Biochemical and Biophysical Research Communications 372 (2008) 121–125

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

# Role of hydration in collagen triple helix stabilization

Alfonso De Simone *, Luigi Vitagliano *, Rita Berisio

*Istituto di Biostrutture e Bioimmagini, C.N.R., via Mezzocannone 16, I-80134 Napoli, Italy*

ARTICLE INFO

*Article history:*
Received 9 April 2008
Available online 15 May 2008

*Keywords:*
Hydroxyproline
Water structure
Molecular dynamics
Fibrous protein
Solvent and stability

ABSTRACT

Collagen is the most abundant protein in higher vertebrates. Despite collagen repetitive sequence, several aspects of its structure and stability are controversial. Here we performed molecular dynamics simulations to analyze triple helix hydration in regions characterized by different imino/aminoacid contents. Data emerged from MD simulations show that (a) MD simulations can reliably reproduce the hydration sites identified experimentally, (b) water molecules bound to regions with a different amino/iminoacid content exhibit diversified residence times, and (c) in the aminoacid-rich region the binding of water molecules is strongly influenced by the local sequence of the peptide. MD results also suggest that, in aminoacid-rich regions, the stabilizing effects of Arg and Hyp residues on collagen triple helix also depend on water-mediated interactions. On this basis, we propose that the mechanism of triple helix stabilization is sequence-dependent.



Collagen is the most abundant protein in higher vertebrates [1]. Collagen sequences are characterized by the repetition of triplets of the type Gly-X-Y. Although all types of aminoacids may be located at positions X and Y of the triplets, they are frequently occupied by iminoacids (Pro and 4*R*-Hydroxy-2*S*-proline (Hyp)).

The complexity of collagen molecule and its fibrous nature prevent detailed investigations on the full-length protein. Due to the repetitive nature of collagen sequence/structure, the use of peptide models embedding specific motifs has been quite successful [1,2]. Nevertheless, the structural bases of collagen stability are still highly disputed [3–7]. The initially proposed model, based on the role of water [5], has progressively declined as hypotheses relying on stereoelectronic effect [4] and/or intrinsic iminoacid propensities have been proposed [3]. However, the great majority of these studies have been conducted using peptides with an over-represented iminoacid content. Indeed, the percentage of triplets with iminoacids in both X and Y is only 13% in real collagen.

Here we report a molecular dynamics investigation aimed at defining the role of solvent in the stabilization of collagen triple helix regions with a varied content of iminoacids. These studies were conducted on the peptide model T3-785, which contains three aminoacid triplets capped by Gly-Pro-Hyp triplets on both sides [8,9].

## Materials and methods

*System and simulation procedure.* All simulations were conducted using the X-ray structure of the peptide T3-785 (PDB code 1BKV) as a starting model [8,9]. This peptide consists of three aminoacids triples (Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala-Gly) capped by the iminoacid-rich $(Pro\text{-}Hyp\text{-}Gly)_3$ and $(Pro\text{-}Hyp\text{-}Gly)_4$ fragments at the N- and C-termini, respectively. Crystallographic water molecules were not included in the starting model in order to avoid any bias on the MD hydration sites. MD simulations were performed with the GROMACS software 3.3 [10]. To reproduce a high preference of Hyp residues for the so-called "up state", the $\chi 1$ side chain dihedral angles of Hyp residues were restrained to an average value of $-25°$ (Figure S1). In all MD simulations the model was immersed in a rectangular box ($68.4 \times 57.5 \times 59.9$ Å$^3$) filled with 7274 water molecules. The GROMOS43a1 force field and the SPCE water model were used in the simulations. The simulations were run with periodic boundary conditions. The electrostatic interactions were calculated using Part Mesh Ewald algorithm with a cutoff of 14 Å. Lennard–Jones interactions were calculated with a 14 Å twin-range cutoff. A dielectric constant of 1 and a time step of 2 fs were used. Systems were simulated in NPT ensemble by keeping constant temperature (300 K) and pressure (1 atm).

The original numbering scheme of peptide residues has been retained in the MD analyses. The hydration analysis is based on the solvent density map whose maxima represent hydration sites associated with the MD sampling [11,12]. The hydration sites bridging the main chain atoms on adjacent peptide chains have been labeled as Wi as reported in Figure S2.

