



Biochemical, biophysical and molecular dynamics studies on the proteoglycan-like domain of carbonic anhydrase IX

Emma Langella¹ · Martina Buonanno¹ · Daniela Vullo² · Nina Dathan³ · Marilisa Leone¹ · Claudiu T. Supuran² · Giuseppina De Simone¹ · Simona Maria Monti¹

Received: 28 November 2017 / Revised: 2 March 2018 / Accepted: 13 March 2018 / Published online: 21 March 2018
© Springer International Publishing AG, part of Springer Nature 2018

Abstract

Human carbonic anhydrase IX (hCA IX) is a tumour-associated enzyme present in a limited number of normal tissues, but overexpressed in several malignant human tumours. It is a transmembrane protein, where the extracellular region consists of a greatly investigated catalytic CA domain and a much less investigated proteoglycan-like (PG) domain. Considering its important role in tumour biology, here, we report for the first time the full characterization of the PG domain, providing insights into its structural and functional features. In particular, this domain has been produced at high yields in bacterial cells and characterized by means of biochemical, biophysical and molecular dynamics studies. Results show that it belongs to the family of intrinsically disordered proteins, being globally unfolded with only some local residual polyproline II secondary structure. The observed conformational flexibility may have several important roles in tumour progression, facilitating interactions of hCA IX with partner proteins assisting tumour spreading and progression.

Keywords Tumour · Polyproline II (PPII) · Intrinsically disordered protein (IDP) · Natively unfolded · Cancer · Hypoxia · Metalloenzyme

Introduction

Carbonic anhydrases (CAs) are ubiquitous metallo-enzymes catalysing the reversible hydration of CO_2 to HCO_3^- and H^+ [1]. In humans, among the 12 catalytically active isoforms

(CA I–IV, VA–VB, VI–VII, IX, and XII–XIV), CA IX has been recognized as a tumour-associated protein [2–5]. In fact, apart its expression in a limited number of normal tissues with an almost total exclusivity in the gastrointestinal tract epithelium [6, 7], CA IX is overexpressed in the cell membrane of several malignant tumour cells, where it is generally associated with the hypoxic phenotype mediated by the transcription factor HIF-1 [7]. In tumours, CA IX modulates growth, survival, proliferation, adhesion, and invasion of malignant cells [8] by means of several mechanisms, such as tumour pH regulation, interference with the Rho/ROCK signaling pathway [9] and interaction of the enzyme with β -catenin, which causes destabilization of intercellular adhesion [3, 5, 10, 11]. Among these mechanisms, the most investigated one regards the pH regulation of cancer cells here summarized. Upon hypoxia, HIF-1 transcription factor activates several specific genes which lead to up-regulation of glycolysis and, therefore, to an over-production of lactate and protons. To maintain a normal intracellular pH (pH_i) [7, 12], these ions are extruded by means of monocarboxylate transporters (MCTs), pumps such as the V-type H^+ ATPase (V-ATPase) and H^+ exchangers as the Na^+/H^+ exchanger (NHE). Alternatively, the formed H^+ ions are titrated by

Emma Langella and Martina Buonanno contributed equally to the work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00018-018-2798-8>) contains supplementary material, which is available to authorized users.

✉ Giuseppina De Simone
gdesimon@unina.it

✉ Simona Maria Monti
marmonti@unina.it

¹ Institute of Biostructures and Bioimaging, CNR, via Mezzocannone, 16, 80134 Naples, Italy

² Neurofarba Department, Section of Pharmaceutical and Nutriceutical Sciences, Università degli Studi di Firenze, Sesto Fiorentino, 50019 Florence, Italy

³ Institute of Protein Biochemistry, CNR, Via Pietro Castellino 111, 80131 Naples, Italy

HCO_3^- , which enters the cell through HCO_3^- transporters, as Na^+ /bicarbonate cotransporters (NBCs) and anion exchangers (AEs) [4, 12]. In this case, the newly formed CO_2 spreads out through the cell membrane. CA IX catalytic domain, expressed on the extracellular membrane of the cell, subtracts the newly spread CO_2 transforming it into protons and bicarbonate ions. As a whole, this process allows the maintenance of a physiologic pH crucial for the proliferation and survival of cancer cells and an acidification of the extracellular pH (pHe 6.9–7.0), which affects cancer progression by promoting invasion and metastasis [13].

Recent studies opened a completely new scenario on this enzyme, demonstrating that it can undergo nuclear translocation through the interaction with proteins involved in nucleocytoplasmic traffic [14]. Furthermore, it has also been shown that it can interact with cullin-associated NEDD8 dissociated protein 1 (CAND1), a protein involved in gene transcription and assembly of SCF ubiquitin ligase complexes. Notably, lower CA IX levels were observed in cells where CAND1 expression is downregulated via shRNA-mediated interference, suggesting that CAND1/CA IX interaction could be required for the enzyme stabilization [14, 15].

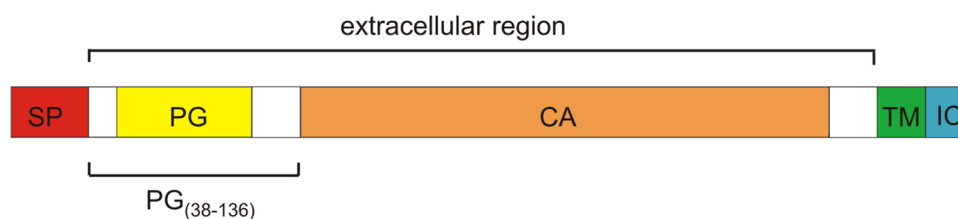
Human (h) CA IX is a multi-domain protein, which consists of an N-terminal signal peptide (SP, residues 1–37), an extracellular part (residues 38–414), a transmembrane (TM) region (residues 415–433) and an intracytoplasmic (IC) tail (residues 434–459) (hereafter CA IX numbering refers to the full-length protein including signal peptide) (Scheme 1) [4]. The extracellular part is constituted by two regions: an N-terminal region (residues 38–136) and a catalytic CA domain followed by a small linker (residues 137–391 and 392–414, respectively). The N-terminal region consists of a small domain (residues 53–111), named PG-like domain due to its high sequence identity with keratan sulfate attachment domain of a large aggregating proteoglycan termed aggrecan [16, 17] and two flanking sequences (residues 38–52 and 112–136). Notably, the region 38–136 is a unique feature of hCA IX with respect to all other hCAs.

The presence of the PG domain makes hCA IX one of the most active enzymes among hCAs [17, 18]. Indeed, kinetic measurements showed that the catalytic activity of the entire extracellular domain was greater than that of the catalytic

domain alone ($k_{\text{cat}}/K_M = 1.5 \times 10^8$ vs $5.4 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$, respectively) [17]. The PG domain was also reported to influence the optimal working pH of the enzyme; indeed, whereas the CA domain alone had an optimal activity at pH 7.0, the entire extracellular domain presented an optimal activity in acidic environment at pH 6.5 [18, 19]. It is worth noting that the slightly acidic pH value of 6.5 is within the typical pH range of solid and hypoxic tumours, where CA IX is generally overexpressed. Thus, it was suggested that the PG domain could be an evolutionarily evolved feature, unique to CA IX, which contributes to the improvement of its catalytic activity at the slightly acidic pH values [3–5].

Due to its role in tumour biology, hCA IX has become an interesting target for the drug design of new diagnostic and therapeutic tools in cancer treatment. Therefore, many studies have been dedicated to the elucidation of its biochemical and structural features. In particular, biochemical studies showed that the enzyme has both an intramolecular (C156–C336) and a symmetric intermolecular (through C174) disulphide bond, with the latter making the protein a dimer on the cell surface [17, 18]. Moreover, two glycosylation sites were identified: an *O*-linked glycosylation in the region immediately flanking the PG domain (T115), and an *N*-linked glycosylation localized on the catalytic domain (N346) [17]. Finally, three phosphorylation sites, namely, T443, S448, and Y449, were recognized on the IC tail [20, 21].

Notably, structural information is only available for the catalytic domain and for the C-terminal part of the protein (residues 418–459). In particular, the catalytic domain was crystallized by our group in 2009, showing the typical α -CA fold with a unique dimeric arrangement [18], whereas information on the secondary structure of the C-terminus has been recently obtained, indicating a predominant helical content for this region [15]. The absence of structural data on the full-length protein or on the PG domain is quite surprising, considering its important role in tumour biology, mediating cell adhesion and intercellular communications [22, 23] in addition to assisting catalysis mediated by the CA domain. Indeed, all attempts to obtain crystallographic structure of the PG domain failed, due to its high propensity to undergo protease degradation [18]. To fill this gap, we



Scheme 1 Schematic domain organization of hCA IX. SP, signal peptide (residues 1–37); PG, proteoglycan-like domain (residues 53–111); CA, catalytic domain (residues 137–391); TM, transmembrane segment (residues 415–433); IC, intracytoplasmic tail (residues 434–459)

hereby report the first detailed investigation on the N-terminal part of the hCA IX protein (residues 38–136), hereafter referred as PG_(38–136) (Scheme 1), by means of a multidisciplinary approach including biochemical, biophysical and molecular dynamics (MD) studies.

