Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2006) 532–542
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Molecular dynamics and circular dichroism studies of human and rat C-peptides

Thiago Rennó Mares-Guia [a], Bernard Maigret [b,c], Natália Florêncio Martins [d], Ana Luiza Turchetti Maia [a], Luciano Vilela [e], Carlos Henrique Inácio Ramos [f], Luiz Juliano Neto [g], Maria Aparecida Juliano [g], Marcos Luiz dos Mares-Guia [a,e,✠], Marcelo Matos Santoro [a,*]

[a] *Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, 31270-901 Belo Horizonte, Brazil*
[b] *eDAM group, UMR CNRS/UHP 7565, Henri Poincaré University, Vandoeuvre les Nancy Cedex, France*
[c] *Universidade de Brasília, Brazil*
[d] *EMBRAPA – Recursos Genéticos e Biotecnologia, Laboratório de Bioinformática, W5 Norte, 70770-900 Brasília, Brazil*
[e] *Biomm S.A., Praça Carlos Chagas, 49, 8th Floor, 30170-020 Belo Horizonte, Brazil*
[f] *Centro de Biologia Molecular Estrutural, Laboratório Nacional de Luz Síncrotron, 13084-971 Campinas, Brazil*
[g] *Departamento de Biofísica, Escola Paulista de Medicina, Universidade Federal de São Paulo, 04044-020 São Paulo, Brazil*

Received 8 July 2005; received in revised form 11 March 2006; accepted 12 March 2006
Available online 5 June 2006

Abstract

Proinsulin C-peptide has been recently described as an endogenous peptide hormone, responsible for important physiological functions others than its role in proinsulin processing. Accumulating evidences that C-peptide exerts beneficial effects in the treatment of long term complications of patients with type 1 diabetes mellitus indicate that this molecule may be administered together with insulin in future therapies. Despite its clear pharmacological interest, the secondary and three-dimensional (3D) structures of human C-peptide are still points of controversy. In the present work we report molecular dynamics (MD) simulations of human, rat I and rat II C-peptides. A common experimental strategy applied to all peptides consisted of homology building followed by multinanosecond MD simulations in vacuum and water. Circular dichroism (CD) experiments of each peptide in the absence and presence of 2,2,2-trifluoroethanol (TFE) were performed to support validation of the theoretical models. A multiple sequence alignment of 23 known mammalian C-peptides was constructed to identify significant conserved sites that would be important for the maintenance of secondary and tertiary structures. The analysis of the molecular dynamics trajectories for the human, rat I and rat II molecules have shown quite different general behavior, being the human C-peptide more flexible than the two others. Human and rat C-peptides exhibit very stable turn-like structures at the middle and C-terminal regions, which have been described as potential active sites of C-peptides. Human C-peptide also presented a short alpha-helix throughout the MD, which was not found in the rat molecules. CD data is in very good agreement with the MD results and both methods were able to identify a greater structural stability and potential in rat C-peptides when compared to the human C-peptide. The simulation results are discussed and validated in the light of multiple sequence alignment, recent experimental data from the literature and our own CD experiments.

Keywords: C-peptide; Homology building; Molecular dynamics; Circular dichroism; Diabetes mellitus

1. Introduction

Human C-peptide is synthesized in the pancreatic beta cells as part of the preproinsulin polypeptide chain [1–3]. At the rough endoplasmic reticulum the N-terminal signal sequence of preproinsulin is cleaved and the resulting proinsulin molecule undergoes folding and disulfide bond formation [4]. In the Golgi apparatus proinsulin is packaged

* Corresponding author. Tel.: +55 31 3499 2627; fax: +55 31 3284 2013.
E-mail addresses: tmguia@uol.com.br (T.R. Mares-Guia), santoro@icb.ufmg.br (M.M. Santoro).
✠ In memoriam.

1093-3263/$ – see front matter © 2006 Published by Elsevier Inc.
doi:10.1016/j.jmgm.2006.03.002

into secretory granules and converted to insulin and C-peptide by specific proteases [5]. The conversion products are stored in mature secretory vesicles and eventually C-peptide and insulin are co-secreted in equimolar amounts into the portal circulation [6].

After the isolation and primary structure elucidation of the 31 amino acid residues human C-peptide [2] many studies focused on possible functions of this molecule. C-peptide plays a functional role in proinsulin folding, linking A and B chains of the insulin moiety and thus helping to provide optimal orientation of sulfhydryl groups for intramolecular disulfide bond formation [4,7]. Prohormone folding is also facilitated by an intramolecular chaperone-like action of C-peptide [8] and it was recently suggested that C-peptide regulates the kinetic folding pathway of proinsulin [9]. Furthermore, C-peptide contributes to the maintenance of the C-peptide/insulin A-chain junction structure in proinsulin, which is a recognition site for type II proinsulin endopeptidase [7].

Except for its clear structural role in proinsulin folding and conversion, C-peptide has long been considered to be a biologically inert molecule [10]. This statement was based on early studies that did not detect insulin-like effects of C-peptide and on the analysis of its primary structure, which varies considerably more among species when compared to insulin [4,10]. However, over the past few years many studies have demonstrated that C-peptide possesses biological activity both in animal models of diabetes and in patients with type 1 diabetes mellitus [11–13].

C-peptide binds specifically to target cells membranes through a putative G-protein-coupled receptor [14], resulting in Ca^{2+} influx and in the activation of the mitogen-activated protein (MAP) kinase signalling cascade, protein kinase C (PKC) isoforms and extracellular signal-regulated kinase (ERK) 1/2 [15]. Subsequently, endothelial nitric oxide (NO) synthase (eNOS) and Na^+, K^+-ATPase are activated [16], enzymes that are reported to be deficient in diabetes mellitus.

Initial information on C-peptide structure came from studies of bovine proinsulin in solution [17,18] and in crystal form, by X-ray diffraction [19]. Crystallographic data in low resolution showed that the tertiary structure of the insulin moiety of bovine proinsulin is very similar to the crystal structure of the insulin molecule and that bovine C-peptide would exhibit two short segments of α-helix (alpha helix) [17,19]. Circular dichroism (CD) spectra of bovine proinsulin described the occurrence of six residues in α-helix and nine residues in β (beta) structure in the C-peptide moiety [17].

Concerning structural information, previous theoretical models of human proinsulin [20,21] and human C-peptide [11] proposed contradictory data about the C-peptide structure. The N-terminal of these C-peptide models was structured as an α-helix [11,20] or a β-turn [21] and the middle segment could be an extended chain [21] or a turn [11,20]. The models partially agreed at the C-terminal region, were an α-helix was proposed to exist. However, molecular dynamics (MD) simulations were not performed to evaluate these C-peptide models.

Data obtained from CD spectra of human C-peptide [22] and infrared spectroscopy of human proinsulin [23] did not confirm the presence of residues in α-helix conformation and showed evidence that the C-peptide molecule is largely unordered except in a few β-turns [23]. These regions of turn structure would be crucial for proinsulin folding and would also serve as sites for enzymatic cleavage during proinsulin conversion [7,21].

It was recently suggested by low-resolution nuclear magnetic resonance (NMR) that the human C-peptide molecule does not possess a stable secondary structure both in aqueous solution and in the presence of lipid vesicles [24]. However, in the presence of 2,2,2-trifluoroethanol (TFE), C-peptide shows an α-helix that comprises its first eleven residues [24].

