Soft Matter

Aggregation propensity of therapeutic fibrinhoming pentapeptides: Insights from experiments and molecular dynamics simulations

Journal:	*Soft Matter*
Manuscript ID	SM-ART-05-2020-000930.R2
Article Type:	Paper
Date Submitted by the Author:	20-Oct-2020
Complete List of Authors:	Zanuy, David; Universitat Politècnica de Catalunya, Chemical Engineering Puiggali Jou, Anna; Universitat Politècnica de Catalunya CONFLITTI, Paolo; Università della Svizzera Italiana Bocchinfuso, Gianfranco; University of Rome Tor Vergata, Chemical Science and Technologies Palleschi, Antonio; University of Rome Tor Vergata, Dept. of Chemical Sciences and Technologies Aleman, Carlos; Universitat Politecnica de Catalunya,

Aggregation Propensity of Therapeutic Fibrin-Homing Pentapeptides: Insights from Experiments and Molecular Dynamics Simulations

David Zanuy,[a*] *Anna Puigalí-Jou,*[a] *Paolo Conflitti,*[b†], *Gianfranco Bocchinfuso,*[b]

Antonio Palleschi,[b] *Carlos Alemán*[a,c]

[a] Department of Chemical Engineering and Barcelona Research Center in Multiscale Science and Engineering, Universitat Polytècnica de Catalunya, 08019 Barcelona, Spain

[b] Dipartimento di Scienze e Tecnologie Chimiche, Universita`di Roma''Tor Vergata'', via della ricerca scientifica 1, 00133, Rome, Italy

[c] Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, 08028 Barcelona Spain

[†] Currently at Università della Svizzera italiana (USI), Faculty of Biomedical Sciences, Institute of Computational Science, Via G. Buffi 13, CH-6900 Lugano, Switzerland

Correspondence: david.zanuy@upc.edu and carlos.aleman@upc.edu

ABSTRACT

CREKA (Cys-Arg-Glu-Lys-Ala) and its engineered analogue CRMeEKA, in which Glu has been replaced by *N*-methyl-Glu to provide resistance against proteolysis, are emerging pentapeptides that were specifically designed to bind fibrin-fibronectin complexes accumulated in walls of tumour vessels. However, many of the intrinsic properties of CREKA and CRMeEKA, which are probably responsible of their different behaviour when combined with other materials (such as polymers) for diagnosis and therapeutics, remain unknown yet. The intrinsic tendency of these pentapeptides to form aggregates has been analysed by combining experimental techniques and atomistic Molecular Dynamics (MD) simulations. Dynamic light scattering assays show the formation of nanoaggregates that increase in size with the peptide concentration, even though aggregation occurs sooner for CRMeEKA, independently of the peptide concentration. FTIR and circular dichroism spectroscopy studies suggest that aggregated pentapeptides do not adopt any secondary structure. Atomistic MD trajectories show that CREKA aggregates faster and forms bigger molecular clusters than CRMeEKA. This behaviour has been explained by stability of the conformations adopted by un-associated peptide strands. While CREKA molecules organize forming intramolecular backbone - side

chain hydrogen bonds, CRMeEKA peptides display main chain - main chain hydrogen bonds closing very stable γ- or β- turns. Besides, energetic analyses reveal that CRMeEKA strands are better solvated in water than CREKA ones, independently if they are assembled or un-associated.

Introduction

Modern medicine requires molecular tools that have environmental stability, high biocompatibility and, particularly, much higher specificity than classical pharmacological approaches. In this context, the development of phage display libraries[1] at the end of the twentieth century allowed researchers to discover naturally occurring protein fragments that presented high affinity and specificity for biomacromolecules with determinant roles in the cellular metabolism.[1] Among them, tumour homing peptides (and related entities such tumour penetrating peptides)[2] stand out due to their exceptional physicochemical features. These short protein segments exhibit an extraordinary capability for specifically interacting with cellular receptors. For this ability, they have already been exploited as key players in nanoconstructs built for specifically targeting tumour cells.[3-6]

Among this ever-increasing palette of homing peptides, CREKA (Cys-Arg-Glu-Lys-Ala) has been the object of many studies as guiding systems to specifically targeting neoplastic tissues.[7] This short peptide binds clotted plasma proteins and selectively homes to tumours that present angiogenesis. Specifically, it is reported that recognizes fibrinand

associated proteins (e.g. fibronectin) accumulated in walls of tumour vessels and in the interstitial spaces within tumours.[7] CREKA activity does not confine to tumour cells but also homes to the surface of atherosclerotic plaques and has been used to deliver and concentrate payloads onto atherosclerotic plaques in genetically modified mouse models.[8] Its conformational preferences have been widely studied and its bioactive conformation characterized.[9] Its most populated conformation consisted in a γ-turn-like structure, when tethered on a metallic surface, with the charged groups of Arg, Glu, and Lys pointing outwards and facilitating the formation of intermolecular interactions.

It is worth noting that the untethered small peptides may present other turn conformations with bioactivity (derivations of either β-turn or γ-turn), because these lightly populated structures still preserve the main structural characteristics for their activity, which correspond to conformations that place all the charged side chains pointing in the same direction[9]. The gathered information was used to suggest several modifications in order to increase its resistance to endogenous proteolysis and to enhance its homing capacities.[10] Peptide analogues that incorporated synthetic alpha-methyl and N-methyl derivatives of Arg, Glu, and Lys were studied. Most of the new compounds showed improved environment stability and enhanced homing activity.[11]

Among them, the analogue with *N*-methyl modification at the Glu position (hereafter referred as CRMeEKA) showed the greater improvement, even retaining most of its anticoagulant power.[12]

The homing ability of CREKA (and derivatives) has been exploited as the means to direct nanostructures towards targeting specific cell lines and to enhance the nano-constructs selectively to specific targets, that have been demonstrated both in vitro and in vivo.[15,16] Most studies were focused on probing and imaging tumours sites in sick organisms, maximizing the accumulation of nano-constructs in the targeted locations.[17-19] However, its use as part of cytotoxic constructs, that were built to kill specific cell lines, revealed that in most cases the destruction of tumours is preceded by the self-assembly of those nanostructures on target. In this context, the inherent hazard of unwanted in vivo aggregation or even cytotoxic fibrogenesis has not been explored.[20] Despite the advantages of accumulating large amounts of active probes on tumoral tissue, the potential pathologic adverse effects due to the complexity of peptide fibrogenesis has not yet been addressed. Even though, some indirect evidence suggests that the aggregation of the peptide core does not correlate to cell

death,[16,19,21] no clear demonstration of its influence over this process has been provided neither its effects to long term exposure.[16,19] Another example of unknown effects of nanoparticles assembly over the cytotoxicity of the constructs is found when building nanoparticles by encapsulating CREKA based peptides in matrices build on the conducting polymer poly(3,4-ethylenedioxythiophene) (PEDOT).[13-14,21] PEDOT nanoparticles (NPs) were loaded in situ with CREKA and CRMeEKA for controlled electro-stimulated release.[21] In all studied cases, NPs cellular uptake led to assembly of those particles either inside vesicles or in inclusion structures in cell cultures. The driving force of this process and the potential implications on the cell's viability are yet to be explored. The evident morphological differences observed in NPs depending on the presence of the wild type peptide or the methylated analogue strongly suggests organizational differences depending on the CREKA analogue that was used.[13] As a first step, we explore some intrinsic properties of the CREKA-based segments. We try to assess if those observed morphological differences[13] stem from diverse aggregation tendencies of both peptide species. Hence, we studied their aggregation using nanomechanical detections and molecular dynamics simulations. The structural properties of the aggregates and their main molecular features are reported.

Methods

Experimental section

Stock solutions. CREKA and CRMeEKA peptides with > 98% of HPLC purity in TFA (trifluoroacetic acid) salt form were purchased from Biomatik (Toronto, ON, Canada). Peptides were dissolved in PBS 1× (pH 7.4) by sonication and vortexing, and the corresponding dilutions were made. Particularly, 10, 5, 1 and 0.5 mg/mL solutions were prepared for dynamic light scattering (DLS) measurements. Samples for FTIR analyses were prepared using D_2O (99.9% D) at 30 mM peptide concentration.