Materials and methods

Materials

Expression host strain *E. coli* BL21(DE3) and engineered plasmid pET28a/SUMO were a kind gift from EMBL, Heidelberg. *E. coli* strain TOP10F' was obtained from Invitrogen (San Diego, CA, USA). QIAprep spin miniprep kit and PCR Clean-Up DNA Purification System were from Qiagen (Germantown, MD, USA). Enzymes and other reagents for DNA manipulation were from New England Biolabs (Ipswich, MA, USA). All other chemicals were from Sigma-Aldrich (Milano, Italy).

Sequence analysis

The primary sequence of the PG_(38–136) protein was analyzed using the program Composition Profiler (<http://www.cprofiler.org/>) [24]. The query sample, analyzed for its intrinsic disorder, was compared with a reference sample which is a standard amino acid dataset (Swissprot) [25]. In the graphical output, the less abundant amino acids have negative values, whereas those more abundant have positive values.

The Charge/Hydrophobicity (CH) relation for PG_(38–136) was obtained as described by Uversky [26]. The CH plot is divided into two regions by a line, which corresponds to the equation $H = (|R| + 1.15)/2.782$, where R is the mean net charge and H is the mean hydrophobicity [26]. Proteins that fall in the left part of the diagram where $H < (R + 1.151)/2.785$ are predicted as disordered, whereas they are predicted as ordered if they fall in the right part. Data regarding the intrinsically disordered proteins were partially taken from Uversky et al. [27], whereas those regarding natively folded proteins were randomly taken from PDB.

Cloning, expression and purification of PG_(38–136)

pET28a/SUMO vector containing SenP2 protease recognition site was chosen for *E. Coli* expression of the PG_(38–136). Briefly, *pg* cDNA was amplified and cloned in the Age I and XhoI site of pET28a/SUMO using the following site-specific primers:

Forward: 5'-TCATCTACCGGTGGTCAGAGGTTGCCCGGATG-3'.

Reverse: 5'-GCGCGCTCGAGTTACTAATCCCCTTC TTTGTCCCTGTGG-3'.

The plasmid generated was verified by appropriate digestion with restriction enzymes and sequencing. The recombinant construct was expressed in *E. coli* BL21(DE3) cells for 16 h at 22 °C with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After centrifugation, the supernatant was resuspended in lysis buffer (20 mM Tris-HCl, 20 mM imidazole, 500 mM NaCl, pH 8.0), in the presence of 1 mM phenylmethanesulfonyl fluoride, 5 mg/ml DNaseI, 0.1 mg/ml lysozyme and 1 μg/ml Aprotinin, 1 μg/ml Leupeptin, 1 μg/ml Pepstatin protease inhibitors and left for 30 min at room temperature before sonication. After centrifugation, the supernatant was loaded onto a nickel-immobilized affinity chromatography column (5 ml His Trap FF column, GE Healthcare) and purified by FPLC according to manufacturer's instruction (GE Healthcare). Fractions containing PG_(38–136) were pooled and dialysed in 20 mM Tris-HCl, 250 mM NaCl, pH 8.0, with a membrane cutoff (MWCO) of 3.500. Tag removal was performed by digesting PG_(38–136) sample with protease enzyme SenP2 in a ratio SenP2/PG 1:25 (w/w) for 3 h at 20 °C and loading the mixture on an affinity His-Trap column according to manufacturer's instruction (GE Healthcare). Purity level was assessed by 15% SDS-PAGE and LC-ESI-MS.

Quaternary structure investigations of PG_(38–136)

The quaternary structure of PG_(38–136) was investigated using SEC-MALS-QELS (Size Exclusion Chromatography-Multi-angle Light-Scattering-Quasi-Elastic Light Scattering) as previously reported [28, 29]. In particular, 50 μl of 1.5 mg/ml protein was loaded onto a Wyatt technology corporation column (WTC 015S5), equilibrated in PBS 1× (10 mM phosphate, 2.7 mM KCl, 137 mM NaCl, pH 7.4) and connected to FPLC ÅKTA, coupled to a light-scattering detector (mini-DAWN TREOS, Wyatt Technology) and a refractive index detector (Shodex RI-101). Data were analyzed using the program ASTRA 5.3.4.14 (Wyatt Technology Corporation).

Dynamic light-scattering (DLS) measurements were carried out using a Malvern nanozetasizer (Malvern, UK) following a procedure previously described [30]. 37 μM PG_(38–136) in 20 mM Tris-HCl, pH 7.5 was placed in a disposable cuvette and held at 25 °C during analysis. Spectra were recorded six times with 11 sub-runs using the multimodal mode. Only monodisperse peaks (% polydispersity lower than 20%) were considered. The Z-average diameter of the monodisperse peak was calculated from the correlation function using the Malvern technology software.

Effect of PG_(38–136) on CA IX catalytic activity

Measurements of the catalytic activity of CA IX were performed by stopped-flow spectrophotometric measurements (Applied Photophysics Model SX.18MV). The solutions containing CO₂ in a concentration range between 1.7 and 17 mM were obtained by bubbling CO₂ in water. The maximum absorbance was observed at 557 nm, using 0.2 mM red phenol as pH indicator. CA domain was used at a concentration of 1×10^{-7} M in 10 mM Hepes, 10 mM Tris–HCl and 100 mM Na₂SO₄, pH 7.5. PG_(38–136) was added at different concentrations ranging from 1×10^{-9} to 1×10^{-6} M with an incubation time of 10 min before reading the absorbance.

Chemico-physical characterization

Circular dichroism (CD)

CD measurements were performed on a Jasco J815 spectropolarimeter (Jasco, Essex, UK), equipped with a temperature control system, using a 1-mm quartz cell in the far UV range 190–260 nm (20 nm/min scan speed). Each spectrum was the average of three scans with the background of the buffer solution subtracted. Measurements were performed at 20 °C at a protein concentration of 14 μM in buffers such as 10 mM Tris–HCl pH 8.0 or 10 mM phosphate buffer, pH 7.5. The effect of temperature on the secondary structure content of PG_(38–136) was investigated using a 18 μM protein in 10 mM Tris–HCl buffer at pH 8.0. CD spectra ranging from 5 to 90 °C were taken every 5 °C, keeping temperature, set manually, within ± 0.1 °C by a peltier device. For the pH titration, PG domain was diluted in 10 mM Citrate phosphate buffer at different pHs (pH 2.6, 3.0, 4.0, 5.0, 6.0, 7.0, and 7.6) and the resulting curves were obtained keeping a fixed temperature of 5 °C. The effect of urea on the secondary structure of the protein was evaluated recording the spectra at 20 °C with different concentrations of the denaturant (0, 2, 4, 6, and 8 M) in 10 mM sodium phosphate pH 7.4. The wavelength range was set from 250 to 215 nm due to the absorbance properties of urea which prevented the acquisition of spectra below 215 nm [31]. For all the CD experiments, raw spectra were corrected for buffer contribution and converted to mean molar ellipticity per residue (Θ) ($\text{deg cm}^2 \text{dmol}^{-1}$) [15].

NMR spectroscopy

NMR spectra were recorded at 303 K on a Varian Unity Inova 600 MHz spectrometer provided with a cold probe. To prepare the NMR sample, PG_(38–136) was dissolved in 600 μl (concentration equal to 0.7 mg/ml) of a buffer containing 20 mM sodium phosphate pH 6.6, 100 mM NaCl, 40 μl D₂O

(98% D, Armar Chemicals, Dottingen, Switzerland) and 0.01% sodium azide. The following NMR experiments were collected: 1D [¹H], 2D [¹H, ¹H] TOCSY [32] (70 ms mixing time), 2D [¹H, ¹H] NOESY [33] (300 ms mixing time). The 1D [¹H] spectrum was acquired with a relaxation delay d1 of 1.5 s and 128 scans; 2D experiments were acquired with 32 scans, 128–256 FIDs in t1, 1024 or 2048 data points in t2. Water suppression was achieved through Excitation Sculpting [34]. Chemical shifts were referenced to the water signal (4.75 ppm). The software VNMRJ (Varian by Agilent Technologies, Italy) was used for spectra processing, whereas NEASY [35], that is comprised in Computer Aided Resonance Assignment (CARA) package (<http://cara.nmr.ch/doku.php>), was implemented for spectra analysis.

Protease sensitivity of PG_(38–136)

Protease sensitivity of PG_(38–136) was evaluated by incubating the protein with TPCK-treated trypsin (Sigma-Aldrich, Milan) protease at different ratios such as 1:100 and 1:200 (w/w) at 26 °C. The reaction was monitored by 15% SDS-PAGE after incubation for 1, 3, 6, and 16 h with the proteolytic enzyme. hCA II as standard was incubated with trypsin in the same conditions.

Modelling and molecular dynamics studies

To build the model of the entire hCA IX extracellular region (residues 38–391), I-TASSER [36] was employed, using the X-ray structure available for the CA domain as reference [18].

Among the five good quality models built by I-TASSER, the fifth one (*C*-score = −2.20) was chosen for subsequent studies, being the only one compatible with the X-ray dimeric structure of CA IX. Indeed, in the other models, PG_(38–136) partially occupied the dimeric interface. The quality of the selected I-TASSER model was further assessed by means of PROSA [37, 38] and PROCHECK [39] software. According to these analyses the I-TASSER model shows 73% of residues in the allowed regions of the Ramachandran plot and an energetic *Z*-score of −7.42 indicating the good quality of the model. Subsequently, the hCA IX (38–391) dimeric model was built by superimposing two identical monomeric models obtained by I-TASSER to the crystallographic dimer of the catalytic domain. The final dimeric model was energy minimized by 1000 steps of Conjugate Gradient using Discover module of InsightII package.

