Published in final edited form as:

J Phys Chem B. 2009 June 4; 113(22): 7879–7889. doi:10.1021/jp9006119.

***In silico* molecular engineering for a targeted replacement in a tumor-homing peptide**

David Zanuy^{1,§}, Alejandra Flores-Ortega^{1,§}, Ana I. Jiménez^{2,*}, M. Isabel Calaza², Carlos Cativiela², Ruth Nussinov^{3,4}, Erkki Ruoslahti^{5,6}, and Carlos Alemán^{1,7,*}

¹Departament d'Enginyeria Química, E. T. S. d'Enginyeria Industrial de Barcelona, Universitat Politècnica de Catalunya, Diagonal 647, Barcelona E-08028, Spain

²Departamento de Química Orgánica, Instituto de Ciencia de Materiales de Aragón, Universidad de Zaragoza - CSIC, 50009 Zaragoza, Spain

³Basic Research Program, SAIC-Frederick, Inc. Center for Cancer Research Nanobiology Program, NCI, Frederick, MD 21702, USA

⁴Department of Human Genetics Sackler, Medical School, Tel Aviv University, Tel Aviv 69978, Israel

⁵Vascular Mapping Center, The Burnham Institute for Medical Research at UCSB, Santa Barbara, California 93106, USA

⁶Cancer Research Center, The Burnham Institute for Medical Research, La Jolla, California 92037, USA

⁷Center for Research in Nano-Engineering, Universitat Politècnica de Catalunya, Campus Sud, Edifici C', C/Pasqual i Vila s/n, Barcelona E-08028, Spain

Abstract

A new amino acid has been designed as a replacement for arginine (Arg, R) to protect the tumor-homing pentapeptide CREKA from proteases. This amino acid, denoted (Pro)hArg, is characterized by a proline skeleton bearing a specifically oriented guanidinium side chain. This residue combines the ability of Pro to induce turn-like conformations with the Arg side-chain functionality. The conformational profile of the CREKA analogue incorporating this Arg substitute has been investigated by a combination of simulated annealing and Molecular Dynamics. Comparison of the results with those previously obtained for the natural CREKA shows that (Pro)hArg significantly reduces the conformational flexibility of the peptide. Although some changes are observed in the backbone...backbone and side chain...side chain interactions, the modified peptide exhibits a strong tendency to accommodate turn conformations centered at the (Pro)hArg residue and the overall shape of the molecule in the lowest energy conformations characterized for the natural and the modified peptide exhibit a high degree of similarity. In particular, the turn orients the backbone such that the Arg, Glu and Lys side chains face the same side of the molecule, which is considered essential for bioactivity. These results suggest that replacement of Arg by (Pro)hArg in CREKA may be useful in providing resistance against proteolytic enzymes while retaining conformational features which are essential for tumor-homing activity.

* Corresponding authors: anisjim@unizar.es and carlos.aleman@upc.edu.

§ Both authors contributed equally to this work

Introduction

The linear pentapeptide CREKA (Cys-Arg-Glu-Lys-Ala) was recently discovered¹ by *in vivo* screening of phage-display peptide libraries² for tumor-homing in tumor-bearing MMTV-PyMT (mouse mammary tumor virus - polyomavirus middle-T antigen) transgenic breast cancer mice.³ After intravenous injection of synthetic CREKA, the peptide was detected in human tumors; but not in normal tissues.⁴ *In vivo* experiments revealed that this tumor-homing pentapeptide binds to clotted plasma proteins, thus establishing its behavior as a clot-binding peptide. In addition, experiments carried out with CREKA linked to amino dextran-coated iron oxide (SPIO) nanoparticles indicated that these systems not only bind to blood and plasma clots but also effectively induce further localized tumor clotting, thus amplifying the nanoparticle homing.¹ It has been established that the chemical nature of the nanoparticle is not essential for this activity since both CREKA-SPIO nanoparticles and CREKA-coated liposomes were found to cause clotting in tumor vessels.¹

In a recent study,⁵ we determined the energy landscape and proposed the bioactive conformation of CREKA by using a multiple conformational search strategy based on Molecular Dynamics (MD) simulations. Two experimentally tested environments were considered: (i) the free peptide immersed in either a dilute aqueous solution or an ionized aqueous solution mimicking the physiological medium; (ii) the peptide attached to a nanoparticle through the sulfhydryl group of cysteine. In addition, we considered a molecular system that mimics the CREKA peptide when inserted in a phage display protein, although in this case the search strategy did not converge completely and the conformational space was not explored exhaustively. Results indicated⁵ that the conformational profile of the REKA sequence is very similar in all cases; significant differences were observed only for the Cys residue. Complete clustering analyses based on the number and distribution of intramolecular interactions showed that such coincidence was particularly remarkable for the accessible minima, *i.e.* those with relative energies lower than 2.0 kcal/mol.

Despite the high potential of CREKA in cancer diagnostics and therapeutics, the application of this and other tumor-homing peptide sequences may be hampered by short half-life times since endogenous proteases rapidly digest such peptides. Protection from proteolytic cleavage is therefore crucial for tumor-homing peptide applications, similar to most peptides exhibiting biological activity. Several strategies have been proposed for this purpose, targeted replacements with non-coded amino acids being among the most successful ones.⁶⁻¹³

Within the framework of a project aimed at improving the biological performance and pharmacological profile of CREKA, we have undertaken the *in silico* study of a CREKA analogue incorporating a non-proteinogenic amino acid. This residue has been conceived to retain the most relevant characteristics of the conformational profile of the natural peptide and simultaneously impart stability against proteolytic cleavage. One should bear in mind that any possible analogue able to overcome the limitations of CREKA while retaining its tumor-homing activity should exhibit a conformational profile with low-energy regions close to those characterizing the natural peptide.⁵ Computer simulations based on the application of theoretical methodologies have proven to be powerful techniques not only for sampling the conformational space of biomolecules^{5,14-16} as small peptides, but also for designing, engineering and testing non-proteinogenic amino acids with potential applications in nanobiology.¹⁷⁻¹⁹

In this work, we report the explicit design and subsequent conformational study of a CREKA analogue generated by replacing an amino acid in the sequence by a non-proteinogenic counterpart. The work is divided into three parts. First, the new amino acid is engineered. Then, force-field parameters for this non-coded residue are determined using quantum mechanical

calculations. Finally, the conformational profile of the modified CREKA sequence attached to a nanoparticle is examined using as the sampling technique a computational procedure that combines a modified simulated annealing^{20,21} with molecular dynamics (SA-MD). A detailed comparison between the conformational properties exhibited by the natural and modified CREKA sequences is provided.

Computational Methods

Quantum mechanical calculations

Density Functional Theory (DFT) methods were applied for quantum mechanical calculations, which were carried out using the Gaussian 03 computer program.²² Specifically, calculations were performed by combining the Becke's three-parameter hybrid functional (B3)²³ with the Lee, Yang and Parr (LYP)²⁴ expression for the nonlocal correlation (B3LYP). Accordingly, all the quantum mechanical calculations presented in this work were performed using the unrestricted formalism of the B3LYP method combined with the 6-31+G(d,p) basis set.²⁵ Frequency analyses were carried out to verify the nature of the minimum state of all the stationary points provided by geometry optimizations and to calculate the zero-point vibrational energies (ZPVE) with both thermal and entropic corrections, the latter statistical terms being used to compute the conformational Gibbs free energies in the gas phase (ΔG^{gP}).