* Corresponding authors. Present address: Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW Cambridge, UK (A. De Simone).
*E-mail addresses:* ad491@cam.ac.uk (A. De Simone), luigi.vitagliano@unina.it (L. Vitagliano).

0006-291X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2008.04.190

## Results

The diversified motifs, in terms of imino/aminoacid content, present in T3-785 allow analyses of triple helix hydration patterns that occur in real collagen. The standard unrestrained MD simulation (UN_MD) of the peptide was preceded by simulations carried out by applying restraints on the peptide coordinates. For a proper comparison of triple helix hydration to that observed in the crystal state, the initial MD simulation was conducted by restraining atomic positions to the crystallographic coordinates (FR_MD). The impact of side chain mobility on the water occupancy/residence time was investigated by running an MD simulation with only peptide backbone Cα atoms restrained to their initial positions (CA_MD).

*Restrained (FR_MD and CA_MD) simulations*

The analysis of the solvent distribution in the FR_MD reveals the presence of several peaks located in the proximity of the peptide chain (Fig. 1), in line with the experimental reports on T3-785 [9].

The dual nature of T3-785 with an aminoacid-rich region embedded in two iminoacid-rich fragments reflects into distinct hydration patterns [9]. The presence of an amino acid at the X position in the central region makes an extra main chain nitrogen atom available for hydrogen bonding interactions. This favors the insertion of a water molecule that bridges the main chain oxygen atom of Gly to the nitrogen atom of the amino acid located in the X position of an adjacent chain [8,13] (ζ water-bridge according to [5]). This bridge may involve polar side chains (Thr, Hyp) located in the Y position (Figure S2). By contrast, the hydration sites of the iminoacid-rich region are essentially external since they bind Hyp side chains and exposed carbonyl groups.

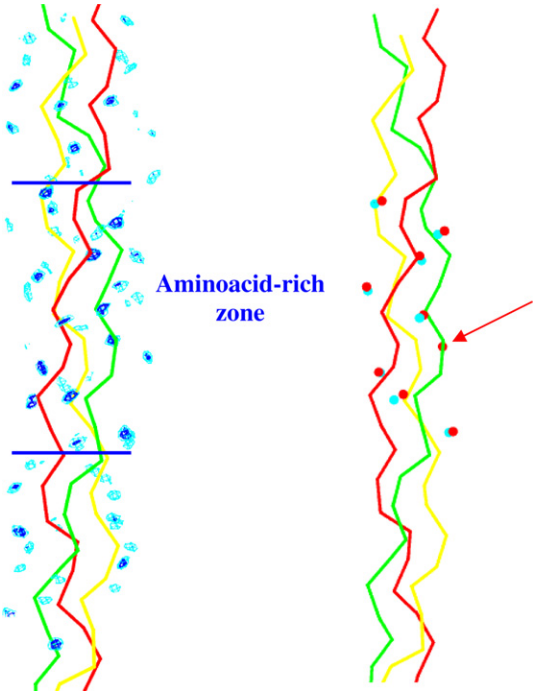


**Fig. 1.** (A) Water density of sites in the FR_MD simulation. Highest peaks in the map ($7\sigma$) are colored in blue. Peaks in cyan are countered at $4\sigma$. (B) A comparison between the location of the highest water density peaks (red) and crystallographic sites (cyan). The arrow indicates the site W6, which is not occupied by a water molecule in the crystal structure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

From the hydration analysis in the FR_MD simulation emerges that highest peaks in the water density map correspond to the main chain bridging water sites (Fig. 1A). The location of these sites is in close agreement with those observed crystallographically (Fig. 1B) [9]. Also the location of the hydration sites of the iminoacid region closely resembles that found experimentally [5,14] since peaks are located in proximity of exposed carbonyl and Hyp hydroxyl groups (Figure S3). The simulation is also able to identify water–water-bridges, evidenced experimentally [5].

The analysis of water occupancy (Figures S4A-B) indicates that, when a sufficiently large size for the box is used (radius 1.4 Å) to detect the presence of the water molecule, both external and ζ bridge sites are highly occupied. The lower intensity in the water maps of the peaks corresponding the external waters is due to the fact they are spread on a larger volume, as indicated by their lower occupancy when a smaller box is used (Figure S5). This is expected since the position of these sites is restrained by a single H-bond interaction with the peptide.