The obtained dimeric model was subjected to all-atom MD simulations using the GROMACS simulation package [40]. CHARMM22* force field [41] was used for simulations, since it was proven to be accurate for the simulation of IDPs, producing conformational ensemble consistent with experimental data [42, 43]. The model was solvated in a

dodecahedral box filled with TIP3P water molecules with at least 12 Å distance to the border adding counterions to neutralize the system (reaching a concentration of 0.1 M). The simulations were run under NPT conditions (300 K and $P=1$ bar) using the V-rescale thermostat [44] and Berendsen barostat, respectively. Periodic boundary conditions were employed and the LINCS algorithm [45] was used to constrain bond lengths. The particle mesh Ewald method was applied to treat electrostatic interactions [46] and a non-bonded cutoff of 1.4 nm was used for the Lennard–Jones potential. Water molecules were relaxed by energy minimization, followed by 50 ps of simulations at 300 K, restraining the protein atomic positions with a harmonic potential. Then, the system was heated up gradually to 300 K and equilibrated as described elsewhere [47]. After equilibration, the system was simulated in NPT standard conditions for 100 ns using positional restraints for backbone atoms of the core-structured part of the CA domains, whereas the

rest of the system was free to move. The analysis of the MD trajectory was carried out using GROMACS tools as well as MOLMOL [48] and DSSP [49] program. PROSS server was used for assignment of Polyproline II conformation [50].

Results and discussion

Sequence analysis

The amino acid sequence of PG_(38–136) is reported in Fig. 1a. The sequence is formed by the PG domain and two short flanking sequences. Interestingly, the PG domain is characterized by the presence of a sixfold tandem repeat of six amino acids, four of which are identical (GEEDLP), whereas the remaining two contain two exchanged amino acids (SEEDSP and REEDPP) [51]. The amino acid sequence of PG_(38–136) is highly acidic, with a theoretical pI of 3.8 and

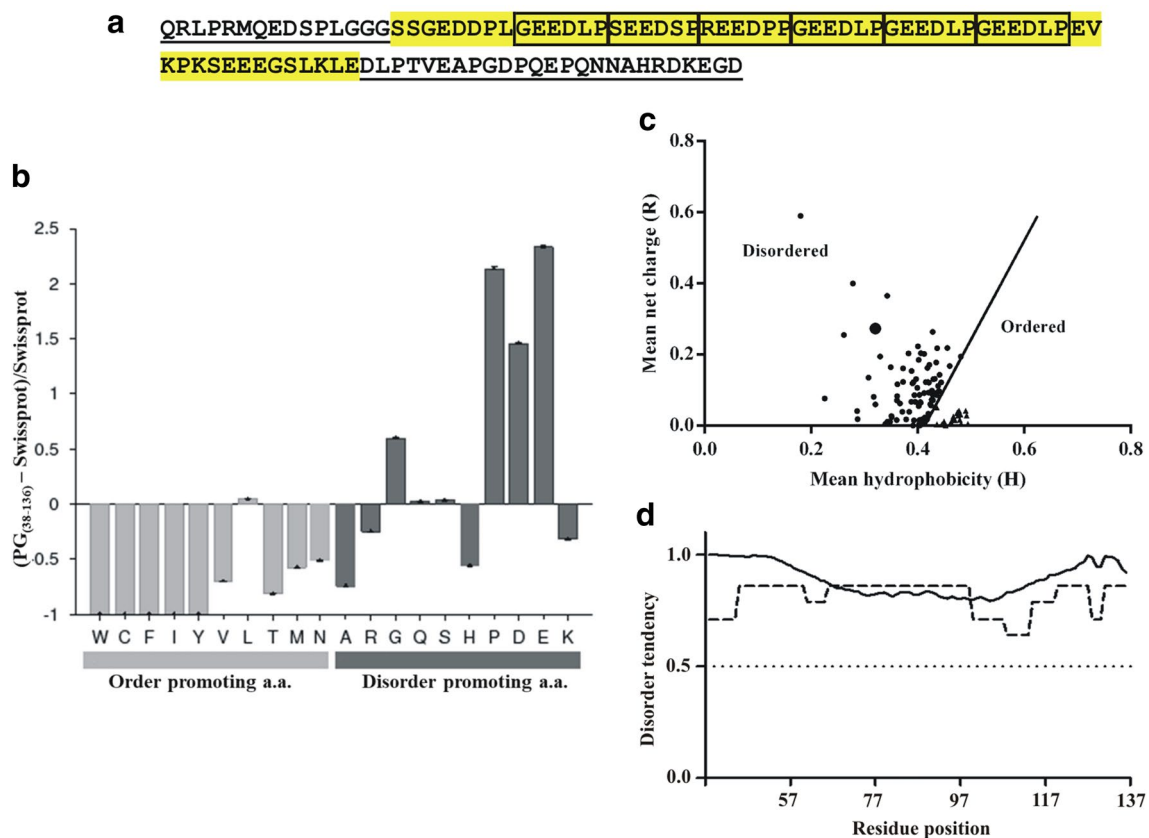


Fig. 1 Sequence properties of PG_(38–136). **a** Amino acidic sequence of PG_(38–136). The PG domain is highlighted in yellow, the six repeats are boxed, and the two flanking regions are underlined. **b** Amino acid compositional analysis performed by means of Composition Profiler tool. PG_(38–136) sequence is compared to the reference value of the average amino acid frequencies of the Swiss-Prot database [25]. Bar heights indicate enrichment or depletion of indicated residue. **c** Charge–hydrophobicity plot generated as described by Uversky [26]. Black dots, IDPs reported in the literature (data partially taken from

Uversky et al. [27]) black triangles, natively folded proteins randomly taken from PDB. The solid black line, that is the border between the ID and the natively folded proteins, is described by the equation $H=(R+1.151)/2.785$, where H and R are the mean hydrophobicity and the mean net charge, respectively. Fully black circle, PG_(38–136). **d** Predictions of intrinsic disorder by PONDR-fit (red line) and DisMETA (blue line) predictors. Values higher than 0.5 indicate a propensity for disorder

contains many structure breaking Pro (15%) and disorder-promoting amino acids such as Asp (13%), Glu (22%) and Gly (11%). Notably, order-promoting residues, such as the aromatic Phe, Trp, Tyr, the bulky hydrophobic Ile, and Cys (which may contribute to protein conformational stability via disulfide bond formation), are absent (Fig. 1a) [52]. To compare the content of order- and disorder-promoting residues of proteins within the Swiss-Prot database, the web-based tool Composition Profiler (<http://www.cprofiler.org/>) [24, 25] was used, confirming the high representation of disorder-promoting residues with respect to the order-promoting ones (Fig. 1b) [53–55]. Concurrently, the CH plot [26], which correlates the net charge of a protein against its mean hydrophobicity, showed the occurrence of $PG_{(38-136)}$ in the region of the intrinsically disordered proteins (IDPs) (Fig. 1c), which are biologically active proteins lacking of a stable and well-defined three-dimensional structure [26, 56]. In agreement with these data, the two disorder predictors PONDR-FIT [57] and DisMeta [58] scored a disorder tendency always above 0.5 (Fig. 1d). Altogether, these results strongly indicate that $PG_{(38-136)}$ possesses typical features of IDPs.

The significant presence of Pro and Glu within the $PG_{(38-136)}$ sequence, as well as its putative belonging to the IDP family, prompted us to investigate whether this domain contained PEST motifs. These sequences, enriched in Pro (P), Glu (E), Ser (S) and Thr (T), frequently located within unfolded protein regions [56, 59], serve as specific degradation signals [56, 59]; therefore, they play an important role in rapid turnover of regulatory proteins involved in signaling pathways that control cell growth, differentiation, stress responses, and physiological cell death [59, 60]. Using the Pestfindalgorithm [61], a PEST sequence (residues 43–72) was identified with a very high score of + 18.55. Studies will be carried out in our lab to explore the exact role of the PEST sequence in PG domain and how it might affect CA IX stability.

Expression and biochemical characterization of $PG_{(38-136)}$

Results obtained by the sequence analysis reported above, led us to believe that $PG_{(38-136)}$ belongs to the IDP family. To corroborate this hypothesis, we cloned this fragment in pET28a/SUMO, expressed it heterologously in *E. coli* and extensively characterized the recombinant product. After digestion with SenP2 protease, $PG_{(38-136)}$ was obtained as a highly purified protein with a yield of 30 mg/L. Since the beginning, the protein showed an unusual behaviour likely related to its peculiar amino acid composition. For instance, the high number of acidic Glu and Asp residues in the primary sequence caused a scarce denaturation of the protein in SDS [25, 62], which resulted in an aberrant

migration on SDS-PAGE, leading to an apparent molecular mass between 15 and 20 kDa instead of the expected 10.8 kDa (Fig. 2a). Likewise, in size exclusion chromatography (SEC), $PG_{(38-136)}$ eluted with an anomalous retention volume (10.68 ml), corresponding to an apparent molecular mass of 50 kDa much greater than the expected one (Fig. 2b plus inset). The observed high retention volume could be ascribed to the formation of an oligomer, to an extended conformation or to a low compactness of the protein. To clarify this point, light-scattering (LS) experiments were performed showing that $PG_{(38-136)}$ is present in solution as a monomer (Fig. 2c). By DLS analysis, a monodisperse peak (17% of polydispersity) was evident, indicative of a species homogeneous in size distribution with a rather large apparent hydrodynamic radius ($4.1 \text{ nm} \pm 0.7$). This result revealed that the high retention volume observed by SEC was a consequence of a non-globular structure. These findings are in line with previously reported studies by our group, which showed that the entire extracellular domain of CA IX, expressed in the baculovirus–insect cell system, had an anomalous behaviour by SDS-PAGE and SEC due to the presence of the PG domain [17]. In the same paper, our finding proved that the PG domain was able to assist the catalysis mediated by the CA domain. To verify whether $PG_{(38-136)}$ was able to modulate the catalytic activity of the CA domain, although not being covalently linked to it, CA domain was titrated with different concentrations of $PG_{(38-136)}$ and the catalytic activity was evaluated. In agreement with the previous results, it was observed that $10 \mu\text{M } PG_{(38-136)}$ was able to increase the CA catalytic activity of about 63% (Fig. S1).