To obtain an estimation of the solvation effects on the relative stability of the different minima, single point calculations were also conducted using a Self-Consistent Reaction Field (SCRF) model. SCRF methods treat the solute at the quantum mechanical level, while the solvent is represented as a dielectric continuum. Specifically, the Polarizable Continuum Model (PCM) developed by Tomasi and coworkers was used to describe the bulk solvent.²⁶ This method involves the generation of a solvent cavity from spheres centered at each atom in the molecule and the calculation of virtual point charges on the cavity surface representing the polarization of the solvent. The magnitude of these charges is proportional to the derivative of the solute electrostatic potential at each point calculated from the molecular wave function. The point charges may then be included in the one-electron Hamiltonian, thus inducing polarization of the solute. An iterative calculation is carried out until the wave function and the surface charges are self-consistent. PCM calculations were performed using the standard protocol and considering the dielectric constants of carbon tetrachloride ($\epsilon = 2.228$), chloroform ($\epsilon = 4.9$) and water ($\epsilon = 78.4$). The conformational free energies in solution (ΔG^{sol} , where *sol* refers to the solvent) were computed using the classical thermodynamics scheme, that is, the free energies of solvation provided by the PCM model were added to the ΔG^{gP} values.

Conformational search and force-field calculations

The conformational preferences of the CREKA analogue studied in this work were explored using the sampling strategy previously employed for the natural peptide.⁵ This procedure is based on the minimization of structures generated at the initial and intermediate states of several SA-MD cycles. The starting temperature is gradually reduced during the simulation, thus allowing the system to surmount energy barriers. In spite of this, in practice, this sampling technique does not always lead the system to the most stable region at the end of the simulation. However, recent studies showed that very low energies are obtained by minimizing the structures extracted from SA-MD processes.^{27,28} According to our previous study,⁵ this sampling technique, which was found to be robust enough to obtain low-energy structures that may be quasi-degenerate with the global minimum but situated in different valleys of the peptide landscape, was applied by submitting 500 structures to energy minimization from each SA-MD cycle.

The simulated system consisted of the CREKA analogue 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 (47.5 Å²), with van der Waals parameters $R = 2.35$ Å and $\epsilon = 0.90$ kcal/mol and no electric charge. This system is identical to that considered for natural CREKA in our previous work⁵ and mimics the experimental conditions^{1,4} in which the peptide is bound to the surface of a nanoparticle. Therefore, the results obtained in this work for the CREKA analogue are compared with those reported for natural CREKA attached to an identical surface (system III in ref.⁵), unless otherwise indicated.

The CREKA analogue attached to the surface was placed in the center of a cubic simulation box ($a = 47.5$ Å) filled with 3405 explicit water molecules, which were represented using the TIP3 model.²⁹ Two negatively charged chloride atoms and one positively charged sodium atom were added to the simulation box to reach electric neutrality (net charges were considered for Arg, Lys and Glu at neutral pH). Force-field parameters for natural residues were extrapolated from AMBER libraries,³⁰ while those for the non-coded amino acid proposed in this work were explicitly developed (see below).

Prior to the production cycles with the modified SA-MD, the simulation box was equilibrated. Thus, 0.5 ns of NVT-MD at 500 K were used to homogeneously distribute the solvent and ions in the box. Next, 0.5 ns of NVT-MD at 298 K (thermal equilibration) and 0.5 ns of NPT-MD at 298 K (density relaxation) were run. 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 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 per 25 ps. 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 minimum generated in a modified SA-MD cycle was used as starting conformation of the next cycle.

The energy was calculated using the AMBER potential.³⁰ Atom pair distance cutoffs 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,³¹ using a time constant for heat bath coupling and a pressure relaxation time of 1 ps. Bond lengths were constrained using the *SHAKE* algorithm³² with a numerical integration step of 2 fs. All MD simulations were performed using the NAMD program.³³

Conformation classification and clustering analysis

The list of unique minimum energy conformations was generated by comparing the 500 minimized structures provided by each cycle of modified SA-MD not only among themselves but also with unique minima generated in previous cycles. The list was organized by rank ordering all the unique minimum energy conformations found in an increasing energy, with the previously listed conformations discarded. Unique minimum energy conformations were identified based on the virtual dihedral angles which characterize the peptide backbone conformation and on hydrogen-bond and salt-bridge interactions. Five virtual dihedral angles were defined considering the α -carbon atoms of the five residues, the methyl carbon atom of the acetyl and methylamide capping groups and one acetyl hydrogen atom (Figure 1a). The existence of interactions was accepted on the basis of the following geometric criteria: a) for salt bridges: distance between the centers of the interacting groups shorter than 4.50 Å; b) for hydrogen bonds: H...O distance shorter than 2.50 Å and N-H...O angle higher than 120.0°. Two structures were considered different when differing in at least one of their virtual dihedral

angles by more than 60° or in at least one of the above interactions. All the structures classified as different were subsequently clustered based on salt bridges and hydrogen bonds.

Results and Discussion

Design of the Chemical Modifications in CREKA

The bioactive conformation proposed for natural CREKA on the basis of theoretical calculations⁵ is depicted in Figure 1b. In this structure, the backbone defines a β type turn motif. The functionalized side chains of the central residues (Arg, Glu and Lys) face the same side of the molecule, and the backbone Cys CO and Lys NH groups are hydrogen bonded. This structural motif was identified in the global minimum of the free peptide, when inserted in a phage display protein, and in many of the accessible minima. Inspection of the side chains reveals salt bridges involving the negatively charged carboxylate group of Glu and the contiguous positively charged side chains of Arg and Lys. These side-chain interactions are made possible by the peptide backbone turn conformation. An extended backbone would place the positively and negatively charged side chains pointing towards opposite sides (Figure 1a). The salt-bridges presented multiple interaction patterns, with the specific atoms involved being dependent on the chemical environment considered, *i.e.* the peptide in the free state, attached to a nanoparticle or inserted in a phage display protein. Accordingly, the salt bridges identified in many of the significant minima were found to differ in the atoms involved in these interactions.

The structural features observed⁵ for CREKA therefore lie in the β -turn conformation adopted by the peptide backbone. In this turn motif, Arg occupies the $i+1$ position. Proline (Pro) is known to exhibit a high propensity to induce turns (of either the β - or the γ -type) and, in particular, to occupy position $i+1$ of β -turns.³⁴⁻³⁶ Moreover, the cyclic structure of Pro, which is unique among naturally-occurring amino acids, is known to impart stability against enzymatic hydrolysis.³⁷⁻⁴⁰ Thus, replacing the Arg in CREKA by a proline-like derivative of this amino acid could lead to an increase in stability with a simultaneous retention, or even enhancement, of the propensity to adopt a folded turn-like conformation.

We therefore considered the replacement of Arg in CREKA by a proline derivative bearing the guanidinium group present in the Arg side chain at the γ -position of the pyrrolidine ring. Proline derivatives of this type may be synthetically accessible by transformation of γ -hydroxyproline, a starting material available in enantiomerically pure form from commercial sources. The exact chemical structure of the newly designed amino acid needs however to be selected: an Arg side chain incorporated at the C^γ in proline can exhibit either a *cis* or a *trans* conformation with respect to the carbonyl function. In addition, the guanidinium group may be attached to the C^γ atom of the pyrrolidine ring through a variable number (n) of methylene units, whose optimal value for formation of salt-bridge interactions with the proximal Glu side chain should be determined.