Dynamic Light Scattering (DLS). DLS studies were performed using a NanoBrook Omni Zeta Potential Analyzer from Brookheaven Instruments Corporation. Measurements consisted of 3 runs each of 120 s duration; then averaged to give the effective diameter (D_{eff}). Samples were analysed at 25 ºC at a scattering angle of 173º. Peptides were dissolved in PBS 1x (NaCl 137 mM; KCl 2.7 mM, Na_2HPO_4 10 mM and KH_2PO_4 1.8 mM, pH 7.4) by sonication and vortexing, and the corresponding dilutions were made.

FTIR Spectroscopy. IR absorption spectra were recorded on a FTIR Jasco 4100 spectrophotometer. Liquids were deposited in an attenuated total reflection accessory (Top-plate) with a diamond crystal (Specac model MKII Golden Gate Heated Single Reflection Diamond ATR). For each sample 400 scans were performed between 4000 and 600 cm^{-1} with a resolution of 4 cm^{-1}. Peptides were dissolved in D_2O pD 7.4 at 30 mM concentration. This condition was specially used for FTIR experiments in order to see the bands of the amide I region, which are hindered by H_2O

Circular Dichroism (CD). CD spectra were recorded between 180 and 260 nm at 25 ºC using a Chirascan equipment with 1 cm cell path and peptide concentrations of 0.005, 0.01 and 0.1 mg/ml in PBS x1 (see above for details). Spectra were acquired with a 1 nm data pitch

using a 1 nm bandwidth and a 5 s digital integration time. Spectra were averaged after three accumulations and corrected by subtraction of the background spectrum.

Computational methods

Force Field and Molecular Details. All bonding and non-bonding parameters for standard amino acids were obtained from Amber03 force field.[22] The parameters of the non-coded residue *N*-methyl Glu had previously been computed and fitted into Amber03.[10] After trying several peptide concentrations (details in E.S.I., Table S1.) we finally obtained a system size suitable to observe aggregation events in a feasible timeframe, a $9.5 \times 8.5 \times 9.0$ nm^3 simulation box . For each studied peptide, 15 identical molecules were randomly placed in a with an intermolecular average distance of about 1.8 nm, *i.e.* molecules were mostly non-interacting with each other ("minimum-bias" approach). At the simulated conditions each studied peptide presented a positive net charge, which was neutralized by adding a chloride ion per strand, for a total of 15 anions per studied model. Finally, the simulation box was filled with approximately 23000 TIP3P water molecules,[23] where all the overlapping water molecules were removed (hereafter referred as *MD-I*). After 70 ns of simulation, in order to speed up the studied process, the formed aggregates were placed in a smaller box ($7.5 \times 7.5 \times 7.5$ nm^3). The new systems were simulated for another 60 ns (*MD-II).* At the end of this second production run the box was shrunk further ($6.5 \times 6.5 \times 6.5$ nm^3) and another 100 ns simulation were run (*MD-III*). For each step of re-scaling dimensions, the necessary number of water molecules was removed in order to keep the solution density.

Molecular Dynamics: Simulation Details. All MD series were performed with NAMD 2.10 software package.[24] The time step was set at 2 fs and the distances of all bonds involving hydrogen atoms were kept at their equilibrium values with the RATLLE algorithm.[25] Atom pair distance cut-offs were applied at 1.4 nm to compute all van der Waals interactions. To

avoid discontinuities in this energy component, the van der Waals energy term was forced to slowly converge to zero by applying a smoothing factor from 1.0 nm. Electrostatic interactions were extensively computed by means of Ewald summations. The real space term was defined by the van der Waals cutoff (1.4 nm), while the reciprocal space was computed by interpolation of the effective charge into a charge mesh with a grid thickness of 1 point per $Å^3$ (particle mesh Ewald).[26] In all MD simulations, both temperature and pressure were controlled by the weak coupling method, the Berendsen thermo-barostat[27] and a time constant of 1 ps was applied for heat bath coupling and pressure relaxation.

Equilibration and production details. The equilibration of the simulated systems has been carried out according the references 28 and 29 with slight differences. Briefly, in all studied cases the equilibration protocol was divided in the following steps: first, the energy of each system was relaxed by 10^4 steps of energy minimization using Newton Raphson method. Then, the solvent was equilibrated using a 1 ns long trajectory with NVT conditions at 500 K while the peptides were kept frozen. Following, the temperature was set at 298 K and another 1 ns NVT trajectory was run, unfreezing the peptide chains for thermal equilibration. Finally, 1 ns in NPT conditions, pressure set at 1.034 bar, and keeping the former temperature in order to relax the density of the solution. This later step is the beginning of the production runs of each trajectory series, keeping identical simulation conditions to those of the NPT equilibration cycle.

Analysis Details. Following similar criteria to those used in early works,[9-10] the existence of different interactions was accounted based on geometric criteria: salt bridges were accounted if the distance between the centres of the interacting groups shortens below 0.450 nm. We computed hydrogen bonds for distances H···O shorter than 0.3 nm and angles ∠N-H···O higher than 120.0°. On this basis, the existence of a strands cluster (*i.e.* an aggregate with 2 or more chains) was established only if one or more of those assessed interactions were present between

two pairs of strands. Chained assembly was considered as an associative feature: if *chain A* presented interactions with *chain B,* and *chain B* presented interactions with *chain C,* all three *A*, *B* and *C chains* would be computed to belong to the same aggregate. Chains that remained "interaction less" with respect to other strands present in the simulation box would be named "lonely chains" or "lonely strands".

Results and Discussion

A huge challenge for the pharmaceutical industry is to achieve physical stability of peptide-based and protein-based therapeutics. However, aggregation in physiological conditions can obscure their usefulness, or also can affect the efficacy of peptide-based materials to be part of composite formulations. Herein, we compare the ability of CREKA to form aggregates *vs* one of its analogue, the pentapeptide containing *N*-methylation on Glu (CRMeEKA).[10]. Aggregate size was studied by DLS after one day in PBS 1× at 37 ºC and 80 rpm for different peptide concentrations (0.5, 1, 5 and 10 mg/mL). Results displayed in Figure 1a-b show that at 0.5 mg/mL CREKA and CRMeEKA aggregates present different diameter sizes 255 ± 55 nm and 59 ± 21 nm, respectively. This difference is maintained at 1 mg/mL (Figure 1c-d) but when the concentration increases both present similar diameters. Thus, the two systems aggregate already at the lower investigated concentrations, but in this regime CRMeEKA forms aggregates smaller than CREKA. A potential explanation would be that aggregation occurs sooner in the case of CREKA(Figure 1a-b**)**. Despite non direct prove has been obtained, these results combined with the behaviour observed in the simulation sections suggests this interpretation (see below).

Information about peptide secondary structure[30,31] can be obtained by studying the absorption changes in the FTIR amide I region (1600-1700 cm^{-1}). Figure 2 shows CREKA and CRMeEKA spectra characterized by a broad absorption band centred around 1676 cm^{-1}, which is related to the TFA carboxylate groups and possibly hiding the β-turn structure (1671 cm^{-1}),

and a smaller peak at 1650 cm^{-1} corresponding to a random coil structure.[32,33] (full FTIR spectra are shown in Figure S1). Unfortunately, the TFA carboxylate groups also interfere with the bands beyond 3000 cm-1, and no more detailed information about molecular arrangements could be obtained (both peptides show the same peak at this region hampering, corresponding to OH groups of TFA interacting with NH amide groups and none detailed information about the strands association patterns could be extracted). On the other hand, bioactive conformation for CREKA tethered to a metallic nanoparticle was a γ-type turn motif with a hydrogen bond between the C=O(Cys) and N–H(Lys) and the side chains of the central charged residues protruding out of the bended main chain in the same direction (Arg, Glu and Lys).[10] These results suggest that the pentapeptides may adopt a mixt between β-turn and other folded conformations when they are aggregating. Small Peptides present more than one stable conformations in solution and small conformation rearrangements allow fast interconversion from between β-turn type II' and γ-turn. The main feature of CREKA bioactive active was the relative position of the charged side chains with respect to the bent main chain, which is identical in both types of turns. Furthermore, differentiation between turn-types when studying small peptides is not a simple task, even when using amino acids with very restricted conformational space.[34] Thus, experimental limitations did not allow to differentiate beyond the present of folded arrangements with intramolecular hydrogen bonds.