Chemico-physical characterization

The nature of the secondary structure of $PG_{(38-136)}$ was evaluated by far-UV-circular dichroism (CD) experiments. In agreement with the hypothesis that $PG_{(38-136)}$ is an IDP, collected spectrum showed a strong negative molar ellipticity value at 198 nm and a negative band between 210 and 230 nm (Fig. 3a), indicative of a protein in a largely disordered conformation. Different temperatures were also investigated to get more insights into the conformational behaviour of the protein. Significant changes in the CD spectrum were observed in a temperature range from 5 to 90 °C. Indeed, increasing the temperature led to a decrease in the negative signal at 198 nm, as well as an increase in the negative band centred between 210 and 230 nm (Fig. 3b). These spectroscopic changes could be indicative of an increase of alpha-helical content, suggesting that raising the temperature could induced folding of the protein as often occurs in IDPs [26]. However, the difference spectrum (Delta 5–90 °C) showed a large negative CD band at 198 nm, and a positive CD signal centred between 210 and 230 nm (Fig. 3b inset), indicating no contribution of alpha-helical structure in the

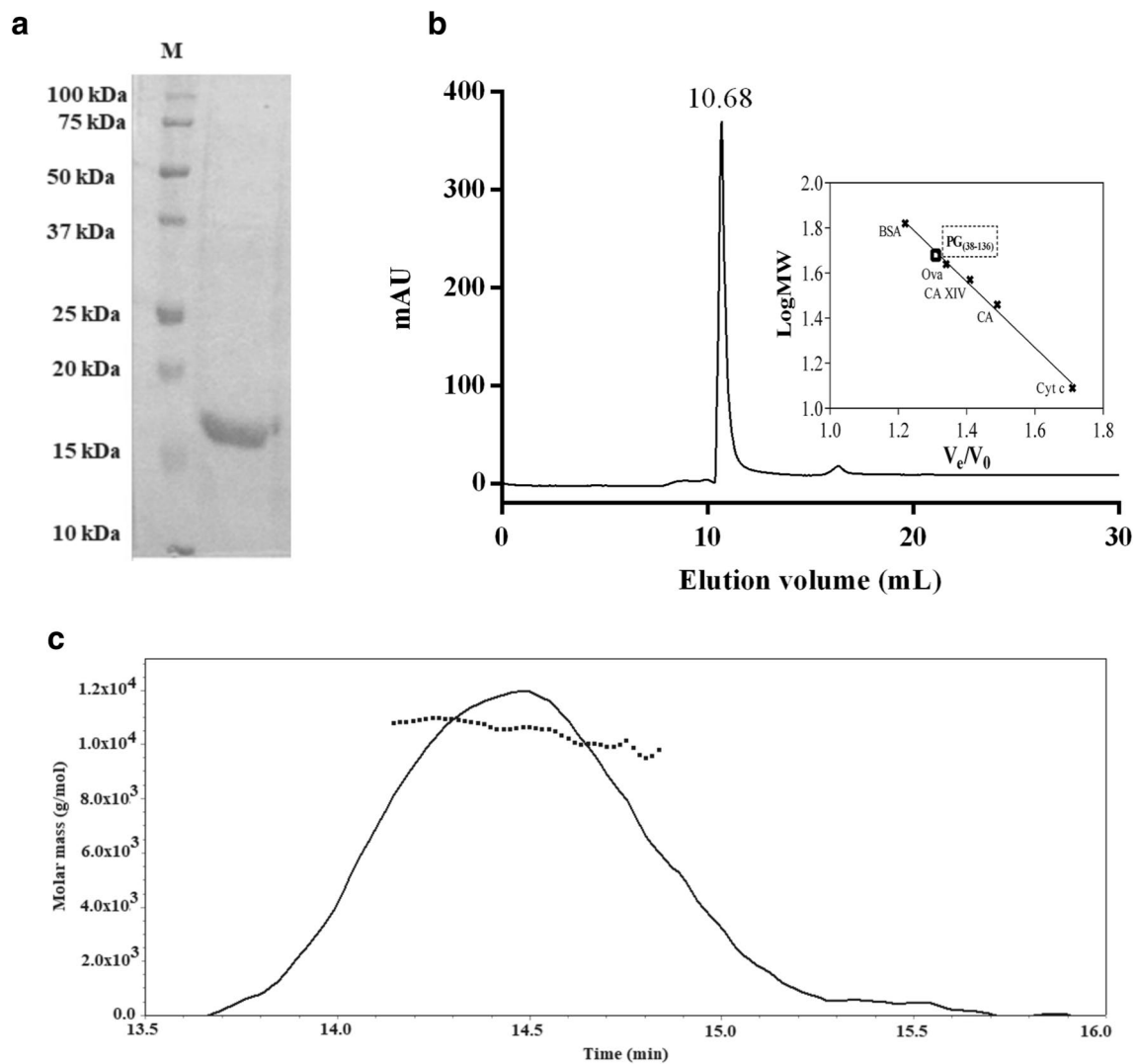


Fig. 2 PG₍₃₈₋₁₃₆₎ has an hydrodynamic dimension typical of IDPs. **a** 15% SDS-PAGE stained with Coomassie Brilliant Blue. Molecular masses (M) of broad range protein marker (20–250 kDa) (BIORAD) are indicated in kDa. **b** Elution profile of PG₍₃₈₋₁₃₆₎ on a Superdex

75 10/16 size exclusion chromatography column. Inset: molecular mass deduced from the calibration curve. **c** Molecular mass value of PG₍₃₈₋₁₃₆₎ determined by light-scattering analysis

spectrum. On the contrary, the spectroscopic features of the difference spectrum were typical of a polyproline II (PPII)-like left handed helical conformation [63]. Since it is known that apart prolines also other residues may adopt a PPII-like conformation [64, 65], we hypothesized that PG₍₃₈₋₁₃₆₎ could contain some regions in PPII conformation, which are disrupted upon increasing the temperature. Accordingly, the presence of a well-defined isodichroic point at 209 nm (Fig. 3b) indicates a conformational transition within the random coil ensemble [66], likely consisting of PPII and unordered conformations with the former prevailing at low temperature [65]. The denaturation process, followed at 222 nm, showed a linear curve (Fig. S2) confirming that PG₍₃₈₋₁₃₆₎ exhibited a PPII-to-unordered equilibrium showing a non-cooperative disruption, in contrast to the sigmoidal

curve typical of the alpha-helix-to-unordered transition [67]. Since pH can drive structural transition easy to follow by circular dichroism and PPII type structures are pH sensitive, we investigated the behaviour of PG₍₃₈₋₁₃₆₎ in a range of different pHs at 5 °C. In agreement with literature data [68], when lowering pH from neutral values to 2.6, a spectroscopic change is observed (Fig. 3c and inset), indicative of a decrease in PPII content due to decreasing pH. Finally, since PPII helical structures are stabilized upon addition of denaturants which shift the equilibrium towards PPII population, we investigated these effects on the local PG backbone. Following the addition of urea to PG₍₃₈₋₁₃₆₎, an increase of ellipticity in the range of 215–230 nm was monitored. Noticeably, a positive band appeared upon addition of 6 M urea pointing out the gain of PPII content (Fig. 3d) [31,

