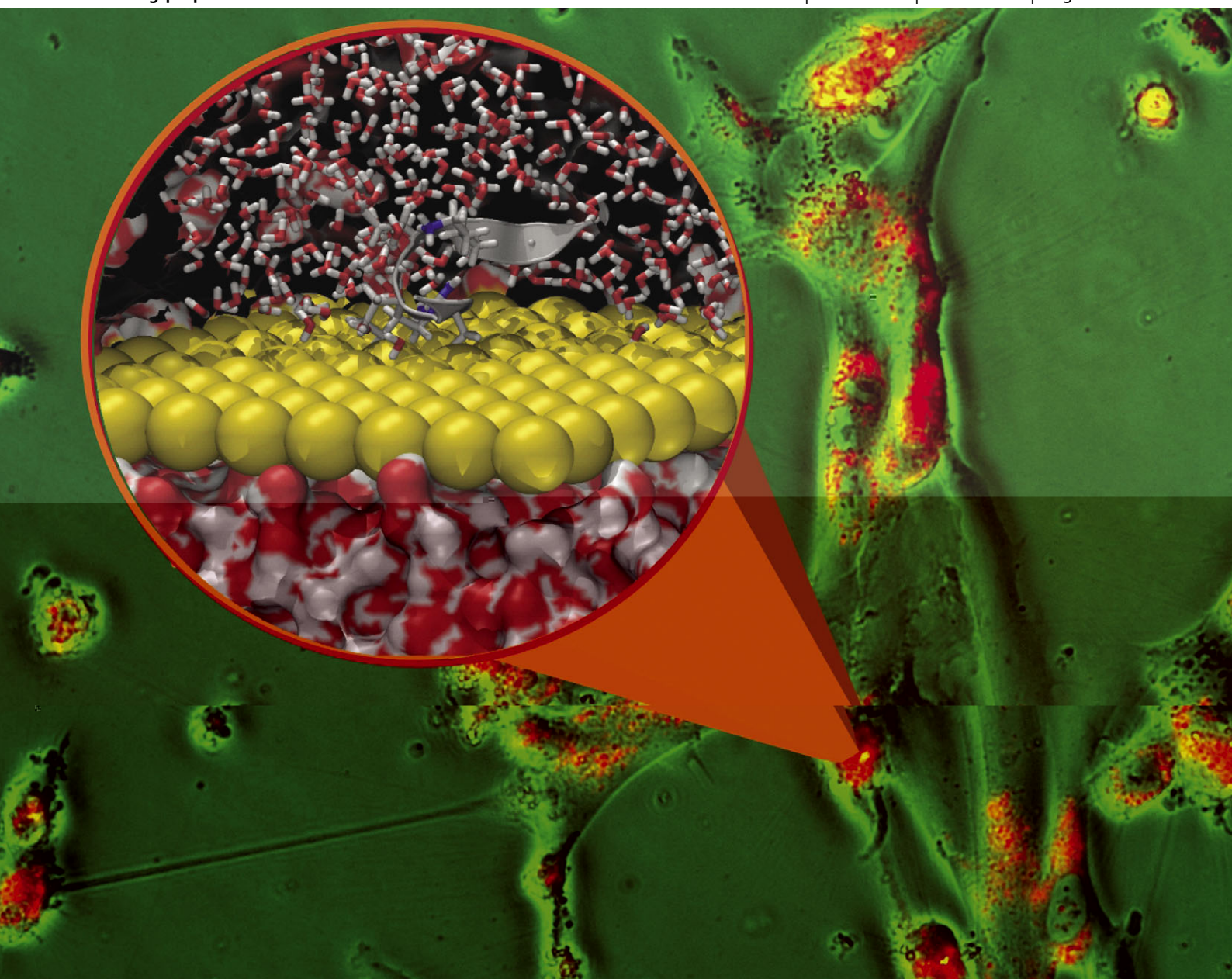
PCCP

Physical Chemistry Chemical Physics

www.rsc.org/pccp

Volume 13 | Number 21 | 7 June 2011 | Pages 9897–10380



Includes a collection of articles on the theme of Nano-bio: The interface between bio-systems and nano-devices

ISSN 1463-9076

COVER ARTICLE

Zanuy *et al.*

Exploring the energy landscape of a molecular engineered analog of a tumor-homing peptide

1463-9076(2011)13:21;1-Z

PCCP

Dynamic Article Links

Cite this: *Phys. Chem. Chem. Phys.*, 2011, **13**, 9986–9994

www.rsc.org/pccp

PAPER

Exploring the energy landscape of a molecular engineered analog of a tumor-homing peptide

Guillem Revilla-López,[a] Juan Torras,*[b] Ruth Nussinov,[cd] Carlos Alemán[ae] and David Zanuy*[a]

Received 17th November 2010, Accepted 5th January 2011
DOI: 10.1039/c0cp02572k

Recently a new non-coded amino acid was designed as a replacement for Arg, to protect the tumor-homing pentapeptide CREKA (Cys-Arg-Glu-Lys-Ala) from proteases. This constrained Arg analog, denoted c_5Arg, was engineered to also promote the stability of the CREKA bioactive conformation. The conformational profile of the CREKA analog obtained by replacing Arg by c_5Arg has been extensively investigated in this work. Two molecular dynamics simulations-based strategies have been employed: a modified simulated annealing and replica exchange. Results obtained using both techniques show that the conformational features of the new analog fulfill the purpose of its design. The new CREKA analog not only preserves the main structural attributes found for the bioactive conformation of the parent peptide but also shows lower flexibility. Moreover, the conformational profile of the mutated peptide narrows towards the most stable structures previously observed for the parent CREKA peptide.

Introduction

The systematic use of nanostructures for clinical applications is rapidly becoming a central milestone in modern nanomedicine. Among the possible nano-approaches, treatment systems based on nanoparticles able to target tumor cells rapidly took the lead.[1] Even though the first generation of nanoparticles relied on vessel "leakiness" for preferential accumulation in tumors—which limited the efficiency of such particles because of their low penetration capability into the targeted tumors—they were trapped by the sustained high pressure of the interstitial tumor fluid.[2]

An alternative approach is targeting the nanoparticles to specific molecular receptors in blood vessels: tumors express many molecules that are not significantly populated in normal tissues and the receptors are available for direct binding of species in the blood stream.[3–5] Different nanosystems have been proposed, such as the super-paramagnetic iron oxide (SPIO) nanoparticles coated with dextran, by Ruoslahti and co-workers. These nanoparticles are able to home to tumors while amplifying their homing activity through a mechanism that resembles formation of blood clots by platelets.[6] The bioactive part of these nanoparticles is the coating peptide, which can recognize selectively clotted plasma proteins.[6] This short linear peptide (CREKA, Cys-Arg-Glu-Lys-Ala) was recently discovered by *in vivo* screening of phage-display peptide libraries[3,7] for tumor-homing peptides in tumor-bearing MMTV-PyMT transgenic breast cancer mice.[8] Synthesized CREKA labeled with the fluorescent dye 5(6)-carboxyfluorescein (FMA) was detectable in human tumors from minutes to hours after intravenous injection, whereas it was essentially undetectable in normal tissues.[9]

