J Comput Aided Mol Des (2013) 27:31–43
DOI 10.1007/s10822-012-9623-5

Engineering strategy to improve peptide analogs: from structure-based computational design to tumor homing

David Zanuy · Francisco J. Sayago · Guillem Revilla-López · Gema Ballano · Lilach Agemy · Venkata Ramana Kotamraju · Ana I. Jiménez · Carlos Cativiela · Ruth Nussinov · April M. Sawvel · Galen Stucky · Erkki Ruoslahti · Carlos Alemán

Received: 9 October 2012 / Accepted: 4 December 2012 / Published online: 14 December 2012

Abstract We present a chemical strategy to engineer analogs of the tumor-homing peptide CREKA (Cys-Arg-Glu-Lys-Ala), which binds to fibrin and fibrin-associated clotted plasma proteins in tumor vessels (Simberg et al. in Proc Natl Acad Sci USA 104:932–936, 2007) with improved ability to inhibit tumor growth. Computer modeling using a combination of simulated annealing and molecular dynamics were carried out to design targeted replacements aimed at enhancing the stability of the bioactive conformation of CREKA. Because this conformation presents a pocket-like shape with the charged groups of Arg, Glu and Lys pointing outward, non-proteinogenic amino acids α-methyl and *N*-methyl derivatives of Arg, Glu and Lys were selected, rationally designed and incorporated into CREKA analogs. The stabilization of the bioactive conformation predicted by the modeling for the different CREKA analogs matched the tumor fluorescence results, with tumor accumulation increasing with stabilization. Here we report the modeling, synthetic procedures, and new biological assays used to test the efficacy and utility of the analogs. Combined, our results show how studies based on multi-disciplinary collaboration can converge and lead to useful biomedical advances.

Keywords Computational design · Bioactive conformation dynamics · Tumor-homing peptide · Peptide synthesis · Tumor growth inhibitors

Electronic supplementary material The online version of this article (doi:10.1007/s10822-012-9623-5) contains supplementary material, which is available to authorized users.

D. Zanuy (✉) · G. Revilla-López · C. Alemán (✉)
Department of Chemical Engineering, ETSEIB, Universitat Politècnica de Catalunya, Diagonal 647, 08028 Barcelona, Spain
e-mail: david.zanuy@upc.edu

C. Alemán
e-mail: carlos.aleman@upc.edu

F. J. Sayago · G. Ballano · A. I. Jiménez · C. Cativiela
Department of Organic Chemistry, ISQCH, University of Zaragoza–CSIC, 50009 Zaragoza, Spain

L. Agemy · V. R. Kotamraju · E. Ruoslahti
Center for Nanomedicine, Sanford-Burnham Medical Research Institute, Burnham Institute for Medical Research at UCSB, University of California, 3119 Biology II Bldg., Santa Barbara, CA 93106-9610, USA

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

R. Nussinov
Department of Human Genetics, Sackler, Medical School, Tel Aviv University, 69978 Tel Aviv, Israel

A. M. Sawvel · G. Stucky
Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106, USA

E. Ruoslahti
Cancer Center, Sanford-Burnham Medical Research Institute, La Jolla, CA 92037, USA

C. Alemán
Center for Research in Nano-Engineering, Universitat Politècnica de Catalunya, Campus Sud, Edifici C, C/Pasqual I Vila s/n, 08028 Barcelona, Spain

Introduction

The tumor-homing pentapeptide CREKA (Cys-Arg-Glu-Lys-Asp) specifically homes to tumors by binding to fibrin and fibrin-associated clotted plasma proteins (e.g. fibronectin) in tumor vessels [1]. The peptide, which was identified by screening phage-display libraries to discover specific targets in tumor vessels [2], does not recognize tumor growth in mutant mice null for fibrogen or in mice lacking plasma fibronectin, which becomes covalently bound to fibrin during blood clotting. In addition, CREKA linked to amino dextran-coated superparamagnetic iron oxide (SPIO) nanoparticles not only binds to blood and plasma but also induces further localized tumor clotting [1]. This amplification system enhanced the homing of the nanoparticles in a mouse tumor model, and the tumor imaging, without causing clotting or other obvious side effects elsewhere in the body.

Recently, we determined the energy landscape and bioactive conformation of CREKA on its own and labeled with fluorescein (FAM), with the fluorescent dye attached to the Cys through a flexible linker, using a multiple conformational search strategy based on molecular dynamics (MD) simulations [3, 4]. Even though the straightforward approach would have been modeling the peptide dynamics within its targeted receptor, the lack of crystallographic information about the receptor structure made us to follow an alternative path. In addition to the previous simulated conditions, we considered a system that mimicked the peptide attached to the nanoparticle by tethering it to the surface of spherical particles [3]. Results indicated that the conformational profile of the REKA sequence is similar in all simulated cases, with the Cys residue providing the distinctive features in each case. The bioactive conformation is displayed in Fig. 1a and features both a β-turn motif and strong interactions involving the side groups of Arg, Glu and Lys. Additional MD simulations showed that this conformation is very stable, and its pocket-like shape with the charged groups pointing outwards allows formation of intermolecular interactions.

Next, we used this bioactive conformation of CREKA to design some analogs using molecular modeling, which were generated by replacing an amino acid in the sequence by a non-proteinogenic counterpart [5]. Our initial aims were to enhance the tumor imaging activity of CREKA by stabilizing the bioactive conformation and to impart protection against proteolytic resistance by incorporating non-proteinogenic amino acids. The combination of the self-amplifying homing system formed by elongated iron oxide nanoparticles (hereafter denoted "nanoworms", NWs) individually coated with CREKA and another tumor-homing peptide allowed not only effective imaging of the tumors but also produced extensive inhibition of tumor growth by obstructing tumor circulation through blood clotting. The efficacy of this theranostic nanosystem was increased by selectively replacing some residues in CREKA by their non-proteinogenic counterparts, which increased the stability of the peptide. The CREKA analogs substantially improved the tumor-homing efficiency and increased the blockage of tumor vessels. Treatment of prostate cancer mice with multiple doses of the theranostic nanoparticles with CREKA analogs induced tumor necrosis and significant reduction in tumor growth [5].

After reporting these biomedical achievements [5], herein we describe the strategy, its rationale and the findings obtained for the CREKA analogs from a chemical

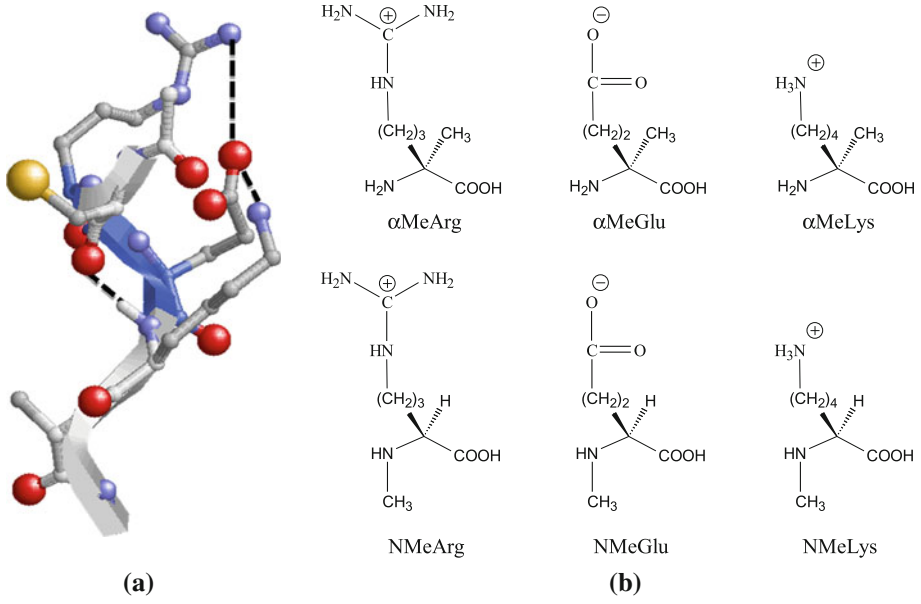


Fig. 1 **a** Bioactive conformation of CREKA. **b** Chemical structure and name to identify the α-methyl (*top row*) and *N*-methyl (*bottom row*) amino acids used to replace Arg, Glu and Lys in CREKA analogs

point of view. Below, we discuss the design of the chemical modifications in CREKA, the intensive modeling to determine the energy landscape and bioactive conformations of the resulting CREKA analogs, the synthetic methods, and the application of the modified peptides as imaging agents. These are described in the framework of our efforts to increase the activity of natural CREKA and its resistance to endogenous proteases. Therefore, in order to overcome the inexistent crystal structure of the CREKA receptors, we use experimental binding assays to corroborate the predicted bioactivity for the different proposed CREKA analogs.

