JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Identifying Residual Structure in Intrinsically Disordered Systems: A 2D IR Spectroscopic Study of the GVGXPGVG Peptide

Joshua Lessing,[†] Santanu Roy,[‡] Mike Reppert,[†] Marcel Baer,[§,∥] Dominik Marx,[§] Thomas La Cour Jansen,[‡] Jasper Knoester,[‡] and Andrei Tokmakoff*[,†]

[†]Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States

[‡]Center for Theoretical Physics and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

[§]Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany

Supporting Information

ABSTRACT: The peptide amide-I vibration of a proline turn encodes information on the turn structure. In this study, FTIR, two-dimensional IR spectroscopy and molecular dynamics simulations were employed to characterize the varying turn conformations that exist in the GVGX^{L}PGVG family of disordered peptides. This analysis revealed that changing the size of the side chain at the **X** amino acid site from Gly to Ala to Val substantially alters the conformation of the peptide. To quantify this effect, proline peak shifts and intensity changes were compared to a structure-based spectroscopic model. These simulated spectra were used to assign the population of type-II β turns, bulged turns, and irregular β turns for each peptide. Of particular interest was the Val variant commonly found in the protein elastin, which contained a 25% population of irregular β turns containing two peptide hydrogen bonds to the proline C═O.

Extensive study of small peptides has revealed the local forces that contribute to protein secondary structure.[1−10] In contrast, far less is known about the forces that contribute to the structure of disordered proteins, where local effects dominate.[11−17] This dearth of molecular-level information on disordered systems results from the lack of detailed structural tools. Disordered proteins cannot be crystallized, and X-ray diffraction or scattering studies provide limited structural information. Although NMR spectroscopy is selective for long-lived conformational states, since conformational exchange typically occurs on picosecond to microsecond time scales, NMR analysis cannot be used to resolve conformational heterogeneity fully.[15,18] Two-dimensional IR spectroscopy (2D IR), with its subpicosecond time resolution, can be used to characterize distinct conformers and structural variations in intrinsically disordered systems.[10,19−21]

In the present study, we investigated the role played by the amino acid preceding a Pro-Gly (PG) turn in determining the turn structure of an intrinsically disordered peptide. The XPG sequence is commonly found in fibrous and elastomeric proteins, in which **X** influences the structural and mechanical properties. Examples include collagen (**X** = Pro),[22] elastin (**X** = Val),[23] mussel byssus (**X** = Gly),[24] dragline spider silk (**X** = Gly),[25] and wheat glutenin (**X** = Gln).[26] To provide a better understanding of the structure and folding of an XPG peptide turn, we investigated how the side-chain size of the N-terminal amino acid of Pro modulates the structure of the PG turn in the disordered peptides GVGX^{L}PGVG (**X** = Gly, Ala, Val). These peptides were characterized with amide-I 2D IR in combination with structural modeling of the spectra.[27−29]

While protein secondary structure assignments from amide-I spectra have been used for years, it is only recently that structure-based spectral calculations have become available for detailed interpretation.[28,30] The amide-I vibration is primarily a C═O stretching motion of a peptide unit; however, the normal modes of vibration are delocalized and can involve all of the backbone amide units of a peptide. The frequency and line shape of the amide-I vibration depend on conformation-dependent vibrational couplings between neighboring amide units as well as the interactions of peptide groups with the local electrostatic environment, such as hydrogen bonding to the amide's C═O or N−H moiety. As a result, the amide-I spectrum encodes information about protein secondary structure.

Unlike other proteinogenic amino acids, the tertiary amide structure of proline changes the amide-I vibration of the peptide linkage preceding proline in sequence. The three heavy atoms bound to its nitrogen give the proline amide-I vibration a larger reduced mass, shifting it to a lower frequency. This frequency shift largely decouples the proline amide-I vibration from those of its neighbors, moving its resonance to a predominantly background-free region on the red side of the absorption spectrum. As a result, proline amide-I spectroscopy can also be used as a local probe of protein structure. Because the proline amide-I vibration has almost all C═O stretch character, a frequency shift of the proline amide-I vibration is either predominantly the result of a change in hydrogen bonding to the amide oxygen or other changes in the electric field experienced by the amide C═O. (For spectroscopic reasons, we refer to the main-chain amide or carbonyl preceding proline as the proline C═O, although it derives

Received: December 6, 2011
Published: February 22, 2012

   dx.doi.org/10.1021/ja2114135 | J. Am. Chem. Soc. 2012, 134, 5032−5035

from the preceding amino acid). With the help of a recently parametrized structural model for proline IR spectroscopy,[29] we were able to use this intrinsically local vibration as a sensitive probe of proline turn structure.