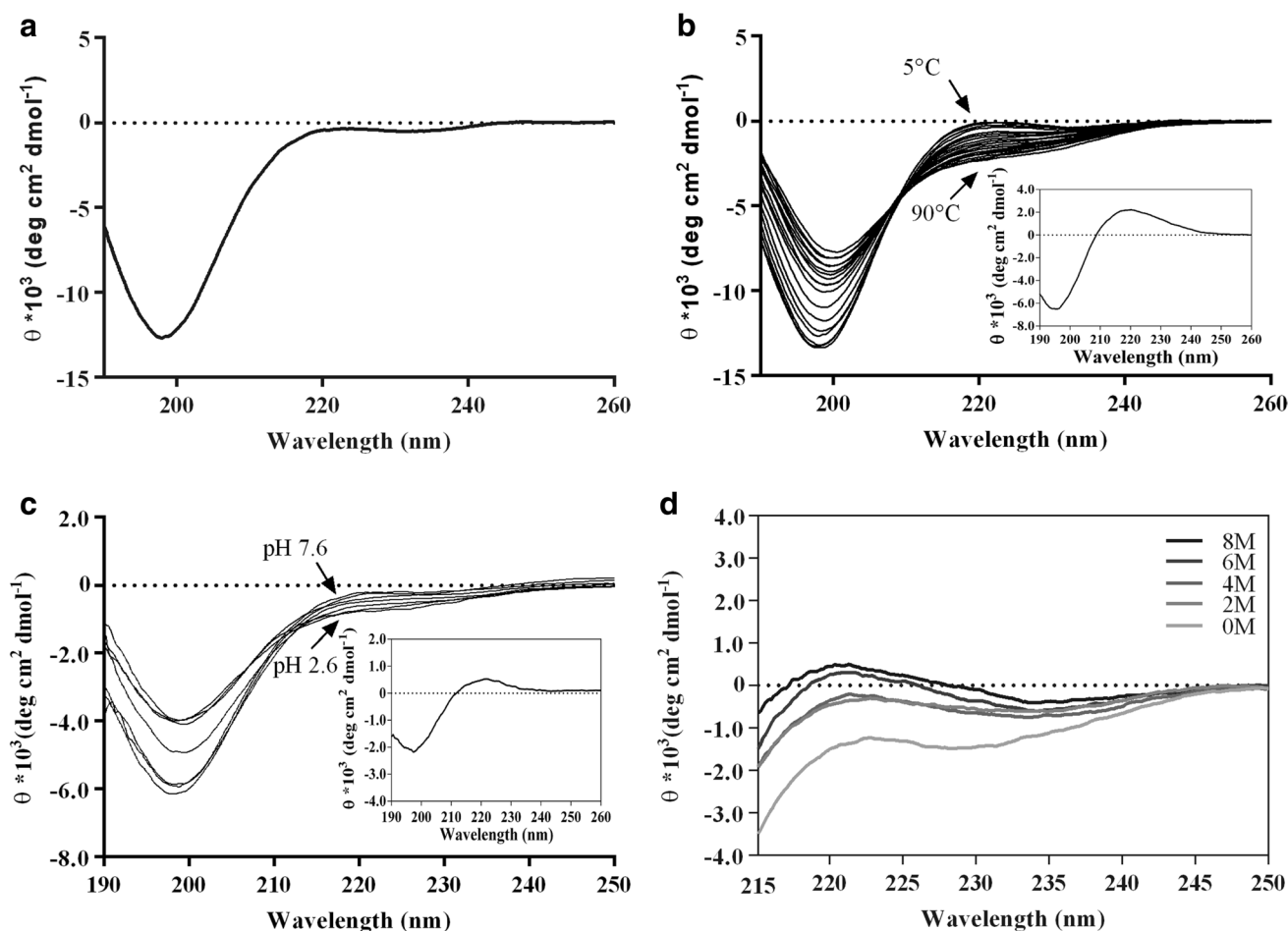


Fig. 3 Far-UV-CD spectra of PG₍₃₈₋₁₃₆₎. **a** CD spectrum was recorded in 10 mM phosphate buffer, pH 8.0 at a protein concentration of 14 μ M. **b** Effect of temperature on PG₍₃₈₋₁₃₆₎ from 5 to 90 °C taken every 5 °C; inset: differential spectrum of PG₍₃₈₋₁₃₆₎ between 5 and

90 °C. **c** Effect of pH on PG₍₃₈₋₁₃₆₎ in 10 mM Citrate phosphate from pH 7.6 to 2.6; inset: differential spectrum of PG₍₃₈₋₁₃₆₎ between pH 7.6 and 2.6. **d** Effect of different concentration of Urea (0, 2, 4, 6, 8 M) in 10 mM sodium phosphate pH 7.4 on PG₍₃₈₋₁₃₆₎

69]. In conclusion, collected far-UV-CD data give strong evidence that PG₍₃₈₋₁₃₆₎ is an IDP possessing some residues in PPII conformation. PPII regions in proteins are reported to play a role in several cellular processes including partner recognition, thus it is reasonable to hypothesize that PPII regions in PG could be involved in cell adhesion and inter-cellular communications.

Further structural insights into PG₍₃₈₋₁₃₆₎ were obtained by means of 1D [¹H] and 2D [¹H, ¹H] NMR spectroscopy [70]. The 1D [¹H] NMR spectrum (Fig. S3a) together with the 2D [¹H, ¹H] TOCSY (Total Correlation Spectroscopy) [32] and 2D [¹H, ¹H] NOESY (Nuclear Overhauser Enhancement Spectroscopy) [33] experiments (Fig. S3b) appear typical of IDPs. In particular, the 1D [¹H] (Fig. S3a) and 2D [¹H, ¹H] TOCSY (Fig. S3b left panel) spectra present low chemical shift dispersion with the backbone amide H_N protons resonating in the narrow random coil range between 8 and 8.6 ppm. Moreover, methyl protons from side chains of

Leu and Val residues give rise to a strong peak at the random coil chemical shift (i.e., 0.9 ppm) (Fig. S3a), thus highlighting the absence of the hydrophobic core of a folded protein. The NOESY spectrum (Fig. S3c) contains a few very weak inter-residue H_N-H_N contacts pointing out the presence of a rather small population of more ordered conformations. Thus, in agreement with the above-described results from other biophysical techniques, NMR spectroscopy further shows the largely disordered nature of PG₍₃₈₋₁₃₆₎.

Protease sensitivity

Due to the lack of a hydrophobic packed core and to the wide solvent accessibility, IDPs are prone to be easily degraded in the presence of a protease with broad substrate specificity such as trypsin [26]. This is one of the main differences compared to structured proteins with well-defined secondary structure elements, which are preferentially cleaved at

exposed and flexible loops [25]. The incubation of PG_(38–136) with trypsin protease at different ratios showed a complete cleavage in the early hours of the reaction (Fig. S4a). The same experiment performed on hCA II, a structured globular protein, showed a strong resistance to proteolysis even after 16 h of incubation (Fig. S4b). These data are in agreement with those obtained by CD, SEC, and LS confirming that PG_(38–136) is a largely disordered and flexible protein.

Molecular modelling and molecular dynamics simulations

To get more details into the structural and functional features of PG_(38–136), a comprehensive structural study of the whole extracellular part of CA IX, inclusive of both PG_(38–136) and CA domain, was undertaken by means of MD simulations. Indeed, MD is able to describe conformational and dynamics properties of highly fluctuating and flexible systems [42, 71]. Therefore, it fits appropriately for the characterization of IDPs, which lack a unique stable globular structure fixed in time and exist as a conformational ensemble [53]. As a first step, a model of the entire CA IX extracellular region, namely, residues 38–391, was built by means of I-TASSER server [36], which used the X-ray structure available for the catalytic domain as a reference (see “Materials and methods” paragraph for details) [18]. In agreement with previously reported data [17, 18], this model is dimeric (Fig. 4a) and shows that PG_(38–136) regions, belonging to the two monomeric units (hereafter indicated as MonA and MonB), are very well solvent-exposed and mainly unfolded, consistently with above reported biophysical experimental data. This model was used as starting structure for an all-atom MD simulation of 100 ns in explicit water using GROMACS

simulation package [40] with special attention to the choice of an appropriate force field for IDP simulations (see “Materials and methods” section) [41, 42]. Positional restraints were used for backbone atoms of the core-structured part of the CA domains, whereas the rest of the system was free to move. Due to the big size of the system, which contains around 400,000 atoms including water molecules, an extensive conformational sampling resulted prohibitive. However, the presence of two independent PG_(38–136) in the simulated dimeric system allowed to double the explored conformational space.

Root mean square fluctuations (RMSF) of C α atom positions were evaluated during simulation time, showing high values for residues 38–136 in both monomeric units, indicative of the high flexibility of this region (Fig. 4b). Interestingly, the RMSF curves of the two monomers in this region are diverse, since the two PG_(38–136) behave differently due to their inherent conformational plasticity.

Structures extrapolated at different time steps (0, 40, 60, and 100 ns) during simulation are shown in Fig. 5. It is worth noting that the two PG_(38–136) do not interact with each other and, although they assume different conformations during simulation, common behaviours can be highlighted: (1) the N-terminal region (residues 38–87) in both monomers, initially completely exposed to the solvent, moves closer to the globular catalytic domain making contacts with its superficial residues and assumes an extended conformation and (2) the C-terminal region (residues 88–136) is slightly more compact, making some self-interactions.