Rational design of chemical modifications in CREKA

Stabilization of structural motifs is made possible by specific interactions and constraints. In particular, incorporation of conformationally constrained amino acids into a peptide has been shown to be a powerful method to confer selectivity and stability [6–19]. Appropriate constraints may not only promote the stability of the bioactive conformation with respect to other structural motifs present in the potential energy landscape of the natural peptide, but also reduce the entropy change during the binding to the active site of a receptor or enzyme [14–19]. Constraints limit the degrees of freedom, reducing the conformational flexibility and the dynamic motion of peptide analogs prior to their complexation. An implicit reasoning is that, although the solvent and proteins may interact in a similar way with the natural peptide and the constrained analogs (i.e. no significant change in the binding enthalpy is expected), the pre-organization required by the parent peptide to reach the biologically active conformation reduces its efficacy and binding affinity, resulting in a less favorable entropy of binding.

In the bioactive conformation proposed for CREKA (Fig. 1a), the backbone defines a β-type turn motif with a hydrogen bond between the C=O(Cys) and N–H(Lys) and the side chains of the central charged residues (Arg, Glu and Lys) face the same side of the molecule forming salt bridges. It should be noted that the turn conformation of the backbone makes possible salt bridges between the carboxylate of the Glu and the positively charged side chains of Arg and Lys. Other conformations (e.g. extended) would place the charged side groups pointing towards opposite sides. These salt bridges present multiple interaction patterns, the atoms involved in such interactions changing with the environment and the dynamic motion. This bioactive conformation was identified not only for the free peptide but also for the peptide attached to the surface of a nanoparticle and for the peptide inserted in a phage display protein.

The pocket shape and the orientation of the charged side groups in the bioactive conformation of CREKA suggest that the three central residues (REK) are crucial for the biological activity of this peptide. After examining the conformational characteristics of different families of non-coded amino acids (i.e. α-tetrasubstituted, dehydro, *N*-substituted and thio-α-amino acids), we selected the α-methyl and *N*-methyl derivatives of Arg, Glu and Lys as the most appropriate for the rational design of CREKA analogs. The name and chemical structure of these non-coded amino acids are displayed in Fig. 1b.

Recent studies indicated that α-methyl α-amino acids, in which the α-hydrogen atom has been replaced by a methyl group, tend to overwhelmingly stabilize folded β-turn-like secondary structures versus extended conformations [20–23]. The incorporation of the methyl group not only produces steric hindrance but also reduces the flexibility of the backbone, the conformational freedom of the α-methyl α-amino acid being more restricted than that of its natural counterpart. Accordingly, the incorporation of α-methyl Arg, Glu and Lys (αMeArg, αMeGlu and αMeLys, respectively) is expected to enhance significantly the stability of the bioactive conformation. On the other hand, the conformational preferences of *N*-methyl α-amino acids are dictated by increased steric hindrance, which produces a destabilization of the extended conformations, and lack of hydrogen bond donor ability associated with the replacement of the hydrogen atom by a methyl group [24–26]. Therefore, replacement of Arg and Glu by their corresponding *N*-methylated counterparts (NMeArg and NMeGlu, respectively) in CREKA are not expected to alter those structural features considered to be essential for its tumor-homing activity. A different scenario can be expected when *N*-methyl Lys (NMeLys) is incorporated into CREKA as a Lys substitute, since the backbone hydrogen bond of the β-turn motif is not possible in the analog. In spite of this, we have also considered this targeted replacement to obtain further insight into the influence of such a hydrogen bond on the pocked-like shape conformation of CREKA.

Methods

Computational methods

The sequences of the six CREKA analogs investigated in this work are: Cys-αMeArg-Glu-Lys-Ala (CR$^{\alpha}$EKA), Cys-Arg-αMeGlu-Lys-Ala (CRE$^{\alpha}$KA), Cys-Arg-Glu-αMeLys-Ala (CREK$^{\alpha}$A), Cys-NMeArg-Glu-Lys-Ala (CR^{N}EKA), Cys-Arg-NMeGlu-Lys-Ala (CRE^{N}KA) and Cys-Arg-Glu-NMeLys-Ala (CREK^{N}A), where the superscripts α and N refer to the α-methyl and *N*-methyl residues, respectively. The energy landscape of the six peptides was examined

using the sampling technique previously employed for CREKA [3], which consists of modified simulated annealing combined with MD (SA-MD) [27, 28]. This procedure is based on the minimization of structures generated in 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. Early studies showed that very low energies are obtained by minimizing the structures extracted from SA-MD processes [29, 30]. In our previous study with CREKA [3] this sampling technique was found to be robust when 500 structures were submitted to energy minimization from each SA-MD cycle.

The sulfhydryl group of the Cys residue was used to tether each analog to a surface formed by 100 spherical particles, which were distributed in a 10×10 square (47.5 Å^2), with van der Waals parameters [$R = 2.35$ Å and $\varepsilon = 0.90$ kcal/mol] and without electric charge. The whole system, which is identical to that used for mimicking CREKA bound to the surface of a nanoparticle, was placed in the center of a cubic simulation box ($a = 47.5$ Å) filled with at least 3,217 explicit water molecules, that were represented using the TIP3 model [31]. One negatively charged chloride and two positively charged sodium atoms were added to the simulation box to reach electric neutrality (net charges were considered for side groups of the α-methyl and *N*-methyl Arg, Lys and Glu at neutral pH).

Prior to the production cycles with the modified SA-MD, the simulation box was equilibrated. Each system was initially minimized to relax the conformational and structural tensions using the conjugate gradient method. After this, 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 minimization process was performed under the same simulation conditions, i.e., using the existing water box. 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 the starting conformation of the next cycle.

The 100 spherical particles used to construct the surface were kept fixed at the initial positions in all MD simulations and energy minimizations. The energy was calculated using the AMBER force-field [32, 33], with the required parameters taken from the AMBER libraries. Atom pair distance cutoffs were applied at 14.0 Å to compute the van der Waals and electrostatic interactions. Both temperature and pressure were controlled by the weak coupling method, the Berendsen thermo-barostat [34], using a time constant for heat bath coupling and a pressure relaxation time of 1 ps. Bond lengths were constrained using the *SHAKE* algorithm [35] with a numerical integration step of 2 fs.

The list of unique minimum energy conformations was generated by comparing the 500 minimized structures provided by each cycle of modified SA-MD among themselves and 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, discarding only those conformations that had already appeared. 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 and the end capping atoms. 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.

Amino acids synthesis

All non-proteinogenic amino acids were prepared as optically pure L enantiomers suitably protected for use in Fmoc-based solid-phase peptide synthesis. The *N*-methyl amino acids were prepared following reported procedures [36–39], they are also commercially available. The oxazolidinone used as starting material for the synthesis of α-methyl amino acids was obtained from Z-D-Ala-OH as described by Akaji et al. [40]. All synthetic intermediates provided satisfactory physical and spectroscopic data. Melting points were determined on a Gallenkamp apparatus. Optical rotations were measured at room temperature using a JASCO P-1020 polarimeter. High-resolution mass spectra were obtained on a Bruker Microtof-Q spectrometer. IR spectra were registered on a Nicolet Avatar 360 FTIR spectrophotometer; ν_{max} is given for the main absorption bands. ^{1}H and ^{13}C NMR spectra were recorded on a Bruker AV-500 or AV-400 instrument at room temperature using the residual solvent signal as internal standard; chemical shifts (δ) are expressed in ppm and coupling constants (*J*) in Hz.