Definitive information on the secondary and tertiary structures of human C-peptide remains to be obtained, both in its free form and as part of the proinsulin molecule. Part of this problem resides in the fact that the C-peptide is very flexible, disallowing the obtainment of proinsulin crystals suitable for high-resolution crystallography studies [7]. High-resolution NMR studies of C-peptide structure in solution also have not been published until the present work was submitted for publication.

The present work intends to study the secondary and tertiary structures of human and rat C-peptides by means of sequence alignment analysis, homology modeling and long-range molecular dynamics simulations. This extensive simulation protocol, data obtained from our own CD studies and experimental data from the literature will be confronted and used for the validation of the C-peptide models.

2. Methodology

2.1. Peptide samples

Recombinant human C-peptide samples were kindly donated by Biomm S.A. (Belo Horizonte, MG, Brazil). Rat I and rat II C-peptides were synthesized by the solid-phase Fmoc (*N*-(9-fluorenyl)methoxycarbonyl) procedure on an automated bench-top simultaneous multiple solid-phase peptide synthesizer (PSSM 8 system from Shimadzu). The final peptides were deprotected in trifluoroacetid acid (TFA) and purified by semipreparative high performance liquid chromatography (HPLC) using an Econosil C-18 column (10 μm, 22.5 mm × 250 mm) and a two-solvent system: (A) TFA/H_2O (1:1000) and (B) TFA/acetonitrile (ACN)/H_2O (1:900:100). The column was eluted at a flow rate of 5 mL min^{-1} with a 10 (or 30)–50 (or 60)% gradient of solvent B over 30 or 45 min. Analytical HPLC was performed using a binary HPLC system from Shimadzu with a SPD-10AV Shimadzu UV–Vis detector, coupled to an Ultrasphere C-18 column (5 μm, 4.6 mm × 150 mm) which was eluted with solvent systems A_1 (H_3PO_4/H_2O, 1:1000) and B_1 (ACN/H_2O/H_3PO_4, 900:100:1) at a flow rate of 1.0 mL min^{-1} and a 10–80% gradient of B_1 over 20 min. The HPLC column eluates were monitored by their absorbance at 220 nm. The molecular weight and purity of synthesized peptides were checked by MALDI-TOF mass spectrometry (TofSpec-E, Micromass).

2.2. Sequence analysis and multiple alignment

Primary sequences of proinsulins from 21 mammal species were retrieved from the GenBank [25] and Swiss-Prot [26] databases. A multiple sequence alignment was built using ClustalX software [27], with default parameters. After alignment completion, N- and C-terminal parts of the sequences corresponding to insulin and the basic flanking residues were manually removed for simplification purposes, remaining only the C-peptide sequences. Identity and similarity scores were calculated using MatGAT application [28]. Accession numbers of the sequences used in the alignment are listed in Fig. 1.

2.3. Secondary structure prediction and template molecules selection

The human C-peptide sequence was submitted to the NPS@ web server [29] looking for a consensual secondary structure pattern. Sequence homology searches of the Protein Data Bank (PDB) [30], using the Fasta3 algorithm [31] were carried out for the complete human C-peptide and fragments of its sequence to identify possible template molecules. Resulting PDB entries homologous to the same C-peptide fragment were grouped together. A 3D template was selected inside each group based on tertiary structure consensus. The same protocol was applied independently to rat I and rat II C-peptides sequences.

2.4. Starting models building

Preliminary 3D ($3D^0$) models of human, rat I and rat II C-peptides were built using the Insight II molecular modeling software [32] and its Builder, Homology, Modeler and Discover modules. Template molecules selected in the previous step were aligned with their correspondent C-peptide targets and homology models for each peptide were constructed. Several energy minimization rounds refined each C-peptide model structure (1000 steps of steepest descents followed by conjugate gradients until convergence). The side-chain positions were first optimized keeping the full backbone fixed. This constraint was then removed until the conjugate gradient algorithm converged. The Insight II Discover module with the CVFF force field [32] was used for that purpose with a distance-dependent dielectric constant and no cut-off. No charged groups (N- and C-terminal groups, Asp and Glu side-chains) were considered at this point, to avoid unexpected ionic intramolecular interactions. These $3D^0$ models were used as the entry points for molecular dynamics (MD) simulations.

2.5. Molecular dynamics simulations

After the crude energy minimized $3D^0$ models for the three C-peptides were obtained, they were validated by means of a protocol based on MD simulations in vacuum and water using the Insight II Discover module with the force field CFF91 [32]. Initially, 6 ns MD simulations of each C-peptide were performed in vacuum in order to test their intrinsic stabilities. The most stable structures emerging from this procedure were energy-minimized and clustered to avoid redundancies, becoming the $3D^1$ models. Human, rat I and rat II $3D^1$ models were placed in periodic cubic boxes of 40 Å side length, each containing 1942, 1929 and 1930 water molecules, respectively. At this point, all the partial charges of ionizable groups were

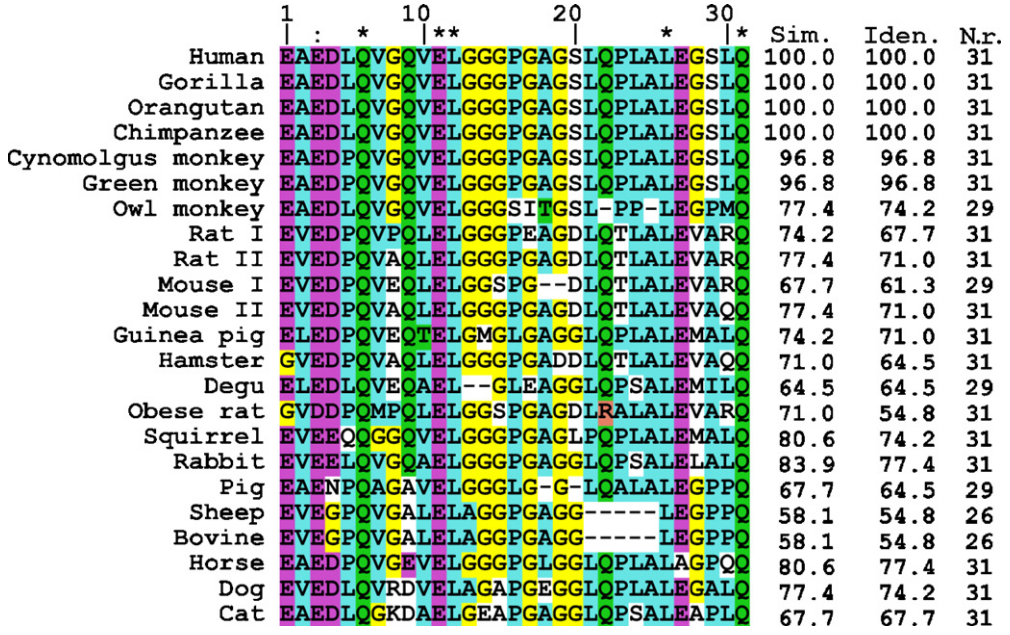

Fig. 1. Multiple sequence alignment of proinsulin C-peptides from 21 mammalian species. Asterisks indicate fully conserved positions. Two dots indicate an acidic character conservation. Similarity (Sim.) and identity (Iden.) scores and number of residues (N.r.) of each peptide are indicated. Swiss-Prot accession numbers: human (P01308), gorilla (Q8HZ81), orangutan (Q8HXV2), chimpanzee (P30410), cynomolgus monkey (P30406), green monkey (P30407), owl monkey (P10604), rat I (P01322), rat II (P01323), mouse I (P01325), mouse II (P01326), guinea pig (P01329), hamster (P01313), degu (P17715), obese rat (Q62587), squirrel (Q91XI3), rabbit (P01311), pig (P01315), sheep (P01318), bovine (P01317), horse (P01310), dog (P01321), cat (Q8WNW6).

introduced, the dielectric constant was set to 1 and a switching function was used to truncate the long-range nonbonded interactions with a cutoff of 10 Å. These $3D^1$ starting models were stabilized by an energy minimization step (conjugate gradients) and became the starting points for longer 10 or 15 ns MD simulations of the human and rat C-peptides in water. The shapes of the boxes were kept cubic and the average temperature and pressure were maintained at 300 K and 1 bar, respectively employing Langevin dynamics and the Langevin piston. At the completion of this step, configurations were extracted every 10 ps, resulting in samples of 1500, 993 and 983 configurations for human, rat I and rat II C-peptides, respectively. All MD simulations were performed at the CINES supercomputer center, Montpellier, France.