The analysis of the residence time evidences lower residence time for external water molecules, compared to ζ bridge waters (Figures S4C–D). Indeed, while an average residence time of the external water is 50 ps, all ζ bridge waters have residence time larger than 100 ps. The low residence time of external sites combined with their high occupancy observed in the present analysis is in line with the proposed hopping model for the hydration of iminoacid-rich peptides [15].

A deeper analysis (Figures S4C–D) also suggests that ζ bridge waters exhibit a diversified spectrum of residence times, whereas residence time distribution is approximately normal for the external sites. Our data indicate that the local environment of these sites (Figure S2) have a significant impact of the residence time (Fig. 2). The proximity of a polar/charged side chain significantly increases the residence time of the water site, with the lowest residence times observed for sites (W5, W6, and W7) in which no extra polar/H-bond interactions are formed by the water with the peptide. Although Thr70 is close to W4, its side chain is not properly oriented for H-bond formation in the starting crystal structure. Accordingly, the residence time of this site is similar to that observed for the sites W5–W7. Higher residence times are displayed by sites (W1, W2, and W3) in which the water molecules are further stabilized by H-bond interactions with the side chain of the residue located at the Y position. High residence times were also observed for ζ sites whose overall architecture involves the guanidinium moiety of Arg side chains (W8, and W9) (Fig. 2). While W9 site exhibits a residence time comparable to that of W1–W3, the W8 site is particularly suited for the binding

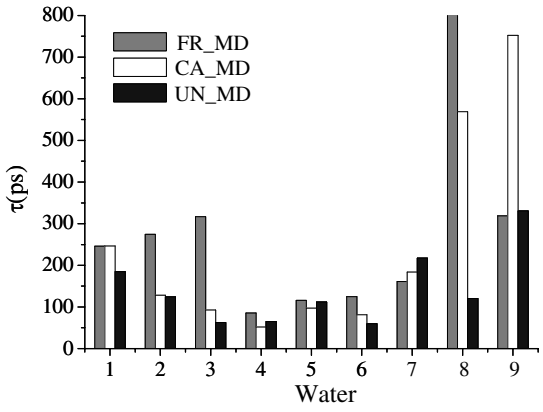


**Fig. 2.** Residence time in FR_MD (gray), CA_MD (white), and UN_MD (black) simulations for all ζ bridge water sites. Water sites have been numbered according to their locations along the peptide sequence.

of a water molecule. Indeed, the water molecules bound to W8 barely exchanges during the FR_MD dynamics. Obviously, the restrictions applied to Arg13 side chain strongly contribute to this effect.

Data emerged from FR_MD indicate that the procedure applied provides a reliable description of solvent distribution around a collagen-like peptide. To measure the effects of side chain mobility, the simulation was repeated by restraining only the coordinates of the Cα atoms to their initial positions (CA_MD). The restriction applied helped the system, whose central region was expected to be rather flexible, to keep the canonical triple helix structure. In CA_MD all major water sites retained very high occupancies (data not shown). The mobility of the side chains of the central region did not change the general picture obtained from the FR_MD simulation, since the highest residence times were displayed by ζ bridge waters. However, an impact of the mobility of side chains is observed on the residence time of water peaks in the central region. In particular, residence times of water sites which form interactions with residues in the Y position of an adjacent chain are affected by the mobility of these side chains (Fig. 2). In the case of water sites interacting with Hyp in Y (W1), there is no substantial variation of water residence time, consistent with the rigidity of this residue. These findings suggest that this type of water-bridge may contribute to the triple helix stability. On this basis it is also possible to rationalize the experimental finding that the *N*-methylation of Ala in the X position (Gly-Ala-Hyp triplets) leads to more pronounced destabilization effects (Tm decrease of 8.6 °C), compared to those observed upon *N*-methylation of Ala in the Y position (Gly-Pro-Ala triplets) (Tm decrease of 3.3 °C) [16]. Indeed, *N*-methylation of the X aminoacid likely hampers the binding of W1 type water molecule. Residence times of sites bound to Thr in Y (W2–W3) decrease significantly in the CA_MD (Fig. 2). The inspection of water densities confirms these observations (Figure S6).