For the last three years we have devoted our efforts to devise protection strategies of CREKA against proteolytic cleavage with concomitant stabilization of the bioactive folded conformation.[10–12] Following determination of the bioactive conformation, specific replacements were designed using non-coded amino acids. This strategy is expected to grant proteolytic protection and, simultaneously, allow control of the conformational preferences. Exploration of CREKA conformational preferences was performed under conditions in which the homing peptide was experimentally shown to bind to the clotted plasma proteins: as a free peptide, attached to a nanoparticle, and inserted into a viral capsid protein (in a phage display library).[10] Independently of the simulation conditions, the most favored organizations often featured turn motifs, which generally involved Cys and either Glu or Lys. Among this variety of turn organizations, a common pattern was observed in all simulated cases: a β-turn arrangement that

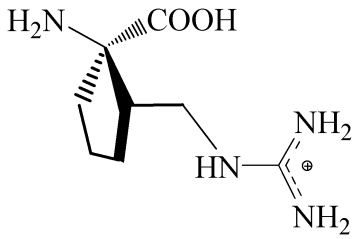
[a] *Departament d'Enginyeria Química, E. T. S. d'Enginyeria Industrial de Barcelona, Universitat Politècnica de Catalunya, Diagonal 647, Barcelona E-08028, Spain. E-mail: david.zanuy@upc.edu*
[b] *Departament d'Enginyeria Química, EEI, Universitat Politècnica de Catalunya, Pça Rei 15, Igualada 08700, Spain. E-mail: joan.torras@upc.edu*
[c] *Basic Science Program, SAIC-Frederick, Inc. Center for Cancer Research Nanobiology Program, NCI, Frederick, MD 21702, USA*
[d] *Department of Human Genetics Sackler, Medical School, Tel Aviv University, Tel Aviv 69978, Israel*
[e] *Center for Research in Nano-Engineering, Universitat Politècnica de Catalunya, Campus Sud, Edifici C', C/Pasqual i Vila s/n, Barcelona E-08028, Spain*

involves the formation of a hydrogen bond between Cys and Lys residues. This organization was identified as the bioactive motif, which enabled the formation of salt bridges between the side-chains of all ionized amino acids, since the curvature site made them protrude outward. Once the bioactive structural profile was delineated, non-coded amino acids were used in order to enhance the protection against proteolytic cleavage and the homing activity.[13]

A key question relates to the final choice of the non-coded amino acid in the substituted position which should satisfy the bioactive conformational requirements. Examination of the conformation adopted by Arg in the bioactive organization shows that it is located at position $i + 1$ of a β-turn.[10] If one searches the conformational preferences of small peptides and proteins, this position is usually occupied by Pro.[14–17] Hence, initially, a new amino acid with a Pro scaffold and Arg side chain, named ($^{\gamma}$Pro)Arg, was modeled.[12] Subsequent work[12] showed that inclusion of this residue significantly reduced the conformational flexibility of the peptide, enhancing its tendency to accommodate partially folded conformations centered at the Arg site. In addition, inspection of the lowest energy conformation indicated that the ionized side chains faced the same side of the molecule, a feature that seems to have a role in the peptide bioactivity. However, insertion of this Arg surrogate presented drawbacks: despite the similarities described above, the turn conformations adopted when ($^{\gamma}$Pro)Arg replaced Arg did not match the most favored conformations previously described,[10] since it shifted the conformational preferences of the CREKA analog towards γ-turn formation. Therefore, the search for appropriate amino acid backbones continued.

Within that context, a specific family of non-coded amino acids, the 1-aminocycloalkane-1-carboxylic acids[18,19] (known in the abbreviated form as Ac_nc, with n referring to the ring size), have been recently proposed as a structural tool for conformational control and enhancement of naturally occurring structural motifs.[19–23] This series of residues exhibits a restricted conformational space characterized by a high propensity to adopt φ, ψ backbone angles typical of the 3_{10}-/α-helix (with some distortion in the case of Ac_3c).[23–25] As we previously mentioned, in the majority of the most favored CREKA conformations Arg is at position $i + 1$ of a β-turn motif.[10] In such arrangements, the backbone dihedral angles adopted by the residue are in the regions corresponding to the aforementioned helical structures. Hence, in order to take advantage of Ac_nc conformational preferences, a new amino acid was designed by attaching the Arg functionalized side-chain to the cycloalkane moiety in Ac_5c, the resulting compound being denoted c_5Arg (see Scheme 1).[26] c_5Arg is a substituted Ac_5c-like derivative of *nor*-arginine, where *nor* refers to a chain-length reduction of one carbon atom with respect to the Arg side-chain.[26] A complete conformational characterization showed that the new residue preserved the conformational preferences corresponding to those expected for its Ac_5c skeleton,[25] which makes it suitable for tests as a CREKA homing peptide-enhancing modification.

This work reports the conformational preferences of a new CREKA analog, hereafter denoted Cc_5*R*EKA, obtained by substituting Arg by the *cis* enantiomer of c_5Arg, where *cis*



Scheme 1

refers to the relative position of the guanidinium group in the cycloalkyl segment. Specifically, results obtained in previous quantum mechanical calculations on both the *cis* and *trans* isomers of c_5Arg suggest that, *a priori*, the former is the best to fulfill the structural requirements of the peptide.[26]

The energy landscape of the new peptide analog has been explored by techniques based on molecular dynamics (MD) simulations, *i.e.*, modified simulated annealing MD (SA-MD)[27] and replica exchange MD (REMD).[28] The conformational features identified in Cc_5*R*EKA are compared with those previously reported of CREKA[10] and the advantages of using c_5Arg as Arg replacement are inferred from such a comparison.

Methods

Conformational search and force-field calculations

The conformational preferences of Cc_5*R*EKA were explored using both modified SA-MD and REMD sampling strategies. The first methodology, which was previously used to study the conformational preferences of the parent peptide,[10,11] is based on the minimization of conformations generated at the initial and intermediate states of several SA-MD cycles. In the SA-MD technique the high starting temperature is gradually reduced during the simulation, allowing the system to surmount energy barriers. In spite of this, in practice, it is known that such a sampling technique does not always lead the system to the most stable region at the end of the simulation. However, recent studies showed that this limitation can be overcome, since very low energy structures can be reached by minimizing structures generated during the SA-MD cycles.[27] Accordingly, such modification of this sampling technique was found to be robust enough to obtain conformations close in energy to the global minimum but located in different regions of the potential energy hypersurface, when the extracted conformations from each SA-MD cycle are minimized.[10–12]

The same conformational profile was studied using REMD.[28] This MD based technique speeds convergence relative to brute force conventional MD.[27] The method is based on the generation of a number of copies (“replicas”) of the system that span from the temperature of interest (*e.g.*, physiological temperature) to heated states, which facilitates overcoming the free energy barriers. Periodic swaps of neighboring replicas, which are performed while preserving an overall Boltzmann-weighted ensemble at each temperature (Monte Carlo based criterion of swap acceptance), enable conformations to heat up and cool down. Nonetheless, recent work has indicated that for short peptides the optimum

convergence of the sampling can be obstructed by a combination of two factors:[29] the intrinsically high flexibility of these systems and the limited thermal agitation associated with the reference temperature. In this work we show that under normal conditions flexible systems may not reach some conformations of medium energy rank, affecting significantly the final bioactive ensemble of conformations.

