.
25. J. Suh, I. S. Scarpa and I. M. Klotz, *J. Am. Chem. Soc*. **1976**, *98*, 7060−7064.
26. A. J. Kirby, *Enzyme Mimics*, In: *Stimulating Concepts in Chemistry*, Ed. by F. Vögtle, J. F. Stoddart and M. Shibasaki, Wiley-VCH Verlag, Weinheim, **2000**, pp. 341−353.
27. R.N. Karmalkar, M.G. Kulkarni, R.A. Mashelkar, *Macromolecules* **1996**, *29*, 1366–1368.
28. S. Yu, Y. Yin, J. Zhu, X. Huang, Q. Luo, J. Xu, J. Shena and J. Liu, *Soft Matter* **2010**, *6*, 5342–5350.
29. H. Dugas, *Bioorganic Chemistry: A Chemical Approach to Enzyme Action*, Springer-Verlag, N.Y.-Berlin-Heidelberg, **2012**.
30. C. T. Nguyen and R. M. Kasi, *Chem. Commun*. **2015**, *51*, 12174–12177.
31. B. Garg 1, T. Bisht and Y.-C. Ling, *Molecules* **2015**, *20*, 14155–14190.
32. A. I. Taylor, V. B. Pinheiro, M. J. Smola, A. S. Morgunov, S. Peak-Chew, C. Cozens, K. M. Weeks, P. Herdewijn and P. Holliger, *Nature* **2015**, *518*, 427–430.
33. R. Qu, L. Shen, A. Qu, R. Wang, Y. An and L. Shi, *ACS Appl. Mater. Interfaces* **2015**, *7*, 16694–16705.
34. C. Buron, K. Sénéchal-David1, R. Ricoux, J.-P. Le Caër, V. Guérineau, P. Méjanelle, R. Guillot, C. Herrero, J.-P. Mahy and F. Banse, *Chem. Eur. J*. **2015**, *21*, 12188-12193.
35. Y. Yin, S. Jiao, Y. Wang, R. Zhang, Z. Shi, X. Hu, *ChemBioChem* **2015**, *16*, 670-676.
36. K. A. Dill, S. Bromberg, K. Z. Yue, K. M .Fiebig, D. P. Yee, P. D. Thomas and H. S. Chan, *Protein Sci*. **1995**, *4*, 561−602.
37. M. H. Hecht, A. Das, A. Go, L. H. Bradley and Y. N. Wei, *Protein Sci*. **2004**, *13*, 1711−1723.
38. A. R. Khokhlov and P. G. Khalatur, *Physica A* **1998**, *249*, 253−261.
39. A. R. Khokhlov and P. G. Khalatur, *Phys. Rev. Lett*. **1999**, *82*, 3456−3459.
40. V. I. Lozinsky, I. A. Simenel, V. K. Kulakova, E. A. Kurskaya, T. A. Babushkina, T. P. Klimova, T. V. Burova, A. S. Dubovik, V. Ya. Grinberg, I. Yu. Galaev, B. Mattiasson and A. R. Khokhlov, *Macromolecules* **2003**, *36*, 7308−7323; V. I. Lozinsky, I. A. Simenel, V. G. Semenova, L. E. Belyakova, V. V. Il'in, V. Ya. Grinberg, A. S. Dubovik and A. R. Khokhlov, *Polym. Sci. A* **2006**, *48*, 435-443; V. I. Lozinsky, *Adv. Polym. Sci*. **2006**, *196*, 87−127.