In order to obtain additional information regarding the pentapeptides conformation when forming aggregates, CD spectra were recorded at different peptide concentrations (0.001, 0.005 and 0.01 mg/mL). As it can be seen in Figure S2, both peptide's spectra present a negative band near 200 nm suggesting random structures, whereas in the case of CRMeEKA there is another minimum around 230 nm. Although the latter may be assigned to the chiral optical absorption of cysteine, [35,36] which is related to a higher cysteine exposure on the peptide analogue, it can also be related to folded structures. For instances, in a recent study on

amphiphilic peptides the two negative bands combined with the positive shoulder at 197 nm was associated to the formation of bended structures.[37] It has also been reported that a band in 230 nm corresponds to the β-turns.[38-41] However, the spectra obtained at really small concentrations (0.001 mg/mL) of CRMeEKA could resemble to those encountered for alpha-helix structures. We believe that the two peptides at high concentrations aggregate in a mix of random coiled and bended structured conformation that are different with respect to each other and might variate with concentration. Concretely, the CREKA spectrum is more like that of random coil with respect to that of CRMeEKA. The presence of the methyl group not only produces steric hindrance but also reduces the flexibility of the backbone significantly increasing the stability of the bioactive conformation.

A molecular description of the aggregation of CREKA and CRMeEKA has been obtained by using atomistic MD simulations, which were conducted using the procedure described in the Methods section. Even though the conditions were not identical to those in which experiments were performed, the framework of molecular mechanics allows to compare experimental observations with the simulated models, the concentration differences mainly acting as an accelerated rate of aggregation. In this context, three different and consecutive box size reductions (previously referred as MD-I, MD-II and MD-III respectively), allowed us to study the events that take place along these trajectories both as a whole vision of the assembly capabilities inherent to each studied peptide. Figure 3 reveals the first differences between the CREKA and CRMeEKA systems at the molecular level. The non-methylated peptide chains quickly aggregate into large clusters of associated chains that keep steadily growing when the simulation box is reduced. At the end of MD-III most of the present chains are associated in a single big cluster, whose size oscillates between 13 and 15 chains (each simulated model presents a total of 15 peptide chains). On the other hand, CRMeEKA shows a seemly opposite behaviour. Independently of the box dimensions, the maximum number of chains in the largest

cluster is 10, being the average value of this parameter lower than 4 for most of the simulated time. Despite these differences at molecular level, the total number of detected clusters remains relatively low in both species (Figures S3 and S4). Therefore, the main differences must stem from the relative size of the smaller clusters formed by each studied peptide. CREKA strands either remain isolated or associated into a single large cluster, as can be seen in the similar values obtained in Figure 3a and Figure S4a. This cluster steadily grows until engulfing almost all the chains present in the simulation box. On the contrary CRMeEKA clusters are smaller (in number of strands) and tend to dissociate quite often. As depicted in Figure S5b, the total number of strands that are clustered oscillates from 15 strands to 2 strands every 10 to 15 ns. However, the size of their largest cluster remains steady throughout all the simulated timeframe (Figure 3b). Hence, small clusters of 2 or 3 chains of CRMeEKA continuously aggregate and rapidly fall apart. DLS observations showed that at low peptide concentrations (0.5 mg/mL) CREKA aggregates are systematically larger than those observed for CRMeEKA aggregates Due to technical limitations the exact experimental conditions cannot be reproduced. Under these conditions, trajectories should extend to the order of seconds (2 orders of magnitude larger than what is currently feasible). Nonetheless, it must be noted that the effect of peptide concentration on the aggregation process in the framework of MD technique does not manifest in the same mode as in macroscopic observations: when simulating aggregation of peptides using MD, increasing concentration mainly accelerates the evolution towards a near-equilibrium state. Thus, comparison with the experiments still provides some support to the differences shown in both simulated

cases. Combining both techniques point towards differences in the molecular organization in which both species aggregate.

Next, we studied the influence of weak interactions on the association preferences for both species. Figure 4 depicts what type of interactions are present as function of the aggregate size for the entire accumulated time (i.e. combining the results of MD-I, MD-II and MD-III as a whole). For each analogue, these values are presented as the percentage of each kind of detected weak interaction (see method section for details) with respect to the total number of detected ones. Noticeable differences appear when focusing on the preferred intra-strand interactions on those strands do not form part of an aggregate. Lonely strands of CREKA tend to stabilize their structure via hydrogen bonds between amide groups of the backbone and some of the polar groups of their own side chains. Despite having been described as an excellent former of turns when tethered to metallic surfaces,[9,10] only 15% of the CREKA lonely strands show coherent hydrogen bonds between backbone functional groups, whereas in most cases putative interaction patterns are preferred: 22% of detected interactions in lonely strands corresponds to hydrogen bonds between side chain groups and backbone amides. The second most common pattern are salt bridges between the charged termini and charged groups of side chains, which represent 20% of the total observed interactions, and are responsible of stabilizing turn-like arrangements (similar structure but devoid of hydrogen bonds).

On the other hand, lonely strands of CRMeEKA show structural patterns closer to those described for this peptide when tethered to a metallic particle.[10] 40% of all interactions correspond to the canonical main chain – main chain hydrogen bonds closing 7 or 10 atoms rings (depending on being γ- or β- turn, respectively). Moreover, up to 20% of the rest of interactions corresponds to salt bridges between both termini, which enable this pentapeptide to adopt conformational arrangements like those of standard β- and γ-turns. The higher number

of stable conformers adopted by the lonely strands of CRMeEKA suggests less propensity to form large aggregates compared with CREKA.

Similar differences between both analogues can be described in the intra-molecular interactions when strands cluster in aggregates. These differences can be traced through all different sizes of aggregates observed in each peptide. While the amount of intra-strand hydrogen bonds between backbone amide groups remains steadily below 10% in all aggregates formed by CREKA, this kind of interaction is present in more than 20% of clusters formed by CRMeEKA. Moreover, this intra-strand interaction pattern is the most populated almost in all studied aggregates of the latter analogue, independently of the aggregate size. In contrast, CREKA segments have a lower presence of intra-strand interactions when assembled, most case below 15%. Most polar interactions detected in CREKA aggregates are between strands. Despite this, inter-strand hydrogen bonds are not the predominant interaction in CREKA assemblies and this feature could explain why aggregates of this peptide are not organized in consistent structures but rather randomly, as qualitative comparison with IR and CD experiments pointed out.

However, independently of the cluster size, there is always a small proportion of inter-strand interactions (around 20% with respect to the total number) that corresponds to hydrogen bonds between backbone amide groups (i.e. those that would be majority in cross beta-structures) that apparently are crucial to retain strands associated once a cluster is formed. Regardless of the cluster size, their presence facilitates the formation of many other inter-strand interactions, mainly salt bridge types (above 45% of the detected inter-strand interactions), that critically contribute to the formation larger clusters in CREKA aggregates than CRMeEKA assemblies (Figure 5a depicts a typical organization with this arrangement).

In latter case, the most remarkable feature is the absence of inter-strand hydrogen bonds: amide – amide hydrogen bonds never come over 10% of the total, dropping below 5% in the larger clusters. In this analogue, the strong tendency to form intra-strand hydrogen bonds favours the formation of smaller and less ordered aggregates. Under such conditions, the main role of keeping the assemblies together rests on salt bridges between different charged side chains and/or the strand terminal charged groups. The combination of tending to retain turn-like conformation with the high population of inter-strand salt bridges leads to assemblies based on two or three strands, each of one with independent folded conformations, bonded via multi-centred salt bridges (Figure 5b shows an example of this sort of arrangement). All these factors favour the formation of dimers/trimers small aggregates) more than wide extended aggregates, where each molecule needs to interact with more than one partner.

Figure 6 confirms the different behaviour observed in both sequences. Both pentapeptides adopt preferably random coil organizations when forming aggregates in water. However, great differences can be observed when studying the minority conformations under these conditions. Whereas in CREKA assemblies the second most populated is the inter-strand β-sheet conformation, in CRMeEKA assemblies this role is mainly monopolized by turn conformations, which confirms the previous structural statements solely based on the percentage of intra/inter–strands differences. A thorough examination of these tendencies shows interesting features Firstly, lonely strand of CREKA in solution do not present a clear propensity to form turn based conformations, in contrast with the previous investigations, in which its conformational freedom was always reduced because CREKA strands were tethered to a metallic surface. In this work, the pentapetides are untethered at the N-terminal site, which increases its conformational flexibility. This behaviour can be visualised by the following data: only 15% of all the observed conformers in solution adopt turn organizations when not forming aggregates in solution. This feature can be observed in the accumulated Ramachandran plot of

the Glu residue (Figure S6), which shows a high degree of conformational freedom, especially when the CREKA strands are part of an aggregate (Figure S6b). It is also noticeable the much lower conformational diversity associated to lonely strands, turn compatible conformations being the most populated in those cases (Figures S6a). Finally, the absence of voluminous groups in the backbone (*i.e.* amide groups are not modified in the parent peptide) allows a significant number of strands to form β-sheet like organization motifs in all detected assemblies. Notice that extended conformations and "*pseudo-β-sheet*" arrangements (Figure 7a) also contribute to the organization of these assemblies, since in these cases the assembled strands present random-like main chain conformation, while preserving the interaction patterns of canonical β-sheets.