2.6. Circular dichroism measurements

Circular dichroism (CD) spectra were recorded on a JASCO J-810 spectropolarimeter equipped with a Peltier type temperature controller. The instrument was calibrated using (+)-10-camphorsulfonic acid [33]. Spectra were obtained from 190 to 250 nm at 25 °C using 0.1 cm path length cells. Stock solutions of recombinant human C-peptide and synthetic rat C-peptides were prepared in water to which appropriate volumes of phosphate buffer (to a final concentration of 10 mM) and TFE (0–90% v/v) were added. Following buffer and TFE addition, the final peptide concentrations were 50, 46 and 47 μM for human, rat I and rat II C-peptides, respectively. The solutions were stirred and incubated at room temperature 10 min before the spectrum was recorded. The pH of all samples was 7.4. The spectra were an average of eight scans recorded at a speed of 50 nm min^{-1}, with a bandwidth of 1.0 at 0.5 nm step size and a 2 s time constant. After background subtraction and smoothing, all of the CD data were converted to mean residue molar ellipticity [θ] ($^\circ$ cm^2 $dmol^{-1}$). The secondary structure content of each peptide was estimated from CD spectra using the programs PEPFIT [34] and K2D [35]. Percentage helicity was calculated as described by Forood et al. [36]. Peptide concentrations were determined by quantitative amino acid analysis.

3. Results

3.1. Sequence analysis and multiple alignment

Multiple sequence alignment of 23 C-peptides from 21 mammalian species (Fig. 1) revealed clusters of conserved residues. The N-terminal pentapeptide of the C-peptides constitute a highly conserved cluster that contains, with few exceptions, acidic residues in positions 1, 3 and 4 and hydrophobic residues in positions 2 and 5. The first four residues are critical for proinsulin conversion by endoprotease Type I (PC1/PC3) [37]. Alanine replacement mutation or deletion of the acidic residues in this region resulted in aggregation of human proinsulin during refolding experiments [8,9].

Gln6 is fully conserved in all mammalian sequences and it has been suggested that it could be involved in interactions with the insulin moiety of proinsulin [20].

Positions 10–12 comprise a conserved sequence X-Glu-Leu, where X is a hydrophobic residue, except for the Thr10 found in the guinea pig peptide. Following this region, the reasonably well conserved middle residues 13–19 of the majority of the peptides have a Pro16 flanked by glycine residues.

Human C-peptide residues 21–26 (LQPLAL) are also significantly preserved among mammal species. Almost all substitutions comprise substitutions of hydrophobic by uncharged residues. Leu26 is fully conserved in all species shown.

Glu27 is found in almost all species and, together with fully conserved Gln31, completes a C-terminal pentapeptide that has been demonstrated to bind specifically to human cells [14] and to respond for the whole activity of the entire molecule [12].

The lowest similarity and identity scores with human C-peptide were 58.1 and 54.8, respectively, obtained for the identical sheep and bovine C-peptides. The highest values were 96.8 of both similarity and identity for the cynomolgus and green monkeys. Excluding the identical sequences, average values of similarity and identity were 76.0 and 71.5, respectively.

A particular comparison between human and rat C-peptides reveals that the human sequence differ from rats I and II in 10 and 9 residues, respectively. In contrast, rats I and II sequences differ only in two residues from each other, at positions 8 and 17.

3.2. Template molecules selection and homology model building

Fasta searches in the PDB with several fragments of human, rat I or II C-peptides returned 26 PDB molecules which served as templates. Identities of templates were between 67% and 100% relative to the peptides fragments.

Human C-peptide templates found: sequence EDLQV returned entry 1B0M; sequence VGQVEL returned entries 1TNR, 3BTA, 1SKY, 1E20 and 5MSF; sequence LGGGPGAG returned 1B8M, 1B98 and 1UKY; sequence EGSLQ returned 1AX8, 1ELK, 1B7E, 1DI2, 4GSS, 1PVD and 1EDQ. Rat C-peptides templates found: sequence EDPQV returned entries 1FRT and 1ADQ; sequence LQTLAL returned 1POH and 1BUL; sequence EVARQ returned 1ZPD and 1BLX.

After a careful analysis, templates representing consensual structural motifs were used to build the homology models. Rat C-peptide II model was constructed using five different structural templates. The N-terminal segment EDPQV 3D coordinates were aligned with those of the 1FRT template; the following segment PQLEL was aligned with 3BTA; GGPEAG was aligned to 1UKY; LQTLAL was aligned to 1BUL and LEVARQ was aligned to 1BLX. This preliminary rat C-peptide II model $3D^0$ is shown in Fig. 2.

An α-helix was obtained from the 1BUL/1BLX 3D coordinates alignment at the C-terminal portion of rat C-peptide II. A loop in the middle segment of the model brings the N- and C-terminuses to a closer position. A turn-like structure was found in all three templates found for this segment. Template 3BTA caused the adjacent segment to adopt an extended conformation. The N-terminal portion of the model shows another turn-like configuration, determined by the 1FRT template.

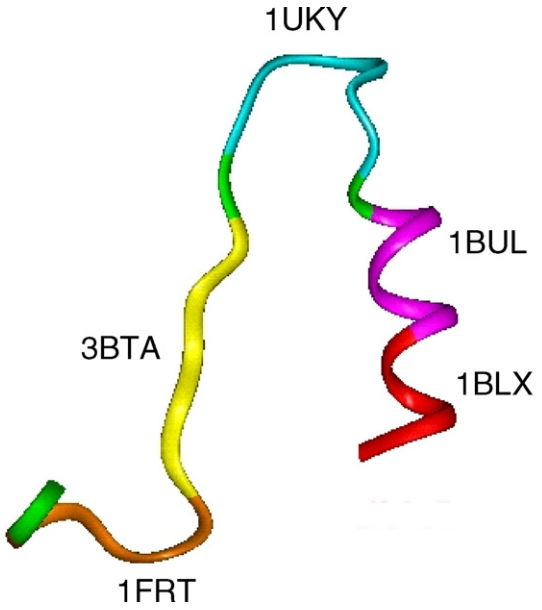

Fig. 2. Preliminary rat C-peptide II model, constructed using five different structural templates. N- and C-terminuses are on the left and right sides of the picture, respectively. The PDB codes for each template are indicated nearby their correspondent structural motif.

Rat C-peptide I model was built from the rat C-peptide II structure. Amino acid residues Ala8 and Gly17 of the rat II molecule were replaced by Pro8 and Glu17 to obtain the rat C-peptide I model $3D^0$.

