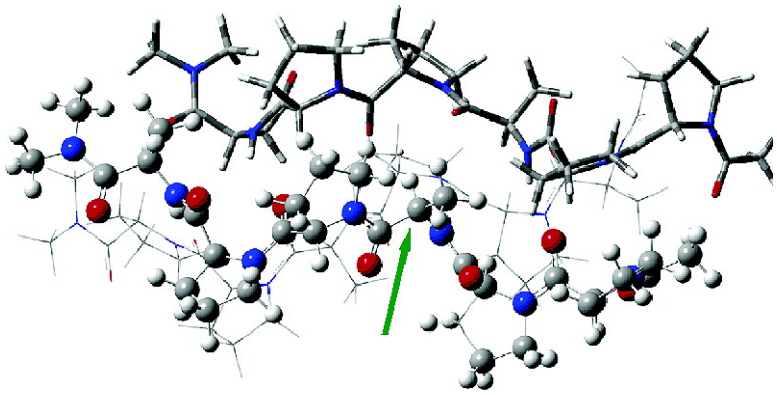
Communication

Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProProAla, and ProProSer Triads

, Yujia Xu, and J. J. Dannenberg

J. Am. Chem. Soc., **2005**, 127 (41), 14130-14131• DOI: 10.1021/ja053768y • Publication Date (Web): 27 September 2005



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 1 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
COMMUNICATIONS

Published on Web 09/27/2005

Completely Geometrically Optimized DFT/ONIOM Triple-Helical Collagen-like Structures Containing the ProProGly, ProProAla, ProPro^{D}Ala, and ProPro^{D}Ser Triads

Midas (I-Hsien) Tsai, Yujia Xu, and J. J. Dannenberg*

Department of Chemistry, City University of New York—Hunter College and The Graduate School, 695 Park Avenue, New York, New York 10021

Received June 8, 2005; E-mail: jdannenberg@gc.cuny.edu

The structural protein, collagen, recently reviewed,[1] consists of triple helical peptides whose structure was proposed by Rich and Crick,[2] in which each of three peptide strands contains the repeating amino acid triad XaaYaaGly, where Xaa and Yaa can be any amino acid residues. In this structure, each peptide strand H-bonds to one of the other two using its Gly N−H's as donors, and to the third, using C═O's on other amino acid residues as acceptors. These triple helices are rich in proline and 4-(*R*)-hydroxyproline (Hyp), which constitute about 20% of the amino acid content, while the most common repeating unit is Pro-Hyp-Gly.[3] Various diseases, such as osteogenesis imperfecta, Ehlers−Danlos syndrome, Alport syndrome, Schmid metaphyseal chondrodysplasia, and dystrophic epidermolysis bullosa, result from amino acid residue mutations in collagens, mostly mutations of Gly.[4,5] These lead to misfolding of the collagen, which can be modeled with short synthetic peptide chains.[6]

To probe the effect of Gly mutations, we present DFT and ONIOM[7,8] calculations using B3LYP/D95(d,p) and AM1 as the high (peptide backbone) and low (all side chains) layers on completely geometrically optimized structures of several triple helices. These structures all contain 18 amino acid residues, which qualify them as among the largest peptide structures, as well as the first triple helical peptides that have been fully optimized by this or other DFT methods. All calculations used the GAUSSIAN 03 suite of programs[9] using a procedure fully described elsewhere.[10] The triple helical structures were corrected for basis set superposition error using the a posteri counterpoise correction (CP). We initially considered the triple helix made from the repeating unit ProProGly. The triple helical structure contains two repeats of the ProProGly triad in each strand. One strand had its sequence dislocated by one amino acid from the other two so that six H-bonds could form (ProGlyPro instead of ProProGly). The triple helices reported here are formed from strands that are capped with acetyl and dimethyl amido groups to prevent the formation of H-bonds involving the COOH's and NH_2's of the peptide strands. We use the simple triad (i.e., ProProGly) to designate the strand containing two repeats and the capping groups or the triple helix of three strands in the following discussion. We compare the energies of the triple helices with those of (a) separated individual peptide strands; (b) the component amino acids (by considering the appropriate condensation reactions: $(CH_3)_2NCOCH_3$ + N amino acids → peptide + N waters); and (c) the energies of the individual strands frozen in their triple helical geometries to illustrate the effect of the reference states.