Fmoc-(αMe)Arg(Mtr)-OH

Mp 116–118 °C. $[\alpha]_D = +1.2$ (*c* 0.41, MeOH). IR (KBr) ν 3,700–2,700, 1,718, 1,641 cm^{-1}. ^1H NMR (400 MHz, DMSO-d_6) δ (major rotamer) 12.23 (bs, 1H), 7.92–7.83 (m, 2H), 7.71–7.62 (m, 2H), 7.45–7.27 (m, 4H), 7.11–6.91 (m, 2H), 6.69–6.63 (m, 1H), 6.56 (bs, 2H), 4.35–4.16 (m, 3H), 3.77 (s, 3H), 3.11–2.92 (m, 2H), 2.59 (s, 3H), 2.52 (s, 3H), 2.04 (s, 3H), 1.86–1.61 (m, 2H), 1.41–1.17 (m, 2H) overlapped with 1.29 (s, 3H). ^{13}C NMR (100 MHz, DMSO-d_6) δ (major rotamer) 176.10, 157.39, 156.19, 153.99, 143.88, 140.69, 137.60, 135.53, 134.58, 127.57, 127.04, 125.10, 123.49, 120.09, 111.69, 64.87, 58.49, 55.43, 46.79, 40.96, 33.55, 25.19, 23.62, 23.42, 17.99, 11.73. HRMS (ESI) $C_{32}H_{39}N_4O_7S$ $[M + H]^+$: calcd. 623.2534, found 623.2559.

Fmoc-(αMe)Glu(OtBu)-OH

Mp 85–87 °C. $[\alpha]_D = +4.2$ (*c* 0.41, $CHCl_3$). IR (KBr) ν 3,650–2,800, 1,724 cm^{-1}. ^1H NMR (400 MHz, DMSO-d_6) δ 12.51 (bs, 1H), 7.92–7.85 (m, 2H), 7.70–7.62 (m, 2H), 7.44–7.29 (m, 4H), 7.13 (bs, 1H), 4.31–4.17 (m, 3H), 2.12–1.85 (m, 4H), 1.38 (s, 9H), 1.30 (s, 3H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 175.20, 172.23, 153.87, 143.90, 140.71, 127.57, 127.02, 125.09, 120.11, 79.36, 64.87, 58.09, 46.81, 31.22, 30.47, 27.75, 23.27. HRMS (ESI) $C_{25}H_{28}NO_6$ $[M - H]^-$: calcd. 438.1922, found 438.1923.

Fmoc-(αMe)Lys(Boc)-OH

Mp 75–77 °C. $[\alpha]_D = +1.9$ (*c* 0.62, MeOH). IR (Nujol) ν 3,700–2,790, 1,707 cm^{-1}. ^1H NMR (500 MHz, DMSO-d_6) δ 12.36 (bs, 1H), 7.92–7.87 (m, 2H), 7.75–7.69 (m, 2H), 7.44–7.30 (m, 5H), 6.77 (bs, 1H), 4.32–4.15 (m, 3H), 2.89 (m, 2H), 1.83–1.70 (m, 1H), 1.70–1.58 (m, 1H), 1.40–1.29 (m, 2H) overlapped with 1.37 (s, 9H), 1.32 (s, 3H), 1.25–1.10 (m, 2H). ^{13}C NMR (100 MHz, DMSO-d_6) δ 175.37, 155.59, 154.80, 143.85, 140.71, 127.62, 127.07, 125.29, 120.09, 77.34, 65.25, 58.27, 46.72, 36.55, 29.77, 28.28, 27.61, 22.25, 20.53. HRMS (ESI) $C_{27}H_{33}N_2O_6$ $[M - H]^-$: calcd. 481.2344, found 481.2349.

Peptide synthesis

Peptides were synthesized on solid phase using the Fmoc/*t*-Bu strategy on a microwave assisted automated peptide synthesizer (Liberty, CEM corporation, Matthews, NC, 28105). Couplings were performed using HATU-collidine reagent system in *N*,*N*′-dimethyl formamide (DMF) with fivefold excess of appropriately protected amino acid derivatives. Fmoc deprotection was performed with 5 % piperazine in NMP containing 0.1 M HOBt with two microwave energy cycles for 30 and 180 s at 75 °C under N_2. The coupling of amino acids except Arg and Cys was done for 5 min with 25 W microwave energy at 75 °C under nitrogen. Fmoc-Arg(Pbf)-OH was double coupled with 0 W microwave for 25 min and for 5 min with 20 W microwave energy at 75 °C. All the residues on the N-terminus of the α-methylated residues were coupled twice for 5 min with 25 W microwave energy at 75 °C under N_2. All the Cys residues were double coupled with a total of six minutes at 50 °C with 0 W and 20 W microwave energy for 2 and 4 min, respectively, each time. The protected peptide resins were then cleaved with TFA:TIS:water (95 %:2.5 %:2.5 %) to yield the crude peptides which were analyzed by reverse phase (RP) HPLC using Gilson Analytical HPLC (Phenomenex Jupiter Proteo, 250 × 4.60 mm column) and purified on Gilson Semi-Preparative HPLC (Phenomenex, Jupiter Proteo, 250 × 15 mm column). The solvents used were A: 0.1 % TFA in water and B: 0.1 % TFA in 60 % acetonitrile in water. The peptides were characterized by electron spray ionization mass spectroscopy (ESI-MS) on a Waters micromass QTOF2 tandem mass spectrometer.

Biological assays

Nanoworms (NWs) and nanoworms coated with peptides (i.e. CREKA-NWs, CR^NEKA-NWs, CRE^NKA-NWs and $CREK^{\alpha}A$-NWs) were prepared as described previously [41]. Synthetic peptides labeled with fluorescein (FAM) were injected intravenously into mice bearing B16F1 tumors, the protocol being the same as in previous work [5]. Similarly, analyses, protocols and procedures to examine the nanoworm biodistribution were identical to those described in Ref. [5].

Results and discussion

The energy landscape of CREKA analogs

The evolution of the number of unique minimum energy conformations against the number of SA-MD cycles for CREKA and its analogs, which are compared in Fig. 2a, indicates that the conformational search converges after only 7 cycles. The number of unique minima is very similar for CREKA and its α-methylated analogs (i.e. 1,323, 1,325, 1,316 and 1,310 for CREKA, $CR^{\alpha}EKA$, $CRE^{\alpha}KA$ and $CREK^{\alpha}A$, respectively), whereas this number decreases by 13–22 % for the *N*-methylated ones (i.e. 1,038, 1,163 and 1,113 for CR^NEKA, CRE^NKA and $CREK^NA$, respectively) reflecting a reduction in the conformational flexibility. Figure 2b indicates that the energy distribution of the unique minima identified for CREKA is largely influenced by the C^{α}- and N-methylation of the charged residues. More specifically, the increment in the

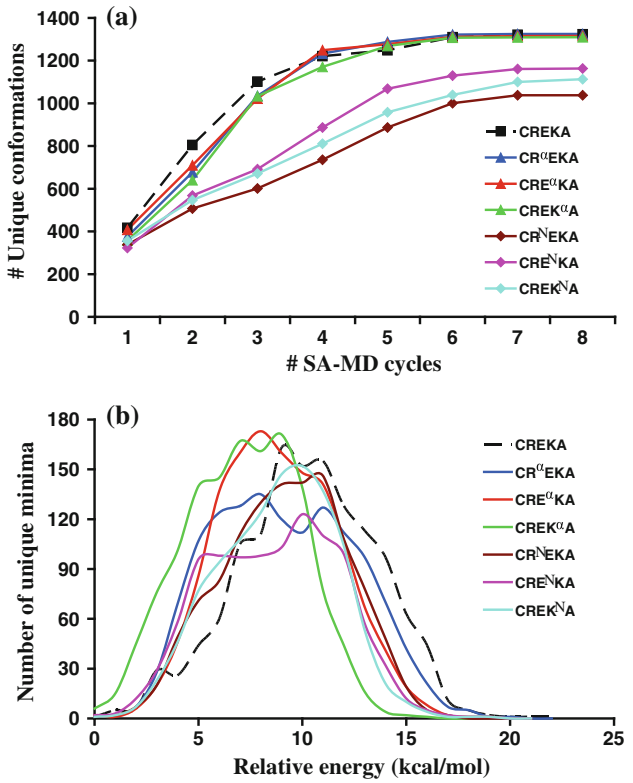


Fig. 2 **a** Number of unique minimum energy conformations found for CREKA and its α- and *N*-methyl analogs against the number of SA-MD cycles used in the conformational search. **b** Distribution of relative energies for the minimum energy conformations found for CREKA and its analogs

steric hindrance produced by the replacement of the hydrogen atom by a methyl group not only alters the shape of the profile but also shifts the distribution towards lower energy values (i.e. the highest energy bordering decreases by about 2–6 kcal/mol).