The human $3D^0$ model was also constructed from the rat II preliminary model. The rat II molecule had residues at positions 2, 5, 8, 10, 20 and 23 replaced by the respective residues from the human C-peptide. The C-terminal pentapeptide EVARQ was replaced by the EGSLQ structure obtained from the PDB template. Pro23 disarranged the C-terminal α-helix in the human $3D^0$ model. This helix was rearranged manually and the structure was minimized and checked for quality. Considering the backbone C^α atoms, the distance root mean square deviation (RMSD) between the minimized and the original structures was always below 1.5 Å.

The human, rat I and rat II C-peptides homology models were then submitted to several rounds of minimization in order to eliminate the structural artifacts introduced into the models during the substitution of the residues and manipulations. The energy-minimized $3D^0$ structures were considered as the starting points for the respective MD simulations of the systems.

3.3. Molecular dynamics simulations

After recording the MD trajectories, a matrix of root mean square deviations (RMSD) was constructed from the structures generated during the conformational evolution of human C-peptide (Fig. 3A). This cluster graph compiles the RMSD between all the heavy atoms of C-peptide among 1500 conformations obtained from the MD trajectory. Structures close in time and with a small RMSD may belong to the same conformational family, where no substantial change in the structure occurred. The structures of the same conformational family tend to group in the shape of a square, as can be seen in Fig. 3A, where plotted points in pink colour corresponding to a RMSD between 4 and 6 Å highlight the existence of, broadly speaking, four structural families. A representative conformation was extracted from each structural family at 3720, 7980, 10 230 and 13 620 ps (Fig. 3A). Examination of the backbone for the four resulting conformations shows some regular secondary structure. All human C-peptide models have a short α-helix comprehending residues 20–23 (SLQP). Another short α-helical segment was formed only in the last representative model and this helix comprehends residues 3–6 (EDLQ) (Fig. 3A). All human models show turn-like structures comprising residues 14–18 (GGPGA) and 26–30 (LEGSL).

Two representative conformations were extracted from each rat C-peptide MD trajectory, based on their cluster graphics (Fig. 3B and C). The RMSD values for both rat trajectories are significantly lower than the human, indicating a higher structural stability of rat C-peptides I and II. Rat C-peptide I models are essentially representatives of the same conformational family. The representative models for rat I C-peptide exhibit turn-like motifs in the segments 14–19 (GGPEAG), 22–25 (QTLA) and in the C-terminus 27–31 (EVARQ), which were very stable during the MD (Fig. 3B). Rat C-peptide II models also presented turns in these same positions (Fig. 3C). The middle segment turn of rat C-peptide II seems to be more flexible than rat C-peptide I, which has a glutamic acid in position 17 instead of a glycine. Both rats I and II C-peptides did not form any α-helix throughout the MD.

Both the human and rat C-peptides underwent a conformational transition in the beginning of their dynamics which can be observed by the distance of the α-carbons of their terminal residues (Cα1–Cα31) during the simulation (Fig. 4). Rat C-peptides I and II maintained Cα1–Cα31 distances of 15.6 and 12.5 Å after 4034 and 1074 ps, respectively. In contrast, human C-peptide reached a stable Cα1–Cα31 distance of 7.5 Å after 2565 ps. This smaller Cα1–Cα31 distance in the human molecule is maintained by a very stable salt bridge between the NH_3^+ of the N-terminus and the COO^- of the C-terminus with a water molecule bridge between (not shown). Nothing like this was observed for the rat C-peptides.

The evolution of the dipole moment (Fig. 4) shows larger fluctuations for the human (between 41 and 169 Debyes) than for the rats (13–93 for rat I, 25–111 for rat II). This shows clearly a completely different electrostatic behavior between the human and the rats, due to the different charged residues. Evolution of the radius of gyration (Fig. 4) revealed the same essential behavior for the three peptides, with fluctuations between 9–13 Å for human and 9–10 Å for rats I and II molecules. This result indicates that the human structure is more elongated than the rat C-peptides.

3.4. Circular dichroism measurements

Circular dichroism spectra of human and rat C-peptides shows that these molecules have very low content of secondary structure in the absence of TFE (Fig. 5A). However, the molar

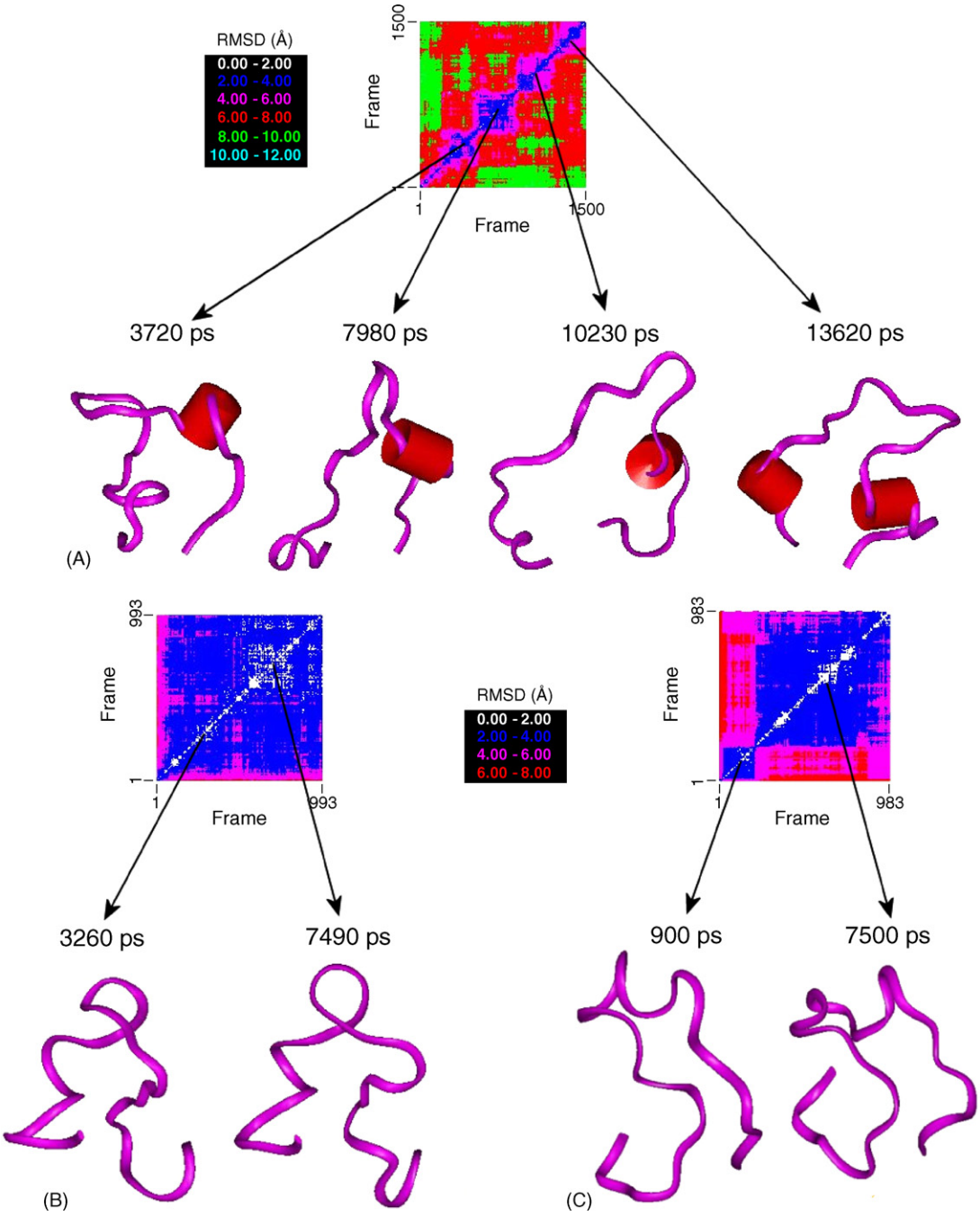

Fig. 3. Evolution of the conformations of human (A), rat C-peptide I (B) and rat C-peptide II (C) during the molecular dynamics simulation in solution. Each model is a snapshot of representative conformational families. Red cylinders represent alpha helix segments. All molecules are oriented with the N-terminal in the left side of the models.