On the other hand, the conformational profile drew for CRMeEKA aggregates points toward an absence of specific conformational changes when forming assemblies. As mentioned above, the strands association follow an opposite trend: in water solution and in absences of aggregation, the second most important organization is based on turns motifs (which includes conformation arrangements for the central residues of the peptide close to those featured by helical conformations). In this case, the engineered peptide favours the formation of the bioactive conformation for this family of homing peptide analogues, being the turn conformers more than 40% of all explored conformations. Remarkably, this feature remains almost unaltered when several strands of the pentapeptide associate, remaining the second preferred arrangement of each strand. This is reflected in Figure S6c-d, which shows that the conformational space found for the *N*-methyl Glu residue is the same for both lonely strands and strands contained in clusters. This behaviour is in absolute agreement with the conformational preferences of this amino acid when CRMeEKA chains are tethered to a metallic surface.[10] Accordingly, the pulsar behaviour of CRMeEKA assemblies can be rationalized as follows: losing the association with other strands does not greatly affect the

molecular conformation. This assessment is confirmed in the Supplementary Information by examining the correlation between conformational motifs and specific types of interaction (Figures S7 and S8).

The last step of the characterization of the structural preferences of these two pentapeptides has been focused on the energetics of the aggregates formed by each analogue. Following similar investigations on peptides dynamics,[42] the energetics characterization of the generated trajectories has been centred on the analysis of the main contributions to the interaction energy (hereafter named E_{tot}). The latter has been computed considering the following contributions: *(i)* the peptide-peptide non-bonding interactions, referred as E_{pept}, (*i.e.*, electrostatic and van der Waals interactions beyond three consecutively bonded atoms or between different peptide chains), which takes into account the most relevant contribution to the internal energy changes associated with conformation changes and inter-strand associations; and *(ii)* the interaction of the peptide strands with solvent molecules, E_{wat}. These two contributions are expected to play the role in the evaluation of variations of free energy upon assembly and disassembly of strands. Preliminary calculations with a reduced set of snapshots revealed that counter ions – water interactions were not necessary to explain the assembly properties of the studied pentapeptides, their role being negligible. In addition, all intra-molecular contribution associated with the peptides bonding interactions (*i.e.* stretching, bending and torsional strain) have been found to be minor when peptides located at the same environment are compared and, consequently, have been omitted for clarity.

At first, we searched for relationship between the actual size of each aggregate and variations of the interaction energy. In order to obtain the broadest possible perspective, for each (temporal) snapshot we considered the potential presence of more than one individual solute. In other words, each aggregate (if any) and any lonely strand (if any) will be considered independent solutes. Within this context, the variation of interaction energy as function of the

apparent size of each solute over was evaluated. The size of each considered solute was estimated by computing their radius of gyration. In this way, it is possible to evaluate the energy differences only as function of the solutes size, independently of the amount of chains that constitute each solute (either lonely strands or clusters of assembled strands). Despite offering significant information (Figures S9 and S10) this analysis only provided a general vision of the differential behaviour between both peptides, yet no clear information about the aggregation preferences could be extracted. This initial overview indicates that CRMeEKA strands, independently if they are assembled or lonely, are more prompt to be solvated in water than CREKA, especially when their structural features imply low values of radius of gyration (see below). It is tempting to assess that the main reason why this analogue prefers to organize as small clusters or as lonely strands under aggregation conditions is due to its better solvation with respect to that when larger aggregates appear. However, there is not a clear separation in terms of radius of gyration with respect the number of strands that constituted each solvated particle. These questions are addressed in Figure 7 (with the information extracted from Figures S11 and S12).

The two plots displayed in Figure 7a correlate the radius of gyration of each solute with the quantity of strands that constitutes it. Although the series for CRMeEKA are not completed (the larger aggregate presenting 10 strands), it is visible that both lonely strands and aggregates adopt organizations that are more voluminous than those constituted by CREKA. A very rough approach, with a linear regression for each case (this is a qualitative estimation due to the large dispersion of data), provides a numerical approximation of these trends: the slope of correlating the number of strands with the corresponding gyration radius for CRMeEKA solutes is $1.3 \cdot 10^{-1}$ strands$\cdot$nm^{-1} whereas the same magnitude for CREKA solutes is $9.0 \cdot 10^{-2}$ strands$\cdot$nm^{-1}, which implies that the actual size of CREKA strands in water is about 7 times smaller than those of CRMeEKA. With these correlations in mind, the values provided by Figures 7b and 7c bring

shape to the chemical interpretation of the differences shown earlier. In all the range of sizes of aggregates, CREKA structures show lower interaction peptide-peptide energies than those of CREMeKA, which correlates with the higher number of inter strand interactions involving polar backbone groups in the respective assemblies (Figure 6a). This feature allows CREKA based assemblies to grow tight and compact, facilitating the solubility of the formed aggregates in water. Albeit CREKA assemblies present higher interaction energies with the solvent than CRMeEKA ones (worse solvation of CREKA assemblies than those of CRMeEKA), the better peptide-peptide interaction energies of assemblies constituted by CREKA strands compensate the better interaction energy of CRMeEKA assemblies with the solvent.

Within this context, if only those aggregates that present the lowest E_{pept} values are studied (bottom section of ordinate axis of both Figure 7b plots), two more features can be inferred. First, for CREKA assemblies, the addition of new strands to a formed aggregate does not alter the interaction energy per strand, whereas in the case of CRMeEKA this energy component tends to increase with the size of aggregates. If this trend is combined with the behaviour observed for the interaction energy with the solvent (Figure 7c), in which the gap between presenting a single strand in water or an assembly of two or more strands is much larger for the latter than for CREKA strands, the outcome of the trajectories can be understood. For CREKA strands the growth of an assembly maintains the interaction energy steady and the penalty of solvating a larger solute is not big enough to hamper the formation of large assemblies. On the other side, when studying the interaction energy components in CRMeEKA models, the increase in the size of an assembly tends to increase the inter/intra-strand interaction energy, while water-peptide interaction energy does not decrease enough to compensate the slight destabilization induced by the increased size. This feature facilitates the constant detachment of strands from already formed assemblies (phenomenon already mentioned earlier, Figure S5).

Conclusions

The aggregation of CREKA and CRMeEKA has been systematically studied using a combination of both experimental and computer simulations techniques. DLS results indicate that CREKA aggregates before CRMeEKA. Although FTIR spectra do not allow to identify well-defined secondary structures for any pentapeptide, CD spectra suggest that CRMeEKA may form folded structures. As understanding of the aggregation tendencies of therapeutic peptides is of fundamental importance for their clinical usage, MD computer simulations have been extensively used to provide microscopic insights of the aggregation mechanism of CREKA and CRMeEKA. Simulations show that CREKA quickly aggregates in large clusters, which is fully consistent with experimental observations (*i.e.* slower rate of aggregation for CRMeEKA than for CREKA). This behaviour has been rationalized in terms of interactions and energy contributions, which have revealed that CRMeEKA has a higher tendency to form stable β- and γ-turns and to be better solvated than CREKA. In summary, our study not only provides the understanding of CREKA and CRMeEKA aggregation, but also provides critical insight for the future development of rational therapeutics using such pentapeptides.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Authors acknowledge MINECO-FEDER (RTI2018-098951-B-I00) and Agència de Gestió d'Ajuts Universitaris i de Recerca (2017SGR359). Support for the research of C.A. was received through the prize “ICREA Academia” for excellence in research funded by the Generalitat de Catalunya.