Figure 3 compares the distribution of the virtual dihedral angles, used to define the conformation in all the identified unique minima, of CREKA and its Cα- and *N*-methyl analogs. As can be seen, the conformational profiles of the three central residues in CREKA are similar or very similar to those obtained in CR$^{\alpha}$EKA, CRE$^{\alpha}$KA, CR^{N}EKA and CRE^{N}KA. Incorporation of αMeLys produces some changes in the conformational profile of the Glu residue. Thus, although the preferences of the other four residues of CREKA remain practically the same as in CREK$^{\alpha}$A, the interactions required to stabilize the bioactive conformation may be affected by such variation. On the other hand, the conformational profile obtained for CREK^{N}A deserves special attention. Even though the replacement of Lys by NMeLys was initially expected to perturb the conformational preferences of the peptide because of the impossibility to form the intramolecular hydrogen bond found in the bioactive β-turn motif, i.e. (Cys)C = O···H–N(Lys), the conformational profiles of both CREKA and CREK^{N}A are remarkably similar. Thus, the conformational agreement between the two peptides extends not only to the positively charged residues but also to the two residues located at the ends, Cys and Ala. Indeed, the only difference affects the central Glu, which does not seem significant.

Table 1 summarize the results of the clustering analysis based on the presence of specific interactions (i.e. hydrogen bonds and salt bridges). Of the 1,323 unique minima of CREKA, 90.52 % involves a total of 3,115 specific interactions (i.e. 1,198 minima with an average of 2.60 interactions per minimum), which are distributed as follows: 57.6 % main chain···main chain (mc–mc) hydrogen bonds, 17.4 % side chain···main chain (sc···mc) hydrogen bonds, 13.3 % side chain···side chain (sc–sc) hydrogen bonds and 11.7 % sc–sc salt bridges. According to the similarities in the interaction pattern, the 1,198 minima with specific interactions were categorized into 687 clusters, even though only 17 of them contained more than 10 minima.

Considering only the unique minima containing specific interactions, the average number of interactions per minimum in α-methyl analogs is similar to that in CREKA with the exception of CR$^{\alpha}$EKA, which shows a higher value (i.e. 3.26 interactions per minimum). Moreover, the distribution of interactions in such analogs also resembles that of CREKA. Thus, mc–mc hydrogen bonds are the most frequent interactions with populations ranging between 58.3 and 61.8 %. The only quantitative discrepancy between the distributions found for the α-methylated analogs and CREKA corresponds to CR$^{\alpha}$EKA, in which the population of mc-sc is smaller than those of sc–sc salt bridges and sc–sc hydrogen bonds. The unique minima of CR$^{\alpha}$EKA, CRE$^{\alpha}$KA and CREK$^{\alpha}$A with specific interactions are distributed in 909, 792 and 767 clusters, respectively, even though only 10, 15 and 12 clusters involve more than 10 structures. Moreover, the largest number of unique minima in a given cluster is 34, 35 and 46 for CR$^{\alpha}$EKA, CRE$^{\alpha}$KA and CREK$^{\alpha}$A, respectively, while it was only 27 for CREKA. Although the number of unique minimum with specific interactions in α-methyl analogs and CREKA is very similar, the number of clusters in CREKA is 222, 105 and 80 lower than in CR$^{\alpha}$EKA, CRE$^{\alpha}$KA and CREK$^{\alpha}$A, respectively. This feature suggests that C$^{\alpha}$-methylation does not restrict the conformational flexibility or the ability to form intramolecular interactions of the peptide.

Inspection of results derived from the clustering analysis of *N*-methyl analogs reveals significant differences. In all cases both the number of unique minima with specific interactions (i.e. 879, 967 and 913 for CR^{N}EKA, CRE^{N}KA and CREK^{N}A, respectively) and the average number of interactions per minimum (i.e. 2.34, 2.32 and 2.20 for CR^{N}EKA, CRE^{N}KA and CREK^{N}A, respectively) are smaller in the *N*-methyl analogs than in the parent peptide.

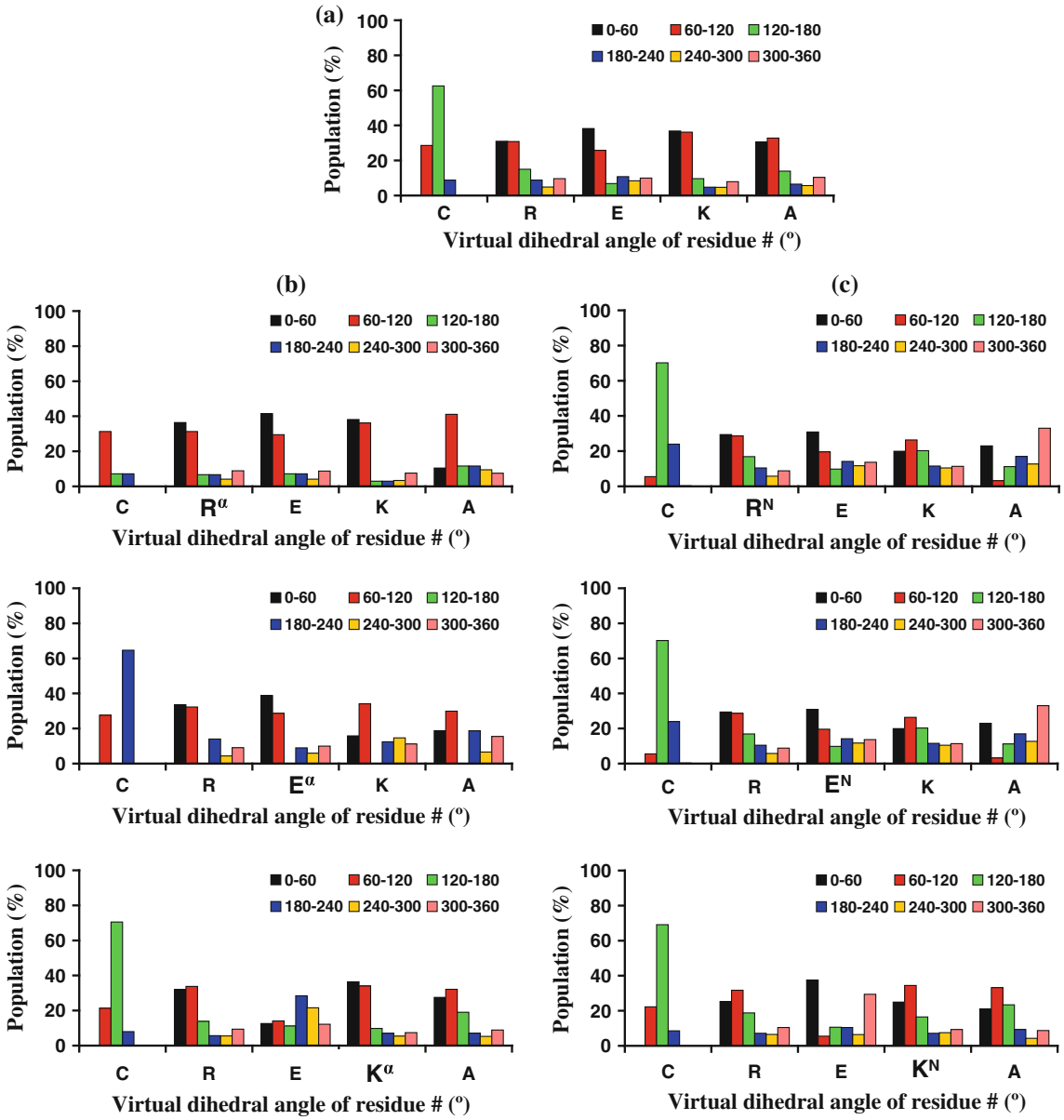


Fig. 3 Distribution of virtual dihedral angles used to define the conformation in the unique minima obtained for: **a** CREKA; **b** C^{α}-methyl CREKA analogs; and **c** *N*-methyl CREKA analogs