Computational details

Energy calculation. The conformational energy for all the simulated systems was calculated using the Amber force field.[30] All bonding and non-bonding parameters were extracted from Amber libraries[30] except for those describing the non-coded residue which were previously optimized and tested.[26]

Molecular models. The simulated system consisted of the Cc_5REKA peptide attached to a surface through the sulfhydryl group of the Cys residue. The N- and C-termini of the peptide backbone were capped with acetyl (Ac, MeCO–) and methylamide (–NHMe) groups, respectively. The surface was formed by 100 spherical particles distributed in a 10 □ 10 square (45.90 □ 45.90 Å^2), with van der Waals parameters R = 2.35 Å and ε = 0.90 kcal mol$^{\square 1}$ and no electric charge. This system, which is identical to that considered for CREKA in our previous work,[10] mimics the experimental conditions,[3,7] *i.e.* the peptide linked to the surface of a nanoparticle. Therefore, the results obtained in this work for the CREKA analog have been compared with those reported for natural CREKA attached to an identical surface (system III in ref. 5), unless otherwise indicated. The Cc_5REKA analog attached to the surface was placed in the center of a cubic simulation box (a = 45.90 Å) filled with 2663 explicit water molecules, which were represented using the TIP3 model.[31] Two chloride ions and one sodium ion were added to the simulation box to reach electric neutrality (net charges were considered for Arg, Lys, and Glu at neutral pH).

Simulated annealing

Prior to the production cycles with the modified SA-MD, the system was equilibrated. 0.5 ns of NVT-MD at 500 K were used to homogeneously distribute the solvent and ions in the box. Next, thermal equilibration for 0.5 ns in the constant-NVT ensemble at 298 K, followed by density relaxation for 0.5 ns in the constant-NPT ensemble at 298 K were performed. The last snapshot of the NPT-MD was used as the starting point for the conformational search process. This initial structure was quickly heated to 900 K at a rate of 50 K ps$^{\square 1}$ to force the molecule to jump to a different region of the conformational space. Along 10 ns, the 900 K structure was slowly cooled to 500 K at a rate of 1 K ps$^{\square 1}$. A total of 500 structures were selected and subsequently minimized during the first cycle of modified SA-MD. The resulting minimum energy conformations were stored in a rank-ordered library of low energy structures. The lowest energy structure generated in a modified SA-MD cycle was used as starting conformation of the next cycle.

Atom pair distance cut-offs were applied at 14.0 Å to compute the van der Waals and electrostatic interactions. In order to avoid discontinuities in the potential energy function, non-bonding energy terms were forced to slowly converge to zero, by applying a smoothing factor from a distance of 12.0 Å. Both temperature and pressure were controlled by the weak coupling method, the Berendsen thermobarostat,[32] using a time constant for heat bath coupling and a pressure relaxation time of 1 ps. Bond lengths were constrained using the SHAKE algorithm[33] with a numerical integration step of 2 fs. All MD simulations were performed using the NAMD program.[34]

Replica exchange molecular dynamics

Simulations were performed using the AMBER 10 program.[35] In order to use the low complexity of the system to ensure a reasonable replica swaps ratio of acceptance (approaching a Partial Replica Exchange Molecular Dynamics model), 8 replicas were exponentially distributed in the temperature range from 283.8 K to 418.7 K.[36] Exchanges were attempted every 40 ps between all neighboring replicas with an average acceptance rate of 16%, above the minimal acceptance rate for the complexity of the studied peptide.[36] The REMD trajectories resulted in a cumulative simulation time of 41 ns. Between replica exchanges, the system was evolved using NVT Langevin MD[37] with a damping coefficient of γ = 2.5 ps$^{\square 1}$ and an integration step of 2 fs. The replicas were previously equilibrated by a set of short runs (isothermal and isobaric equilibration), and completed with a final NVT run of 0.5 ns to ensure that each replica reached the target temperature. In all cases the surface particles were fixed at the initial positions and only peptide, water molecules, and ions were allowed to move. Atom pair distance cut-offs were applied at 14.0 Å to compute the van der Waals and electrostatic interactions. Bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm[33] and explicit water molecules of the TIP3 model[31] were used in every replica except for evaluating the replicas energies before the swapping process. In order to increase the global efficiency of the technique, a hybrid solvent model was used to evaluate each replica total energy.[38] Specifically, 118 explicit water molecules were taken into account as a first hydration shell while the rest of solvent contribution was computed using Generalized Born Model implemented by Hawkins and coworkers,[39] using the late parametrization of Tsui and Case.[40] The REMD data were collected from the last 10 ns of simulation at the targeted temperature of 300 K to perform clustering analysis.

Conformation classification and clustering analysis

In order to construct a list of unique minimum energy conformations, each set of structures provided by either modified SA-MD or REMD were compared among them. The list was organized by rank ordering all the unique minimum energy conformations found following an increasing order of energy. In the case of SA-MD, previously listed conformations that appear at a new cycle were discarded. The criterion to identify unique minimum energy conformations was already developed in previous work, and it is based on defining virtual dihedral angles combined with the computation of the interaction pattern, *i.e.* identification of salt bridges, hydrogen bonds and dipole–dipole interactions.[10]

Five virtual dihedral angles were defined considering the α-carbon atoms of the five residues, the methyl carbon atom of the acetyl and *N*-methylamine capping groups, and one acetyl hydrogen atom. The existence of different interactions is accepted on the basis of the following geometric criteria: (a) salt bridges: distance between the centers of the interacting groups shorter than 4.50 Å; (b) hydrogen bonds: H⋯O distance ($d_{H\cdots O}$) shorter than 2.50 Å and >N–H⋯O angle higher than 120.0°; and (c) dipole–dipole: distance between dipoles shorter than 3.00 Å and the interaction has not been counted as a hydrogen bond. Two structures were considered different if they differ in at least one of their virtual dihedral angles by more than 60° or in at least one of the interactions counted. All the structures categorized as different were subsequently clustered according to a criterion based on the presence of the intramolecular interactions mentioned above.