41. V. I. Lozinskii, I. A. Simenel and A. R. Khokhlov, *Dokl. Chem*. **2006**, *410*, 170−173.
42. V. I. Lozinskii, O. E. Zaborina, T. P. Klimova, T. A. Babushkina, A. S. Kovaleva, E. B. Boltuchina, B. P. Chernyshov, T. V. Burova, N. V. Grinberg, V. Ya. Grinberg and A. R. Khokhlov, *Polym. Sci. A* **2015** (in press).
43. A. V. Berezkin, P. G. Khalatur and A. R. Khokhlov, *J. Chem. Phys*. **2003**, *118*, 8049–8060.
44. P. J. Hoogerbrugge and J. M. V. A. Koelman, *Europhys. Lett*. **1992**, *19*, 155−160.
45. A. A. Gavrilov, P. V. Komarov and P. G. Khalatur, *Macromolecules* **2015**, *48*, 206–212.
46. P. V. Komarov, C. Yu-Tsung, C. Shih-Ming, P. G. Khalatur and P. Reineker, *Macromolecules* **2007**, *40*, 8104−8113.
47. A. A. Gavrilov, A. V. Chertovich, P. G. Khalatur, A. R. Khokhlov, *Soft Matter* **2013**, *9*, 4067−4072.
48. A. A. Gavrilov, A. V. Chertovich, P. G. Khalatur and A. R. Khokhlov, *Macromolecules,* **2014***, 47*, 5400−5408.
49. Askadskii, A. A. *Computation in the Material Science of Polymers*; Cambridge International Science, Cambridge, **2003**, 650 pp.
50. C. F. Fan, B. D. Olafson, M. Blanco and S. L. Hsu, *Macromolecules* **1992**, *25*, 3667.
51. *DPDchem, a software package for simulation of polymers and liquids using dissipative particle dynamics*: http://polymer.physik.uni-ulm.de/~khalatur/exchange/DPD_Chem (accessed October 2015); *DPD_dd, Multiscale computational approach to the design of polymer-matrix nanocomposites*: http://www.compnanocomp.eu (accessed October 2015).
52. H. Sun, *Macromolecules* **1995**, *28*, 701−712.
53. LAMMPS package; http://lammps.sandia.gov (accessed October 2015).
54. J. Genzer, P.G. Khalatur and A.R. Khokhlov, Conformation-Dependent Design of Synthetic Functional Copolymers. In: *Polymer Science: A Comprehensive Reference*, Volume 6: *Macromolecular Architectures and Soft Nano-Objects*. Editors-in-Chief: K. Matyjaszewski and M. Möller, Elsevier B.V., **2012**, pp. 689−723.
55. *Handbook of Radical Polymerization*, Editors: K. Matyjaszewski and T. P. Davis, J. Wiley & Sons, **2002**.
56. A. V. Berezkin, P. G. Khalatur and A. R. Khokhlov, *Macromolecules* **2006**, *39*, 8808−8815.
57. M. Ohgushi and A. Wada, *FEBS Lett*. **1983**,*164*, 21–24; V.S. Pande and D.S. Rokhsar, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1490–1494.

58. K. Vamvaca, B. Vögeli, P. Kast, K. Pervushin and D. Hilvert, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12860–12864.
59. K. Pervushin, K. Vamvaca, B. Vögeli and D. Hilvert, *Nature Struct. Molec. Biol*. **2007**, *14*, 1202–1206.
60. M. Karplus and J. Kuriyan, *Proc. Natl. Acad. Sci. USA* **2005**, 102, 6679–6685.
61. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33–38.
62. G. Nemethy, M.S. Pottle and H.A. Scheraga, *J. Phys. Chem*. **1983**, *87*, 1883–1887; J.A. Ippolito, R.S. Alexander and D.W. Christianson, *J. Mol. Biol*. **1990**, *215*, 457–471.
63. A. Laskar, E. J. Rodger, A. Chatterjee and C. Mandal, *BMC Res. Notes* **2012**, *5*, 256.
64. I. M. De la Fuente, F. Vadillo, M.-B. Pérez-Pinilla, A. Vera-López and J. Veguillas, *PLoS One* **2009**, *4*, e7510.
65. F. M Raushel, J. B. Thoden and H. M. Holden, *Acc. Chem. Res*. **2003**, *36*, 539−548.
66. B. Erman, *Proteins* **2015**, *83*, 805–808.

Table of contents



We have designed, for the first time, a functional analog of chymotrypsin from synthetic monomers imitating protein amino acid residues.