Moreover, the unique minima of $CR^{N}EKA$, $CRE^{N}KA$ and $CREK^{N}A$ are distributed in 537, 453 and 493 clusters, respectively, which represent a drastic decrease with respect to the C^{α}-methyl analogs. Similarly, the number of clusters with more than ten unique minima goes from 17 in CREKA to 11, 13 and 11 in $CR^{N}EKA$, $CRE^{N}KA$ and $CREK^{N}A$, respectively. Combined, these results indicate that N-methylation reduces significantly the conformational flexibility of the peptides and, consequently, the variability of hydrogen bonding and salt bridges interactions. On the other hand, it is worth noting that sc–sc salt bridges are more frequent than sc–sc hydrogen bonds in the three *N*-methyl analogs (Table 1), which represents a remarkable change in the interaction pattern with respect to CREKA and the C^{α}-methyl analogs. Indeed, the percentage of sc–sc salt bridges shown by $CREK^{N}A$ is even larger than that of the mc-sc hydrogen bonds. These results suggest that the side groups of the three charged residues may interact more favorably with the receptor in the *N*-methyl analogs, especially in $CREK^{N}A$, than in CREKA and the C^{α}-methyl analogs.

Figure 4a shows the distribution of the backbone dihedral angles in the φ,ψ-Ramachandran map of the five residues of CREKA considering all the minima included in the clustered analyses, whereas Figs. 4b, c display the distribution for the three α-methylated and *N*-methylated

Table 1 Results of the clustering analysis for CREKA and its C^{α}- and *N*-methyl analogs (see text)

	CREKA	$CR^{\alpha}EKA$	$CRE^{\alpha}KA$	$CREK^{\alpha}A$	$CR^{N}EKA$	$CRE^{N}KA$	$CREK^{N}A$
Total number of unique minima	1,323	1,325	1,316	1,310	1,038	1,163	1,113
Number of minima with hydrogen bonds and/or salt bridges ($N_{hb/sb}$)	1,198	1,266	1,232	1,220	879	967	913
Number of interactions per minimum included in $N_{hb/sb}$	2.60	3.26	2.73	2.65	2.34	2.32	2.20
Number of clusters found in the $N_{hb/sb}$ minima	687	909	792	767	537	463	493
Number of clusters with more than 10 unique minima	17	10	15	12	11	13	11
Percentage of main chain–main chain hydrogen bonds (%)	57.6	58.3	61.8	58.9	55.6	57.3	50.9
Percentage of main chain–side chain hydrogen bonds	17.4	8.2	17.8	18.7	24.1	12.6	14.0
Percentage of side chain–side chain hydrogen bonds (%)	13.3	19.9	10.1	13.0	8.4	13.8	15.0
Percentage of side chain–side chain salt bridges (%)	11.7	13.6	10.3	9.4	11.9	16.3	20.1

residues, respectively, as derived from the minimum energy conformations of the corresponding analogs. The distributions obtained for Arg, Glu and Lys are very similar to those obtained for their α-methyl and *N*-methyl counterparts. Thus, the substitutions engineered in this work are not expected to damage the biological activity of CREKA.

Synthesis of non-proteinogenic amino acids

Preparation of CREKA analogs required the availability of α-methylated and *N*-methylated Arg, Glu and Lys. Because of their wide use in biochemistry and medicinal chemistry, great efforts have been devoted to the development of methods for the synthesis of α-methyl [42, 43] and *N*-methyl [36–39, 44] α-amino acids. For the present work, the target amino acids have been prepared in enantiomerically pure form (L configuration) and suitable protected for use in solid phase peptide synthesis using Fmoc chemistry, that is, with the main-chain amino moiety bearing an Fmoc group and the side chain functionality protected with an acid-labile group. Efficient procedures to access *N*-methyl amino acids have been reported [36–39, 44] and were applied to synthesize those used in the present work (they are also available from commercial sources). To obtain α-methylated Arg, Glu, and Lys, synthetic routes were developed using oxazolidinone **1** as a common substrate (Scheme 1). This oxazolidinone [45] was prepared in enantiomerically pure form by condensation of Z-D-Ala-OH with benzaldehyde dimethyl acetal as described by Akaji et al. [40] Reaction of **1** with bromide **2**, **3**, or **4** afforded the corresponding α-methylated oxazolidinones **5**, **6**, and **7** as single stereoisomers (Scheme 1), which were used as precursors of αMeGlu, αMeLys, and αMeArg, respectively. These synthetic intermediates bear the main-chain amino and carboxylic acid functions masked as an oxazolidinone while the substituent introduced in each case can be elaborated to generate the desired side chain (Scheme 1). Thus, catalytic hydrogenation of **5** resulted in cleavage of the oxazolidinone ring with concomitant deprotection of the amino and carboxylic acid groups and simultaneous reduction of the double bond in the side chain. Protection of the amino function with Fmoc under standard conditions provided the desired amino acid, Fmoc-αMeGlu(OtBu)-OH (**9**). A similar protocol was used to transform **6** into Fmoc-αMeLys(Boc)-OH (**11**), which involved an additional step to eliminate one of the Boc side-chain protecting groups. The synthesis of αMeArg from **7** required the introduction of a guanidinium function. To this end, a side-chain amino moiety was first incorporated. The terminal alkene in **7** was submitted to hydroboration and the hydroxyl group thus introduced was displaced with bis(*tert*-butoxycarbonyl)amide. The oxazolidinone ring was then cleaved under basic conditions and the Boc groups were eliminated by acidic treatment to yield **12**, which bears a free side-chain amino function adequate to introduce the guanidinium moiety. The latter was accomplished by reaction with *N*,*N*′-bis(*tert*-butoxycarboyl)-*N*″-trifylguanidine. Subsequent replacement of side-chain Boc protection by Mtr (4-methoxy-2,3,6-trimethylbenzenesulfonyl) followed by Z-to-Fmoc exchange afforded the target amino acid, Fmoc-αMeArg(Mtr)-OH (**13**).

Biological activity of CREKA analogs

The biological activity and resistance against proteolytic degradation of CREKA analogs was reported in our previous study [5]. FAM-labeled peptides incorporating α-methyl and *N*-methyl analogs showed higher tumor fluorescence than the FAM-labeled parent peptide. Quantitative comparison with CREKA indicated that accumulation in nude mice bearing orthotropic 22Rv1 xenograft tumors was around threefold higher for $CREK^{\alpha}A$ and $CRE^{N}KA$, and more than twofold higher for $CRE^{\alpha}KA$. In particular, $CR^{N}EKA$ was the only analog with tumor fluorescence weaker than CREKA. On the other hand, it was reported that the fluorescence in the tumors 3 h after injection remained practically unaltered for