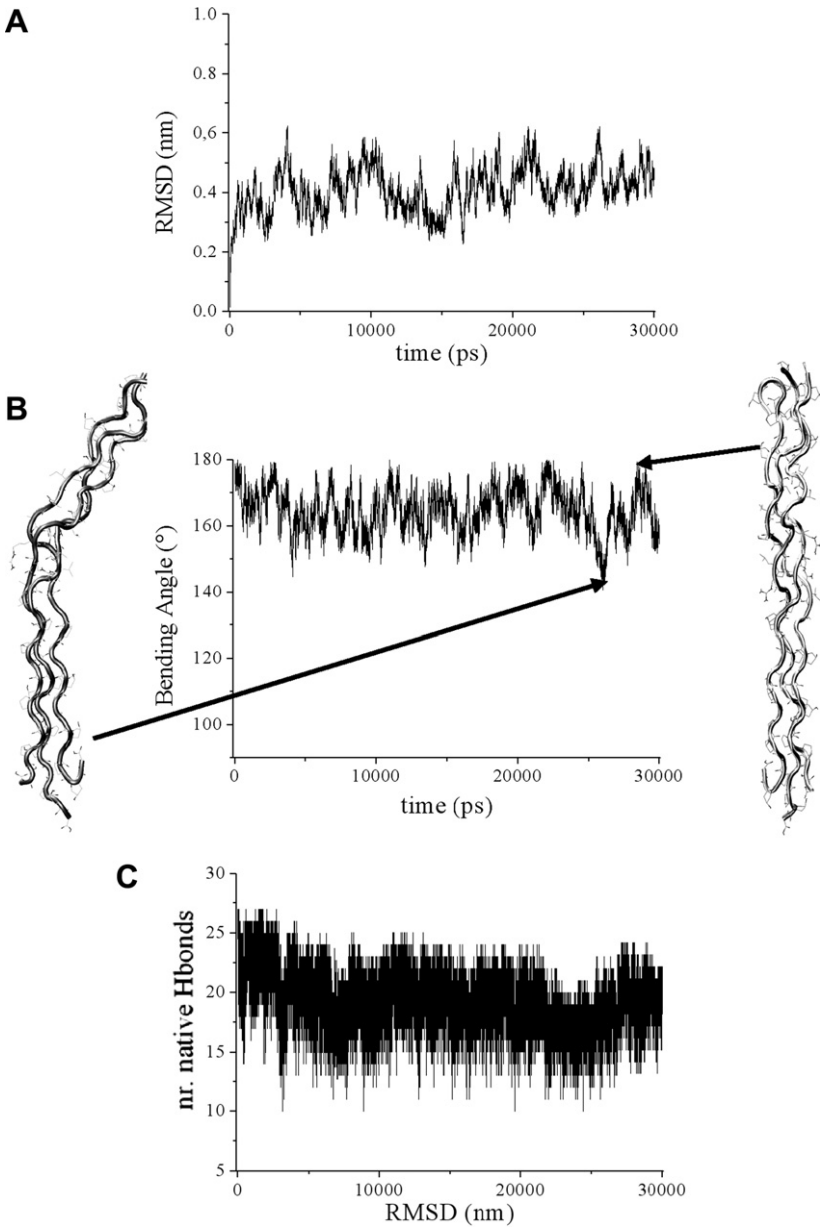


**Fig. 3.** Evolution of the peptide along the trajectory. (A) RMSD of the trajectory structures versus the X-ray model; (B) Triple helix bending angle of trajectory structures. (C) Evolution of total main chain–main chain hydrogen bonds. Representative models of linear and bent peptide structures are shown.

CA_MD simulation also provides information on Arg side chain mobility within a triple helix scaffold. As shown in Figure S7, a number of discrete rotameric states are observed for the Arg residues. Among these, one of the most populated, also detected in the crystal state, is characterized by direct H-bond interactions of the guanidinium group with the carbonyl moiety of an adjacent chain. The increased mobility of Arg side chains detected in the simulation is not surprising since these residues are involved in packing interactions in the crystalline structure [9]. Despite their relative flexibility, Arg side chains play a major role in the stabilization of the water sites W8 and W9. This is clear from the comparison (Fig. 2) of residence times of these water sites (~700 ps) with those of sites W4–W7 (~100 ps), which are only stabilized by main chain interactions.