Results and discussion

Clustering analysis

Following the previously developed analysis strategy, we classified the microstructures generated by both SA-MD and REMD following a hierarchical clustering.[10] In this process, each structure is compared with the rest of the structures obtained using the same strategy and included in a cluster that presents the same main-chain conformation and the same polar interaction types. This analysis provides important information about the population of conformations generated for the studied system and, additionally, it serves as a simple and efficient criterion of convergence for SA-MD (*i.e.* when no new conformations are obtained in the annealing cycles the search is concluded). In Fig. 1 the convergence of the exploration clearly shows that after 9 cycles the number of new conformations does not increase. Moreover, the number of different structures produced for the peptide analog, 1125, is close to the number of different structures described for the wild type peptide, 1306, after an equivalent number of SA-MD cycles.[10] However, the significant reduction of the number of different minima, 14%, seems to indicate that the replacement has succeeded, at least in terms of reduction of the conformational flexibility. On the other hand, much faster convergence was reached by REMD considering the same criterion (*i.e.* the number of newly explored conformations), a plateau being obtained after only 8 ns of cumulative simulation time, and the maximum number of explored conformations was found to be only 550. Below, we further show that this relatively small number of different structures may be enough to represent the most relevant conformational features of the bioactive organization for the Cc_5REKA peptide.

Although the main goal of the current study was assessing the suitability of c_5Arg as a homing enhancer, comparison between two different conformational search strategies for short peptides is necessary. Despite the general perception that REMD is the most suitable strategy to explore the conformational space of any given molecule, recent works have pointed out the flaws of the method when dealing with highly flexible peptides.[41] These limitations were attributed to the inherent efficiency of REMD for locating deep basins. Thus, when the energy profile that separates the lowest energy structure and other low energy arrangements is not sufficiently rough, the exploration can be confined to those structures of lower conformation energy. In this study, the rapid convergence is apparently reached by biasing the conformational ensemble towards those structures that presented lower potential energy.

Fig. 2, which depicts the energy distribution of all the generated structures, shows a Gaussian and a bimodal function for the SA-MD and REMD methods, respectively. These profiles indicate that for REMD the lower energy basins are overpopulated while zones that might represent intermediate potential energy ranges are not explored. Despite the apparent independent evolution of each trajectory, these results show that replicas close to the target temperature may present some degree of structural correlation. Thus, in cases of very flexible peptides longer cumulative time should be used to complete a proper exploration of the conformation space (*i.e.*, higher computational effort). Taking into account that only the last 10 ns of simulation were collected to obtain statistical information about the conformational distribution at the given temperature, the computational resources used by REMD for obtaining 40 ns of cumulative simulated time were significantly larger than those committed for reaching the search convergence using SA-MD. It is worth noting that in

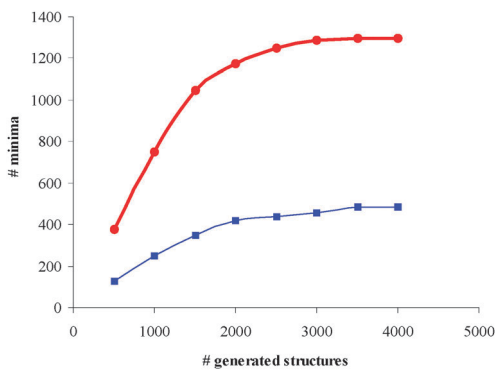


Fig. 1 Variation of the number of unique minimum energy conformations found for Cc_5REKA against the number of conformations generated using the SA-MD (red line, solid circles) and REMD (blue line, solid squares) exploration methods.

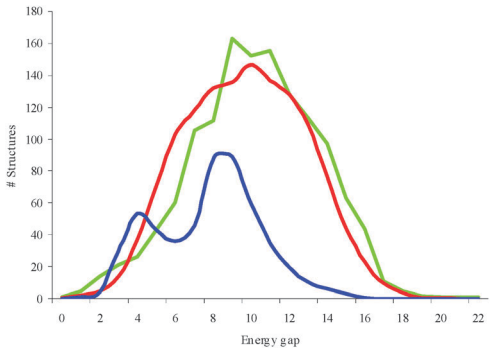


Fig. 2 Distribution of energies for the unique minimum energy conformations found for Cc_5REKA using SA-MD (red line) and REMD (blue line). The profile previously obtained for CREKA (ref. 10) using SA-MD (green line) is displayed for comparison. Energies are relative to the corresponding global minimum.

our case this methodology has allowed us to obtain an energy distribution closer to a canonical ensemble than the one reached by REMD. The main difference between the two methods is the introduction of the temperature as a variable in the conformational search process, which imposes a heavier thermal hindrance for reaching those conformations that are kinetically less favored. As we show later, the first peak of the bimodal distribution derived from the REMD results corresponds to those conformations that are similar to the wild type bioactive motif, while the second peak corresponds to organizations that do not resemble those of lower energy. The central void in the population of medium energy structures may denote kinetic traps that preclude the formation of transient structures at 300 K. These structures, which are needed to reach the conformations that belong to the lowest energy rank, are only described by the SA-MD method.

It was previously demonstrated that the bioactive profile of CREKA mainly corresponds to the lowest energy conformations.[12] Because both the brute force-based simulations (SA-MD) and the statistically rigorous exploration (REMD) techniques drove toward the most energetically relevant structures, they should be equally efficient in assessing the conformational preferences of the Cc_5*R*EKA analog. Our REMD results still might suggest the presence of some degree of deficient exploration. These apparent restrictions though can be surpassed by considering a higher number of replicas, which would allow overcoming the REMD limitations when it is used to search the energy landscape of small flexible peptides. Nevertheless, it should be emphasized that this option increases remarkably the computer resources devoted to the conformational search.[41]

Conformational features

A detailed analysis of the conformational exploration performed by both SA-MD and REMD shows interesting features and differences that are inherent to these techniques. Fig. 3, which shows the Ramachandran plots for the five residues of Cc_5*R*EKA, suggests that SA-MD covers wider extent of the main chain conformational space than REMD. However, the former method explores conformations that are accessible, without representing the effect of the temperature agitation (*i.e.* the kinetics accessibility of these conformations at a determined temperature). Thus, most of the conformations obtained by SA-MD might not be reachable at the studied reference temperature. As we have previously indicated, this is not a surprising feature since this technique was designed to speed up the search of the most relevant conformational minima at a specific temperature. Despite this difference, the lower energy structures derived from the two techniques are almost coincident (Fig. 3). In other words, at 300 K the most relevant conformations for the studied pentapeptide analog are well represented with relatively short trajectories and with a limited amount of replicas. It is worth noting that at such a temperature some conformations that are separated by 4 or 5 kcal mol$^{\square 1}$ from the lowest energy structure, which were detected *via* SA-MD, are not found by REMD. These conformations that had a meaningful contribution in the conformational profile of the wild type segment[10–12] require higher thermal agitation to be reached. This question though does not invalidate our bioactive motif of both wild type and the peptide analog, since the physiological conditions are represented by temperatures that feature almost 10 degrees above the studied conditions.

