Circular Dichroism Constrains NMR-Derived Structures of a Folded Trinitrophenylated Hexapeptide in Solution

Thomas P. Burghardt,*,† Nenad Juranić,† Slobodan Macura,† Andras Muhlrad,‡ and Katalin Ajtai†

Contribution from the Department of Biochemistry, Mayo Foundation, Rochester, MN, and Department of Oral Biology, Hebrew University, Jerusalem, Israel

Received May 3, 1999. Revised Manuscript Received August 19, 1999

Abstract: A trinitrophenylated hexapeptide, with sequence imitating that near the reactive lysyl residue (Lys[84] or RLR) in skeletal muscle myosin, has an induced circular dichroism (CD) signal in the absorption band of the trinitrophenyl group (TNP) characteristic to the TNP in a structured environment. Nuclear Overhauser Effect (NOE) and coupling constant data obtained with ^{1}H NMR confirm that the TNP-hexapeptide (TNP-6p) exists as an ensemble of closely related 3-dimensional structures. A simulated annealing procedure constrained by the NOE distances produced a solution set of 47 structures for the TNP-6p with potential energies less than or approximately equal to the root-mean-squared energy fluctuation expected for this peptide. The CD signals induced in the three lowest-energy electronic transitions of the TNP absorption bands were computed for each structure in the solution set using the matrix method implemented for TNP as the signal donor group. The computed CD signals distinguish two subsets of structures with opposite chirality. One structural isomer subset produces an ensemble averaged CD signal in agreement with experimental results. The other subset or the total set of structures produce ensemble-averaged CD signals that disagree with the experimental results. These findings demonstrate the importance of CD constraints in the refinement of NMR derived structures of small proteins and peptides and that the matrix method is a reliable predictor of CD signals. The TNP-6p can now serve as a practical test case for new theoretical methods for computing CD signals because of its strong and detailed CD spectrum and known solution structure.

Introduction

A small peptide of known structure is a useful construct for studying protein characteristics related to the elementary determiners of conformation or folding pathway. A significant problem for solving small peptide structures with proton nuclear magnetic resonance (^{1}H NMR) is ambiguity in assigning correct chirality to all or a part of the molecule since handedness is not constrained by the atomic distances. In addition, there is a question of the uniqueness of the NMR-derived structures because small peptides may exist as a large ensemble of conformers. The chiral ambiguity in the NMR structures might be removed, and the number of conformers in the ensemble reduced, by comparison of the peptide circular dichroism (CD) to a theoretical signal prediction computed from trial peptide structures. Essential for successful combination of CD with NMR data is a reliable CD signal calculation method applicable to peptides under experimental conditions.

The near-ultraviolet (UV) CD spectrum is sensitive to secondary and tertiary structure of a polypeptide. Characteristics of this signal are calculable from the nature of the electromagnetic interactions using the polypeptide atomic structure.[1,2] Known approximate and tractable methods produce calculated near UV CD spectra from protein crystallographic coordinates that agree with spectra observed from the protein in solution.[3,4] Assumptions implicit in these calculations are that (i) crystal coordinates provide the predominant solution structure and that (ii) interactions contributing significantly to the signal are from residues near the signal source (e.g., all residues within a certain radius of a tryptophan). These assumptions are relaxed when considering a small polypeptide that folds into a well-defined structure in solution and contains a suitable signal donor. Then ^{1}H NMR can characterize the peptide's solution structure, notwithstanding the chiral ambiguity, and the spectroscopic signal calculation can include interactions from all residues.

A peptide suitable for this application has six residues taken from the chicken skeletal myosin sequence surrounding the reactive lysine residue (RLR or Lys[84]).[5] The hexapeptide (6p) is trinitrophenylated at the side chain ϵ-amino group of Lys[3], the lysine residue corresponding to RLR. The trinitrophenyl (TNP) group in the hexapeptide (TNP-6p), acting as the signal donor in our calculation, has an observable induced CD signal in the three lowest-energy transitions of its absorption band. This folding peptide provides the means to test our implementation of the methods for calculating the near-UV CD spectrum of a peptide and to utilize suggestions from the CD to refine the ^{1}H NMR structure determination.

Materials and Methods

Chemicals. Dimethyl sulfoxide (DMSO) and HEPES were from Sigma (St. Louis, MO). The 2,4,6-trinitrobenzene sulfonic acid (TNBS) was from Fluka (Milwaukee, WI), DMSO-d_6 was from Isotec Inc.

* Corresponding author.
† Mayo Foundation.
‡ Hebrew University.