The series of model peptides studied here differ in the volume of the side chain on the N-terminal side of proline. The peptide in the series with the smallest volume is Gly, with volume increases $\Delta V(\mathbf{X} - \text{Gly})$ of 25.1 and 75.3 Å^3 for **X** = Ala and Val, respectively.[31] The resulting changes in turn conformation were detected in the experimental IR spectra as red shifts of the low-frequency proline amide-I vibration with increasing side-chain size, from 1625 cm^{-1} for Gly to 1614 cm^{-1} for Val (Figure 1). The features in β-turn peptides at 1670

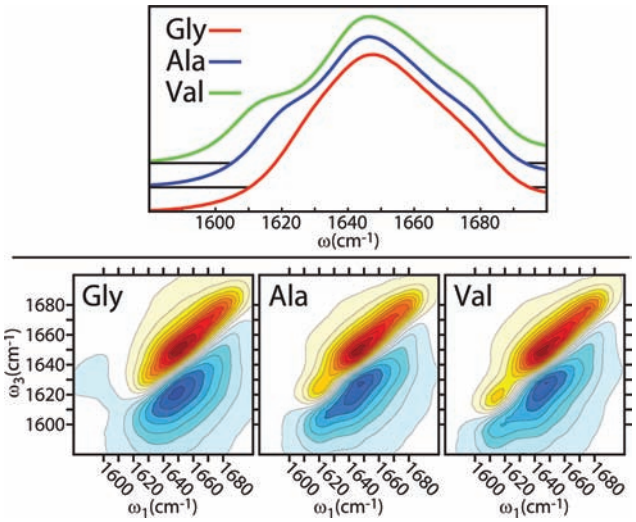

Figure 1. (top) FTIR and (bottom) 2D IR spectra of GVG**X**^{L}PGVG at T = 10 °C in pH 1.0 DCl in D_2O. The proline resonance frequencies are 1625 cm^{-1} for Gly, 1619 cm^{-1} for Ala, and 1614 cm^{-1} for Val. The maximum optical densities were 0.23−0.25. Contours for all of the 2D IR surfaces presented in this paper are plotted from −1.0 to 1.0 in equally spaced steps of 0.08.

and 1640 cm^{-1} are commonly assigned to the parallel and perpendicular bands of an antiparallel β sheet.[9] Since the proline amide-I vibration is only weakly coupled to the remaining amide vibrations of the peptide, the observed frequency shift primarily reflects variations in the configuration of the hydrogen bond to the oxygen of the C═O moiety preceding proline.

As a rule of thumb for a time-averaged observation, the amide frequency is red-shifted by roughly 16 cm^{-1} for each hydrogen bond accepted, and on average, the oxygen can accept up to two hydrogen bonds.[10] Therefore, these shifts encode considerable changes in the turn hydrogen-bond configuration and possible changes in solvent exposure. We note that although these spectra were acquired at pH 1.0 in order to shift the C-terminal COO^- asymmetric stretch at 1590 cm^{-1} out of the spectral window, our findings are consistent with spectra measured at neutral pH [see the Supporting Information (SI)].

To extract information on the peptide structure from the IR data, FTIR and 2D IR spectra were calculated from molecular dynamics (MD) simulations of the three peptides using a mixed quantum−classical model in which classical structures are mapped onto elements of a local-mode Hamiltonian for the coupled amide-I vibrations. Our model assigns frequencies to the individual amide-I oscillators on the basis of the electric fields experienced by individual amide units and allows for through-bond and through-space coupling effects.[27−29] This model is accurate enough that general features such as peak splitting and relative amplitudes can be used for structure interpretation, but exact frequencies are less reliable metrics.