References

1. J. McCafferty and D. Schofield, *Curr. Opin. Chem. Biol.*, 2015, **26**, 16–24.
2. E. Ruoslahti, *Adv. Drug Deliv. Rev.*, 2017, **110**, 3–12.
3. V. A. Petrenko and J. W. *Expert. Opin. Drug. Deliv.*, 2017, **14**, 373–384.
4. L. Lu, H. Qi, J. Zhu, W. X. Sun, B. Zhang, C. Y. Tang and Q. Cheng, *Biomed. Pharmacothe.*, 2017, **92**, 187–195.
5. F. Biscaglia, G. Ripani, S. Rajendran, C. Benna, S. Mocellin, G. Bocchinfuso, M. Meneghetti, A. Palleschi and M. Gobbo, *ACS Appl. Nano Mater.*, 2019, **2**, 6436–6444.
6. F. Biscaglia, S. Rajendran, P. Conflitti, C. Benna, R. Sommaggio, L. Litti, S. Mocellin, G. Bocchinfuso, A. Rosato, A. Palleschi, D. Nitti, M. Gobbo and I. M. Meneghetti *Adv Healthc Mater.*, 2017, **6**, 1700596.
7. D. Simberga, T.; Duzaa, J.H. Park, M. Esslera, J. Pilcha, L. Zhang, A. M. Derfusa, M. Yang, R. M. Hoffman, S. Bhatia, M. J. Sailor and E. Ruoslahti, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 932–936.
8. D. Peters, M. Kastantin, V. R. Kotamraju, P. P Karmali, K. Gujraty, M. Tirrell and E. Ruoslahti, *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 9815–9819.
9. D. Zanuy, A. Flores-Ortega, J. Casanovas, D. Curco, R. Nussinov and C. Alemán, *J.Phys. Chem. B*, 2008, **112**, 8692–8700.
10. D. Zanuy, F. J. Sayago, G. Revilla-López, G. Ballano, L. Agemy, V.R. Kotamraju, A.I. Jiménez, C. Cativiela, R. Nussinov, A. M. Sawvel, G. Stucky, E. Ruoslahti and C. Aleman, *J. Comput. Aided Mol. Des.*, 2013, **27**, 31–43.
11. L. Agemy, K. N. Sugahara, V. R. Kotamraju, K. Gujraty, O. M.; Girard, Y. Kono, R. F. Mattrey, J. H. Park, M. J. Sailor, A. I. Jimenez, C. Cativiela, D. Zanuy, F.J. Sayago, C. Aleman, R. Nussinov and E. Ruoslahti, *Blood*, 2010, **116**, 2847–2856.
12. Z. G. She, X. Liu, V. R. Kotamraju and E. Ruoslahti, *ACS Nano*, 2014, **8**, 10139–10149.
13. A. Puiggalí-Jou, L.J. del Valle, E. Armelin and C. Alemán, *Macromol. Biosci.*, 2016, **16**, 1461−1474.
14. A. Puiggalí-Jou, L. J. del Valle, C. Alemán and M. M. Pérez-Madrigal, *J. Pept. Sci.*, 2017, **23**, 162–171.
15. X. Zeng, Z. Chen, L. Tang, H. Yang, N. Liu, H. Zhou, Y. Li, J. Wu, Z. Deng, Y. Yu, H. Deng, X. Hong, Yuling Xiao *Chem. Commun.* 2019, **55**, 2541-2544.

16. Y. Zhong, Y Zhang, J. Xu, J. Zhou, J. Liu, M. Ye, L. Zhang, B. Qiao, Z.Wang, H. Ran, D Guo Low-Intensity Focused Ultrasound-Responsive Phase-Transitional Nanoparticles for Thrombolysis without Vascular Damage: A Synergistic Nonpharmaceutical Strategy *ACS Nano* 2019, **13**, 3387−3403.

17. V.S. Perera, G. Covarrubias, M. Lorkowski, P. Atukorale, A. Rao, S. Raghunathan, R. Gopalakrishnan, B.O. Erokwu, Y. Liu, D. Dixit S.M. Brady-Kalnay, D. Wilson, C. Flask, J. Rich, P.M. Peirisa, E. Karathanasis *Nanoscale*, 2017, **9**, 9659–9667.

18. L. B. de Oliveira Freitas, L. de Melo Corgosinho, J. Araújo Quintao A. Faria, V. Mateus dos Santos, J. Magalhaes Resende, A. Soares Leal, D. Assis Gomes, E. Martins Barros de Sousa *Microporous Mesoporous Mater.* 2017, **242**, 271-283

19. S. Yang, D. Yao, Y. Wang, W. Yang, B. Zhang, D. *Chem. Commun.*, 2018, **54**, 9841-9844.
20. R. Gallardo, N.A. Ranson, S. E. Radford Amyloid structures: much more than just a cross-β-fold Curr. Opin. Struct. Biol. 2020, 60, 7–16.

21. A. Puiggalí-Jou, L. J. del Valle and C. Alemán *ACS Biomater. Eng. Sci.* 2020, **6**, 2135–2145.

22. Y. Duan, C. Wu., S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999–2012.

23. W. L. Jorgensen, J. Chandrasekhar, J.D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

24. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.

25. H. C. Andersen, *J. Comput. Phys.*, 1983, **52**, 24–34.
26. A. Toukmaji, C. Sagui, J. Board and T. Darden, *J. Chem. Phys.*, 2000, **113**, 10913–10927.
27. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

28. D. Zanuy, J. Poater, M. Sola, I. W. Hamley and C. Aleman, *Phys. Chem. Chem. Phys*., 2016, **18**, 1265–1278.

29. J. Triguero, C. Alemán and D. Zanuy, *Chem. Phys.*, 2019, **522**, 163–170.
AKI

30. S. Fleming, P. W. J. M. Frederix, I. Ramos Sasselli, N. T. Hunt, R. V. Ulijn and T. Tuttle, *Langmuir*, 2013, **29**, 9510–9515.

31. A. Barth and C. Zscherp, *Q. Rev. Biophys*., 2002, **35**, 369–430.
32. J. Kong, S. Yu. *Acta Biochim. Biophys. Sin. (Shanghai)* 2007 **39**, 549-559.
33. J. Le Coutre, H.R. Kaback, C.K.N. Patel, L. Heginbotham, C. *Proc. Natl. Acad. Sci. U. S. A.* 1998, **95**, 6114–6117.
34. E. Vass, M. Hollosi F. Besson, R. Buchet *Chem. Rev.* 2003**, 103***,* 1917-1954.
35. F. Zhu, X. Li, Y. Li, M. Yan and S. Liu, *Anal. Chem.*, 2015, **87**, 357–361.
36. C. Krittanai and W. C. Johnson, *Anal. Biochem.*, 1997, **253**, 57–64.
37. J. Casanovas, E. Mayans, A. Díaz, A. Gil, A. I. Jiménez, C. Cativiela, J. Puiggalí and C. Alemán, *Chem. Commun.*, 2019, **55**, 8556–8559.
38. A. Iyer, S. J. Roeter, V. Kogan, S. Woutersen, M. M. A. E. Claessens and V. Subramaniamm, *J. Am. Chem. Soc.*, 2017, **139**, 15392–15400.
39. Y. Momotani, R. Arie and T. Takagic, *BBA Rep.*, 1981, **668**, 193–196
40. J. A. Smith, L. G.Pease and K. D. Kopple, *Crit. Rev. Biochem. Mol. Biol.*, 1980, **8**, 315–399.
41. K. Kornmueller, I. Letofsky-Papst, K. Gradauer, C. Mikl., F. Cacho-Nerin, M. Leypold, W. Keller, G. Leitinger, H. Amenitsch and R. Prassl, *Nano Res.*, 2015, **8**, 1822–1833.
42. S. Maity, D. Zanuy, Y. Razvag, P. Das, C. Aleman and M. Reches, *Phys. Chem. Chem. Phys.*, 2015, **17**, 15305–15315.

Captions to Figures

Figure 1. (a) Effective diameter (D_{eff}) *vs* peptide concentration graph and (b) the corresponding value table. (c) DLS intensity *vs* diameter graph and (d) the corresponding autocorrelation functions for CREKA and CRMeEKA (1 mg/mL in filtered PBS 1x).

Figure 2. (a) FTIR absorption spectra of 30 mM peptide solutions in D_2O and (b) peak table.

Figure 3. Maximum size of the pentapetide aggregates versus time for (a) CREKA and (b) CRMeEKA assemblies, respectively. Notice that in each plot the vertical dashed lines indicate the precise moment in which each box size reduction takes place.