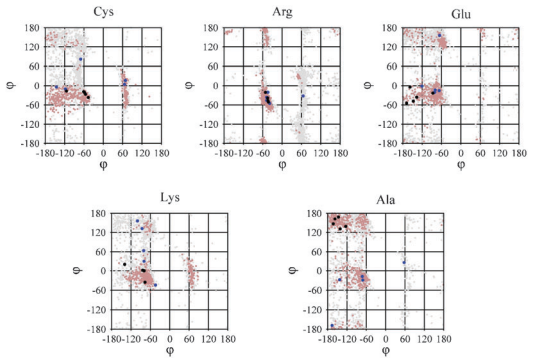


Fig. 3 Ramachandran plot of the main chain dihedral angles (φ, ψ) for the five residues of Cc_5*R*EKA considering the unique minimum energy conformations obtained using the SA-MD and REMD methods. Grey and dark-brown dots represent all the conformations derived from SA-MD and REMD, respectively. The 5 conformations of lower energy obtained by SA-MD and REMD are indicated by blue and black dots, respectively.

Regarding the conformational space accessible by c_5Arg (position 2 in Cc_5*R*EKA), Fig. 3 shows that the non-coded amino acid has achieved its primary goal: it restricts the conformational freedom of the peptide and, simultaneously, favors specific conformation profiles. If comparing the accessible conformations of both CREKA[10–12] and Cc_5*R*EKA peptides, it is noticeable that the latter peptide explores smaller number of different main chain conformations than the former one, with roughly the same amount of generated minima. In other words, the number of backbone positions that are accessible to the peptide when c_5Arg is present is significantly reduced, *i.e.*, the global flexibility of the homing peptide analog is lower than the parent wild type segment.

At the same time, such flexibility reduction is selective because the cyclic nature of the Arg surrogate favors the adoption of the conformation that the coded Arg presents in the bioactive conformation.[10] Independently of the exploration method, the main-chain conformations found for the second residue of the Cc_5*R*EKA analog mainly clustered at the two α regions of the Ramachandran plot, which correspond to the more favored conformations found for a single c_5Arg residue in aqueous solution.[26] In both cases the structures of lower energy (blue and black dots in Fig. 3) present conformations for Cc_5*R*EKA's 2nd position that are located at the α_L region, which corresponds to that adopted by Arg in the parent CREKA peptide. On the other hand, the lower efficiency exhibited by the SA-MD method for locating the more stable conformations widens the range of explored conformations for c_5Arg, with a significant number of Cc_5*R*EKA arrangements with relative energies higher than 5 kcal mol$^{\square 1}$ *versus* the tight conformation distribution observed using REMD. However, in the case of the SA-MD exploration, those Cc_5*R*EKA structures that were within 2–3 kcal mol$^{\square 1}$ from the absolute minimum again showed conformations at

the α_L region, which was the main objective of introducing c_5Arg as Arg surrogate.

Despite the initial success, in order to enhance the homing activity, the biased conformation of c_5Arg in the peptide analog should also favor the formation of turn conformations beyond the residue position, in order to adopt the same conformation described as bioactive. As we mentioned before, we had already tried enhancing the stability of the β-turn featured by the bioactive conformation of CREKA replacing the Arg by another surrogate of low conformational flexibility, ($^{\gamma}$Pro)Arg.[12] In fact, the incorporation of this Pro derivative into CREKA also reduced the conformational flexibility, especially at the position 2. However, it caused the disruption of the desired β-turn conformation, apparently because of the drastic structural changes that the Pro skeleton introduced in the dynamics of the adjacent residues. In the case of c_5Arg, despite fixing the conformational preferences of position 2, the higher flexibility of the Ac_5c skeleton[25] facilitates the adjacent Glu adopting the proper conformation for the continuation of the turn (Fig. 3). Thus, in the minimum energy conformations found for the C*c*$_5$*R*EKA analog, Glu usually falls in the same region that was featured for the parent peptide, making possible the formation of the β-turn.[10–12]

Structural comparison between the lowest energy conformations

Examination of the backbone conformations for the most favored arrangements of C*c*$_5$*R*EKA reveals several features that are closely related with the exploration efficiency and the effect of c_5Arg on the conformational properties of CREKA. Fig. 4 plots the Root Mean Square Difference (RMSD) correlation between the 10 lowest energy structures obtained in each landscape exploration (C*c*$_5$*R*EKA generated by SA-MD and REMD, named *SAn* and *REn*, respectively, where *n* refers to the rank position and number 01 the lowest energy structure) and their structural correlation with the previously determined bioactive conformational pool of the wild type homing peptide[10] (named *WTn* following an analogous nomenclature). As can be seen, the most favored conformations generated by REMD are significantly similar between themselves. The largest main-chain difference between the structures is 1.83 Å, which corresponds to a pair of almost iso-energetic conformations, RE04 and RE05. Again, this feature indicates that the conformational space explored by REMD is essentially restricted to peptide organizations with low energy, which is consistent with the results showed in the previous sections.

On the other hand, SA-MD appears a more efficient tool when exploring the conformation of small peptides since high structural diversity is observed among those structures of lower energy. This feature becomes especially remarkable for the *SA02* conformation, which features a γ-turn that does not correspond to the β-turn motif frequently found in C*c*$_5$*R*EKA. However, differences among the rest of the conformations of C*c*$_5$*R*EKA obtained using both methodologies are relatively small, essentially reflecting small variations in the β-turn motif with respect to the most favored organization. Hence, the use of c_5Arg has achieved another primary goal: its incorporation in CREKA instead of Arg drastically restricts the potential

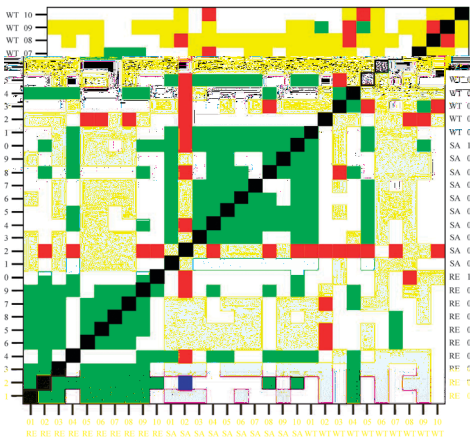


Fig. 4 Backbone structural pair-correlation among the 10 unique conformations of lower energy obtained using SA-MD for CREKA (*WTn*, ref. 10) and C*c*$_5$*R*EKA (*SA*n), and using REMD for C*c*$_5$*R*EKA (*REn*). The correlation was computed as the Root Mean Square Deviation (RMSD) between two conformations. Each color represents a specific RMSD interval of values in which the deviation value of two intersecting structures is situated: green color corresponds to a RMSD lower than 1.3 Å, white color defines a RMSD comprised between 1.3 Å and 2.0 Å, yellow color refers to the interval between 2.0 Å and 3.0 Å and, finally, orange higher than 3.0 Å.

organizations to those that characterized the bioactive structure for the parent pentapeptide (*i.e.* the β-turn).