Eleven structurally distinct initial configurations were selected from a 200 ns MD trajectory of the Val peptide carried out previously.[13−16] These initial configurations were representative of the main structural and dynamical motifs observed, including extended and collapsed states, and a variety of PG turn configurations. This judicious choice of initial conformations enabled us to carry out rather short simulations to compute 2D IR spectra of local structural motifs, since the picosecond time scale of the experiments is far shorter than the (multi)nanosecond relaxation times of the peptide. These 11 configurations were then mutated to obtain the initial configurations of the Ala and Gly variants. New MD trajectories were generated starting from these initial configurations (for the protocol, see the SI) and subsequently used to simulate 2D IR spectra. The simulations of these disordered peptides sampled a variety of configurations, including variations in turn structures, fraying or extension, and side-chain packing. Although these simulations may not have provided a complete sampling of all configurations, the sampling was extensive enough to provide a diversity of trial structures from which structure/spectrum correlations could be created.

We distinguished turn conformers on the basis of their hydrogen-bonding patterns to the **X** carbonyl preceding the proline. Six hydrogen-bond bins were investigated: 0/0, 1/0, 0/1, 1/1, 2/0 and 0/2, where the first digit corresponds to the number of hydrogen bonds the CO accepts from other amide groups in the peptide and the second digit to the number of hydrogen bonds the CO accepts from water. Dynamic FTIR and 2D IR spectra were then calculated for each hydrogen-bond bin by ensemble-averaging only the parts of the trajectories belonging to a particular bin (see the SI for a detailed description). As expected, the calculated FTIR spectra showed that the lowest-frequency amide-I peak red-shifted with increasing number of hydrogen bonds to the proline. From the simulated FTIR spectra of Gly in Figure 2 we see that (1) the proline resonance is not resolved for the 0/0 bin (Figure 2 - Left), which consists of disordered turns or condensed states that do not accommodate hydrogen bonds; (2) the proline resonance has a weakly red-shifted tail and a small increase in intensity for the 1/0 bin (Figure 2 - Center), which

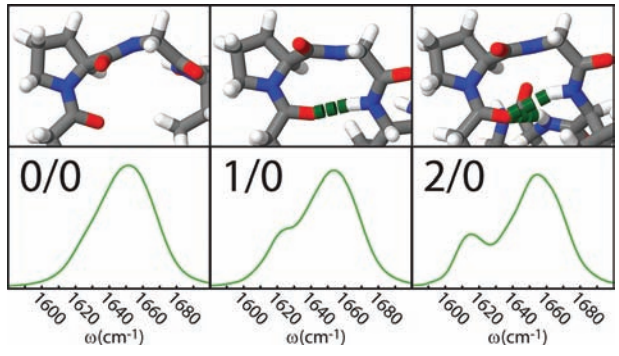

Figure 2. (top) Visualization of representative turn structures in the 0/0, 1/0, and 2/0 bins. (bottom) Simulated FTIR spectra for the Gly peptide in the 0/0, 1/0, and 2/0 bins.

dx.doi.org/10.1021/ja2114135 | J. Am. Chem. Soc. 2012, 134, 5032−5035

predominantly consists of type-II β turns ($i \rightarrow i + 3$ hydrogen bonds) or bulged turns ($i \rightarrow i + 4$ hydrogen bonds); and (3) the proline resonance has both a large red shift and an increase in intensity for the 2/0 bin (Figure 2 - Right), which predominantly consists of an irregular β turn with dual hydrogen bonds ($i \rightarrow i + 3$ and $i \rightarrow i + 4$) corresponding to ϕ/ψ angles for the VPGV sequence of (i) −79°/151°, (i + 1) −59°/127°, (i + 2) 118°/−29°, and (i + 3) −128°/−30°. These simulated proline peak shifts were also observed for the lowest-frequency amide-I peak in the experimental spectra. For example, a high population of the irregular β turn was observed in the experimental spectra for the Val peptide, which has a proline peak with a red shift relative to the band maximum of $\omega_{max} - \omega_p = 31$ cm^{-1}. This model reproduced the features in the experimental spectra except for the parallel mode of the antiparallel β sheet, which was found to be more intense in the experimental data, where it was observed as a subtle shoulder on the blue side of the amide-I band (Figure 1). Since the proline red shift was the primary observable used to predict the structure in this study, a small difference in intensity on the blue side of the spectra did not diminish the value of the model.