For this purpose, graphical molecular modeling was performed using the proposed bioactive conformation⁵ of natural CREKA (Figure 1b) as a template. This qualitative analysis provided the best fitting when the Arg side chain was arranged in *cis* with respect to the proline carbonyl group and for a chain length corresponding to $n = 2$ (Figure 1c). The amino acid thus designed was denoted as (Pro)hArg, where hArg stands for *homo*arginine, that is, the Arg homologue containing one more methylene unit. Figure 2 shows the structures of Arg, hArg and their corresponding *cis* proline-like derivatives (Pro)Arg and (Pro)hArg, respectively, for comparison.

Conformational Profile and Force-Field Parametrization

Before analyzing the conformational impact derived from the incorporation of (Pro)hArg into CREKA, parametrization of this non-proteinogenic residue is necessary. For this purpose, the conformational properties of its *N*-acetyl-*N'*-methylamide derivative, Ac-(Pro)hArg-NHMe (Figure 3b), have been analyzed. Given the large number of dihedral angles in this molecule and the subsequent huge number of starting geometries to be considered, a simplified methodology has been applied. Thus, the minimum energy conformations of this compound have been derived from those obtained for the residue containing one less methylene unit, Ac-(Pro)Arg-NHMe (Figure 3a).

The backbone ($\omega_0, \phi, \psi, \omega$) and side chain (χ^i , endocyclic; ξ^i exocyclic) dihedral angles considered for Ac-(Pro)Arg-NHMe and Ac-(Pro)hArg-NHMe are defined in Figure 3. The minimum energy conformations of these compounds have been denoted using a three-label code describing the backbone conformation, the puckering of the five-membered ring and the conformation of the exocyclic substituent. Specifically, the first label identifies the backbone conformation, defined by the ϕ, ψ dihedral angles, using Perczel's nomenclature.⁴¹ Accordingly, nine different backbone conformations can be distinguished in the potential energy surface $E=E(\phi, \psi)$ of an α -amino acid: $\gamma_D, \delta_D, \alpha_D, \varepsilon_D, \beta_L, \varepsilon_L, \alpha_L, \delta_L$ and γ_L . In the case of proline, only the γ_L (γ -turn or C_7), α_L (α -helical), and ε_L (polypyrroline II) conformations are possible because ϕ is fixed around -60° .³⁴⁻³⁶ The *up* or *down* puckering of the five-membered pyrrolidine ring is next indicated using the [u] and [d] labels, respectively. These conformational states are defined as those in which the C^γ atom and the carbonyl group of proline (or the proline-like residue) lie on the same or opposite sides, respectively, of the plane defined by the C^δ, N and C^α atoms. In particular, the *down* ring puckering is identified when χ^1 and χ^3 are positive while χ^2 and χ^4 are negative. Conversely, the *up* ring puckering is characterized by negative values of χ^1 and χ^3 and positive values of χ^2 and χ^4 . Finally, the last set of letters indicates the *gauche*⁺ (g^+), *skew*⁺ (s^+), *trans* (t), *skew*⁻ (s^-) or *gauche*⁻ (g^-) arrangement of each exocyclic dihedral angle ξ^i .

In a first step, the intrinsic conformational properties of Ac-(Pro)Arg-NHMe (Figure 3a) were investigated, using DFT calculations at the B3LYP/6-31+G(d,p) level. The conformational search was performed considering that this compound retains the restrictions imposed by the five-membered pyrrolidine ring in proline. Thus, the three minimum energy conformations previously characterized for Ac-Pro-NHMe⁴² with all *trans* amide bonds (ω_0 and $\omega \approx 180^\circ$), namely $\gamma_L[d], \gamma_L[u]$ and $\alpha_L[u]$, were used to generate the starting structures of Ac-(Pro)Arg-NHMe. The arrangement of the side group in (Pro)Arg is defined by the flexible dihedral angles ξ^1 and ξ^2 (Figure 3a), which are expected to exhibit three different minima: *trans* (180°), *gauche*⁺ (60°) and *gauche*⁻ (-60°). Accordingly, 3 (minima of Ac-Pro-NHMe) $\times 3$ (minima of ξ^1) $\times 3$ (minima of ξ^2) = 27 minima were anticipated for the potential energy hypersurface (PEH) $E=E(\phi, \psi, \chi^i, \xi^i)$ of Ac-(Pro)Arg-NHMe. All these structures were used as starting points for subsequent full geometry optimizations.

Table 1 lists the geometric parameters and the relative energies (ΔE^{SP}) of the five minima characterized for this compound in the gas phase, which are displayed in Figure 4. In the lowest energy minimum ($\gamma_L[u]g^+t$, Figure 4a), the backbone acetyl CO and methylamide NH sites form an intramolecular hydrogen bond defining a seven-membered cycle (γ -turn or C_7 conformation), while the pyrrolidine ring exhibits an *up* puckering and the exocyclic dihedral angles ξ^1 and ξ^2 are arranged in *gauche*⁺ and *trans*, respectively. This side chain disposition allows formation of a strong hydrogen bond involving the (Pro)Arg carbonyl oxygen and the NH moiety in the guanidinium group. Similar backbone...backbone and backbone...side chain hydrogen-bond interactions are present in the third minimum ($\gamma_L[d]s^+s^+$, Figure 4c) although, in this case, a *down* puckering of the pyrrolidine ring and a *skew*⁺ conformation of both ξ^1 and

ξ^2 are required, and this side chain rearrangement is associated with an energy penalty of 3.6 kcal/mol.

Indeed, for a γ_L backbone conformation, the most favorable backbone...side chain interaction when the pyrrolidine ring is *down*-puckered seems to involve the NH_2 rather than the NH site in the guanidinium side chain. This is inferred from the geometry of the second conformer in Table 1 ($\gamma_L[\text{d}]\text{ts}^-$, Figure 4b), which is destabilized with respect to the global minimum by only 0.4 kcal/mol. Interestingly, deviation of the ψ angle in this second conformer to values around 100° results in a new minimum energy structure ($\gamma_L[\text{d}]\text{g}^-s^+$, Figure 4e) where the backbone...backbone and backbone...side chain hydrogen-bond interactions are retained. However, the large ψ angle leads to a much less favorable geometry for the hydrogen bond stabilizing the γ -turn conformation and this results in a ΔE^{gp} value of 6.4 kcal/mol.

The only conformer in Table 1 with a backbone structure other than γ_L is the fourth minimum ($\epsilon_L[\text{d}]\text{g}^-s^+$, Figure 4d), which corresponds to a polyproline II conformation. Despite the presence of a strong interaction involving the (Pro)Arg carbonyl and one guanidinium NH_2 moiety, this conformer is disfavored by 5.6 kcal/mol, due to the absence of hydrogen bonds linking the backbone amide groups.

The free energies in the gas phase (ΔG^{gp}) calculated for the five minimum energy conformations of Ac-(Pro)Arg-NHMe are displayed in Table 2. As can be seen, consideration of the ZPVE, thermal and entropic corrections for the transformation of ΔE^{gp} into ΔG^{gp} represents relative variations lower than 0.2 kcal/mol in all cases with the exception of the $\gamma_L[\text{d}]\text{ts}^-$ disposition, for which these statistical corrections produce a destabilization of 1.2 kcal/mol. Accordingly, $\gamma_L[\text{u}]\text{g}^+t$ is the only conformation significantly populated in the gas phase at room temperature according to a Boltzmann distribution since all the local minima exhibit ΔG^{gp} values higher than 1.5 kcal/mol.