Thus, Fig. 4 reveals that the majority of backbone conformations are very similar to that described for the bioactive organization of CREKA, with the exception of the aforementioned structure *SA02*. These results indicate that the most accessible arrangements for the C*c*$_5$*R*EKA analog correspond to those describe as bioactive for the CREKA peptide.

Table 1 lists the main electrostatic interactions detected in the four arrangements of lower energy for both the parent peptide and C*c*$_5$*R*EKA. It is worth noting that *WT01*, *WT03* and *WT04*, despite their evident differences, share important features that make them part of the bioactive profile of CREKA. The three conformations are folded into a β-turn and all expose the charged side chain to the solvent on the same side of the peptide chain.[10–12] When comparing their interaction patterns, both present remarkable similarities in the central segment, –REK–. Specifically, in *WT01* (Fig. 5a) and *WT04* the backbones of Lys (N–H) and Cys (C═O) interact through a hydrogen bond, closing a type II β-turn. On the other hand, in *WT03* the Lys backbone (N–H) forms a hydrogen bond with the terminal acetyl blocker (Ac), defining an α-turn that may easily transform into a β-turn if the competing Ala C═O forces tightening the turn (Ac and Ala are mutually blocked by both amide groups forming two interlocked hydrogen bonds). The topological outcome of this complex arrangement resembles *WT01* and places the ionized side-chains in the proper position for the formation of salt bridges. In summary, with the exception of *WT02*, the structures mainly differ in the conformation adopted by both terminal groups.

Analysis of the four lower energy conformations obtained for C*c*$_5$*R*EKA using REMD (Table 1) demonstrates that they all have identifiable counterparts in the structures

Table 1 Comparison among the interaction patterns[a] of the four unique conformations of lower energy obtained for CREKA[b] and Cc_5REKA using SA-MD and REMD[c]

	Cc_5REKA (SA-MD)		Cc_5REKA (REMD)		CREKA (SA-MD)
SA01	HB Lys(N–H)⋯(O═C)Cys[1] HB Ala(N–H)⋯(O═C)Glu[2] HB c_5Arg(C═O)⋯(^+H_3N, side chain)Lys SB Lys⋯Glu	RE01	HB Glu(N–H)⋯(O═C)Ace[1] HB Lys(N–H)⋯($OOC^□$-,side chain)Glu SB Lys⋯Glu	WT01	HB Lys(N–H)⋯(O═C)Cys[1] SB Arg⋯Glu SB Lys⋯Glu
SA02	HB c_5Arg(N–H)⋯(O═C)Ace[2] HB Glu(N–H)⋯(O═C)Cys[2] HB c_5Arg(N–H_2 guanidinium)⋯(O═C)Glu HB Ala(N–H)⋯($OOC^□$–,side chain)Glu HB NMe(N–H)⋯($OOC^□$–, side chain)Glu SB Lys⋯Glu	RE02	HB Glu(N–H)⋯(O═C)Ace[1] HB Lys(N–H)⋯(O═C)Ace[3] HB c_5Arg(N–H_2 guanidinium)⋯(O═C)Ala SB Lys⋯Glu	WT02	HB Ala(N–H)⋯(O═C)Arg[1] HB NMe(N–H)⋯(O═C)Glu[1] SB Arg⋯Glu
SA03	HB Lys(N–H)⋯(O═C)Cys[1] HB c_5Arg(N–H)⋯(O═C)Lys[2] SB Lys⋯Glu	RE03	HB Glu(N–H)⋯(O═C)Ace[1] HB Lys(N–H)⋯(O═C)Ace[3] SB Lys⋯Glu	WT03	HB Lys(N–H)⋯(O═C)Ace[3] HB Ala(N–H)⋯(O═C)Ace[4] HB NMe(N–H)⋯(O═C)Cys[4] SB Arg⋯Glu
SA04	HB Lys(N–H)⋯(O═C)Cys[1] HB c_5Arg(N–H_2 guanidinium)⋯(O═C)Cys HB c_5Arg(N–H_2 guanidinium)⋯(O═C)Lys SB Lys⋯Glu	RE04	HB Ala(N–H)⋯(O═C)Cys[3] SB Lys⋯Glu	WT04	HB Lys(N–H)⋯(O═C)Cys[1] HB Ala(N–H)⋯(O═C)Cys[3] HB NMe(N–H)⋯(O═C)Cys[4] SB Arg⋯Glu SB Lys⋯Glu

[a] Interaction pattern: hydrogen bonds and salt bridges are labeled as HB and SB, respectively. [b] Four minima generated according to ref. 10. [c] Classification of tight turns as a function of the distance (in number of residues) between the hydrogen bonding donor and acceptor: [1]β-turn (*i*, *i* + 3), [2]γ-turn(*i*, *i* + 2), [3]α-turn (*i*, *i* + 4) and [4]π-turn, (*i*, *i* + 5).

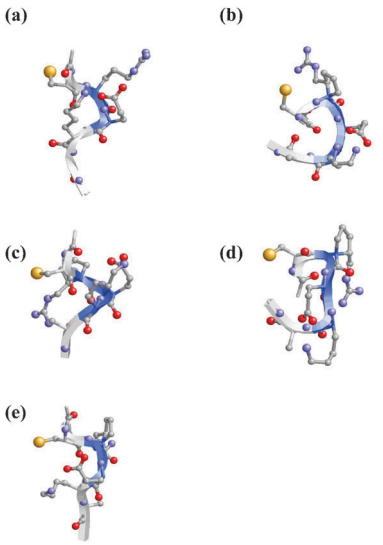


Fig. 5 Detailed representation of selected low energy conformations obtained for CREKA and Cc_5REKA. The main chain is represented by solid cartoons, with standard protein conformational colors (blue and white for turn motifs). In the side chains heavy atoms are represented with solid scaled balls and their bonds with solid sticks following the CPK colors convention. Hydrogen atoms of the side groups are omitted for clarity. The depicted structures are: (a) *WT01* (extracted from ref. 10); (b) *RE01*; (c) *RE04*; (d) *SA02*; and (e) *SA03* (the meaning of the codes is described in the text).

characterized for the parent peptide, with the exception of structure *RE01* (Fig. 5b). The latter conformation features a single β-turn at the C terminus, which involves the N–H of Glu and the C═O of the acetyl blocking group that was not observed in any CREKA conformation. This organization is partially due to the interaction between the guanidinium group of c_5Arg and the C═O of Ala, which precludes the formation of other hydrogen bonding patterns more similar to those observed in CREKA. *R02* and *R03* define arrangements intermediate between the α- and β-turns, leading to turns at the C termini with geometries slightly wider than that typically expected for a conventional β-turn. However, a simple rearrangement of the N terminal group would easily facilitate