The structure of the extended single strand of ProProGly is quite tortured (Figure 1). Consequently, the formation of the three ProProGly strands from the component amino acids is calculated to be energetically unfavorable by 53.0 kcal/mol. The value is consistent with other calculations from our laboratory, which indicate that the substitution of a Pro for an Ala in $Ac(Ala)_{17}NH_2$ destabilizes the peptide by about 7.5 kcal/mol (there are 12 P's in the combined three strands).[11]

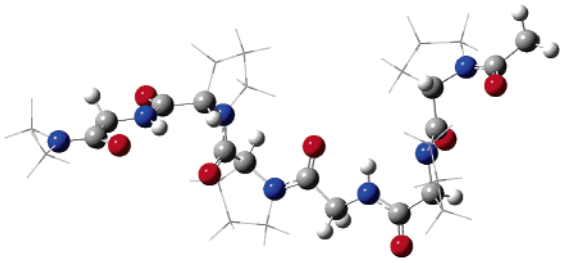


Figure 1. Optimized ProProGly single strand.

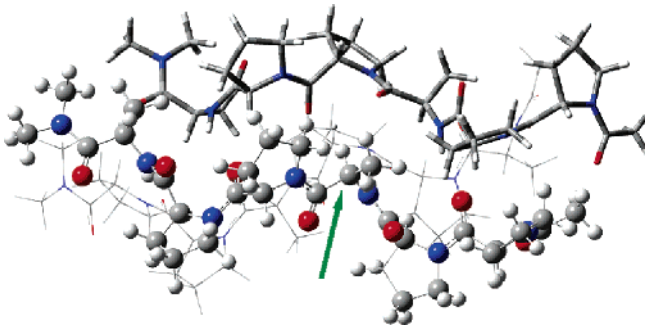


Figure 2. The ProProGly triple helix with one ^{D}Ala (indicated by green arrow). Each of the three strands is rendered differently (ball and stick, tube, and wire frame).

The H-bonding interaction in the triple helix is −20.8 kcal/mol after counterpoise (CP) but no vibrational correction, while the average O···N distance across the six H-bonds is 2.961 Å compared to 3.012, 2.960, 2.971, and 2.968 Å for protein data bank structures 1A3I, 1A3J, 1G9W, and 1ITT, respectively. We note that our calculations were optimized before CP correction, thus they provide a lower limit for both the H-bonding stabilizations and the H-bonding distances for this level of calculation. On the other hand, the experimental H-bond distances come from much longer chains, so they are less influenced by end effects (where the H-bonds are longer).

From inspection of Figures 1 and 2, one can easily see that the requirement for Gly as every third amino acid in collagen and collage-like triple helices derives from its enantiomorphic property, unique among naturally occurring amino acids. Had an ^{L}amino acid been substituted for Gly, the side chain would sit in the center of the triple helix, causing considerable strain and weakening the structure. However, if a ^{D}amino acid were substituted for Gly, its side chain would extend peripherally from the helix, seemingly avoiding this steric problem. To test this hypothesis, we considered ProProGly triple helices, where a single Gly near the

10.1021/ja053768y CCC: $30.25

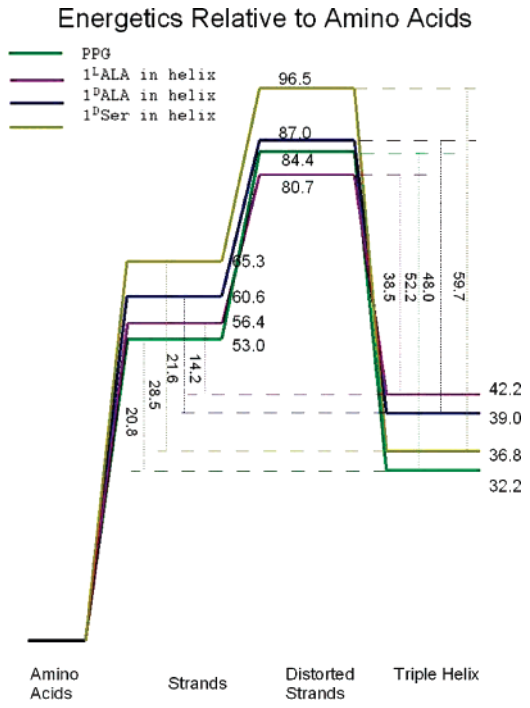


Figure 3. Energies of the (combined) three optimized strands, strands distorted to their triple helical geometries, and optimized triple helices (including CP correction). PPG represent ProProGly.