In order to obtain an estimation of the solvation effects on the relative stability of the different minima, single point calculations were conducted on the optimized structures using the PCM method. Table 2 includes the relative free energies in carbon tetrachloride, chloroform and water solutions (ΔG^{CCl_4} , ΔG^{CHCl_3} and $\Delta G^{\text{H}_2\text{O}}$, respectively). The solvent introduces significant changes in the relative stability of the different minima characterized for Ac-(Pro)Arg-NHMe. Carbon tetrachloride was found to considerably stabilize the $\gamma_L[\text{d}]\text{ts}^-$, $\epsilon_L[\text{d}]\text{g}^-s^+$ and $\gamma_L[\text{d}]\text{g}^-s^+$ conformations, to the point that the former becomes the most stable structure and the two latter are stabilized by 3.3 and 2.3 kcal/mol, respectively, with respect to the gas phase. The higher polarity of chloroform results in a further stabilization of the $\epsilon_L[\text{d}]\text{g}^-s^+$ conformation, even though the lowest energy structure remains $\gamma_L[\text{d}]\text{ts}^-$, as in carbon tetrachloride. Finally, $\epsilon_L[\text{d}]\text{g}^-s^+$ becomes the most stable structure in aqueous solution, the $\Delta G^{\text{H}_2\text{O}}$ values of the other four conformers being higher than 1.5 kcal/mol. The stabilization detected for this conformation, which increases significantly with the polarity of the solvent, should be attributed to the accessibility of the peptide bonds to the solvent. Thus, the amide groups are better solvated when the backbone adopts a ϵ_L conformation than when it is arranged in γ_L . Furthermore, the strength of the attractive amide...solvent interactions increases with the polarity of the solvent, which should be attributed to the crucial role played by the electrostatic contribution. On the other hand, the relative stabilities of the $\gamma_L[\text{d}]\text{s}^+s^+$ and $\gamma_L[\text{d}]\text{g}^-s^+$ remains essentially unaltered upon solvation.

The $\gamma_L[\text{u}]\text{g}^+t$ and $\epsilon_L[\text{d}]\text{g}^-s^+$ arrangements, corresponding to the lowest free energy conformations of Ac-(Pro)Arg-NHMe in the gas phase and in aqueous solution, respectively (Table 2), were used as starting structures for the conformational study of Ac-(Pro)hArg-NHMe (Figure 3b). It is worth noting that this decision was taken on the basis of the following considerations: (i) the $\epsilon_L[\text{d}]\text{g}^-s^+$ was the only conformation found for Ac-(Pro)Arg-NHMe with

a non-negligible population in aqueous solution, which is the environment that will be considered in the study of the pentapeptide; and (ii) consideration in the force-field parametrization process of the two backbone conformations identified in Tables 1 and 2, γ_L and ϵ_L , is *a priori* highly desirable. Thus, the starting geometries for the latter compound were prepared by including an additional methylene group in the side chain of (Pro)Arg for such two representative structures. The conformation of the new methylene group is defined by the dihedral angle ξ^3 (Figure 3b), for which three different arrangements were considered: *gauche*⁺, *trans* and *gauche*⁻. Accordingly, the conformational search of Ac-(Pro)hArg-NHMe was carried out considering 2 [representative minima of Ac-(Pro)Arg-NHMe] $\times$ 3 (minima of ξ^3) = 6 starting geometries. Energy minimizations at the B3LYP/6-31+G(d,p) level led to five different minimum energy structures, the three most stable being given in Table 3 and Figure 5. The remaining two are not included since their relative energies were found to be above 10 kcal/mol and therefore were not considered representative.

The lowest energy conformation ($\gamma_L[d]tg^+g^-$, Figure 5a) corresponds to a γ -turn conformation stabilized by a hydrogen bond linking the backbone terminal CO and NH sites. The (Pro)hArg CO group and one of the NH₂ moieties in the guanidinium side chain are also involved in a strong hydrogen-bond interaction. The latter is lost in the $\gamma_L[d]g^-tt$ minimum (Figure 5c), which explains its high relative energy. Conversely, the second conformer in Table 3 ($\epsilon_L[d]g^-g^+g^+$, Figure 5b) exhibits no backbone...backbone hydrogen bond, as expected for an ϵ_L conformation, while being stabilized by a strong side chain...backbone interaction. Accordingly, it is destabilized by 5.5 kcal/mol with respect to the lowest energy minimum.

Table 4 lists the ΔG^{gp} , ΔG^{CCl_4} , ΔG^{CHCl_3} and ΔG^{H_2O} values calculated for the three Ac-(Pro)hArg-NHMe minima described above. Notably, the ZPVE, thermal and entropic corrections stabilize the $\gamma_L[d]g^-tt$ structure in the gas phase by 3.5 kcal/mol. In spite of this, the $\gamma_L[d]tg^+g^-$ remains the only accessible conformation both in the gas phase and in carbon tetrachloride solution. In contrast, $\gamma_L[d]g^-tt$ becomes the most stable structure in the presence of chloroform or water. This significant variation should be attributed to the charged nature of the side group. Accordingly, the impact of the environment on the arrangement of the side chain increases with the polarity of the solvent and, therefore, the largest change is detected in aqueous solution.

These results indicate that (Pro)hArg retains the most important structural features of proline,⁴² reflecting a high tendency to induce peptide turns. It should also be noted that, even if the γ_L and ϵ_L conformations are characterized by different values of the ψ dihedral angle, both of them correspond to turn-like backbone conformations, namely the *i*+1 position of a γ - and a β II-turn, respectively.³⁴⁻³⁶

The stretching, bending, torsion and van der Waals parameters used in the AMBER force-field³⁰ to describe Pro and Arg were directly transferred to (Pro)hArg. Atomic centered charges for the minimum energy conformations listed in Table 3 were calculated by fitting the UHF/6-31G(d) quantum mechanical and the Coulombic molecular electrostatic potentials (MEPs) to a large set of points placed outside the nuclear region. It should be noted that the electrostatic parameters derived at this level of theory are fully compatible with the current AMBER force-field.³⁰ On the other hand, electrostatic force-field parametrization using a strategy based on weighted multiple conformations through a Boltzmann distribution in the gas phase, which was originally proposed by different authors,⁴³ has been demonstrated to be especially successful for non-proteinogenic residues.^{43b,44} Accordingly, parameters for (Pro)hArg (Figure 6) have been obtained considering the atomic charges of the lowest energy minimum only ($\gamma_L[d]tg^+g^-$, Table 4) since the other two local minima are unfavored by more than 5.1 kcal/mol and, therefore, their contribution to a normalized Boltzmann distribution can be considered negligible.