Figure 4. Percentage of types of polar interactions present in the studied (a) CREKA and (b) CRMeEKA peptide assemblies as function of the assembly size. Lonely strands are included as clusters of 1 single strand. Colours code meaning are explicitly indicated in the offset legend box (from bottom to the top: intra-strand salt bridges, inter-strand salt bridges, intra-strand hydrogen bonds and inter-strand hydrogen bonds). *S* refers to side chain, *M* to main chain. Thus *S*-*S* means two interacting groups from two different side chains, *S*-*M* indicates that one of the interacting groups is located in the main chain whereas its counterpart is in the side chain, and *M*-*M* means that both interacting groups are part of the peptide main chain.

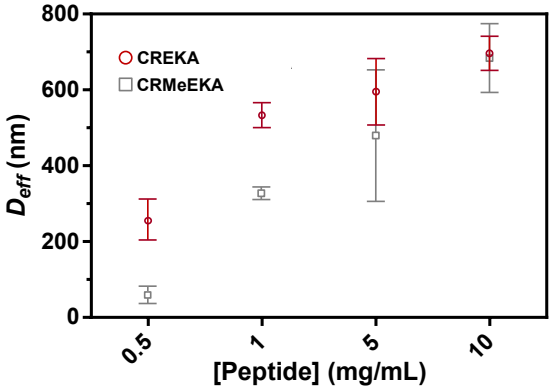
Figure 5. Atomistic representation of most common association modes for (a) CREKA and (b) CRMeEKA assemblies. Atom colours follow the CPK convention. Hydrophobic hydrogen atoms have been removed for clarity. Main chain α-carbons and backbone have been remarked in yellow colour. In CRMeEKA model, the *N*-methyl of Glu residue has been remarked with a green space filled ball. Encircled zones point to interactions mentioned in the text.

Figure 6. Percentage of types of conformational motifs in the (a) CREKA and (b) CRMeEKA pentapeptides assemblies as function of the assembly size. Lonely strands are included as clusters of 1 single strand. Colours code meaning is explicitly indicated in the offset

legend box. Conformations named "*int. pattern X*" refer to those conformers in which the main chain dihedral angles do not correspond to the canonical values, but the ring closed by the interacting amide groups fits the specified conformational motif.

Figure 7. (a) Correlation between radius gyration and the number of strands involved in any given present solute throughout the studied trajectories. (b) Correlation between peptide interaction energy (E_{pept}), in kcal·mol^{-1}·strand^{-1}, and the number of strands. ***(c)*** *Correlation* between water–peptide/s interaction energy (E_{wat}), in kcal·mol^{-1}·strand^{-1}, and the number of strands.

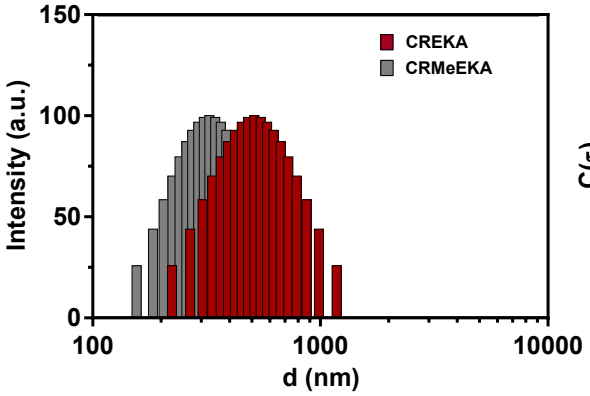
(a)



(b)

[Peptide]	CREKA Averaged	CRMeEKA Averaged
0.5	255 ± 55	59 ± 21
1	533 ± 33	327 ± 12
5	589 ± 93	479 ± 172
10	698 ± 42	685 ± 92

(c)



(d)

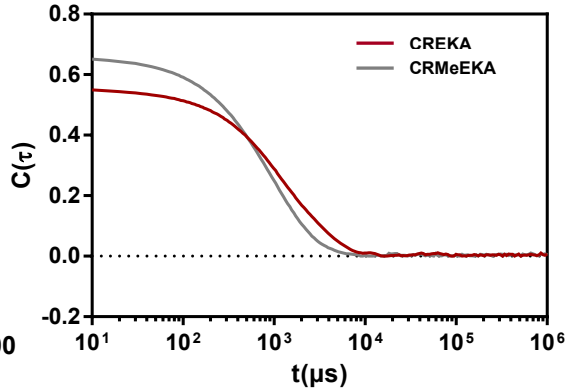
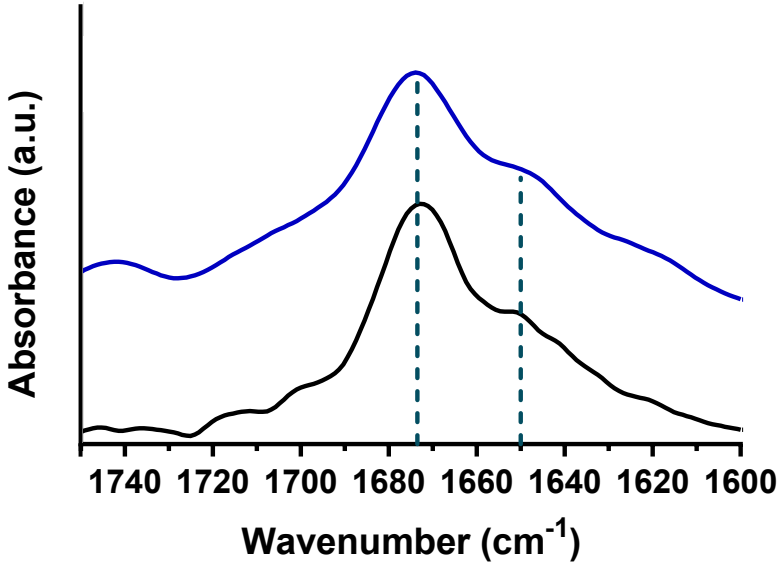


Figure 1

(a)



(b)

Peptide	Amide I band (cm^{-1})	
CREKA	1676	1650
CRMeEKA	1673	1650

Figure 2

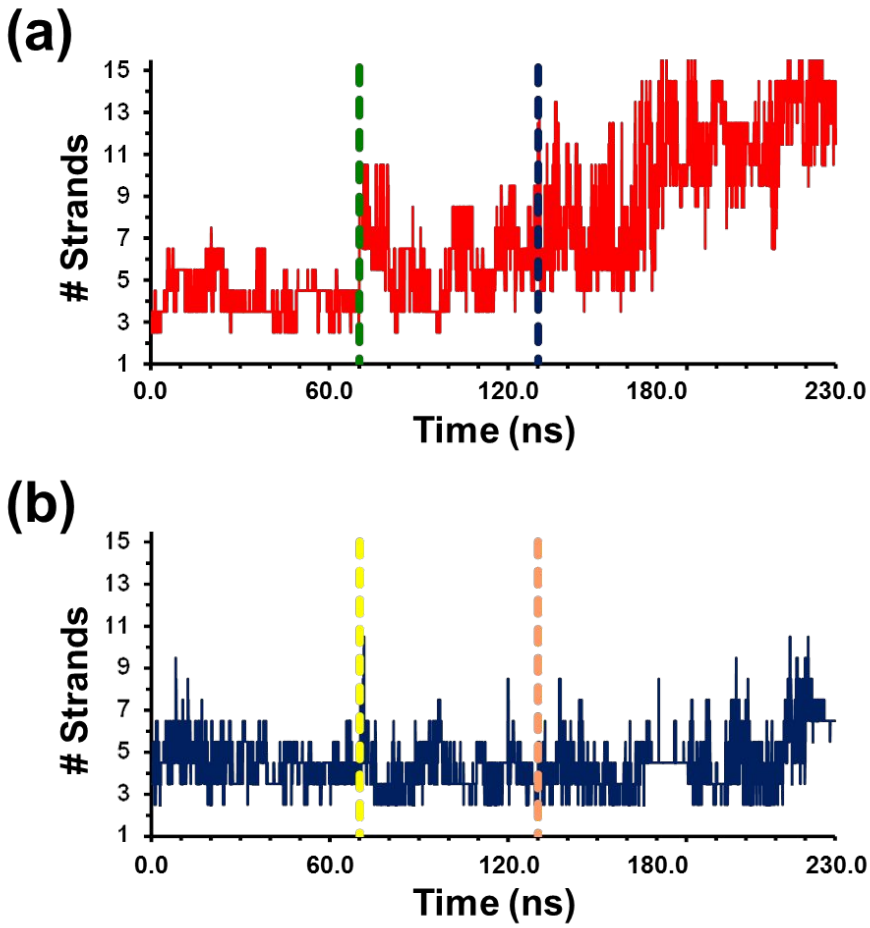

(a)
Strands
15
13
11
9
7
5
3
1
0.0
60.0
120.0
180.0
230.0
Time (ns)
(b)
Strands
15
13
11
9
7
5
3
1
0.0
60.0
120.0
180.0
230.0
Time (ns)