To assign the hydrogen-bond population for each peptide, 2D IR spectra were simulated for the six hydrogen-bond bins. The nonlinear intensity dependence in conjunction with the additional frequency dimension of a 2D IR spectrum reveals information on hydrogen-bond composition not seen in FTIR spectra. Complete simulations for the three peptides are provided in the SI, and the salient features common to all of the peptides simulated are summarized in Figure 3. Now better

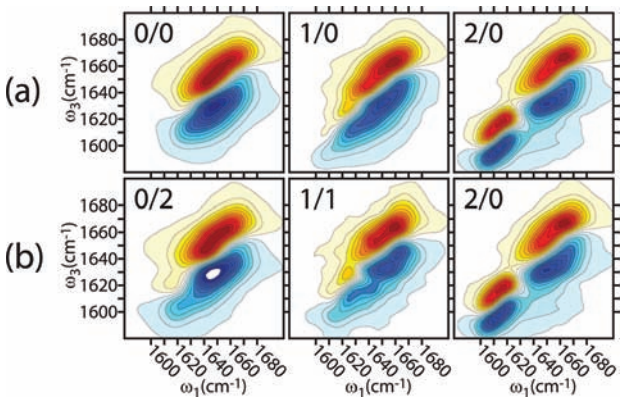

Figure 3. (a) Simulated 2D IR spectra of the Val peptide as a function of number of peptide−peptide hydrogen bonds. (b) Simulated 2D IR spectra of the Val peptide as a function of the number of peptide−peptide vs peptide−water hydrogen bonds.

resolved, the red shift of the proline resonance provides a count of the number of hydrogen bonds to proline (Figure 3a). Fitting the simulated lowest mode peak frequency as a function of the total number of hydrogen bonds to proline (n_{THB}) for all six states gave the equation $\omega_p(\text{cm}^{-1}) \approx -12.6n_{THB} + 1642$.

In addition, spectral modeling of the six bins showed that for these peptides, the proline peak is more intense when proline is hydrogen-bonded to a peptide amide unit than when proline is hydrogen-bonded to water. This effect is amplified by the nonlinear intensity dependence of a 2D IR spectrum, thereby providing a method of distinguishing whether the proline hydrogen bonds are donated by peptide groups or water. Although it remains a challenge to simulate precise line shapes, features such as the relative intensity or frequency splitting of the proline and main amide bands are characteristic enough to distinguish uniquely those states with varying peptide-to-proline hydrogen-bonded configurations. These conformers can be grouped as irregular $i \rightarrow i + 3$ and $i \rightarrow i + 4$ turns [2/0 or two-hydrogen-bond (2HB) turns]; $i \rightarrow i + 3$ β turns or $i \rightarrow i + 4$ bulged turns (1/0 and 1/1 or 1HB turns); compact disordered states (0/0 and 0/1); and the 0/2 states, which consist primarily of extended conformers.

To assign an approximate hydrogen-bond composition for each peptide, the experimental 2D IR surfaces were fit for frequencies <1650 cm^{-1} with the simulated 2D IR spectra for all of the peptide bins using a genetic algorithm containing a least-squares fitness function and no additional constraints (see the fitted spectra in the SI). The Gly peptide was found to contain ∼15% 1HB turns and <1% 2HB turns, with the remainder consisting of highly disordered peptide configurations. The Ala peptide was found to contain ∼70% 1HB turns and <10% 2HB turns, with the remainder in disordered condensed or extended states. This is consistent with a high population of type-II β turns and bulged turns for the folded state of the Ala peptide. Finally, the Val peptide was found to contain ∼60% 1HB turns and ∼25% 2HB turns, which is consistent with a high population of type-II β turns, bulged turns, and irregular β turns.