Conformational Search of the CREKA Analogue Attached to a Nanoparticle. Comparison with the Natural Peptide

The conformational preferences of the CREKA analogue incorporating (Pro)hArg as an Arg substitute, hereafter denoted as CR*EKA, have been explored using the sampling technique previously employed for the study of the natural pentapeptide.⁵ Figure 7a represents the evolution of the number of unique minimum energy conformations against the number of modified SA-MD production cycles necessary for the conformational search of CR*EKA to converge. As can be seen, the exploration is completed after seven cycles, the last one providing only 3 new structures to the list of unique conformations. Interestingly, the replacement of Arg by (Pro)hArg did not lead to a reduction in the amount of modified SA-MD cycles required to complete the conformational search, but the number of accessible low energy conformations diminished dramatically, *i.e.* 612 and 1305 minima were obtained for CR*EKA and CREKA, respectively. Moreover, the energetic distribution of the minima generated was also affected by this targeted replacement, as indicated in Figure 7b. Thus, the conformational restrictions imposed by the presence of (Pro)hArg eliminated a large number of minima with relative energies in the interval between 4 and 11 kcal/mol, which was found to be the most populated for CREKA. This effect is shown by the distribution obtained for CR*EKA, which contrasts with the Gaussian-like profile achieved for the natural peptide.

The conformational preferences of the backbone in the two pentapeptides have been compared by analyzing the virtual dihedral angles used to define the specific arrangement of each residue. Results are given in Figure 8, which represents the distribution of such dihedrals through histograms. As can be seen, the incorporation of (Pro)hArg produces some changes in the general conformational profile of the peptide. The distributions obtained for the (Pro)hArg and Glu residues in CR*EKA are very similar to those found for their Arg and Glu counterparts in CREKA, while important variations are detected for the other three residues, especially Lys and Ala. Indeed, the conformational space of the two latter amino acids is significantly narrower in CR*EKA, indicating that the incorporation of (Pro)hArg reduces the conformational flexibility of the whole peptide. This is also reflected in Figure 9, which compares the distribution of the ϕ/ψ backbone dihedral angles of the five residues for the minima with relative energies lower than 2 kcal/mol.

A clustering analysis based on hydrogen bonds and salt bridges indicated that 68% of the CR*EKA minima (417 structures) present at least one such interaction, while this value reaches 82% (1092 structures) for natural CREKA. The total number of interactions (either of the hydrogen-bond or salt-bridge type) detected in these minima is 667 (1.6 interactions per structure) and 2198 (2.1 interactions per structure) for CR*EKA and CREKA, respectively, which are distributed in 106 and 448 clusters (Figure 10). It is interesting to note that only 3 clusters in CR*EKA and 8 in CREKA contain more than 25 structures, respectively grouping 31% and 25% of the minima. Clusters are organized as follows: (i) backbone...backbone hydrogen bonds: 79.3% (529 interactions) for CR*EKA and 83.6% (1839 interactions) for CREKA; (ii) backbone...side chain hydrogen bonds: 0.0% (no interaction) for CR*EKA and 4.7% (102 interactions) for CREKA; (iii) side chain...side chain salt bridges: 20.7% (138 interactions) for CR*EKA and 11.7% (257 interactions) for CREKA.

The most frequent interactions in the modified peptide are the Glu(N-H)...(O=C)Cys hydrogen bond (219 interactions; 32.8%) and the Glu...Lys salt bridge (138 interactions; 20.7%), while in natural CREKA they are the Glu(N-H)...(O=C)Ac, Lys(N-H)...(O=C)Cys and Ala(N-H)...(O=C)Arg backbone...backbone hydrogen bonds (251, 225 and 207 interactions, respectively; 11.4%, 10.2% and 9.4%, respectively) and the Arg...Glu salt bridge (187 interactions; 8.5%). These distinct interaction patterns suggest important differences between the two peptides, which are confirmed upon a cross-comparison of the frequency in which a particular interaction is observed. Thus, the backbone...backbone hydrogen bond most frequently detected in

CR*EKA, Glu(N-H)⋯(O=C)Cys, accounts for only 3.4% of the CREKA interactions. Conversely, the Glu(N-H)⋯(O=C)Ac, Lys(N-H)⋯(O=C)Cys and Ala(N-H)⋯(O=C)Arg hydrogen bonds are, respectively, 0.8%, 3.6% and 5.1% of the CR*EKA interactions. Regarding side chain⋯side chain contacts, the frequency of the Glu⋯Lys salt bridge is 7.5% in CREKA, while the Arg⋯Glu interaction was not identified in any of the CR*EKA minima.

These results indicate that the conformational restrictions imposed by the presence of (Pro)hArg lead to some alterations in the conformational profile of the whole peptide that affect the turn type generated and the residues involved in this turn, as well as the interactions between adjacent ionized side chains. As expected, the conformational propensities of the (Pro)hArg-containing peptide seem to be more clearly defined than those of the natural sequence, as shown by the smaller variation observed for the preferred backbone⋯backbone and side chain⋯side chain interactions in CR*EKA with respect to CREKA. This is deduced from the cluster analysis described above and it becomes even clearer when the interaction schemes of the three conformations of lowest energy generated for the two peptides are compared (Table 5). Thus, up to three different β -turns are detected in these CREKA conformations, namely those centered at Arg-Glu [stabilized by a Lys(N-H)⋯(O=C)Cys hydrogen bond; minima # 1 and 3], Glu-Lys [stabilized by a Ala(N-H)⋯(O=C)Arg hydrogen bond; minimum # 2] and Lys-Ala [stabilized by a NHMe(N-H)⋯(O=C)Glu hydrogen bond; minimum # 2]. In comparison, all three CR*EKA conformers in Table 5 share a γ -turn centered at the (Pro)hArg residue and stabilized by a Glu(N-H)⋯(O=C)Cys hydrogen bond. This is due to the much higher propensity of proline, when compared to other proteinogenic amino acids, in nucleating turns and therefore occupying the central turn positions.³⁴⁻³⁶ The fact that the turn observed in the lowest energy conformation in CREKA is of the β type whereas it is of the γ type in CR*EKA, suggests that the difference in the overall shapes of the molecules is rather small. In both turns the $i+1$ position is occupied by either Arg (in CREKA) or its substitute (in CR*EKA) and both turn conformations orient the charged side chains of the three central residues toward the same region of the molecule. Thus, the main structural requirements considered to be essential for the bioactivity of CREKA⁵ are also present in CR*EKA.

In spite of this similarity, the natural peptide and its analogue differ in the interactions involving the three charged side chains. Specifically, the only salt-bridge type detected in the three CR*EKA minima in Table 5 involves Glu and Lys, while the natural CREKA presents multiple salt bridges Arg⋯Glu⋯Lys (minimum # 1) or an Arg⋯Glu only (minima # 2 and 3). The Arg⋯Glu salt bridge seems to be disfavored when (Pro)hArg replaces Arg, but the biological consequences of this change are difficult to predict since this specific interaction may not be essential to bioactivity. In the peptide-receptor recognition process, the Arg, Gly and Lys side chains are likely to interact with the complementary groups in the receptor, rather than among themselves.

Overall, these results suggest that, although the targeted replacement of Arg by (Pro)hArg in CREKA induces some changes in the conformational profile of the peptide, the most important structural trends of the natural compound are retained. This general resemblance is reflected in the lowest energy conformations found for CR*EKA and for CREKA, with the backbones within 1.5 Å of each other. Figure 11 shows the superposition of the two structures. This shape similarity between the most stable minima suggests that the incorporation of (Pro)hArg to provide resistance against the proteolytic enzymes would not induce major alterations in the conformational features considered to be essential for the tumor-homing activity of the pentapeptide.