residual ellipticity at 222 nm ($\Theta_{222\ nm}$), which indicates helix content, is quite different between the human (−1696) and the rats C-peptides (−4 744 and −4 040 for rats I and II, respectively). This secondary structure content difference was magnified when 45% TFE was added to the peptide solutions (Fig. 5B). Rat C-peptides I and II show similar potential to acquire secondary structure under influence of TFE, with a $\Theta_{222\ nm}$ of −13 986 and −14 625, respectively. Human C-peptide acquires less helical content than the rat C-peptides in the presence of TFE ($\Theta_{222\ nm}$ of −5304). Deconvolution of the CD data obtained in the absence of TFE by two different programs is compared to secondary structure prediction consensus (Table 1). Compared to CD results by both methods, secondary structure predictions overestimated the α-helical content of the three peptides. β-sheet and random coil content values calculated using PEPFIT were much closer to the predictions than the values generated with K2D. Furthermore, K2D predicted a much higher content of α-helix for rat C-peptide I than rat C-peptide II, a result that was not observed with PEPFIT. PEPFIT allows the calculation of turn content

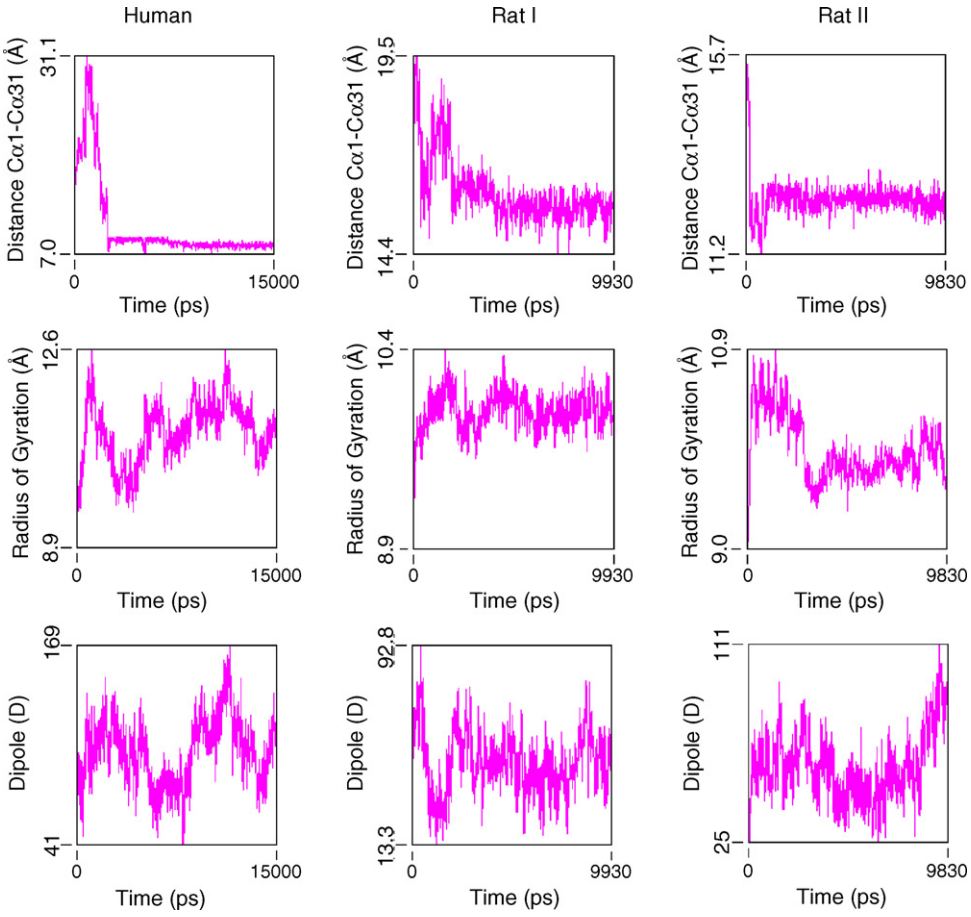

Fig. 4. Trajectory analyses of human, rat I and rat II C-peptides molecular dynamics simulations in solution.

from CD data, an advantage considering the turn-like motifs proposed to exist in the representative C-peptides conformations from MD simulations. PEPFIT indicates two residues in turn conformation in the human and rat C-peptides. When no turn content was considered in the PEPFIT deconvolution, curve-fitting was not satisfactory. It is interesting to comment that two algorithms from the secondary structure consensus predicted four residues in turn conformation, which were not considered in the consensual result.

These spectra show that rat C-peptides have a more stable structure than the human molecule, results that corroborate the MD simulations in which substantially lower RMSD values were found for the rat C-peptides trajectories when compared to human. Beta-turn tendencies and β-strand content detected by CD also agree with the turn motifs formed in all three C-peptides models. Alpha-helices of the human theoretical models were not detected and this may be due to the size of these helices which are quite small and could not be recognized

Table 1
Secondary structure content of human and rat C-peptides determined by circular dichroism experiments and secondary structure prediction consensus

Method	Alpha helix			Beta sheet			Random coil			Turn[d]		
	Human	Rat I	Rat II	Human	Rat I	Rat II	Human	Rat I	Rat II	Human	Rat I	Rat II
Prediction[a]	5	9	9	5	4	6	21	18	16	–	–	–
	16%	29%	29%	16%	13%	19%	68%	58%	52%			
CD K2D[b]	1	6	2	6	9	7	24	16	22	–	–	–
	3%	20%	5%	20%	30%	22%	77%	50%	74%			
CD PEPFIT[c]	1	4	3	6	5	6	22	20	20	2	2	2
	4%	13%	11%	20%	14%	18%	71%	66%	66%	5%	7%	5%

Numbers on top and bottom are number of residues and percentage of secondary structure, respectively.
[a] Data from secondary structure prediction consensus (see Section 2).
[b] CD data calculated using K2D software [35].
[c] CD data calculated using PEPFIT software [34], which provides turn content calculation.
[d] Turn content calculation is not available in the secondary structure prediction consensus and K2D.