Figure 3

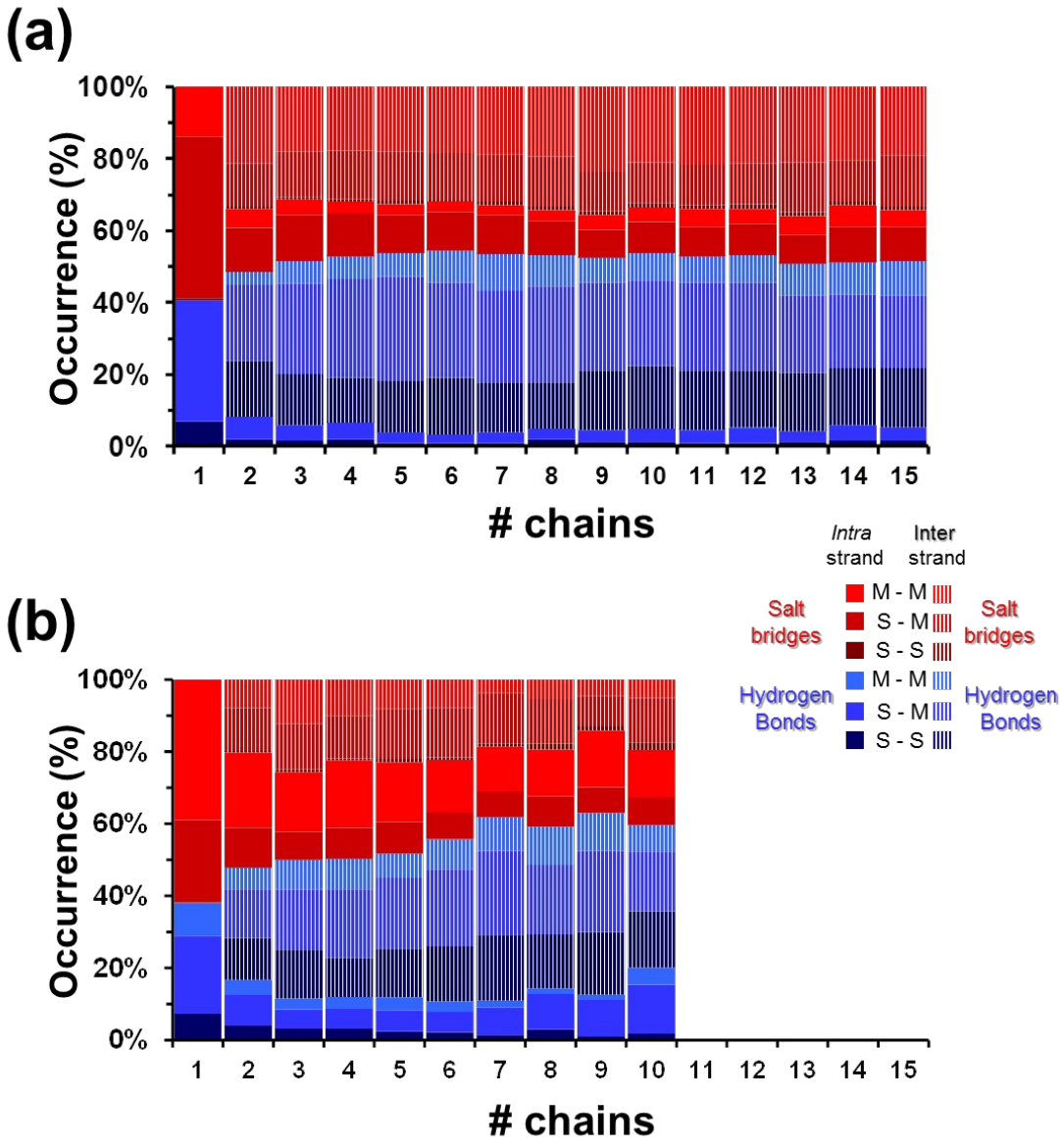

(a)
100%
80%
60%
40%
20%
0%
Occurrence (%)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
chains
(b)
100%
80%
60%
40%
20%
0%
Occurrence (%)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
chains
Intra strand
Inter strand
Salt bridges
M - M
S - M
S - S
Salt bridges
Hydrogen Bonds
M - M
S - M
S - S
Hydrogen Bonds