Within each monomer, PG_(38–136) conformations are stabilized by polar interactions with the aqueous solvent, as well as by intra- and inter-domain interactions (between PG_(38–136) and CA), mainly through the formation of

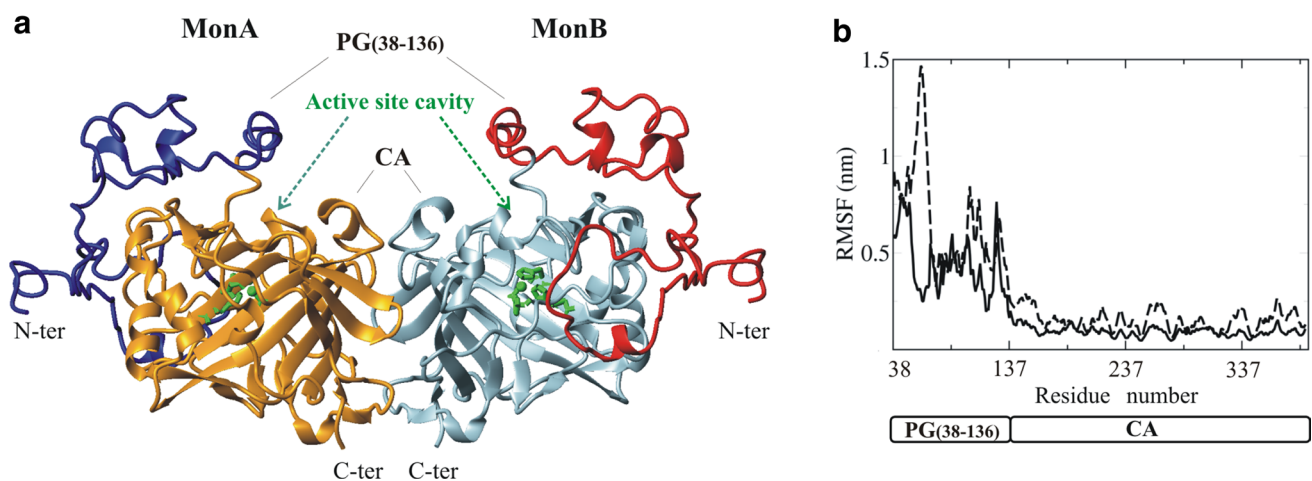


Fig. 4 hCA IX (38–391) dimeric model. **a** Cartoon representation of hCA IX (38–391) dimeric model. PG_(38–136) regions and CA domains belonging to the two monomeric units (MonA and MonB) are shown in different colors. PG_(38–136) and CA are reported in blue and orange

in MonA, and in red and gray in MonB. Active site residues are reported in green sticks. **b** Root mean square fluctuations (RMSF) of each residue of hCA IX (38–391) in MonA (continuous line) and MonB (dotted line)

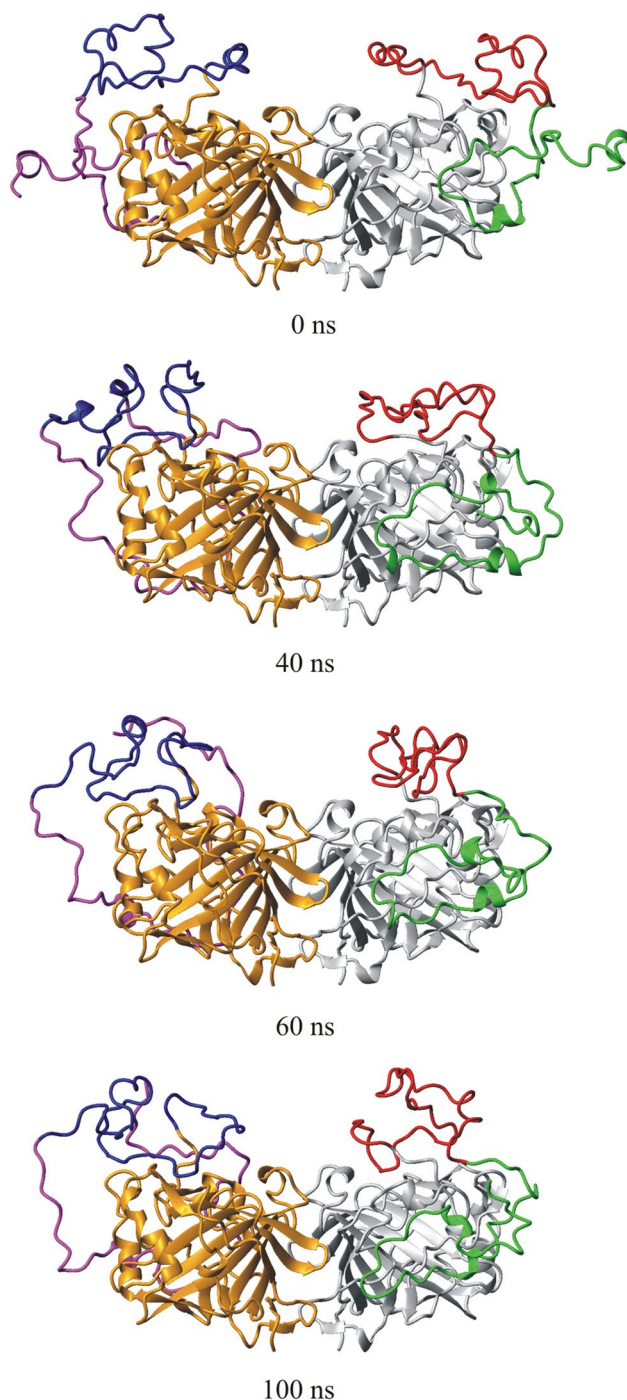


Fig. 5 hCA IX (38–391) snapshots. hCA IX (38–391) structures extrapolated at different time steps (0, 40, 60, 100 ns) during MD simulation. The N-terminal region (residues 38–87) of $PG_{(38-136)}$ is shown in magenta in MonA and in green in MonB, and the C-terminal region (residues 88–136) of $PG_{(38-136)}$ is shown in blue in MonA and in red in MonB

salt-bridges and hydrogen bonds (Table S1). Interestingly, couples of residues involved into hydrogen bonds are different in the two monomers, further indicating the flexibility of $PG_{(38-136)}$. Indeed, this region possesses many polar and

charged residues within its six repeats, which can be alternately involved into stabilizing interactions according to the adopted conformation.

The secondary structure assumed by $PG_{(38-136)}$ residues during the simulation time was analyzed using the DSSP program (Fig. 6). Plots show a high conformational plasticity with residues changing conformations along the trajectory from random coil to turn or bend structures. Despite the presence of a short α -helix of five residues spanning from residues 70 to 75 (blue spots in Fig. 6), the $PG_{(38-136)}$ sequence is mainly random coil (white spots in Fig. 6). The secondary structure analysis of both CA domains was also performed (Fig. S5), showing stable secondary structural elements along the trajectory, different from what was observed for $PG_{(38-136)}$. Summarizing, the MD analysis indicates that $PG_{(38-136)}$ is mainly random coil and possesses a high degree of flexibility, in agreement with the above reported biophysical studies (CD, DLS, and NMR).

Moreover, interesting insights derive from the analysis of the preferential conformations assumed by $PG_{(38-136)}$. To this aim, a cluster analysis was performed on both monomeric units along the trajectory. The representative structures of the two most populated clusters for each monomeric unit (ClusterI and ClusterII) are shown in Fig. S6. Interestingly, in ClusterII of monomer B (Figs. S6 and 7a), the C-terminal region of $PG_{(38-136)}$ arranges itself in a way that partially closes the entrance of the active site. This conformation is mainly stabilized by salt-bridges involving the non-conserved residues of the CA domain Arg196 and Arg261 (Fig. 7b) [18]. Remarkably, the presence of a $PG_{(38-136)}$ region located on the active site border could sterically control the access of the substrate or participate into the proton-transfer reaction. This finding is in agreement with the catalytic assay data reported here and elsewhere [17–19] indicating an involvement of the PG domain in the enzyme catalytic activity.

Finally, since far-UV-CD analysis indicated that $PG_{(38-136)}$ possesses some content of PPII conformation, PPII occurrence along MD trajectory was investigated. To this aim, PROSS server was employed, since differently from most commonly used secondary structure assignment methods, it can assign PPII structures. For PROSS analysis, 20 structures of the two most populated clusters for each monomer were selected and obtained data were reported in terms of frequency of occurrence of PPII structure vs residue number. The results indicate the presence of at least four regions having a significant preference for PPII conformation (frequency > 50%) (Fig. S7). The four short regions (3–5 residues in length) are wide-spread along the $PG_{(38-136)}$ sequence and roughly correspond to regions 58–60, 75–77, 98–100, and 119–121 (Fig. S7). As a consequence, the computed data are consistent with far-UV-CD analysis showing the presence of a residual PPII structure in $PG_{(38-136)}$.