Conclusions

A targeted replacement strategy has been followed to protect CREKA, a promising tumor-homing pentapeptide for cancer diagnostics and treatment, from proteolytic cleavage. For this purpose, a proline-like derivative of arginine, denoted as (Pro)hArg, has been designed, parameterized and incorporated into CREKA, replacing Arg. The conformational flexibility of the CREKA analogue incorporating (Pro)hArg is significantly reduced with respect to the natural CREKA. We attribute this reduction to the conformational restrictions imposed by the cyclic residue. The presence of this arginine surrogate produces some alterations in the interaction patterns as observed in the clustering analysis based on hydrogen bonds and salt bridges and in the detailed comparison of the interactions in the lowest energy conformations in the natural and modified peptide. In spite of this, the most important features proposed for the bioactive conformation of CREKA, *i.e.* a turn motif orienting the charged side chains towards the same side of the molecule, are retained in the (Pro)hArg-containing analogue. The shape of the lowest energy conformation of the modified peptide shows a significant degree of similarity with that proposed for the bioactive conformation of the natural peptide. Thus, our results suggest that the incorporation of (Pro)hArg into CREKA as an arginine substitute could be useful to protect the peptide from proteolytic attack while retaining, or even enhancing, those structural features of the parent peptide considered to be essential for its tumor-homing activity.

Acknowledgements

Gratitude is expressed to the Centre de Supercomputació de Catalunya (CESCA) and to the Barcelona Supercomputer Center (BSC) for computational resources. We also acknowledge 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. Classic calculations were partially formed by utilizing 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 the Ministerio de Educación y Ciencia (project CTQ2007-62245, Ramon y Cajal contract for DZ), Gobierno de Aragón (research group E40) and from the U.S. National Cancer Institute (grants CA124427 and CA 119335 to ER) is gratefully acknowledged. Support for the research of C.A. was received through the prize "ICREA Academia" for excellence in research funded by the Generalitat de Catalunya. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number N01-CO-12400. 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.

References

1. Simberg D, Duza T, Park JH, Essier M, Pilch J, Zhang L, Derfus AM, Yang M, Hoffman RM, Bathia S, Sailor MJ, Ruoslahti E. *Proc. Natl. Acad. Sci. U.S.A* 2007;104:932. [PubMed: 17215365]
2. (a) Hoffman JA, Giraudo E, Singh M, Zhang L, Inoue M, Porkka K, Hanahan D, Ruoslahti E. *Cancer Cell* 2003;4:383. [PubMed: 14667505] (b) Pasqualini R, Ruoslahti E. *Nature* 1996;380:364. [PubMed: 8598934]
3. Hutchinson JN, Muller WJ. *Oncogene* 2000;19:6130. [PubMed: 11156526]
4. Ruoslahti, E. 2007. unpublished results
5. Zanuy D, Flores-Ortega A, Casanovas J, Curcó D, Nussinov R, Alemán C. *J. Phys. Chem. B* 2008;112:8692. [PubMed: 18588341]
6. Adessi C, Soto C. *Curr. Med. Chem* 2002;9:963. [PubMed: 11966456]
7. Lelais G, Seebach D. *Biopolymers* 2004;76:206. [PubMed: 15148683]
8. Yamaguchi H, Kodama H, Osada S, Kato F, Jelokhani-Niaraki M, Kondo M. *Biosci. Biotech. Biochem* 2003;67:2269.
9. Sadowsky JD, Murray JK, Tomita Y, Gellman SH. *ChemBioChem* 2007;8:903. [PubMed: 17503422]
10. Biron E, Chatterjee J, Ovadia O, Langenegger D, Brueggen J, Hoyer D, Schmid HA, Jelinek R, Gilon C, Hoffman A, Kessler H. *Angew. Chem. Int. Ed* 2008;47:2595.

11. Banerjee R, Basu G, Chene P, Roy S. J. Pept. Res 2002;60:88. [PubMed: 12102721]
12. Webb AI, Dunstone MA, Williamson NA, Price JD, de Kauwe A, Chen W, Oakley A, Perlmutter P, McCluskey J, Aguilar MI, Rossjohn J, Purcell AW. J. Immunol 2005;175:3810. [PubMed: 16148127]
13. Horne WS, Gellman SH. Acc. Chem. Res 2008;41:1399. [PubMed: 18590282]
14. Agrafiotis DM, Gibbs AC, Zhu F, Izrailev S, Martin E. J. Chem. Inf. Model 2007;47:1067. [PubMed: 17411028]
15. Steinbach PJ. Proteins 2004;57:665. [PubMed: 15390266]
16. Swaminathan P, Hariharan M, Murali R, Singh CU. J. Med. Chem 1996;39:2141. [PubMed: 8667358]
17. Zheng J, Zanuy D, Haspel N, Tsai C-J, Alemán C, Nussinov R. Biochemistry 2007;46:1205. [PubMed: 17260950]
18. Zanuy D, Jiménez AI, Cativiela C, Nussinov R, Alemán C. J. Phys. Chem. B 2007;111:3236. [PubMed: 17388467]
19. Ballano G, Zanuy D, Jiménez AI, Cativiela C, Nussinov R, Alemán C. J. Phys. Chem. B 2008;112:13101. [PubMed: 18811190]
20. Kirkpatrick S, Gelatt CD Jr. Vecchi MP. Science 1983;220:671. [PubMed: 17813860]
21. Steinbach PJ, Brooks BR. Chem. Phys. Lett 1994;226:447.
22. Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JA.; Vreven, T., Jr.; Kudin, KN.; Burant, JC.; Millam, JM.; Iyengar, SS.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, GA.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, JE.; Hratchian, HP.; Cross, JB.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, RE.; Yazyev, O.; Austin, AJ.; Cammi, R.; Pomelli, C.; Ochterski, JW.; Ayala, PY.; Morokuma, K.; Voth, GA.; Salvador, P.; Dannenberg, JJ.; Zakrzewski, VG.; Dapprich, S.; Daniels, AD.; C. Strain, M.; Farkas, O.; Malick, DK.; Rabuck, AD.; Raghavachari, K.; Foresman, JB.; Ortiz, JV.; Cui, Q.; Baboul, AG.; Clifford, S.; Cioslowski, J.; Stefanov, BB.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, RL.; Fox, DJ.; Keith, T.; Al-Laham, MA.; Peng, CY.; Nanayakkara, A.; Challacombe, M.; Gill, PMW.; Johnson, B.; Chen, W.; Wong, MW.; Gonzalez, C.; Pople, JA. Gaussian 03, Revision B.02. Gaussian, Inc.; Pittsburgh PA: 2003.
23. Becke AD. J. Chem. Phys 1993;98:1372.
24. Lee C, Yang W, Parr RG. Phys. Rev. B 1993;37:785.
25. McLean AD, Chandler GS. J. Chem. Phys 1980;72:5639.
26. (a) Tomasi J, Mennucci B, Cammi R. Chem. Rev 2005;105:2999. [PubMed: 16092826] (b) Tomasi J, Persico M. Chem. Rev 1994;94:2027. (c) Miertus S, Tomasi J. Chem. Phys 1982;65:239. (d) Miertus M, Scrocco E, Tomasi J. Chem. Phys 1981;55:117.
27. Baysal C, Meirovitch H. J. Comput. Chem 1999;20:1659.
28. Simmerling C, Elber R. J. Am. Chem. Soc 1994;116:2534.
29. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. J. Chem. Phys 1983;79:926.
30. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. J. Am. Chem. Soc 1995;117:5179.
31. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. J. Chem. Phys 1984;81:3684.
32. Ryckaert JP, Ciccotti G, Berendsen HJC. J. Comput. Phys 1977;23:327.
33. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. J. Comput. Chem 2005;26:1781. [PubMed: 16222654]
34. MacArthur MW, Thornton JM. J. Mol. Biol 1991;218:397. [PubMed: 2010917]
35. Rose GD, Gierasch LM, Smith JA. Adv. Protein Chem 1985;37:1. [PubMed: 2865874]
36. Chakrabarti P, Pal D. Prog. Biophys. Mol. Biol 2001;76:1. [PubMed: 11389934]
37. Vanhoof G, Goossens F, De Meester I, Hendriks D, Scharpé S, Faseb J 1995;9:736. [PubMed: 7601338]
38. Walker JR, Altman RK, Warren JW, Altman E. J. Pept. Res 2003;62:214. [PubMed: 14531845]
39. Green BD, Gault VA, Irwin N, Mooney MH, Bailey CJ, Harriott P, Greer B, Platt PR, O'Harte FPM. Biol. Chem 2003;384:1543. [PubMed: 14719796]