From the 2D IR data, we observed a systematic change in the turn conformational preference as a function of point mutation. We found that as the size of the side chain of the **X** amino acid decreases, the number of proline hydrogen bonds to the peptide decreases and the number of proline hydrogen bonds to water increases. This indicates that the preference for the **X**-Pro-Gly sequence to form a β turn increases in the order Gly < Ala < Val. The observed stabilization of the **X**-Pro-Gly turn may arise from a number of factors, including steric clashes between the side chain of the point mutation and the proline pyrrolidine ring and hydrophobic clustering of the **X** and Val side chains.

The Val peptide's irregular 2/0 β turn was found to result from the formation of dual hydrogen bonds between proline and the N−H moieties of two sequential amino acids following the PG loop. A search of 75 587 Protein Data Bank entries using the φ/ψ values for the Val peptide VPGV segment in a 2/0 turn and a search tolerance of 30° found 158 protein segments containing a four amino acid sequence with these φ/ψ angles.[32] Of these sequences, 21% have a proline in the i position, 21% have a proline in the i + 1 position, and 87% have a glycine in the i + 2 position. Clearly, the Gly residue plays a crucial role in accommodating the tight 2/0 turn conformation in this sequence. Interestingly, a similar conformation was previously observed by X-ray diffraction in the cyclic peptide $(APGVGV)_2$, where the cyclic constraint generates a closed turn.[33]

To summarize, we have presented three findings that should prove useful in future studies of structure in proline-containing systems. First, the **X** point mutation provides a robust means of tuning the compactness or stability of turns in engineered proteins and biopolymers. Second, the frequency shift of the lowest-frequency amide-I vibration in the FTIR spectrum can be employed as a measure of the number of hydrogen bonds to proline. This finding is particularly interesting because it provides a simple method for interrogating the protein structure in proline-rich tissue samples such as elastin and collagen matrices that are not readily addressed by other methods. Third, through the use of 2D IR, it is possible to assign the hydrogen-bond composition of these turns. With

dx.doi.org/10.1021/ja2114135 | J. Am. Chem. Soc. 2012, 134, 5032−5035

further refinement of the modeling methods, it will be possible to obtain a quantitative assignment of the populations of the different conformations in the equilibrium structural ensemble of each peptide from the 2D IR spectrum. This analysis will thereby provide a new approach for characterizing conformational states in intrinsically disordered proteins. In addition, it can be used in the future to benchmark the accuracy of force fields commonly used to describe structural and dynamical properties of small peptides.

■ ASSOCIATED CONTENT

Supporting Information

Details of the peptide synthesis, 2D IR spectroscopy, MD simulations, hydrogen-bond criteria, binning of peptide structures, and spectral modeling. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

tokmakof@mit.edu

Present Address

[‖]Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.T. acknowledges the National Science Foundation (CHE-0911107) and the Department of Energy (DE-FG02-99ER14988) for supporting this work. T.L.C.J. acknowledges The Netherlands Organization for Scientific Research (NWO) for support through a VIDI Grant. D.M. acknowledges the Research Department "Interfacial Systems Chemistry" for partial financial support as well as computational support by RV-NRW and BOVILAB@RUB. M.R. acknowledges support from a NSF Graduate Research Fellowship.