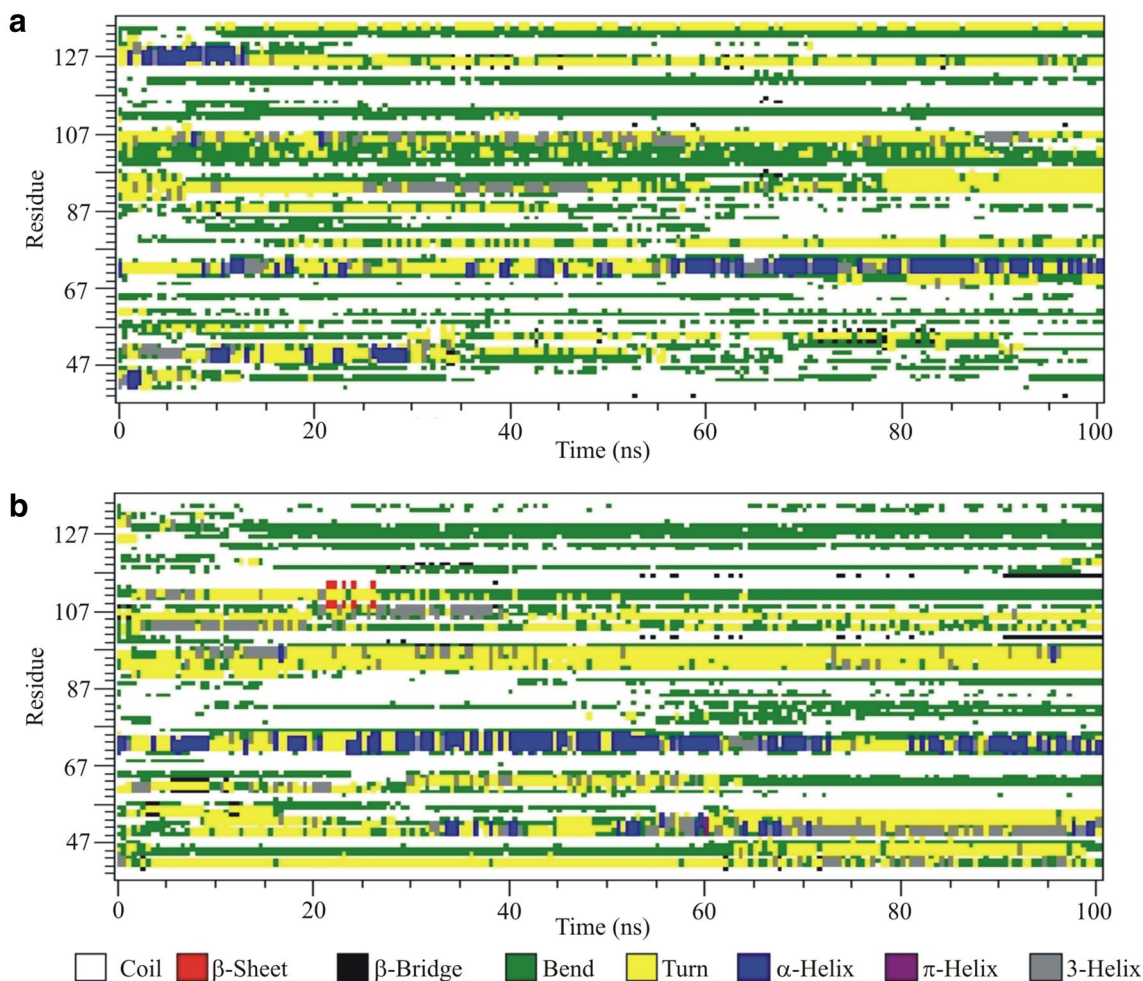


Fig. 6 Secondary structure analysis. Time evolution of the secondary structure of PG₍₃₈₋₁₃₆₎ in MonA (a) and MonB (b) calculated by DSSP [49]. The secondary structure is color-coded according to the legend

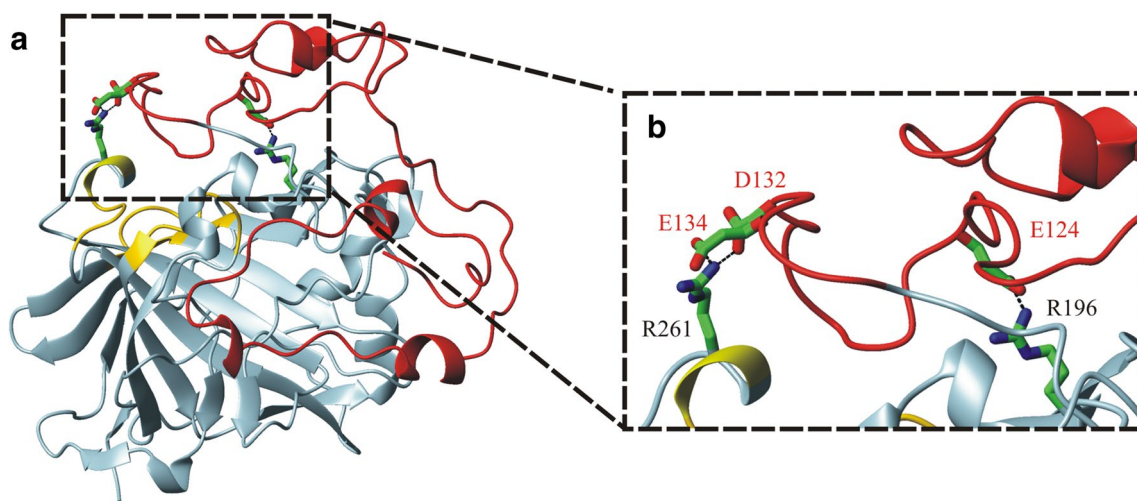


Fig. 7 Structural details of ClusterII of MonB. **a** Representative structure of ClusterII of MonB drawn as a ribbon. PG₍₃₈₋₁₃₆₎ is shown in red and CA in gray apart the border of the active site which is in

yellow. The region responsible for the partial closure of the active site is boxed. **b** Enlarged view of the boxed region with main residues involved into stabilizing salt-bridges shown in sticks

Conclusions

Despite the great amount of studies on tumour-associated protein hCA IX, to date, very little information concerning the biochemical and structural features of its N-terminal region containing the PG domain is available. By means of a multidisciplinary approach, we hereby report for the first time a comprehensive study on PG_(38–136), showing that it belongs to the family of IDPs, being natively highly flexible and mainly unfolded with only local tendencies to assume PPII conformations. Furthermore, the obtained data indicate that N-terminal residues (38–87) show a more extended conformation, being probably involved into partner recognition, whereas C-terminal residues (88–136) adopt a slightly more compact conformation and could have a role in modulating the catalytic activity of the CA domain. These results further extend our previous studies on the structural features of CA IX protein and provides new pieces in the complicated puzzle of CA IX functions in tumour biology.

Acknowledgements We thank Immacolata Ventotto for her help during protein purification; we are grateful to Luca De Luca, Maurizio Amendola and Giosuè Sorrentino for their technical assistance. This work was supported by a Grant from CNR-DSB ProgettoBandiera “InterOmics”.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest with the contents of this article.

References

- Alterio V, Di Fiore A, D'Ambrosio K, Supuran CT, De Simone G (2012) Multiple binding modes of inhibitors to carbonic anhydrases: how to design specific drugs targeting 15 different isoforms? *Chem Rev* 112:4421–4468
- Supuran CT (2008) Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nat Rev Drug Discov* 7:168–181
- Monti SM, Supuran CT, De Simone G (2013) Anticancer carbonic anhydrase inhibitors: a patent review (2008–2013). *Expert Opin Ther Pat* 23:737–749
- Monti SM, Supuran CT, De Simone G (2012) Carbonic anhydrase IX as a target for designing novel anticancer drugs. *Curr Med Chem* 19:821–830
- Supuran CT, Di Fiore A, Alterio V, Monti SM, De Simone G (2010) Recent advances in structural studies of the carbonic anhydrase family: the crystal structure of human CA IX and CA XIII. *Curr Pharm Des* 16:3246–3254
- Ilardi G, Zambrano N, Merolla F, Siano M, Varricchio S, Vecchione M, De Rosa G, Mascolo M, Staibano S (2014) Histopathological determinants of tumor resistance: a special look to the immunohistochemical expression of carbonic anhydrase IX in human cancers. *Curr Med Chem* 21:1569–1582
- Pastorek J, Pastorekova S (2015) Hypoxia-induced carbonic anhydrase IX as a target for cancer therapy: from biology to clinical use. *Semin Cancer Biol* 31:52–64
- Swietach P, Wigfield S, Cobden P, Supuran CT, Harris AL, Vaughan-Jones RD (2008) Tumor-associated carbonic anhydrase 9 spatially coordinates intracellular pH in three-dimensional multicellular growths. *J Biol Chem* 283:20473–20483
- Shin HJ, Rho SB, Jung DC, Han IO, Oh ES, Kim JY (2011) Carbonic anhydrase IX (CA9) modulates tumor-associated cell migration and invasion. *J Cell Sci* 124:1077–1087
- Tafreshi NK, Lloyd MC, Bui MM, Gillies RJ, Morse DL (2014) Carbonic anhydrase IX as an imaging and therapeutic target for tumors and metastases. *Subcell Biochem* 75:221–254
- Ondriskova E, Debreova M, Pastorekova S (2015) Tumor-associated carbonic anhydrases IX and XII. In: Supuran CT, De Simone G (eds) *Carbonic anhydrases as biocatalysts—from theory to medical and industrial applications*. Elsevier, Amsterdam, pp 169–206
- Swietach P, Vaughan-Jones RD, Harris AL (2007) Regulation of tumor pH and the role of carbonic anhydrase 9. *Cancer Metastasis Rev* 26:299–310
- Estrella V, Chen T, Lloyd M, Wojtkowiak J, Cornnell HH, Ibrahim-Hashim A, Bailey K, Balagurunathan Y, Rothberg JM, Sloane BF, Johnson J, Gatenby RA, Gillies RJ (2013) Acidity generated by the tumor microenvironment drives local invasion. *Cancer Res* 73:1524–1535
- Buanno P, Renzone G, Monteleone F, Vitale M, Monti SM, Sandomenico A, Garbi C, Montanaro D, Accardo M, Troncone G, Zatovicova M, Csaderova L, Supuran CT, Pastorekova S, Scaloni A, De Simone G, Zambrano N (2013) Characterization of carbonic anhydrase IX interactome reveals proteins assisting its nuclear localization in hypoxic cells. *J Proteome Res* 12:282–292
- Buonanno M, Langella E, Zambrano N, Succio M, Sasso E, Alterio V, Di Fiore A, Sandomenico A, Supuran CT, Scaloni A, Monti SM, De Simone G (2017) Disclosing the interaction of carbonic anhydrase IX with Cullin-associated NEDD8-dissociated protein 1 by molecular modeling and integrated binding measurements. *ACS Chem Biol* 12:1460–1465
- Pastorekova S, Zatovicova M, Pastorek J (2008) Cancer-associated carbonic anhydrases and their inhibition. *Curr Pharm Des* 14:685–698
- Hilvo M, Baranauskiene L, Salzano AM, Scaloni A, Matulis D, Innocenti A, Scozzafava A, Monti SM, Di Fiore A, De Simone G, Lindfors M, Janis J, Valjakka J, Pastorekova S, Pastorek J, Kulomaa MS, Nordlund HR, Supuran CT, Parkkila S (2008) Biochemical characterization of CA IX, one of the most active carbonic anhydrase isozymes. *J Biol Chem* 283:27799–27809
- Alterio V, Hilvo M, Di Fiore A, Supuran CT, Pan P, Parkkila S, Scaloni A, Pastorek J, Pastorekova S, Pedone C, Scozzafava A, Monti SM, De Simone G (2009) Crystal structure of the catalytic domain of the tumor-associated human carbonic anhydrase IX. *Proc Natl Acad Sci USA* 106:16233–16238
- Innocenti A, Pastorekova S, Pastorek J, Scozzafava A, De Simone G, Supuran CT (2009) The proteoglycan region of the tumor-associated carbonic anhydrase isoform IX acts as an intrinsic buffer optimizing CO₂ hydration at acidic pH values characteristic of solid tumors. *Bioorg Med Chem Lett* 19:5825–5828
- Ditte P, Dequiedt F, Svastova E, Hulikova A, Ohradanova-Repic A, Zatovicova M, Csaderova L, Kopacek J, Supuran CT, Pastorekova S, Pastorek J (2011) Phosphorylation of carbonic anhydrase IX controls its ability to mediate extracellular acidification in hypoxic tumors. *Cancer Res* 71:7558–7567
- Dorai T, Sawczuk IS, Pastorek J, Wiernik PH, Dutcher JP (2005) The role of carbonic anhydrase IX overexpression in kidney cancer. *Eur J Cancer* 41:2935–2947