40. (a) Markert Y, Köditz J, Ulbrich-Hofmann R, Arnold U. *Protein Eng* 2003;16:1041. [PubMed: 14983085] (b) Frenken LGJ, Egmond MR, Batenburg AM, Verrips CT. *Protein Eng* 1993;6:637. [PubMed: 8234234] (c) Reidhaar-Olson JF, Parsell DA, Sauer RT. *Biochemistry* 1990;29:7563. [PubMed: 2148681]
41. Perczel A, Angyan JG, Kajtar M, Viviani W, Rivail J-L, Marcoccia J-F, Csizmadia IG. *J. Am. Chem. Soc* 1991;113:6256.
42. Flores-Ortega A, Jiménez AI, Cativiela C, Nussinov R, Alemán C, Casanovas J. *J. Org. Chem* 2008;73:3418. [PubMed: 18351745]
43. (a) Reynolds CA, Essex JW, Richards WG. *J. Am. Chem. Soc* 1992;114:9075. (b) Alemán C, Casanovas J. *J. Chem. Soc. Perkin Trans 2* 1994:563. (c) Cieplak P, Cornell WD, Bayly CI, Kollman PA. *J. Comput.Chem* 1995;16:1357.
44. (a) Casanovas J, Zanuy D, Nussinov R, Alemán C. *J. Org. Chem* 2007;72:2174. [PubMed: 17291048] (b) Casanovas J, Zanuy D, Nussinov R, Alemán C. *Chem. Phys. Lett* 2006;429:558. (c) Alemán C, Zanuy D, Casanovas J, Cativiela C, Nussinov R. *J. Phys. Chem* 2006;110:21264.

Figure 1.

(a) Schematic representation of the system under study showing the position of the particles (black dots) used to define the virtual dihedral angles (see Methods Section). (b) Bioactive conformation proposed for CREKA (according to ^{ref. 5}). (c) CREKA analogue constructed by replacing the Arg residue by (Pro)hArg within the bioactive conformation. The Arg and (Pro)hArg side chains are depicted in yellow.

Figure 2. Structure of arginine (Arg) and its homologue containing an additional methylene group (hArg), as well as that of their respective proline-like derivatives considered in this work.

Figure 3.

Dihedral angles used to identify the conformations of the *N*-acetyl-*N'*-methanamide derivatives of (Pro)Arg (a) and (Pro)hArg (b) studied in this work. The dihedral angles ω_0 , ϕ , ψ and ω are defined using backbone atoms, whereas the endocyclic dihedral angles (χ^i) and ξ^1 , respectively, are given by the atoms of the five-membered ring. In particular, the sequence of atoms used to define ϕ , χ^0 and ξ^1 are C(O)-N-C $^\alpha$ -C(O), C $^\delta$ -N-C $^\alpha$ -C $^\beta$ and C $^\beta$ -C $^\gamma$ -CH $_2$ -N, respectively.

Figure 4. Minimum energy conformations of Ac-(Pro)Arg-NHMe obtained from B3LYP/6-31+G(d,p) calculations: (a) $\gamma_L[u]g^+t$; (b) $\gamma_L[d]ts^-$; (c) $\gamma_L[d]s^+s^+$; (d) $\epsilon_L[d]g^-s^+$; (e) $\gamma_L[d]g^-s^+$ (see Table 1 for geometries). Distances (H...O) and angles (N-H...O) associated with the backbone...backbone and backbone...side chain interactions (dashed lines) are given.

Figure 5. Minimum energy conformations of Ac-(Pro)hArg-NHMe obtained from B3LYP/6-31+G(d,p) calculations: (a) $\gamma_L[d]tg^+g^-$; (b) $\epsilon_L[d]g^-g^+g^+$; (c) $\gamma_L[d]g^-tt$ (see Table 3 for geometries). Distances (H...O) and angles (N-H...O) associated with the backbone...backbone and backbone...side chain interactions (dashed lines) are given.

Figure 6.
Electrostatic parameters determined for the (Pro)hArg residue.

Figure 7.

(a) Number of unique minimum energy conformations found for CR*EKA (grey line and diamonds) and natural CREKA (black line and circles) against the number of modified SA-MD cycles used for the conformation search. (b) Distribution of energies for the unique minimum energy conformations of CR*EKA and natural CREKA. Energies are computed relative to the corresponding lowest energy minimum.

Figure 8.

Comparison of the distribution of virtual dihedral angles used to define the backbone conformation of CREKA (a) and CR*EKA (b) in each unique minimum energy structure obtained. Color code for the bars: black for dihedral angle values ranging from 0° to 60°, red from 60° to 120°, green from 120° to 180°, blue from 180° to 240°, yellow from 240° to 300° and grey from 300° to 360°.

Figure 9.

Ramachandran plot distribution for the five residues of CR*EKA (open circles) and CREKA (filled black circles) considering the more representative minimum energy structures, *i.e.* those within a relative energy interval of 2 kcal/mol.

Figure 10. Distribution of the unique minimum energy structures generated for CR*EKA (top) and CREKA (bottom) in clusters, which have been grouped on the basis of the formation of hydrogen bonds and salt bridges.

Figure 11. Lowest energy minimum obtained for CR*EKA (a) and CREKA (b) attached to a nanoparticle, and superposition of both structures (c). The surface used to mimic the nanoparticle is represented by a single green ball.

Table 1

Dihedral angles^{a,b} of the backbone and the exocyclic side group, pseudorotational parameters^c of the pyrrolidine ring (A and P), and relative energy^c (ΔE^{sp}) of the minimum energy conformations characterized for Ac-(Pro)Arg-NHMe at the B3LYP/6-31+G(d,p) level in the gas phase.