*Unrestrained (UN) dynamics*

Several structural parameters were used as diagnostics to monitor the structural features of the structures in the trajectory of the unrestrained simulation (Fig. 3). The analysis of RMSDs between the starting X-ray model and the trajectory structures shows that the system rapidly evolves toward states which display RMSD values of about 0.4 nm (Fig. 3A). However, the fluctuations of the RMSD values suggest some structural variability of the system. Possible overall motions of the peptide centered on the aminoacid-rich region of the peptide were evaluated by defining a bending angle formed by center of mass of the N-terminal, the aminoacid-rich and C-terminal regions. As shown in Fig. 3B, this angle oscillates between 150° and 180°, and occasionally decreases to 140°.

The analyses of local variations occurring during the simulation clearly indicate that MD structures are stabilized by a smaller number of backbone–backbone H-bonds, which canonically stabilize triple helical motifs, when compared to the starting X-ray structure (Fig. 3C and Figure S8). Indeed, trajectory structures present an average number of native hydrogen bonds (close to 20) considerably lower than those present in the starting structure (27). This can be attributed both to unfolding effects at the N- and C-termini of the triple helix and to a large conformational freedom of the central aminoacid-rich region.

To quantify the dynamical behavior of the specific peptide regions, the simulation was analyzed by splitting the molecule in three zones, which correspond to the two iminoacid-rich regions (zones 1 and 3) and the central aminoacid-rich region (zone 2). As shown in Figure S9, RMSD values for the three zones are relatively low, compared to overall RMSD values (Fig. 3A). The comparison of RMSD values of the three zones shows that, as expected, zone 2 exhibits the largest deviations from the starting X-ray model.

The observed structural fluctuations of the central region clearly indicate that the ζ type water-bridges are not able to keep this part of the molecule in a rigid state. Typically, investigations on these structural motifs through model peptides usually require additional constraints to keep the triple helix stable. Despite the flexibility observed for the central region of the triple helix, the trends of water residence times emerged from the UN_MD simulation resemble those observed for FR_MD and CA_MD (Fig. 2 and Figure S10). As shown in Figure S10 residence times of the UN_MD well correlate with those found in the CA_MD (linear correlation coefficient of 0.91) for eight out of nine sites. The only significant exception to this trend is displayed by W8 whose residence time declines in the UN_MD simulation as a result of the more pronounced mobility of the local region embedding the site (Figure S8). The distortions of the local structure often lead to the formation of a direct hydrogen bond between the main chain nitrogen of Leu45 and the carbonyl oxygen of Gly14, whose interaction is mediated by the W8 water in the crystal structure (Figure S2). The formation of an additional direct backbone–backbone interaction, also observed in a previous MD analysis carried out using a different force field [17], is responsible for the decrease of the residence time of this water in the UN_MD simulation. The high mobility of the local structure of the peptide also reduces the residence time for the site W9, although to a minor extent.

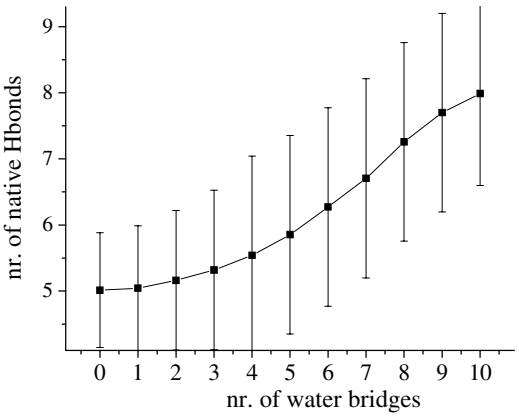


**Fig. 4.** Correlation between triple helix regularity and number of bridge waters. The numbers refer to the central aminoacid-rich zone of the peptide. The bars represent standard deviations.

The distortion of triple helix structure of the peptide in the central region provides an opportunity to investigate the correlation between triple helix regularity and presence of ζ type waters. The criterion used for the assessment of triple helix regularity was the presence and the number of backbone–backbone H-bonds that canonically stabilize this motif. The number of native main chain H-bonds increases with the number of ζ water-bridges present in the structures along the trajectory (Fig. 4). Although the associated standard deviations are rather high, the observed trend suggests that these bridges can occasionally stabilize the local structure of the peptide.