Figure 4

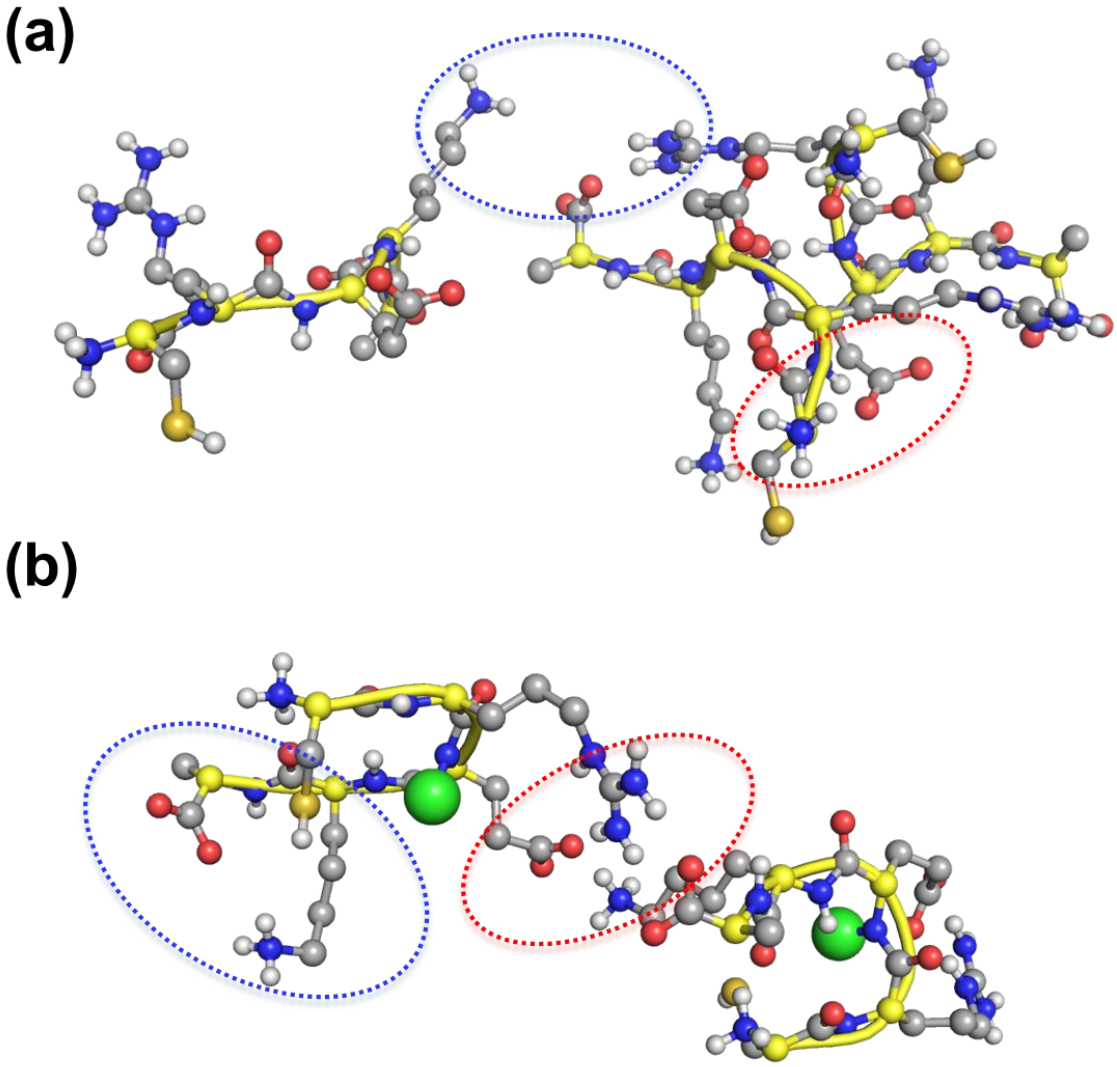


Figure 5

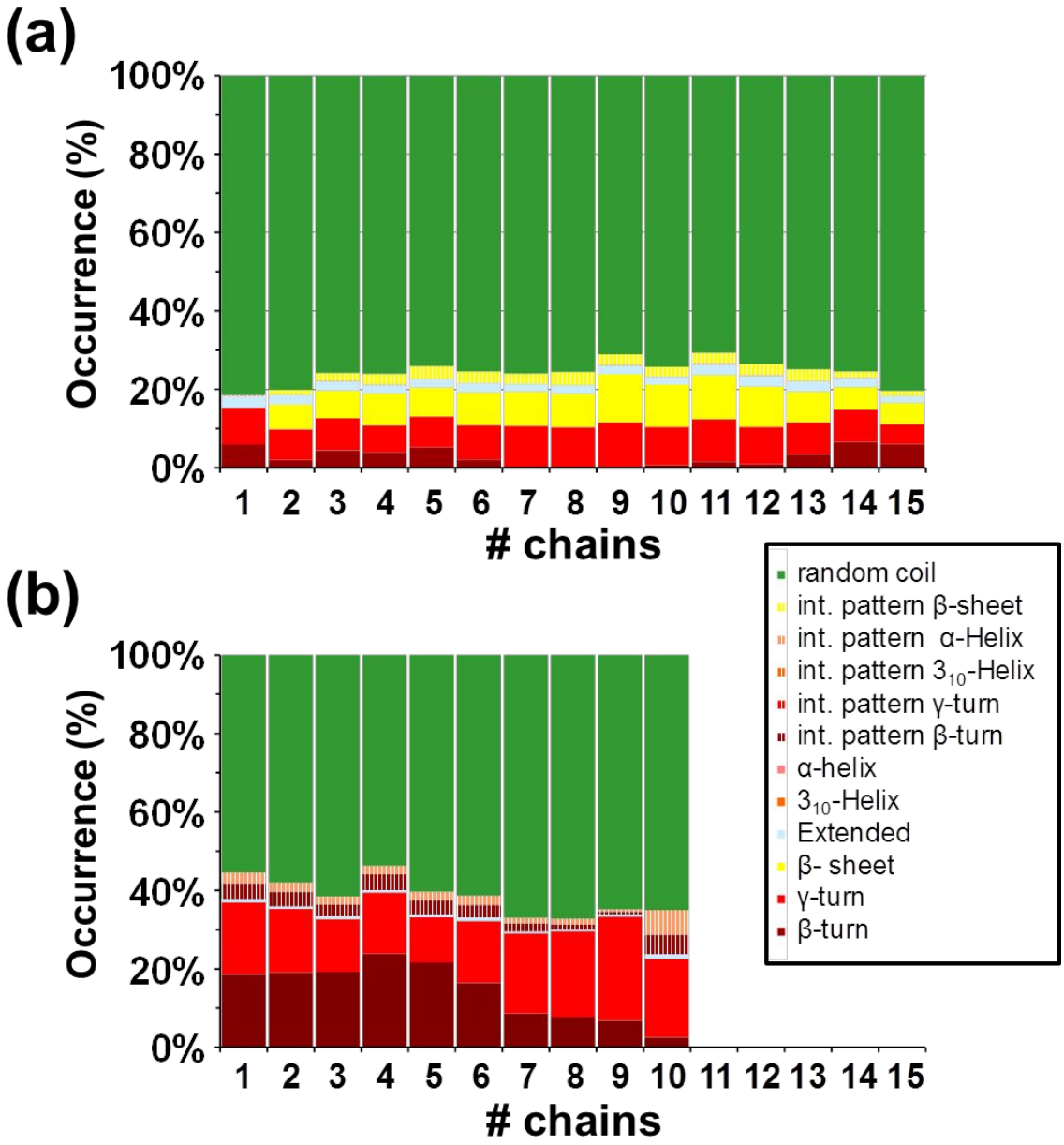

(a)
Occurrence (%)
100%
80%
60%
40%
20%
0%
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
chains
(b)
Occurrence (%)
100%
80%
60%
40%
20%
0%
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
chains
random coil
int. pattern β-sheet
int. pattern α-Helix
int. pattern 3$_{10}$-Helix
int. pattern γ-turn
int. pattern β-turn
α-helix
3$_{10}$-Helix
Extended
β- sheet
γ-turn
β-turn

Figure 6

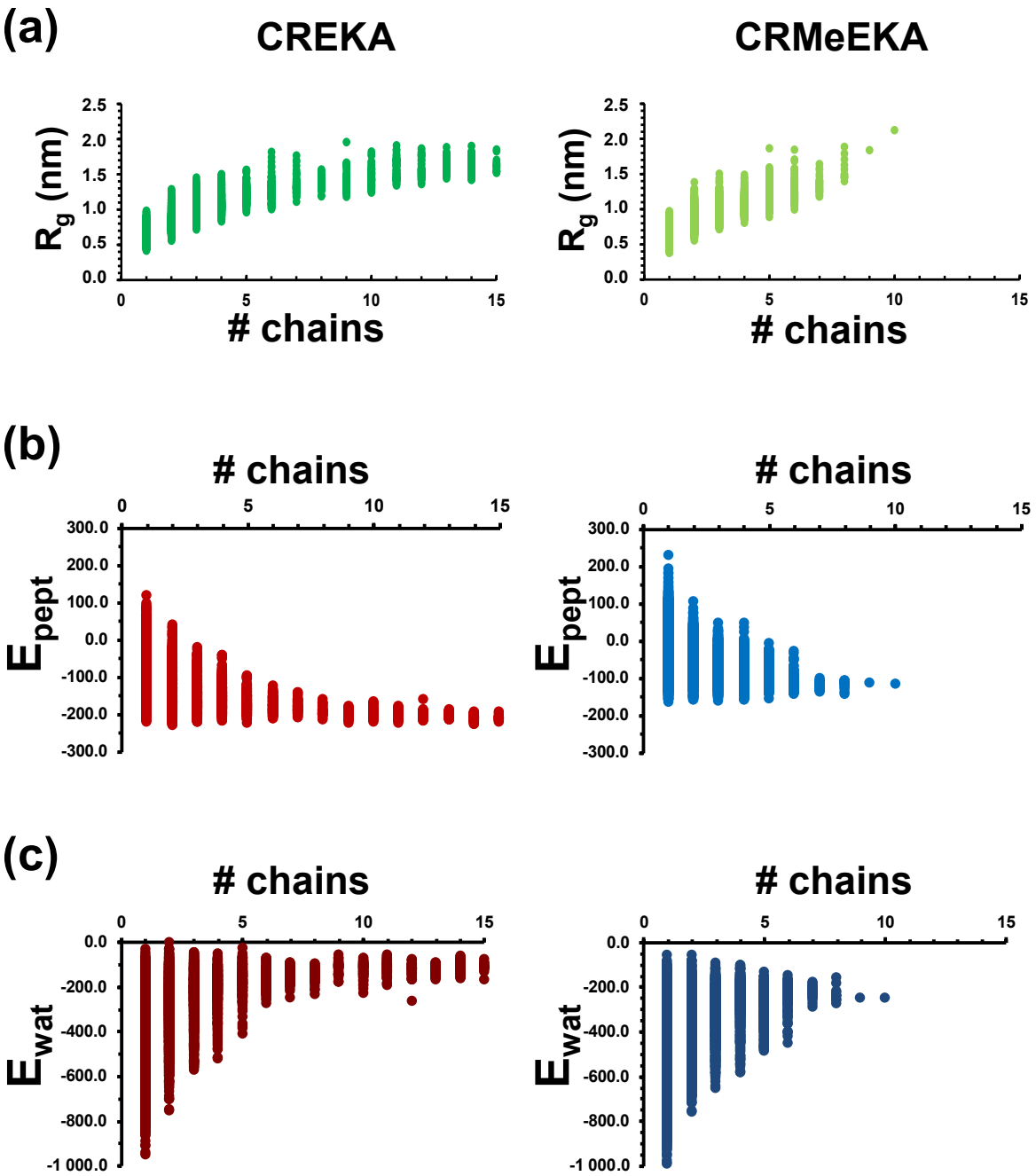


Figure 7

Graphical Abstract