■ REFERENCES

(1) Dyson, H. J.; Wright, P. E. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 519.
(2) Blanco, F. J.; Jiménez, M. A.; Herranz, J.; Rico, M.; Santoro, J.; Nieto, J. L. *J. Am. Chem. Soc.* **1993**, *115*, 5887.
(3) Searle, M. S.; Williams, D. H.; Packman, L. C. *Nat. Struct. Biol.* **1995**, *2*, 999.
(4) Syud, F. A.; Stanger, H. E.; Gellman, S. H. *J. Am. Chem. Soc.* **2001**, *123*, 8667.
(5) Espinosa, J. F.; Syud, F. A.; Gellman, S. H. *Protein Sci.* **2002**, *11*, 1492.
(6) Hilario, J.; Kubelka, J.; Keiderling, T. A. *J. Am. Chem. Soc.* **2003**, *125*, 7562.
(7) Blandl, T.; Cochran, A. G.; Skelton, N. J. *Protein Sci.* **2003**, *12*, 237.
(8) Smith, A. W.; Tokmakoff, A. *Angew. Chem., Int. Ed.* **2007**, *46*, 7984.
(9) Smith, A. W.; Tokmakoff, A. *J. Chem. Phys.* **2007**, *126*, 045109.
(10) Smith, A. W.; Lessing, J.; Roy, S.; Ganim, Z.; Peng, C. S.; Jansen, T. L. C.; Knoester, J.; Tokmakoff, A. *J. Phys. Chem. B* **2010**, *114*, 10913.
(11) Urry, D. W. *J. Phys. Chem. B* **1997**, *101*, 11007.
(12) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J. Biomol. NMR* **1995**, *5*, 67.
(13) Schreiner, E.; Nicolini, C.; Ludolph, B.; Ravindra, R.; Otte, N.; Kohlmeyer, A.; Rousseau, R.; Winter, R.; Marx, D. *Phys. Rev. Lett.* **2004**, *92*, 148101.
(14) Baer, M.; Schreiner, E.; Kohlmeyer, A.; Rousseau, R.; Marx, D. *J. Phys. Chem. B* **2006**, *110*, 3576.
(15) Glaves, R.; Baer, M.; Schreiner, E.; Stoll, R.; Marx, D. *ChemPhysChem* **2008**, *9*, 2759.
(16) Rousseau, R.; Schreiner, E.; Kohlmeyer, A.; Marx, D. *Biophys. J.* **2004**, *86*, 1393.
(17) Ahmed, Z.; Scaffidi, J. P.; Asher, S. A. *Biopolymers* **2008**, *91*, 52.
(18) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. *J. Am. Chem. Soc.* **2007**, *129*, 1179.
(19) Kim, Y. S.; Hochstrasser, R. M. *J. Phys. Chem. B* **2009**, *113*, 8231.
(20) Shim, S.-H.; Gupta, R.; Ling, Y. L.; Strasfeld, D. B.; Raleigh, D. P.; Zanni, M. T. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6614.
(21) Kolano, C.; Helbing, J.; Kozinski, M.; Sander, W.; Hamm, P. *Nature* **2006**, *444*, 469.
(22) Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B. *Biochemistry* **2000**, *39*, 14960.
(23) Gray, W. R.; Sandberg, L. B.; Foster, J. A. *Nature* **1973**, *246*, 461.
(24) Coyne, K. J.; Qin, X.-X.; Waite, J. H. *Science* **1997**, *277*, 1830.
(25) Jenkins, J. E.; Creager, M. S.; Butler, E. B.; Lewis, R. V.; Yarger, J. L.; Holland, G. P. *Chem. Commun.* **2010**, *46*, 6714.
(26) Anderson, O. D.; Greene, F. C.; Yip, R. E.; Halford, N. G.; Shewry, P. R.; Malpica-Romero, J. M. *Nucleic Acids Res.* **1989**, *17*, 461.
(27) Jansen, T. L. C.; Knoester, J. *J. Chem. Phys.* **2006**, *124*, 044502.
(28) Jansen, T. L. C.; Dijkstra, A. G.; Watson, T. M.; Hirst, J. D.; Knoester, J. *J. Chem. Phys.* **2006**, *125*, 044312.
(29) Roy, S.; Lessing, J.; Meisl, G.; Ganim, Z.; Tokmakoff, A.; Knoester, J.; Jansen, T. L. C. *J. Chem. Phys.* **2011**, *135*, 234507.
(30) Woutersen, S.; Hamm, P. *J. Phys.: Condens. Matter* **2002**, *14*, R1035.
(31) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
(32) Golovin, A.; Henrick, K. *BMC Bioinf.* **2008**, *9*, 312.
(33) Karle, I. L.; Urry, D. W. *Biopolymers* **2005**, *77*, 198.