## Discussion

Complexity is one of the intrinsic features of the collagen molecule. Reductionistic approaches, through the use of peptide models, have been widely applied. However, the necessity to mimic collagen triple helix with smaller compounds has been met by enriching these models with iminoacids [1,2,18]. Although this approach has been extremely helpful, it should be pointed out that the extension of the ensuing results to real collagen may be limited. The structural characterization of the peptide T3-785 [8] has been instructive in this sense since it has shown that triple helix symmetry is sequence dependent and that the aminoacid-rich region possesses a specific hydration network (ζ bridges).

Most of the computational studies on collagen-like peptides have been focused on iminoacid-rich sequences [19–21]. To extend our knowledge of the water role for biologically relevant collagen sequences, we carried out MD analyses on the peptide T3-785. The overall validity of the approach was initially checked through a fully restrained MD simulation, which was expected to mimic the behavior of the peptide in the crystal state. The analysis of the hydration of the trajectory structures is in close agreement with that provided by the crystallographic analysis of the peptide. We then gradually increased the freedom of the peptide atoms in the simulation to analyze the role of these effects on collagen triple helix solvation. We carried out an MD

simulation by restraining the position of the Cα atoms (CA_MD) to evaluate the role of the ζ bridges within a stable triple helix scaffold. The analysis of the water residence time suggests that some stabilization to the triple helix is provided by a polar and rigid side chain (Hyp) through water-mediated H-bonds. It is interesting to note that this interaction may only be formed when Hyp is located in Y. Together with other factors [3], this may contribute to the observed preference of Hyp residues for the Y position. On the other hand, mobile polar side chains (e.g. Thr) provide little contribution to the stabilization of the site. Thr-driven stabilizing effects likely arise from post-translational modifications of these residues (i.e. glycosylation) [22].

Interestingly, charged Arg side chains produce the strongest effects on these hydration sites. This suggests that the observed stabilization effects of this residue when located in the Y are not limited to their H-bond interactions of the guanidinium group with the carbonyl moiety of adjacent chains in the triple helix scaffold [9,23]. Since side chains of Arg residues located in Y stabilize the water sites which are H-bonded to the nitrogen main chain atom of the aminoacid present in the X+1 position (Figure S2), we analyzed collagen sequences to check the occurrences of amino or iminoacid in position X+1 when an Arg residue is located in Y. We selected for these investigation sequences of the homotrimeric type III collagen. While sequences of the type Gly-Pro represent 28% of the total Gly-X sequences (95 out of 340) in human type III collagen, sequences Arg-Gly-Pro only represent 15% of the total number of Arg-Gly-X (6 out of 40). The enrichment in aminoacids at position X+1 when an Arg is present at Y is also observed in other mammalian type III collagens. This finding may be correlated to the ability of Arg in Y to stabilize the water site bound to the nitrogen of the aminoacid in X+1.

As observed earlier [17], these water-bridges are not sufficient to maintain the structure of the aminoacid-rich in a rigid state. These findings may suggest that the flexibility of the region is important for the cleavage of this site by metallo-proteases [17].

Despite the observed flexibility of the peptide and the transient nature of the bridge hydration sites, the hydration pattern identified in the UN_MD simulation is similar to that detected in the CA_MD simulations. The correlation between the number of ζ bridge waters and the number of native triple helix H-bonds observed in the UN_MD suggests a role for these bridges in the stabilization of the triple helix. Although, these effects are not sufficient to hold an iminoacid free region in a triple helix, they may be effective in regions with a mixed aminoacid/iminoacid composition. Indeed, consecutive triplets constituted solely by aminoacids are rather rare in real collagen. The local rigidity of the triple helix in real collagen is likely in between those exhibited by models of the simulations CA_MD and UN_MD.

In conclusion, our findings highlight the complexity of collagen triple helix. We suggest that the role of water in collagen stabilization may strongly depend on the local sequence. While the Hyp-induced stabilization in iminoacid-rich regions is likely dictated by its intrinsic conformational preferences [3], the stabilizing effect of this unique residue is also mediated by water molecules in regions with a mixed aminoacid/iminoacid content. Our data also suggest that water-mediated stabilizations also hold for charged residues. Taking into account the very low tendency of fluorine to form H-bonds, we predict that the over-stabilization of 4RFluoroproline (Flp) versus Hyp [4] may be reduced in peptides containing aminoacids in the X position.

## Acknowledgments

We thank Region Campania (L.R. 5 2003) for financial support. CINECA Supercomputing (Project cne0fm4h) is acknowledged for computational support. A.D.S. is supported by European Molecular Biology Organization (EMBO), Fellowship No.: 584-2007.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2008.04.190.