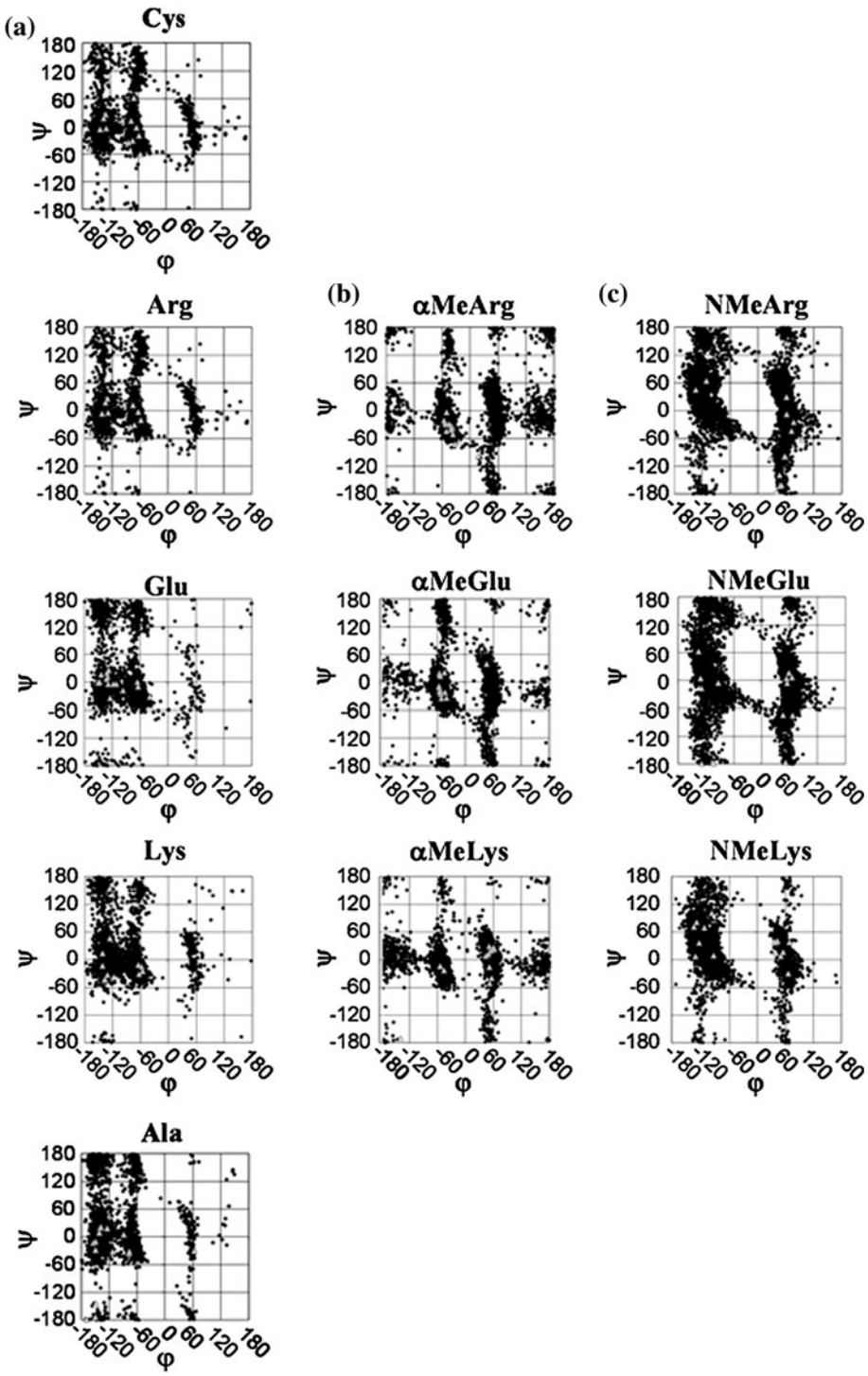
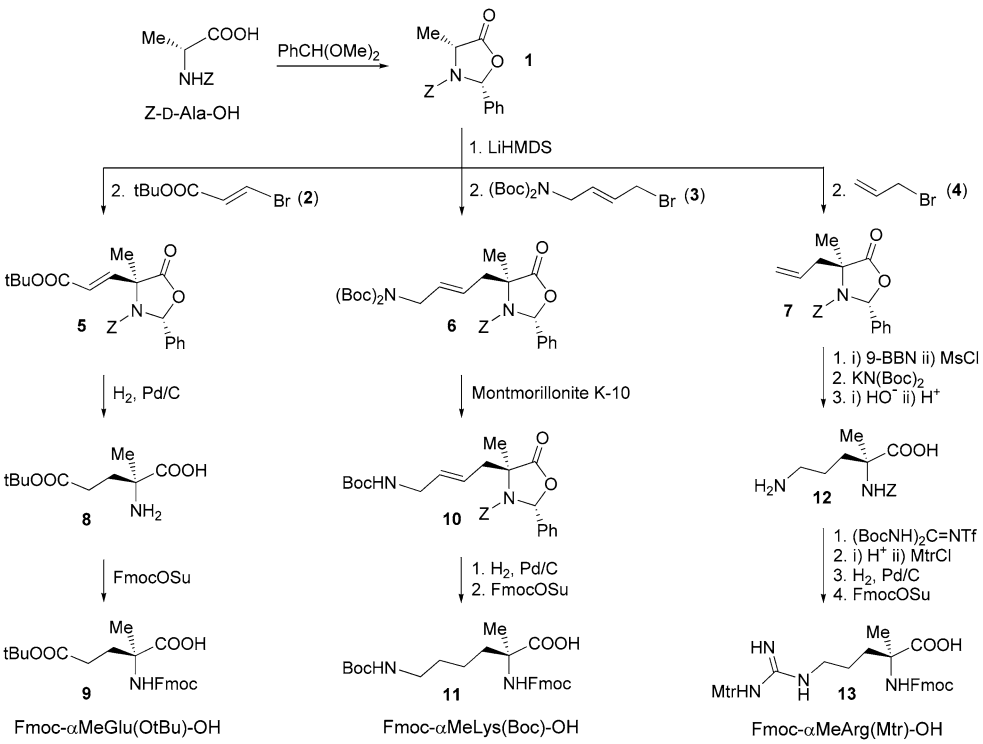


Fig. 4 Ramachandran plot distribution for: **a** the five residues of CREKA; **b** the αMeArg, αMeGlu and αMeLys residues of CR$^{\alpha}$EKA, CRE$^{\alpha}$KA and CREK$^{\alpha}$A, respectively; and **c** the NMeArg, NMeGlu and NMeLys of CR^{N}EKA, CRE^{N}KA and CREK^{N}A, respectively. The unique minimum energy conformations were considered for each case

CREKA analogs whereas decreased 70 % for CREKA, indicating that the stability of the analogs is remarkably higher than that of the unmodified peptide [5].

The activity prediction follows the same rational previously discussed and successfully used for other studied potential analogs [4]: we had indirectly established the active conformation of CREKA, which was among the 5 lowest energy conformations of the free peptide [3]. The comparison of this specific conformation of CREKA with the results obtained for each analog exploration would indicate how far each analog is from the targeted conformation. In the most ideal case, the lowest energy conformer

Scheme 1 Synthetic routes to protected derivatives of enantiopure αMeGlu, αMeLys, and αMeArg



for each analog would provide a similar conformation to the CREKA biological one. The more dissimilar the analogs absolute minimum is with respect the CREKA bioactive conformation, the lower is the chances of obtaining an active analog. We can assess this dissimilarity on basis of the RMSD difference between each analog absolute minimum and the conformation that the most resembles active conformation of CREKA. This assessment will tell how far the targeted conformation is from the lowest energy conformation in each studied analog [46–48]. This strategy is reliable since all the studied conformers in this last step for each analog belong to clusters that present a potential energy that is bellow the margin 5 kcal·mol^{-1} from each respective absolute minimum (Fig. 2b).

Figure 5 shows the superimpositions of two selected structures of CREKA with the lowest energy conformation of each analog. The two structures of the parent peptide are the lowest energy conformation (left side of each pair), which was proposed as the bioactive conformation [3], and the minimum energy conformation of the parent peptide showing the highest resemblance to the lowest energy conformation of each analog (right side of each pair). The root mean square deviation (rmsd) between each pair of conformations is also displayed. The parameter Δ_{rmsd}, which is defined for each analog as the absolute difference between the rmsd values determined for the two pairs of superimposed structures, increases as follows: CREK$^{\alpha}$A ($\Delta_{rmsd} = 0.28$ Å) < CRE^{N}KA ($\Delta_{rmsd} = 0.32$ Å) < CRE$^{\alpha}$KA ($\Delta_{rmsd} = 0.41$ Å) < CREK^{N}A ($\Delta_{rmsd} = 0.74$ Å) < CR$^{\alpha}$EKA ($\Delta_{rmsd} = 0.84$ Å) < CR^{N}EKA ($\Delta_{rmsd} = 0.92$ Å). This order is fully consistent with the tumor fluorescence observed for CREKA analogs, i.e. the tumor accumulation of the peptide increases when Δ_{rmsd} decreases. Thus, the variation in tumor accumulation determined for the different CREKA analogs was ranked as follows: CREK$^{\alpha}$A > CRE^{N}KA > CRE$^{\alpha}$KA > CREK^{N}A > CR^{N}EKA (biological assays were not carried out with CR$^{\alpha}$EKA) [5]. This success allows us to propose Δ_{rmsd} as a useful parameter, which can provide a reliable estimate of the stabilization induced in the bioactive conformation.