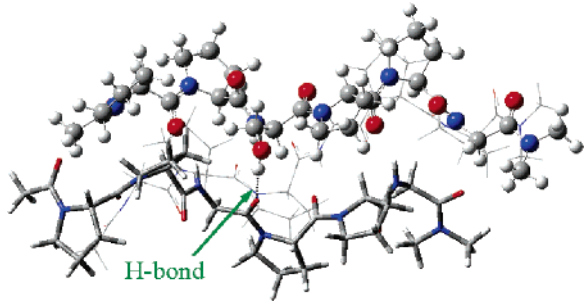


Figure 4. Structure of the triple helix containing one ^{D}Ser in place of Gly. The additional H-bond is indicated.

center of the structure was mutated either for ^{L}Ala or ^{D}Ala. The mutated triple helices had H-bonding interactions of −14.2 and −21.6 kcal/mol for ^{L} and ^{D}Ala, respectively. Thus, ^{L}Ala *reduces* the H-bonding stabilization by 6.6 kcal/mol while increasing the average N···O distance to 3.039 Å, and ^{D}Ala *increases* the stabilization by 0.8 kcal/mol but slightly increases the average N···O distance (to 2.999 Å).

The increased H-bonding in the structure with the Gly → ^{D}Ala mutation encouraged us to evaluate the effect of a Gly → ^{D}Ser mutation. Ser is simply Ala with one methyl H transformed to an OH, which would be near enough to the C═O of a neighboring strand to form an additional H-bond, thus increasing the triple helical stability. In fact, this triple helix had the most stabilizing H-bonding energy of all: −28.5 kcal/mol (−7.7 more than ProProGly) with respect to the optimized single peptide strands with an average N···O of 3.029 Å (see Figure 4).

When considering the interaction energies upon going from the single strands to the triple helix, one must realize that there is a bias inherent in the choice of the optimized single strands as a reference point, and that the interaction energy calculated in this manner can be conceived of as a combination of both a distortion energy from the optimized strand to the conformation it assumes in the triple helix and the interactions of the three distorted strands into the helix. When approached this way, it becomes immediately clear that the closer the geometry of an optimized strand is to that of its distorted geometry, the greater the stabilization upon forming the triple helix (assuming the stabilizations from the distorted strands to the helix remain the same). Further reflection suggests that the energies of the strands might be different with respect to the component amino acids. The capped (single) strand ProPro^{L}AlaProProGly is 4.1 kcal/mol more stable than ProPro^{D}AlaProProGly, its diastereomer, as is reflected in the total energies of the three optimized strands that differ by the configuration of the Ala. However, when comparing nonisomeric peptides, we can use the polycondensation reaction to evaluate their energies relative to the component amino acids. As seen from Figure 3, the energies of the three states considered (optimized and distorted strands and triple helices) assume a different order. This provides an extraordinary example of the importance of the choices of reference state. Thus, the ProProGly triple helix is the most stable with reference to the amino acids, but the triple helix with the ^{D}Ser is the most stable with reference to the optimized strands.

We note that the energy of triple helix containing ^{D}Ala has a significantly smaller (4.7 kcal/mol) CP than that containing ^{L}Ala (due to the greater effect of the orbitals of the methyl buried in the middle of the ^{L}Ala strand compared to those of the exterior methyl in the ^{D}Ala strand), and that this contributes significantly to the difference in interaction energies.

In summary, the enantiomorphic Gly plays the role of a ^{D}amino acid in the repeating triad XYG. Thus, ^{D}amino acids can potentially replace it to energetic advantage.

Acknowledgment. This work was supported in part by grants from the donors of the Petroleum Research Foundation administered by the American Chemical Society and from PSC-CUNY. Some calculations were performed on the CUNY Graduate School research-computing cluster.

Supporting Information Available: Cartesian coordinates of the relevant structures and complete ref 8. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Bhattacharjee, A.; Bansal, M. *IUBMB Life* **2005**, *57*, 161−172.
(2) Rich, A.; Crick, F. H. C. *J. Mol. Biol.* **1961**, *3*, 483−506.
(3) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. *Science* **1994**, *266*, 75−81.
(4) Persikov, A. V.; Pillitteri, R. J.; Amin, P.; Schwarze, U.; Byers, P. H.; Brodsky, B. *Hum. Mutat.* **2004**, *24*, 330−337.
(5) Chan, D.; Ho, M. S. P.; Cheah, K. S. E. *J. Biol. Chem.* **2001**, *276*, 7992−7997.
(6) Baum, J.; Brodsky, B. *Curr. Opin. Struct. Biol.* **1999**, *9*, 122−128.
(7) Morokuma, K. *Philos. Trans. R. Soc. London, Ser. A* **2002**, *360*, 1149−1164.
(8) Dapprich, S.; Komiromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *THEOCHEM* **1999**, 0166−1280.
(9) Frisch, M. J. et al. *GAUSSIAN 03*, revision C01; Gaussian, Inc.: Pittsburgh, PA, 2003.
(10) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2004**, *126*, 14198−14205.
(11) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2005**, submitted for publication.

JA053768Y