## References

[1] B. Brodsky, A.V. Persikov, Molecular structure of the collagen triple-helix, Adv. Protein Chem. 70 (2005) 301–339.
[2] C.L. Jenkins, R.T. Raines, Insights on the conformational stability of collagen, Nat. Prod. Rep. 19 (2002) 49–59.
[3] L. Vitagliano, R. Berisio, L. Mazzarella, A. Zagari, Structural bases of collagen stabilization induced by proline hydroxylation, Biopolymers 58 (2001) 459–464.
[4] S.K. Holmgren, K.M. Taylor, L.E. Bretscher, R.T. Raines, Code for collagen's stability deciphered, Nature 392 (1998) 666–667.
[5] J. Bella, B. Brodsky, H.M. Berman, Hydration structure of a collagen peptide, Structure 3 (1995) 893–906.
[6] R. Berisio, V. Granata, L. Vitagliano, A. Zagari, Imino acids and collagen triple helix stability: characterization of collagen-like polypeptides containing Hyp-Hyp-Gly sequence repeats, J. Am. Chem. Soc. 126 (2004) 11402–11403.
[7] R. Improta, R. Berisio, L. Vitagliano, Contribution of dipole–dipole interactions to the stability of the collagen triple helix, Protein Sci. 17 (2008) 955–961.
[8] R.Z. Kramer, J. Bella, P. Mayville, B. Brodsky, H.M. Berman, Sequence dependent conformational variations of collagen triple-helical structure, Nat. Struct. Biol. 6 (1999) 454–457.
[9] R.Z. Kramer, J. Bella, B. Brodsky, H.M. Berman, The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence, J. Mol. Biol. 311 (2001) 131–147.
[10] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (2005) 1701–1718.
[11] A. De Simone, C. Pedone, L. Vitagliano, Structure, dynamics, and stability of assemblies of the human prion fragment SNQNNF, Biochem. Biophys. Res. Commun. 366 (2008) 800–806.
[12] G. Colombo, M. Meli, A. De Simone, Computational studies of the structure, dynamics and native content of amyloid-like fibrils of ribonuclease A, Proteins 70 (2008) 863–872.
[13] J. Bella, H.M. Berman, Integrin–collagen complex: a metal–glutamate handshake, Struct. Fold Des. 8 (2000) R121–R126.
[14] R. Berisio, L. Vitagliano, L. Mazzarella, A. Zagari, Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)(10)](3) Protein Sci. 11 (2002) 262–270.
[15] G. Melacini, A.M. Bonvin, M. Goodman, R. Boelens, R. Kaptein, Hydration dynamics of the collagen triple helix by NMR, J. Mol. Biol. 300 (2000) 1041–1049.
[16] E.A. Kersteen, R.T. Raines, Contribution of tertiary amides to the conformational stability of collagen triple helices, Biopolymers 59 (2001) 24–28.
[17] C.M. Stultz, Localized unfolding of collagen explains collagenase cleavage near imino-poor sites, J. Mol. Biol. 319 (2002) 997–1003.
[18] R. Berisio, L. Vitagliano, L. Mazzarella, A. Zagari, Recent progress on collagen triple helix, structure, stability, and assembly, Protein Pept. Lett. 9 (2002) 107–116.
[19] R.J. Radmer, T.E. Klein, Severity of osteogenesis imperfecta and structure of a collagen-like peptide modeling a lethal mutation site, Biochemistry 43 (2004) 5314–5323.
[20] S. Park, T.E. Klein, V.S. Pande, Folding and misfolding of the collagen triple helix: Markov analysis of molecular dynamics simulations, Biophys. J. 93 (2007) 4108–4115.
[21] S.S. Raman, R. Parthasarathi, V. Subramanian, T. Ramasami, Role of length-dependent stability of collagen-like peptides, J. Phys. Chem. B. 112 (2008) 1533–1539.
[22] J.G. Bann, H.P. Bachinger, Glycosylation/hydroxylation-induced stabilization of the collagen triple helix, J. Biol. Chem. 275 (2000) 24466–24469.
[23] L. Vitagliano, G. Némethy, A. Zagari, H. Scheraga, Stabilization of the triple-helical structure of natural collagen by side-chain interactions, Biochemistry 32 (1993) 7354–7359.