Complementary data indicate that this molecular engineering procedure helps to target tumor recognition. We compare the efficacy of nanoworms (NWs) coated with CREKA and different analogs homing to tumors by binding to fibrin and fibrin-associated clotted plasma proteins in tumor vessels. Figure 6 shows the quantification of the fibrin(ogen) accumulation obtained for CR^{N}EKA-NWs, CRE^{N}KA-NWs and CREK$^{\alpha}$A-NWs, which is compared with that of CREKA-NWs, results that are complementary to those already shown [5]. Replacement of Glu and Lys by NMeGlu and αMeLys, respectively, enhances clotting by twofold, while substitution of Arg by NMeArg does not alter this activity. Obviously, these results are less striking than those obtained by combining CREKA analogs-NWs with CRKDKC-NWs (i.e. clotting increased four- or even five-times when two types of coated nanoparticles were used) [5].

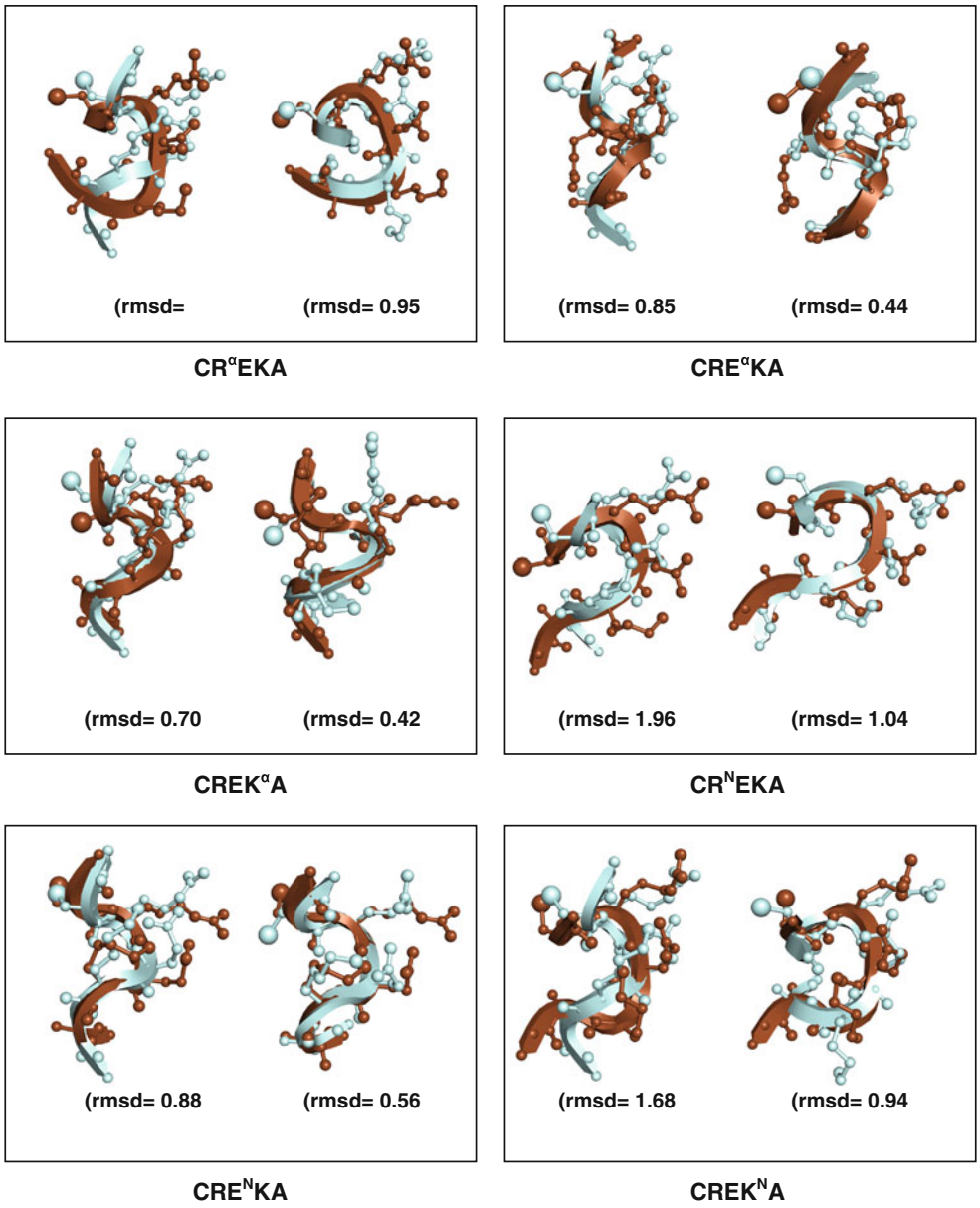


Fig. 5 Superimpositions of two selected structures of CREKA with the lowest energy conformation of each analog: the lowest energy conformation (*left side* of each pair) and the minimum energy conformation showing the highest resemblance with each analog (*right side* of each pair). The root mean square deviation (rmsd) between each pair of conformations is also indicated

Nonetheless, they clearly indicate that many of the targeted AAAAAAAAAA substitutions designed following this engineering protocol improve the biological activity of CREKA. Moreover, the substitution does not deter the clotting, with the results being practically identical to those obtained for CREKA-NWs.

Finally, the observed improvement in the tumor delivery observed for CREKA analogs is not due to the nanoworms but it is an intrinsic property, related to the stabilization of the bioactive conformation through a restriction of the conformational space (Imaging details are provided in Supplementary Material). Comparison of the binding obtained for CREKA and CREKA-NWs with that of CRE^{N}KA and CRE^{N}KA-NWs, respectively, shows that the accumulation of the peptide analog in B16F1 tumors is higher, independently of whether the peptide is presented individually or coating nanoworms. More detailed information about the binding affinity, homing to tumors, lack of homing to normal organs (non-tumor tissues) was reported in our previous work and its corresponding supporting information [5], the biomedical implications of these findings being beyond the scope of this work.

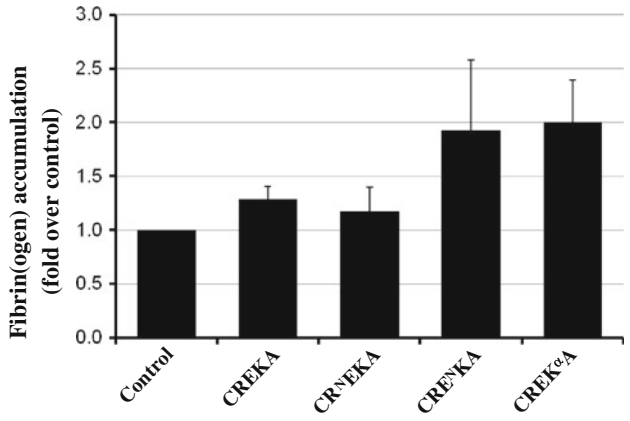


Fig. 6 Iron oxide nanoworms coated with FAM-labeled CREKA peptide and its analogs were intravenously injected (5 mg iron per kg body weight) into nude mice bearing orthotropic 22Rv1, human prostate tumors. The mice had been pre-injected with Ni-liposomes to reduce uptake by the reticulo-endothelial system. Tumors were harvested 5 h later, and tumor sections were stained with anti-fibrin(ogen) antibody. The stained sections were subjected to image analysis using Scanscope to quantify fibrin(ogen)-positive areas

Summary and outlook

A multidisciplinary approach, which includes systematic molecular modeling, synthetic organic chemistry, peptide synthesis and biological assays, has been used for rational design of CREKA analogs. Targeted mono-substitutions of specific natural amino acids by non-proteinogenic counterparts have been found to improve significantly both the biological activity and proteolytic resistance of the CREKA tumor-homing pentapeptide. The in vitro biological assays and the animal models, both using the peptide analogs that incorporate the synthetic α- and *N*-methyl derivatives of Arg, Glu and Lys, show an excellent correlation with the theoretical predictions obtained using a modified simulated annealing combined with MD. Although the combination of computational chemistry, experimental chemistry and biology is still in a nascent state, the three different disciplines can combine to address challenging translational medicine goals.