	ω_0	ϕ	ψ	ω	(P,A)	ξ_1	ξ_2	ΔE^{sp}
$\gamma_L[\text{u}]^{\text{t}}$	-175.9	-82.3	84.4	-176.7	(177.2, 27.7) ^d	67.7	-164.5	0.0 ^e
$\gamma_L[\text{d}]^{\text{s}}$	-170.2	-82.9	74.0	-178.8	(-108.9, 30.7) ^f	-171.4	-115.3	0.4
$\gamma_L[\text{d}]^{\text{s}}^{\text{+}}$	-169.0	-82.0	71.4	179.4	(-97.4, 38.3) ^g	149.8	110.3	3.6
$\epsilon_L[\text{d}]^{\text{g}}^{\text{+}}$	-172.9	-69.2	137.4	-178.5	(-91.1, 33.3) ^h	-38.5	104.7	5.6
$\gamma_L[\text{d}]^{\text{g}}^{\text{+}}$	-176.4	-82.4	98.4	-175.8	(-152.0, 29.0) ⁱ	-36.4	108.8	6.4

^a In degrees.

^b See Figure 3a for definition.

^c In kcal/mol.

^d $\chi^0 = -27.7^\circ$, $\chi^1 = 20.4^\circ$, $\chi^2 = -7.3^\circ$, $\chi^3 = -8.7^\circ$ and $\chi^4 = 23.2^\circ$.

^e $E = -817.237062$ a.u..

^f $\chi^0 = -9.9^\circ$, $\chi^1 = 25.2^\circ$, $\chi^2 = -30.8^\circ$, $\chi^3 = 24.3^\circ$ and $\chi^4 = -9.2^\circ$.

^g $\chi^0 = -4.9^\circ$, $\chi^1 = 26.7^\circ$, $\chi^2 = -37.8^\circ$, $\chi^3 = 33.8^\circ$ and $\chi^4 = -18.6^\circ$.

^h $\chi^0 = -0.6^\circ$, $\chi^1 = 20.4^\circ$, $\chi^2 = -31.7^\circ$, $\chi^3 = 30.8^\circ$ and $\chi^4 = -19.5^\circ$.

ⁱ $\chi^0 = -25.6^\circ$, $\chi^1 = 27.8^\circ$, $\chi^2 = -20.9^\circ$, $\chi^3 = 5.8^\circ$ and $\chi^4 = 12.6^\circ$.

Table 2

Relative free energy^a in the gas phase (ΔG^{gp}) and in carbon tetrachloride, chloroform and aqueous solutions (ΔG^{CCl_4} , ΔG^{CHCl_3} and $\Delta G^{\text{H}_2\text{O}}$, respectively) for the minimum energy conformations of Ac-(Pro)Arg-NHMe at the B3LYP/6-31+G(d,p) level.

	ΔG^{gp}	ΔG^{CCl_4}	ΔG^{CHCl_3}	$\Delta G^{\text{H}_2\text{O}}$
$\gamma_{\text{L}}[\text{u}]\text{g}^+\text{t}$	0.0 ^b	1.1	2.6	6.2
$\gamma_{\text{L}}[\text{d}]\text{ts}^-$	1.6	0.0	0.0	1.7
$\gamma_{\text{L}}[\text{d}]\text{s}^+\text{s}^+$	3.8	3.7	4.3	6.6
$\gamma_{\text{L}}[\text{d}]\text{g}^-\text{s}^+$	5.7	2.4	0.6	0.0
$\gamma_{\text{L}}[\text{d}]\text{g}^-\text{s}^+$	6.4	4.1	3.3	4.2

^a In kcal/mol.

^b $G = -816.963546$ a.u.

Table 3

Dihedral angles^{a,b} of the backbone and the exocyclic side group, pseudorotational parameters^c of the pyrrolidine ring (A and P), and relative energy^c (ΔE^{gp}) of the minimum energy conformations characterized for Ac-(Pro)hArg-NHMe at the B3LYP/6-31+G(d,p) level in the gas phase.

	$\omega 0$	ϕ	ψ	ω	(P,A)	ξ_1	ξ_2	ξ_3	ΔE^{gp}
$\gamma_{\text{L}}[\text{d}]\text{lg}^+\text{g}^-$	-169.4	-82.5	68.8	-179.5	(-105.0, 35.2) ^d	-159.3	52.3	-85.3	0.0 ^e
$\gamma_{\text{L}}[\text{d}]\text{lg}^-\text{g}^+$	-171.0	-68.9	139.9	-179.4	(-86.6, 37.1) ^f	-51.8	70.0	69.8	5.5
$\gamma_{\text{L}}[\text{d}]\text{lg}^+\text{t}^+$	-170.9	-83.7	78.1	-178.4	(-105.5, 32.9) ^g	-72.2	179.3	-176.2	9.6

^aIn degrees.

^bSee Figure 3b for definition.

^cIn kcal/mol.

^d $\chi^0 = -9.1^\circ$, $\chi^1 = 27.6^\circ$, $\chi^2 = -35.1^\circ$, $\chi^3 = 29.0^\circ$ and $\chi^4 = -12.9^\circ$.

^eE = -856.547433 a.u.

^f $\chi^0 = -10.4^\circ$, $\chi^1 = 26.8^\circ$, $\chi^2 = -32.9^\circ$, $\chi^3 = 26.0^\circ$ and $\chi^4 = -10.2^\circ$.

^g $\chi^0 = 2.2^\circ$, $\chi^1 = 20.5^\circ$, $\chi^2 = -34.3^\circ$, $\chi^3 = 35.0^\circ$ and $\chi^4 = -24.1^\circ$.

Table 4

Relative free energy^a in the gas phase (ΔG^{gp}) and in carbon tetrachloride, chloroform and aqueous solutions (ΔG^{CCl_4} , ΔG^{CHCl_3} and $\Delta G^{\text{H}_2\text{O}}$, respectively) for the minimum energy conformations of Ac-(Pro)hArg-NHMe at the B3LYP/6-31+G(d,p) level.

	ΔG^{gp}	ΔG^{CCl_4}	ΔG^{CHCl_3}	$\Delta G^{\text{H}_2\text{O}}$
$\gamma_{\text{L}}[\text{d}]\text{tg}^+\text{g}$	0.0 ^b	0.0	2.1	5.3
$\varepsilon_{\text{L}}[\text{d}]\text{g}^-\text{g}^+\text{g}^+$	5.1	3.8	4.3	5.1
$\gamma_{\text{L}}[\text{d}]\text{g}^-\text{tt}$	6.1	1.7	0.0	0.0

^a In kcal/mol.

^b $G = -856.244850$ a.u.

Table 5

Comparison between the interaction pattern^a of the three minima of lowest energy generated for natural CREKA^b and its analogue CR*EKA. Relative energies^c (ΔE) are also given.

#	ΔE	CR*EKA	ΔE	CREKA
1	0.0	HB Glu(N-H)···(O=C)Cys SB Lys···Glu	0.0	HB Lys(N-H)···(O=C)Cys SB Arg···Glu SB Lys···Glu
2	1.0	HB Glu(N-H)···(O=C)Cys HB Ala(N-H)··· (O=C)Cys HB NHMe (N-H)···(O=C)Glu SB Lys···Glu	1.0	HB Ala(N-H)···(O=C)Arg HB NHMe (N-H)···(O=C)Glu SB Arg···Glu
3	1.0	HB Glu(N-H)···(O=C)Cys HB NHMe(N-H) ···(O=C)Lys SB Lys···Glu	2.0	HB Lys(N-H)···(O=C)Cys HB Ala(N- H)···(O=C)Cys SB Arg···Glu

^aHydrogen bonds and salt bridges are labeled as HB and SB, respectively.

^bAccording to reference 5.

^cIn kcal/mol.