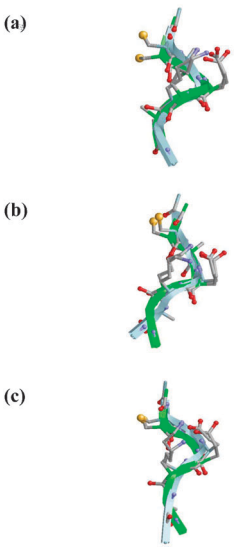


Fig. 6 Superimposition of selected structures of Cc_5REKA with the lowest energy conformation of CREKA (denoted *WT01* in the text; ref. 10). In all cases main chains are represented by solid cartoons, green and cyan colors depicting the *WT01* and each particular Cc_5REKA structure, respectively. In the side chains heavy atoms are represented with solid scaled balls and their bonds with solid sticks following the CPK colors convention. Hydrogen atoms of the side groups were removed for clarity. The superimposed structures are: (a) *RE04*; (b) *SA01*; and (c) *SA03* (the meaning of the codes is described in the text).

the formation of the proper turn since in those two conformations the hydrogen bonding acceptor is the C═O of the terminal Ac, whereas in the bioactive conformation of CREKA it is the C═O of Cys. Finally, *R04* features an α-turn that presents a similar topology to that observed in the parent peptide (Fig. 5c). Thus, c_5Arg favors the formation of a turn that makes possible the correct spatial location of the Glu and Lys ionized side chains allowing their interaction (Fig. 6a).

Results obtained with SA-MD show a better match with the expected arrangements for Cc_5REKA. This search strategy allows locating low energy structures that adopt the β-turn featured by *WT01* and *WT04*. Thus, the N-termini organization adopted by *SA01*, *SA03* and *SA04* (Fig. 6b and c, respectively) correspond to that found for *WT01*, even though some differences are observed at the C terminus (Fig. 5e). Finally, *SA02* presents a double γ-turn conformation (Fig. 5d), which was also predicted for other Ac$_5$c-containing peptides.[21,23]

Conclusions

The conformational preferences of a homing peptide synthetic analog have been widely explored by different techniques based on MD simulations. Independent of the strategy, our results have indicated that the recently engineered residue, c_5Arg,[26] does accomplish its design aim, which is to reduce the conformational freedom of the CREKA peptide,[10] biasing the accessible conformations towards those that feature the bioactive one. The lower energy conformations of Cc_5REKA involve the formation of β-turn motifs centered around the second and third residues of the synthetic analog, similar to those observed in the parent peptide.[10] At the same time, this conformational bias aids the formation of an interaction network between the side-chains of the central residues, which was described as a necessary feature for the bioactivity of CREKA (*i.e.*, the orientation of the ionized central side chains towards the same side of the turn allows their interaction with the tumor vessel receptors). In this context, it is worth noting that even in those conformations that were not described by the parent peptide, such as *RE01* and *SA02*, the final topological distribution of the side chains guarantees the formation of the required interaction pattern.

On the other hand, differences between the two methodologies demonstrate that the temperature is an important variable in the Cc_5REKA conformational preferences. Even though the energy differences among the ten absolute lowest energy conformations are less than 5 kcal mol$^{\square 1}$ some of them are not detected at the REMD reference temperature. Thus, the lower thermal agitation featured under normal conditions precludes the complete coverage of the bioactive conformational profile. In contrast, the SA-MD, which is a technique that explores the most relevant regions of the energy landscape without help from higher temperatures, allows obtaining conformations of Cc_5REKA that were previously reported as relevant for CREKA's bioactivity. Both the lack of thermal distribution in the generated ensembles and the redundant production of conformations starting from those that are already located in low energy regions facilitated the identification of quasi-degenerated energy arrangements that were not detected at the REMD targeted temperature.

In summary, results derived from the exploration of the conformational bioactive profile of Cc_5REKA demonstrate that inclusion of the engineered residue c_5Arg as Arg replacement does achieve the objectives of its design: not only could the new surrogate protect the homing peptide from the protease activity but also increase the stability of the bioactive conformation previously determined for the wild type peptide CREKA. These results confirm our strategy[12,13,19–23,26] that is aimed to re-design natural amino acids to improve properties of biotechnological interest in natural peptides.

Acknowledgements

Computer resources were generously provided by the Centre de Supercomputació de Catalunya (CESCA), the National Cancer Institute for partial allocation of computing time and staff support at the Advanced Biomedical Computing Center of the Frederick Cancer Research and Development Center and the high-performance computational capabilities of the Biowulf PC/Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov). Financial support from Generalitat de Catalunya (research group 2009 SGR 925; XRQTC; ICREA Academia prize for excellence in research to C.A.) is gratefully acknowledged. This project has been funded in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number HHSN261200800001E. The content of this publication does not necessarily reflect the view of the policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. Government. This research was supported [in part] by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

Notes and references

1 R. Weissleder, A. Bogdanov Jr, E. A. Neuwelt and M. Papisov, Long-Circulating Iron-Oxides for MR-Imaging, *Adv. Drug Delivery Rev.*, 1995, **16**, 321–334.
2 N. Desai, V. Trieu, Z. W. Yao, L. Louie, S. Ci, A. Yang, C. L. Tao, T. De, B. Beals, D. Dykes, P. Noker, R. Yao, E. Labao, M. Hawkins and P. Soon-Shiong, Increased antitumor activity, intratumor paclitaxel concentrations, and endothelial cell transport of Cremophor-free, albumin-bound paclitaxel, ABI-007, compared with Cremophor-based paclitaxel, *Clin. Cancer Res.*, 2006, **12**, 1317–1324.
3 J. A. Hoffman, E. Giraudo, M. Singh, L. Zhang, M. Inoue, K. Porkka, D. Hanahan and E. Ruoslahti, Progressive vascular changes in a transgenic mouse model of squamous cell carcinoma, *Cancer Cell*, 2003, **4**, 383–391.
4 P. Oh, Y. Li, J. Yu, E. Durr, K. M. Krasinska, L. A. Carver, J. E. Testa and J. E. Schnitzer, Subtractive proteomic mapping of the endothelial surface in lung and solid tumors for tissue-specific therapy, *Nature*, 2004, **429**, 629–635.
5 E. Ruoslahti, Specialization of tumor vasculature, *Nat. Rev. Cancer*, 2002, **2**, 83–90.
6 D. Simberg, T. Duza, J. H. Park, M. Essler, J. Pilch, L. Zhang, A. M. Derfus, M. Yang, R. M. Hoffman, S. Bhatia, M. J. Sailor and E. Ruoslahti, Biomimetic amplification of nanoparticle homing to tumors, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 932–936.
7 R. Pasqualini and E. Ruoslahti, Organ targeting *in vivo* using phage display peptide libraries, *Nature*, 1996, **380**, 364.
8 J. N. Hutchinson and W. J. Muller, Transgenic mouse models of human breast cancer, *Oncogene*, 2000, **19**, 6130–6137.