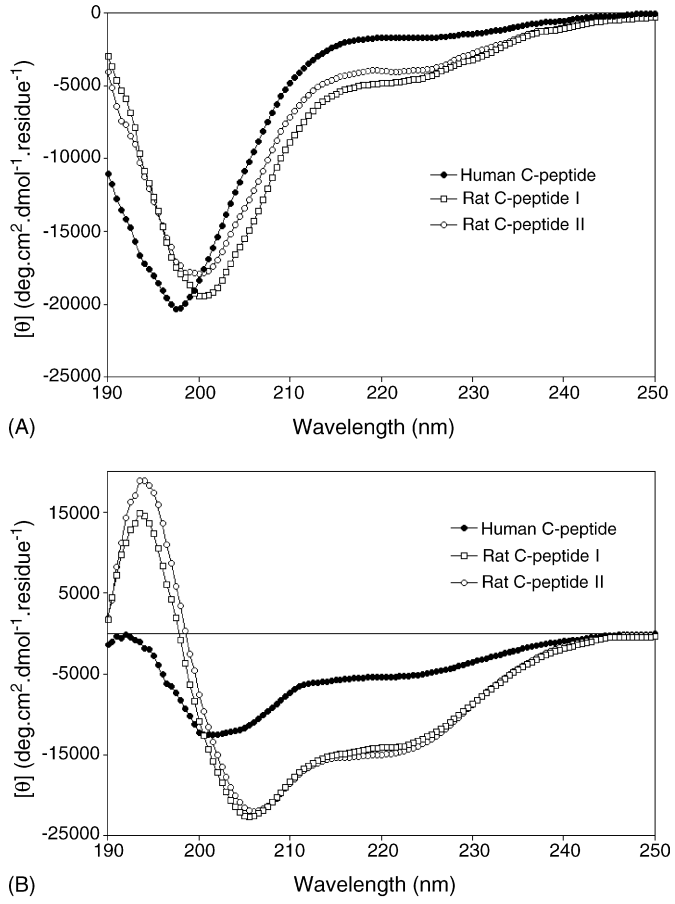

Fig. 5. Circular dichroism spectra of human and rat C-peptides in 10 mM sodium phosphate buffer pH 7.4 at 25 °C in the absence (A) and presence (B) of 45% TFE.

by the CD analysis algorithms, which are based on relatively high amplitude reference curves exhibited by regular secondary structure such as helix and sheet [34]. Nevertheless, human and rat C-peptides present a relatively similar secondary structure but a significantly different structural potential.

4. Discussion

The molecular dynamics simulations and circular dichroism experiments indicate that human and rat C-peptides are not random coil molecules. Local and global ordered structures in the C-peptides are suggested by the analysis of the present results and these findings will be discussed and confronted to the literature.

The high structural flexibility evidenced in the MD trajectories of human, rat I and rat II C-peptides molecules would indicate that the time scale of 10–15 ns employed in their simulations in solution was not sufficient to cover the whole conformational space of the investigated peptides. However, equilibrium conformations will not be obtained even with longer MD runs. The high flexibility is a real characteristic of human and rat C-peptide molecules that has been evidenced by FT-IR [23], low-resolution NMR [7,24] and, more recently, by high-resolution 2D NMR experiments [41].

With a closer analysis of the RMSD graphs and the overall structures of the representative models shown in Fig. 3 it is possible to observe that despite their high flexibility, the main structural features of the three peptides do not change substantially throughout the dynamics. The simulation time scale was long enough for the molecules to achieve and stabilize important turn-like structures in the middle and C-terminal regions, the most important characteristics of the human and rat C-peptides conformers. Considering the evolution of the lengths of the simulations published during the last years, our MD time scale is in the range of what is considered presently as acceptable [38–40]. Khandelia and Kaznessis [40] recently reported 16 ns MDs of a small antimicrobial peptide and two analogues in SDS micelles, with a total time of nearly 50 ns of simulation. In the present work we performed a total simulation time of 18 ns in vacuum followed by a 40 ns total simulation in water for human, rat I and rat II C-peptides. It seems like an experimental validation rather than an extension of the MD time scale needs to be performed to validate the present simulation protocol. In this light, the results will be discussed considering circular dichroism experiments, experimental data from the literature and a recent paper [41], published after the present work was submitted, in which the solution structure of human C-peptide obtained by high-resolution 2D NMR is described (see note added in proof).

The turn-like middle regions of human (Fig. 3A) and rat C-peptides molecules (Fig. 3B and C) would drive a tendency of the peptides to keep a bent conformation in which glycine residues and pro16 exert an important role. Pro16 is conserved in 19 of the 23 mammalian C-peptides aligned and three to six glycine residues are found in the middle region of all sequences (Fig. 1). The occurrence of a turn [2,4,11,20] or a half-β-turn [21] in this segment of human C-peptide has been previously hypothesized. However, this is the first time that a protocol based on long-range MD was employed to validate this hypothesis. The finding that the middle segment of human and rat C-peptides adopt a stable turn-like structure throughout the MD simulations is of particular interest as this glycine-rich sequence has been found to be important for the biological function of human C-peptide, responding for nearly the full activity of the entire molecule [11]. Furthermore, this putative central turn brings the C-peptides models N- and C-terminuses relatively close to each other, a structural condition that has been experimentally demonstrated in previous CD studies of bovine and porcine C-peptides [42,43]. When part of the proinsulin molecule, C-peptide would present this bent conformation in order to be attached to the C-terminus of the B and to the N-terminus of the A chains of insulin, which are only 8–10 Å apart from each other [44,45]. It was recently evidenced by high-resolution NMR studies of human C-peptide in solution that the middle region of the molecule, comprehending residues 15–18, presents a tendency to form a β-bend [41] what is consistent with our results. According to this experimental work the bent conformation is also likely to occur in part of the structures analyzed, but with a longer distance between the N- and C-terminuses [41].

Human and rats C-peptides molecules exhibited during their MD trajectories another very stable turn-like structure at their C-terminal sequences LEGSL and EVARQ, respectively. It is important to observe that these regions correspond to the human pentapeptide EGSLQ, which interacts specifically with human cell membranes [14] and is responsible for up to 100% activity of the whole peptide in some in vitro assays [12]. This same Glu 27 to Gln 31 sequence of human C-peptide was recently reported to be structured in a very stable type-III' β-turn in high-resolution NMR experiments [41]. A β-turn encompassing this segment has also been suggested in a theoretical model of human proinsulin [21]. Simulations and low-resolution NMR experiments proposed that when part of the proinsulin molecule, this C-peptide sequence would adopt a stable local structure and interact with the pancreatic convertases at the CA junction [7,21]. It has been postulated that the physiological functions of C-peptide were acquired during mammalian evolution, since residue conservation is particularly high in this class of animals [13,45]. It is plausible to suppose that this pentapeptide may adopt a turn-like conformation in mammals, since Gln31 is fully conserved and Glu27 is conserved in 22 of 23 mammalian C-peptide sequences (Fig. 1).

Experimental evidences that β-structure exists in C-peptides were demonstrated in CD spectra of human [22], porcine [43,46] and bovine [46] C-peptides and in FT-IR [23] and low-resolution NMR [7] studies of human proinsulin. Our CD data show the occurrence of five to six residues in β-structure and two residues in β-turn conformation, indicating the real existence of such motifs in human, rat I and rat II C-peptides (Table 1). Clear evidence for β-turns in the same middle and C-terminal regions of human C-peptide has recently been obtained by high-resolution 2D-NMR experiments in solution [41], in excellent agreement with the representative conformations of human and rat C-peptides shown in Fig. 3.