22. Zavadova Z, Zavada J (2005) Carbonic anhydrase IX (CA IX) mediates tumor cell interactions with microenvironment. *Oncol Rep* 13:977–982
23. Csaderova L, Debreaova M, Radvak P, Stano M, Vrestiakova M, Kopacek J, Pastorekova S, Svastova E (2013) The effect of carbonic anhydrase IX on focal contacts during cell spreading and migration. *Front Physiol* 4:271
24. Vacic V, Uversky VN, Dunker AK, Lonardi S (2007) Composition profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinf* 8:211
25. Habchi J, Mamelli L, Darbon H, Longhi S (2010) Structural disorder within Henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment. *PLoS One* 5:e11684
26. Habchi J, Tompa P, Longhi S, Uversky VN (2014) Introducing protein intrinsic disorder. *Chem Rev* 114:6561–6588
27. Uversky VN, Gillespie JR, Fink AL (2000) Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 41:415–427
28. D'Ambrosio K, Lopez M, Dathan NA, Ouahrani-Bettache S, Kohler S, Ascione G, Monti SM, Winum JY, De Simone G (2014) Structural basis for the rational design of new anti-Brucella agents: the crystal structure of the C366S mutant of L-histidinol dehydrogenase from *Brucella suis*. *Biochimie* 97:114–120
29. Ascione G, de Pascale D, De Santi C, Pedone C, Dathan NA, Monti SM (2012) Native expression and purification of hormone-sensitive lipase from *Psychrobacter* sp. TA144 enhances protein stability and activity. *Biochem Biophys Res Commun* 420:542–546
30. Di Lelio I, Caccia S, Coppola M, Buonanno M, Di Prisco G, Varicchio P, Franzetti E, Corrado G, Monti SM, Rao R, Casartelli M, Pennacchio F (2014) A virulence factor encoded by a polydnavirus confers tolerance to transgenic tobacco plants against lepidopteran larvae, by impairing nutrient absorption. *PLoS One* 9:e113988
31. Whittington SJ, Chellgren BW, Hermann VM, Creamer TP (2005) Urea promotes polyproline II helix formation: implications for protein denatured states. *Biochemistry* 44:6269–6275
32. Griesinger C, Otting G, Wuthrich K, Ernst RR (1988) Clean TOCSY for proton spin system identification in macromolecules. *J Am Chem Soc* 110:7870–7872
33. Kumar A, Ernst RR, Wuthrich K (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton cross-relaxation networks in biological macromolecules. *Biochem Biophys Res Commun* 95:1–6
34. Hwang T-L, Shaka AJ (1995) Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients. *J Magn Reson A* 112:5
35. Bartels C, Xia TH, Billeter M, Guntert P, Wuthrich K (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* 6:1–10
36. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725–738
37. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35:W407–W410
38. Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins* 17:355–362
39. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291
40. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26:1701–1718
41. Piana S, Lindorff-Larsen K, Shaw DE (2011) How robust are protein folding simulations with respect to force field parameterization? *Biophys J* 100:L47–L49
42. Rauscher S, Gapsys V, Gajda MJ, Zweckstetter M, de Groot BL, Grubmuller H (2015) Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment. *J Chem Theory Comput* 11:5513–5524
43. Carballo-Pacheco M, Strodel B (2017) Comparison of force fields for Alzheimer's A beta42: a case study for intrinsically disordered proteins. *Protein Sci* 26:174–185
44. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126:014101
45. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472
46. Darden T, York D, Pedersen L (1993) Particle Mesh Ewald—an N.Log(N) method for ewald sums in large systems. *J Chem Phys* 98:10089–10092
47. Autiero I, Saviano M, Langella E (2013) In silico investigation and targeting of amyloid beta oligomers of different size. *Mol Biosyst* 9:2118–2124
48. Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 14(51–55):29–32
49. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637
50. Srinivasan R, Rose GD (1999) A physical basis for protein secondary structure. *Proc Natl Acad Sci USA* 96:14258–14263
51. Zavada J, Zavadova Z, Pastorek J, Biesova Z, Jezek J, Velek J (2000) Human tumour-associated cell adhesion protein MN/CA IX: identification of M75 epitope and of the region mediating cell adhesion. *Br J Cancer* 82:1808–1813
52. Uversky AV, Xue B, Peng Z, Kurgan L, Uversky VN (2013) On the intrinsic disorder status of the major players in programmed cell death pathways. *FI000Res* 2:190
53. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z (2001) Intrinsically disordered protein. *J Mol Graph Model* 19:26–59
54. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK (2007) Intrinsic disorder and functional proteomics. *Biophys J* 92:1439–1456
55. Midic U, Obradovic Z (2012) Intrinsic disorder in putative protein sequences. *Proteome Sci* 10(Suppl 1):S19
56. Uversky VN (2013) Unusual biophysics of intrinsically disordered proteins. *Biochim Biophys Acta* 1834:932–951
57. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* 1804:996–1010
58. Huang YJ, Acton TB, Montelione GT (2014) DisMeta: a meta server for construct design and optimization. *Methods Mol Biol* 1091:3–16
59. Singh GP, Ganapathi M, Sandhu KS, Dash D (2006) Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes. *Proteins* 62:309–315
60. Marie N, Lindsay AJ, McCaffrey MW (2005) Rab coupling protein is selectively degraded by calpain in a Ca²⁺-dependent manner. *Biochem J* 389:223–231
61. Rechsteiner M, Rogers SW (1996) PEST sequences and regulation by proteolysis. *Trends Biochem Sci* 21:267–271
62. Alves VS, Pimenta DC, Sattlegger E, Castilho BA (2004) Biophysical characterization of Gir2, a highly acidic protein of *Saccharomyces cerevisiae* with anomalous electrophoretic behavior. *Biochem Biophys Res Commun* 314:229–234

63. Lopes JL, Miles AJ, Whitmore L, Wallace BA (2014) Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: applications in secondary structure analyses. *Protein Sci* 23:1765–1772
64. Narwani TJ, Santuz H, Shinada N, Melarkode Vattekkatte A, Ghouzam Y, Srinivasan N, Gelly JC, de Brevern AG (2017) Recent advances on polyproline II. *Amino Acids* 49:705–713
65. Adzhubei AA, Sternberg MJ, Makarov AA (2013) Polyproline-II helix in proteins: structure and function. *J Mol Biol* 425:2100–2132
66. Kulp JL 3rd, Clark TD (2009) Engineering a beta-helical D,L-peptide for folding in polar media. *Chemistry* 15:11867–11877
67. Delak K, Harcup C, Lakshminarayanan R, Sun Z, Fan Y, Moradian-Oldak J, Evans JS (2009) The tooth enamel protein, porcine amelogenin, is an intrinsically disordered protein with an extended molecular configuration in the monomeric form. *Biochemistry* 48:2272–2281
68. Garcia-Alai MM, Gallo M, Salame M, Wetzler DE, McBride AA, Paci M, Cicero DO, de Prat-Gay G (2006) Molecular basis for phosphorylation-dependent, PEST-mediated protein turnover. *Structure* 14:309–319
69. Hotta K, Ranganathan S, Liu R, Wu F, Machiyama H, Gao R, Hirata H, Soni N, Ohe T, Hogue CW, Madhusudhan MS, Sawada Y (2014) Biophysical properties of intrinsically disordered p130Cas substrate domain—implication in mechanosensing. *PLoS Comput Biol* 10:e1003532
70. Gibbs EB, Cook EC, Showalter SA (2017) Application of NMR to studies of intrinsically disordered proteins. *Arch Biochem Biophys* 628:57–70
71. Naqvi MA, Rauscher S, Pomes R, Rousseau D (2014) The conformational ensemble of the beta-casein phosphopeptide reveals two independent intrinsically disordered segments. *Biochemistry* 53:6402–6408