9 P. P. Karmali, V. R. Kotamraju, M. Kastantin, M. Black, D. Missirlis, M. Tirrell and E. Ruoslahti, Targeting of albumin-embedded paclitaxel nanoparticles to tumors, *Nanomed.: Nanotechnol., Biol. Med.*, 2009, **5**, 73–82.
10 D. Zanuy, A. Flores-Ortega, J. Casanovas, D. Curco, R. Nussinov and C. Aleman, The Energy Landscape of a Selective Tumor-Homing Pentapeptide, *J. Phys. Chem. B*, 2008, **112**, 8692–8700.
11 D. Zanuy, D. Curcó, R. Nussinov and C. Alemán, Influence of the dye presence on the conformational preferences of CREKA, a tumor homing linear pentapeptide, *Pept. Sci.*, 2009, **92**, 83–93.
12 D. Zanuy, A. Flores-Ortega, A. I. Jimenez, M. I. Calaza, C. Cativiela, R. Nussinov, E. Ruoslahti and C. Aleman, *In Silico* Molecular Engineering for a Targeted Replacement in a Tumor-Homing Peptide, *J. Phys. Chem. B*, 2009, **113**, 7879–7889.
13 L. Agemy, K. N. Sugahara, V. R. Kotamraju, K. Gujraty, O. M. Girard, Y. Kono, R. F. Mattrey, J. H. Park, M. Sailor, A. I. Jimenez, C. Cativiela, D. Zanuy, F. J. Sayago, C. Aleman, R. Nussinov and E. Ruoslahti, Nanoparticle-induced vascular blockade in human prostate cancer, *Blood*, 2010, **116**, 2847–2856.
14 P. Chakrabarti and D. Pal, Interrelationships of side-chain and main-chain conformations in proteins, *Prog. Biophys. Mol. Biol.*, 2001, **76**, 1–102.
15 M. Marraud and A. Aubry, Crystal structures of peptides and modified peptides, *Biopolymers*, 1996, **40**, 45–83.
16 M. W. MacArthur and J. M. Thornton, Influence of Proline residues on protein conformation, *J. Mol. Biol.*, 1991, **218**, 397–412.
17 J. Venkatraman, S. C. Shankaramma and P. Balaram, Design of Folded Peptides, *Chem. Rev.*, 2001, **101**, 3131–3152.
18 C. Toniolo, F. Formaggio, B. Kaptein and Q. B. Broxterman, You are sitting on a gold mine!, *Synlett*, 2006, 1295–1310.
19 D. Zanuy, A. I. Jimenez, C. Cativiela, R. Nussinov and C. Aleman, Use of Constrained Synthetic Amino Acids in β-Helix Proteins for Conformational Control, *J. Phys. Chem. B*, 2007, **111**, 3236–3242.
20 J. Zheng, D. Zanuy, N. Haspel, C. J. Tsai, C. Aleman and R. Nussinov, Nanostructure Design Using Protein Building Blocks Enhanced by Conformationally Constrained Synthetic Residues, *Biochemistry*, 2007, **46**, 1205–1218.
21 D. Zanuy, F. Rodriguez-Ropero, N. Haspel, J. Zheng, R. Nussinov and C. Aleman, Stability of Tubular Structures Based on β-Helical Proteins: Self-Assembled *versus* Polymerized Nanoconstructs and Wild-Type *versus* Mutated Sequences, *Biomacromolecules*, 2007, **8**, 3135–3146.
22 F. Rodriguez-Ropero, D. Zanuy, J. Casanovas, R. Nussinov and C. Aleman, Application of 1-Aminocyclohexane Carboxylic Acid to Protein Nanostructure Computer Design, *J. Chem. Inf. Model.*, 2008, **48**, 333–343.
23 D. Zanuy, G. Ballano, A. I. Jimenez, J. Casanovas, N. Haspel, C. Cativiela, D. Curco, R. Nussinov and C. Aleman, Protein Segments with Conformationally Restricted Amino Acids Can Control Supramolecular Organization at the Nanoscale, *J. Chem. Inf. Model.*, 2009, **49**, 1623–1629.
24 C. Alemán, Conformational properties of alpha-amino acids disubstituted at the alpha-carbon, *J. Phys. Chem. B*, 1997, **101**, 5046–5050.
25 C. Alemán, D. Zanuy, J. Casanovas, C. Cativiela and R. Nussinov, Backbone Conformational Preferences and Pseudorotational Ring Puckering of 1-Aminocyclopentane-1-carboxylic Acid, *J. Phys. Chem. B*, 2006, **110**, 21264–21271.
26 G. Revilla-Lopez, J. Torras, A. I. Jimenez, C. Cativiela, R. Nussinov and C. Aleman, Side-Chain to Backbone Interactions Dictate the Conformational Preferences of a Cyclopentane Arginine Analogue, *J. Org. Chem.*, 2009, **74**, 2403–2412.
27 C. Baysal and H. Meirovitch, Efficiency of simulated annealing for peptides with increasing geometrical restrictions, *J. Comput. Chem.*, 1999, **20**, 1659–1670.
28 Y. Sugita and Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, *Chem. Phys. Lett.*, 1999, **314**, 141–151.
29 E. Lin and M. S. Shell, Convergence and Heterogeneity in Peptide Folding with Replica Exchange Molecular Dynamics, *J. Chem. Theory Comput.*, 2009, **5**, 2062–2073.
30 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J. Am. Chem. Soc.*, 1995, **117**, 5179.
31 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of Simple Potential Functions for Simulating Liquid Water, *J. Chem. Phys.*, 1983, **79**, 926–935.
32 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, Molecular Dynamics with Coupling to an External Bath, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
33 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, Numerical-Integration of Cartesian Equations of Motion of a System with Constraints-Molecular-Dynamics of N-Alkanes, *J. Comput. Phys.*, 1977, **23**, 327–341.
34 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, Scalable Molecular Dynamics with NAMD, *J. Comput. Chem.*, 2005, **26**, 1781–1802.
35 D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin and P. A. Kollman, *AMBER 10*, University of California, San Francisco, 2008.
36 X. Cheng, G. Cui, V. Hornak and C. Simmerling, Modified Replica Exchange Simulation Methods for Local Structure Refinement, *J. Phys. Chem. B*, 2005, **109**, 8220–8230.
37 R. W. Pastor, B. R. Brooks and A. Szabo, An analysis of the accuracy of Langevin and molecular dynamics algorithms, *Mol. Phys.*, 1988, **65**, 1409–1419.
38 A. Okur, L. Wickstrom, M. Layten, R. Geney, K. Song, V. Hornak and C. Simmerling, Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model, *J. Chem. Theory Comput.*, 2006, **2**, 420–433.
39 G. D. Hawkins, C. J. Cramer and D. G. Truhlar, Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium, *J. Phys. Chem.*, 1996, **100**, 19824–19839.
40 V. Tsui and D. A. Case, Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations, *Biopolymers*, 2000, **56**, 275–291.
41 E. Lin and M. S. Shell, Convergence and Heterogeneity in Peptide Folding with Replica Exchange Molecular Dynamics, *J. Chem. Theory Comput.*, 2009, **5**, 2062–2073.