The α-helical contents of human and rat C-peptides were overestimated by secondary structure predictions compared to the simulation and CD results. Considering the simulation results, human C-peptide representative models present a short segment of α-helix encompassing residues 20–23, which persisted during the entire simulation. The correspondent regions of rat I and rat II molecules exhibit turn-like structures. Another less stable helix comprehending residues 3–6 was also found in one of the representative human C-peptide conformers. High-resolution NMR studies of human C-peptide indicate that residues in positions 2–5 and 22–27 possess a helical tendency, consisting of short unstable helices or turns coexisting in solution [41]. In this context, the apparent contrast in α-helix estimation considering the C-peptide representative models and CD results could be interpreted as the three peptides having a tendency to form local stable helices or turns in some regions of their sequences. To further explore this tendency, 2,2,2-trifluoroethanol was included in the CD experiments. As observed in Fig. 5B, TFE addition stabilizes local α-helices and induces further helix formation in all three peptides. CD spectra of human, rat I and rat II C-peptides in increasing TFE concentrations exhibit an isodichroic point at 204 nm, typical of peptides undergoing coil-helix transitions [47] (data not shown). CD results in the absence of TFE indicate more helical content in rats than in human C-peptides (Table 1), suggesting a more stable structure of rats C-peptides. In the presence of 45% TFE, rat C-peptides I and II adopt a clear helical conformation, reaching 36% and 37% helicity, respectively, while human C-peptide reaches only 14% helicity at the same TFE concentration. Above 45% TFE, only a small gain in helicity is observed in human C-peptide and no significative gain in helicity is detected in rat C-peptides (data not shown). Rat C-peptides I and II presented a very similar behavior during the MD simulations and no helix was detected in both rat C-peptides. Nevertheless, rat C-peptides trajectories RMSD values were around 3 Å, in a significantly lower level than the values observed for the human molecule, which were in the range from 4 to 6 Å. Studies of human C-peptide in solution by high-resolution 2D-NMR detected a RMSD of 4.7 Å [41], in very good agreement with our simulation results. Both MD and CD methods were able to detect an important difference in structure stability between rats and human C-peptides.

Our final representative models exhibit substantial differences when compared to human and rat C-peptides models proposed by Ido et al. [11], which were constructed based solely on secondary structure predictions and have not been submitted to validation protocols such as molecular dynamics simulations. The main differences reside in the absence of the C-terminal turn-like structure and in the α-helix distribution and content, overestimated in human and rat C-peptides models suggested by these authors [11].

Previous low-resolution NMR studies of human C-peptide at 40 °C, pH 3.6 and 95% TFE proposed its N-terminal 11 residues to adopt an α-helical conformation while residues 12–31 would be in random-coil, which would represent a 35% helicity [24]. CD spectra from these authors also detected a 35–40% helicity of C-peptide in 95% TFE, but at an even lower pH, since human C-peptide was dissolved in a water/TFE solution. In contrast, CD data presented here and collected at 25 °C and pH 7.4 detected 14% helicity in 45% TFE and only 23% helicity in 90% TFE (data not shown). This divergence would be due to the different experimental conditions which would also explain the fact that no turn-like motifs were detected in the experiments performed by Henriksson et al. [24]. Furthermore, only short unstable helices or turns were observed in this N-terminal segment by high-resolution NMR experiments of human C-peptide at 80% TFE and pH 7.0 [41], in good agreement with our CD and simulation results.

The MD protocol and CD experiments employed in the validation of the C-peptides structures constitute an improvement in the fidelity of human and rat C-peptides models. Important features of previously proposed models of C-peptides were not observed after long-range molecular dynamics, such as stable and relatively long α-helices in human and rat molecules [11]. The most significant and physiologically relevant finding detected in our representative structures and not exhibited by previous models was the presence of a very stable turn-like motif in the C-terminal region of human and rat C-peptides. The other turn-like structure situated in the middle region of the C-peptides was also maintained throughout the MD simulations. The main

features of the human, rat I and rat II representative models are these two conserved and turn-like structured sequences that may represent a common motif to all mammalian C-peptides.

5. Conclusions

The analyses of the molecular dynamics trajectories and circular dichroism spectra of human, rat I and rat II C-peptides provided important information on the structures of these molecules. The validation of the main structural features of the representative conformations was achieved by a long-range simulation protocol, in which stable local structures were preserved, and by CD studies which provided experimental evidence to corroborate the MD results. According to our findings human and rat C-peptides possess segments with stable secondary structure and are not random-coil molecules, a fact that was experimentally proved by other authors through different spectroscopic methods. A substantial difference in structural stability and potential was detected between human and rat C-peptides. As demonstrated by MD and CD, rat C-peptides I and II are more stable and acquire more secondary structure in the presence of TFE than human C-peptide. The middle regions of all C-peptides exhibit a turn-like structure, which would drive a tendency of the molecules to adopt a bent conformation. The C-terminal pentapeptide of human and rat C-peptides molecules also acquired a very stable turn-like conformation. CD spectra are in excellent agreement with the proposed structures for all C-peptides. Furthermore, these findings were recently validated by a paper that was published after the present work was submitted, in which the same middle and C-terminal regions of human C-peptide were described to adopt a turn-like conformation (see Section 6). Since both the middle and C-terminal segments are significantly conserved in all mammalian C-peptides and are reported to be responsible for the full activity of the entire molecule through interactions with putative C-peptide receptors, it is plausible to consider that these regions of other mammalian C-peptides may adopt the stable turn-like structures observed in human and rat molecules.

6. Note added in proof

Following submission of this manuscript, Munte and co-workers reported the solution structure of human C-peptide obtained by high-resolution 2D-NMR experiments [41]. Consistent with our MD and CD results, very stable turn-like structures were clearly detected at the same middle and C-terminal regions of human C-peptide, being the C-terminal pentapeptide extremely well-defined. Some of the NMR conformations are very similar to the human C-peptide representative models of the present work, corroborating the structures suggested by our simulation and experimental data.

Acknowledgements

This work was funded by the ‘‘Support of Science and Technology Program’’ (PADCT III), from Brazilian Government. T.R.M.G. was supported by fellowships from FAPEMIG, CAPES and CNPq. B.M. thanks CAPES for providing a researcher fellowship during his stay at Universidade de Brasília. T.R.M.G. would like to thank Jamil Silvano and people at Biomm S.A., for the opportunity of working with them and for the kind donation of recombinant human C-peptide samples. T.R.M.G. dedicates this paper to the memory of the inspiring and unforgettable Dr. Marcos Luiz dos Mares-Guia (1935–2002).