Acknowledgments Computer resources were generously provided by the Centre de Supercomputació de Catalunya (CESCA), the Barcelona Supercomputing Center–Centro Nacional de Supecomputación (BSC–CNS), 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). 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. The work developed by the Spanish groups has been supported by MICINN and FEDER (MAT2009-09138 and CTQ2010-17436), by the Generalitat de Catalunya (research group 2009 SGR 925 and XRQTC), and Gobierno de Aragon-FSE (research group E-40). Support for the research of C.A. was received through the prize "ICREA Academia" for excellence in research funded by the Generalitat de Catalunya. Research in the Ruoslahti laboratory is supported by grants from the NIH/NCI 5 P30 CA30199-28, awarded to the Sanford-Burnham Medical Research Institute, Cancer Center, and a DOD/CDMRP grant W81XWH-10-1-0199.

References

1. Simberg D, Duza T, Park JM, Essier M, Pilch J, Zhang L, Derfus AM, Yang M, Hoffman RM, Bathia S, Sailor MJ, Ruoslahti E (2007) Proc Natl Acad Sci USA 104:932–936
2. Hoffman JA, Laakkonen P, Porkka K, Bernasconi M, Ruoslahti E (2004) In: Clackson T, Lowman HB (eds) In vivo and ex vivo selections using phage-displayed libraries. Oxford University Press, Oxford
3. Zanuy D, Flores-Ortega A, Casanovas J, Curcó D, Nussinov R, Alemán C (2008) J Phys Chem B 112:8692–8700
4. Zanuy D, Curcó D, Nussinov R, Alemán C (2009) Biopolymers (Pept Sci) 92:83–93
5. Agemy L, Sugahara KN, Kotamraju VR, Gujraty K, Girard OM, Kono Y, Mattrey RF, Park J-H, Sailor MJ, Jimenez AI, Cativiela C, Zanuy D, Sayago FJ, Alemán C, Nussinov R, Ruoslahti E (2010) Blood 116:2847–2856
6. Kazmierski WM, Wire WS, Lui GK, Knapp RJ, Shook JE, Burks TF, Yamamura HI, Hruby VJ (1988) J Med Chem 31:2170–2177
7. Kazmierski WM, Yamamura HI, Hruby VJ (1991) J Am Chem Soc 113:2275–2283
8. Kyle DJ, Martin JA, Farmer SG, Burch RM (1991) J Med Chem 34:1230–1233
9. Tóth G, Darula Z, Peter A, Fülop F, Tourwé D, Jaspers H, Verheyden P, Bocskey Z, Tóth Z, Borsodi AJ (1997) Med Chem 40:990–995
10. Cowell SM, Lee YS, Cain JP, Hruby VJ (2004) Curr Med Chem 11:2785–2798
11. Hruby VJ, Balse PM (2000) Curr Med Chem 7:945–970
12. Hruby VJ (2001) Acc Chem Res 34:389–397
13. Lukaszuk A, Demaegdt H, Feytens D, Vanderheyden P, Vauquelin G, Tourwé D (2009) J Med Chem 52:5612–5618
14. Böhm H-J, Klebe G (1996) Angew Chem Int Ed 35:2588–2614
15. Loughlin WA, Tyndall JDA, Glenn MP, Fairlie PD (2004) Chem Rev 104:6085–6117
16. Hansen KK, Grosch B, Greiveldinger-Poenaru S, Bartlett PA (2003) J Org Chem 68:8465–8470
17. Tsantrizos YS, Bolger G, Bonneau P, Cameron DR, Goudreau N, Kukolj G, LaPlante SR, Llinas-Brunet M, Nar H, Lamarre D (2003) Angew Chem Int Ed 42:1355–1360
18. Nam N-H, Ye G, Sun G, Parang K (2004) J Med Chem 47:3131–3141
19. Ghosh AK, Swanson LM, Cho H, Leshchenko S, Hussain KA, Kay D, Walters DE, Koh Y, Mitsuya H (2005) J Med Chem 48:3576–3585
20. Toniolo C, Crisma M, Formaggio F, Valle G, Cavicchioni G, Précigoux G, Aubry A, Kamphuis J (1993) Biopolymers 33: 1061–1072

21. Toniolo C, Benedetti E (1991) Macromolecules 24:4004–4009
22. Toniolo C, Crisma M, Formaggio F, Peggion C (2001) Biopolymers (Pept Sci) 60:396–419
23. Moretto A, De Zotti M, Formaggio F, Kaptein B, Broxterman QB, Toniolo C (2005) Biopolymers (Pept Sci) 80:279–293
24. Teixido M, Albericio F, Giralt E (2005) J Pept Res 65:153–166
25. Rainaldi M, Moretto V, Crisma M, Peggion E, Mammi S, Toniolo C, Cavicchioni G (2002) J Pept Sci 8:241–252
26. Revilla-López G, Rodríguez-Ropero F, Curcó D, Torras J, Calaza MI, Zanuy D, Jiménez AI, Cativiela C, Nussinov R, Alemán C (2011) Proteins 79:1841–1852
27. Kirkpatrick S, Gelatt CD Jr, Vecchi MP (1983) Science 220: 671–680
28. Steinbach PJ, Brooks RB (1994) Chem Phys Lett 226:447–452
29. Baysal C, Meirovitch H (1999) J Comput Chem 20:1659–1670
30. Simmerling C, Elber R (1994) J Am Chem Soc 116:2534–2541
31. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935
32. Wang J, Cieplak P, Kollman PA (2000) J Comput Chem 21: 1049–1074
33. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman P (1995) J Am Chem Soc 117:5179–5197
34. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) J Chem Phys 81:3684–3690
35. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) J Comput Phys 23:327–341
36. Aurelio L, Brownlee RTC, Hughes AB (2004) Chem Rev 104: 5823–5846
37. Aurelio L, Brownlee RTC, Hughes AB (2002) Org Lett 4: 3767–3769
38. Biron E, Kessler H (2005) J Org Chem 70:5183–5189
39. Di Gioia ML, Leggio A, Liguori A, Perri F (2007) J Org Chem 72:3723–3728
40. Akaji K, Kuriyama N, Kiso Y (1996) J Org Chem 61:3350–3357
41. Park JH, von Maltzahn G, Zhang L, Derfus AM, Simberg D, Harris TJ, Ruoslahti E, Bhatia SN, Sailor MJ (2009) Small 5: 694–700
42. Cativiela C, Díaz-de-Villegas MD (1998) Tetrahedron Asymmetry 9:3517–3599
43. Cativiela C, Díaz-de-Villegas MD (2007) Tetrahedron Asymmetry 18:569–623
44. Leggio A, Belsito EL, De Marco R, Liguori A, Perri F, Viscomi MC (2010) J Org Chem 75:1386–1392
45. Karady S, Amato JS, Weinstock LM (1984) Tetrahedron Lett 25:4337–4340
46. Zanuy D, Flores-Ortega A, Jimenez AI, Calaza MI, Cativiela C, Nussinov R, Ruoslahti E, Aleman C (2009) J Phys Chem B 113:7879–7889
47. Revilla-Lopez G, Jimenez AI, Cativiela C, Nussinov R, Aleman C, Zanuy D (2010) J Chem Inf Model 2010(50):1781–1789
48. Revilla-Lopez G, Torras J, Nussinov R, Aleman C, Zanuy D (2011) Phys Chem Chem Phys 13:9986–9994