References

[1] D.F. Steiner, D. Cunningham, L. Spigelman, B. Aten, Insulin biosynthesis: evidence for a precursor, Science 157 (1967) 697–700.
[2] P.E. Oyer, S. Cho, J.D. Peterson, D.F. Steiner, Studies on human proinsulin: Isolation and aminoacid sequence of the human pancreatic C-peptide, J. Biol. Chem. 246 (1971) 1375–1386.
[3] S.J. Chan, P. Keim, D.F. Steiner, Cell-free synthesis of rat preproinsulins: characterization and partial amino acid sequence determination, Proc. Natl. Acad. Sci. U.S.A. 73 (1976) 1964–1968.
[4] D.F. Steiner, On the role of the proinsulin C-peptide, Diabetes 27 (1978) 145–148.
[5] H.W. Davidson, C.J. Rhodes, J.C. Hutton, Intraorganellar calcium and pH control proinsulin cleavage in the pancreatic beta cell via two distinct site-specific endopeptidases, Nature 333 (1988) 93–96.
[6] A.H. Rubenstein, J.L. Clark, F. Melani, D.F. Steiner, Secretion of proinsulin C-peptide by pancreatic beta cells and its circulation in blood, Nature 224 (1969) 697–699.
[7] M.A. Weiss, B.H. Frank, I. Khait, A. Pekar, R. Heiney, S.E. Shoelson, L.J. Neuringer, NMR and photo-CIDNP studies of human proinsulin and prohormone processing intermediates with application to endopeptidase recognition, Biochemistry 29 (1990) 8389–8401.
[8] L.M. Chen, X.W. Yang, J.G. Tang, Acidic residues on the N-terminus of proinsulin C-peptide are important for the folding of insulin precursor, J. Biochem. 131 (2002) 855–859.
[9] Z.S. Qiao, C.Y. Min, Q.X. Hua, M.A. Weiss, Y.M. Feng, In vitro refolding of human proinsulin, J. Biol. Chem. 278 (2003) 17800–17809.
[10] A.E. Kitabchi, Proinsulin and C-peptide: a review, Metabolism 26 (1977) 547–587.
[11] Y. Ido, A. Vindigni, K. Chang, L. Stramm, R. Chance, W.F. Heath, R.D. DiMarchi, E. Di Cera, J.R. Williamson, Prevention of vascular and neural dysfunction in diabetic rats by C-peptide, Science 277 (1997) 563–566.
[12] J. Wahren, K. Ekberg, J. Johansson, M. Henriksson, A. Pramanik, B.L. Johansson, R. Rigler, H. Jörnvall, Role of C-peptide in human physiology, Am. J. Physiol. Endocrinol. Metab. 278 (2000) 759–768.
[13] J. Wahren, C-peptide: new findings and therapeutic implications in diabetes, Clin. Physiol. Funct. Imaging 24 (2004) 180–189.
[14] R. Rigler, A. Pramanik, P. Jonasson, G. Kratz, O.T. Jansson, P.-Å. Nygren, S. Ståhl, K. Ekberg, B.-L. Johansson, S. Uhlén, M. Uhlén, H. Jörnvall, J. Wahren, Specific binding of proinsulin C-peptide to human cell membranes, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 13318–31323.
[15] Z. Zhong, O. Kotova, A. Davidescu, I. Ehrén, K. Ekberg, H. Jörnvall, J. Wahren, A.V. Chibalin, C-peptide stimulates Na^+, K^+-ATPase via activation of ERK1/2 MAP kinases in human renal tubular cells, Cell. Mol. Life Sci. 61 (2004) 2782–2790.
[16] Z. Zhong, A. Davidescu, I. Ehrén, K. Ekberg, H. Jörnvall, J. Wahren, A.V. Chibalin, C-peptide stimulates ERK1/2 and JNK MAP kinases via activation of protein kinase C in human renal tubular cells, Diabetologia 48 (2005) 187–197.
[17] J. Markussen, Structural changes involved in the folding of proinsulin, Int. J. Protein Res. 3 (1971) 201–207.
[18] B.H. Frank, A.H. Pekar, A.J. Veros, Insulin and proinsulin conformation in solution, Diabetes 21 (1972) 486–491.
[19] W.W. Fullerton, R. Potter, B.W. Low, Proinsulin: crystallization and preliminary X-ray diffraction studies, Proc. Natl. Acad. Sci. U.S.A. 66 (1970) 1213–1219.

[20] C.R. Snell, D.G. Smyth, Proinsulin:, A proposed three-dimensional structure, J. Biol. Chem. 250 (1975) 6291–6295.
[21] G. Lipkind, D.F. Steiner, Predicted structural alterations in proinsulin during its interactions with prohormone convertases, Biochemistry 38 (1999) 890–896.
[22] V.K. Naithani, M. Dechesne, J. Markussen, L.G. Heding, Improved synthesis of human proinsulin C-peptide and its benzyloxycarbonyl derivative. Circular dichroism and immunological studies of human C-peptide, Hoppe-Seyler's Z. Physiol. Chem. 356 (1975) 997–1010.
[23] L. Xie, C.-L. Tsou, Comparison of secondary structures of insulin and proinsulin by FT-IR, J. Protein Chem. 12 (1993) 483–487.
[24] M. Henriksson, J. Shafqat, E. Liepinsh, M. Tally, J. Wahren, H. Jörnvall, J. Johansson, Unordered structure of proinsulin C-peptide in aqueous solution and in the presence of lipid vesicles, Cell. Mol. Life Sci. 57 (2000) 337–342.
[25] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, D.L. Wheeler, GenBank: Update, Nucl. Acids Res. 32 (2004) 23–26.
[26] R. Apweiler, A. Bairoch, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi, L.S. Yen, Uniprot: the universal protein knowledgebase, Nucl. Acids Res. 32 (2004) 115–119.
[27] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, J.D. Thompson, Multiple sequence alignment with the Clustal series of programs, Nucl. Acids Res. 31 (2003) 3497–3500.
[28] J.J. Campanella, L. Bitincka, J. Smalley, MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences, BMC Bioinformat. 4 (2003) 29.
[29] C. Combet, C. Blanchet, C. Geourjon, G. Deléage, NPS@: Network protein sequence analysis, Trends Biochem. Sci. 25 (2000) 147–150.
[30] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucl. Acids Res. 28 (2000) 235–242.
[31] W.R. Pearson, D.J. Lipman, Improved tools for biological sequence comparison, Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 2444–2448.
[32] Insight II, Discover force fields CVFF and CFF91, Accelrys Corp., San Diego, CA, USA.
[33] G.C. Chen, J.T. Yang, Two-point calibration of circular dichrometer with d-10-camphorsulfonic acid, Anal. Lett. 10 (1977) 1195–1207.
[34] J. Reed, T.A. Reed, A set of constructed type spectra for the practical estimation of peptide secondary structure from circular dichroism, Anal. Biochem. 254 (1997) 36–40.
[35] M.A. Andrade, P. Chacón, J.J. Merelo, F. Morán, Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network, Prot. Eng. 6 (1993) 383–390.
[36] B. Forood, E.J. Feliciano, K.P. Nambiar, Stabilization of α-helical structures in short peptides via end capping, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 838–842.
[37] J.E. Kaufmann, J.C. Irminger, P.A. Halban, Sequence requirements for proinsulin processing at the B-chain/C-peptide junction, Biochem. J. 310 (1995) 869–874.
[38] G.S. Jas, K. Kuczera, Equilibrium structure and folding of a helix-forming peptide: circular dichroism measurements and replica-exchange molecular dynamics simulations, Biophys. J. 87 (2004) 3786–3798.
[39] V.P. Raut, M.A. Agashe, S.J. Stuart, R.A. Latour, Molecular dynamics simulations of peptide–surface interactions, Langmuir 21 (2005) 1629–1639.
[40] H. Khandelia, Y.N. Kaznessis, Molecular dynamics simulations of helical antimicrobial peptides in SDS micelles: what do point mutations achieve? Peptides 26 (2005) 2037–2049.
[41] C.E. Munte, L. Vilela, H.R. Kalbitzer, R.C. Garratt, Solution structure of human proinsulin C-peptide, FEBS J. 272 (2005) 4284–4293.
[42] J. Markussen, H.E. Schiff, Molecular parameters of C-peptide from bovine proinsulin, Int. J. Peptide Protein Res. 5 (1973) 69–72.
[43] H.-P. Vogt, A. Wollmer, V.K. Naithani, H. Zahn, The conformational potential of porcine proinsulin C-peptide, Hoppe-Seyler's Z. Physiol. Chem. 357 (1976) 107–116.
[44] T.L. Blundell, G.G. Dodson, D.C. Hodgkin, D.A. Mercola, Insulin: the structure in the crystal and its reflection in chemistry and biology, Adv. Prot. Chem. 26 (1972) 279–402.
[45] D.F. Steiner, The proinsulin C-peptide—a multirole model, Exp. Diabesity Res. 5 (2004) 7–14.
[46] J. Markussen, A. Volund, Conformation analysis of circular dichroism spectra of insulin, proinsulin and c-peptides by non-linear regression, Int. J. Peptide Protein Res. 7 (1975) 47–56.
[47] R.W. Woody, Circular dichroism, in: K. Sauer (Ed.), Methods in Enzymology, vol. 246, Academic Press, New York, 1995, pp. 34–71.