(1) Tinoco, I. *J. Chem. Phys.* **1960**, *33*, 1332−1338.
(2) Bayley, P. M.; Nielsen, E. B.; Schellman, J. A. *J. Phys. Chem.* **1969**, *73*, 228−243.
(3) Strickland, E. H. *Biochemistry* **1972**, *11*, 3465−3474.
(4) Goux, W. J.; Hooker, T. M. *J. Am. Chem. Soc.* **1980**, *102*, 7080−7087.
(5) Ajtai, K.; Peyser, Y. M.; Park, S.; Burghardt, T. P.; Muhlrad, A. *Biochemistry* **1999**, *38*, 6428−6440.

10.1021/ja991445u CCC: $18.00
Published on Web 10/21/1999

(Miamisburg, OH), and *N*-Fmoc L-amino acids were from Calbiochem-Novabiochem Corp. (San Diego, CA). All other chemicals were reagent grade.

Synthesis of the TNP-6p Peptide and Modification with TNBS. The peptide (*N*-Acetyl-Pro-Pro-Lys-Tyr-Asp-Lys-NH_2) was synthesized (by Dr. Daniel J. McCormick, Mayo Protein Core Facility) as a peptide amide by solid-phase method on Rink (4-2′,4′-dimethoxyphenyl-Fmoc-aminomethyl-phenoxy) co-polystyrene resin (0.1 mmol) with *N*-Fmoc L-amino acids. The NH_2-terminal amino acid, *N*-acetyl-L-proline, was coupled at position 1. The 4-methyltrityl (Mtt) blocking group was used for the ϵ-amino side chain protection of Lys at position 3 (Lys^3), and *tert*-butyloxycarbonyl (Boc) was the side chain protecting group for Lys^6. The *tert*-butyl (But) group was used for the side chain protection of the amino acids Asp^5 and Tyr^4. Four equivalents (0.4 mmol) of each Fmoc-L-amino acid were activated with PyBop/HoBt/4-methymorpholine and coupled to the resin-linked peptide chain in 1-methyl-2-pyrrolidinone (NMP) for 2 h following deprotection of each *N*-Fmoc protecting group with 20% piperidine in NMP.

After synthesis of the peptide, the Mtt protection on Lys^3 was removed by treatment of the resin-linked protected peptide with 5 washes (2 min each) of a mixture containing 1.0% trifluoroacetic acid (TFA)/2.5% triisopropylsilane in dichloromethane. The trinitrobenzene sulfonyl group was then coupled to the deprotected lysyl side chain (position 3) using 2.5 equiv of TNBS and 2.5 equiv of diisopropylethylamine in dimethylformamide for 2 h in the dark.

The TNP-6p peptide was deprotected and removed from the resin by treatment with a mixture of 90% TFA/5.0% water/2.5% ethanedithiol/2.5% thioanisole for 90 min in the dark at room temperature. The peptide was then washed (three times) by precipitation in 50 volumes of cold methyl *tert*-butyl ether and purified by reversed-phase HPLC on a Vydac C18 column (2.1 × 25 cm) in 0.1% TFA/water with a 60 min gradient of 10−80% acetonitrile in 0.1% TFA. Electrospray ionization (ESI) mass analysis on a PE SCIEX API 165 biospectrometer (Applied Biosystems, Inc.) verified that the TNP-6p had the correct mass of 1001 (MH^+) amu.

Two other peptides, in addition to TNP-6p, were also synthesized by analogous methods. They were the hexapeptide without TNP (6p) and a TNP-modified pentapeptide identical to TNP-6p except without Lys^6 (TNP-5p).

Absorption and Circular Dichroism. Absorption spectra were measured on a Beckman DU650 (Beckman Instruments, Fullerton, CA) spectrophotometer. Near-UV CD spectra were recorded on a JASCO J720 spectropolarimeter (Tokyo, Japan). Measurements were carried out at peptide concentrations of 1.6 mM in aqueous buffer or DMSO at 20 °C.

The absorption dipole, $\mu_{j,0}$, or extinction coefficient, $\epsilon_j(\lambda)$, give the calculated or observed electric transition dipole strength for the j^{th} electronic transition, D_j, by,[6]

$$D_j = \mu_{0,j}\,\mu_{j,0} = 9.18 \times 10^{-3} \int [n\epsilon_j(\lambda)/\beta^2\lambda]\mathrm{d}\lambda \ (\mathrm{Debye}^2) \qquad (1)$$

where λ is wavelength, $\mu_{j,0} = \langle j|\mu|0\rangle$ for $\langle j|$ and $|0\rangle$ the excited and ground-state molecular orbital wave functions, and $\beta = (n^2 + 2)/3$ for the peptide imbedded in a medium of refractive index n. The transition magnetic dipole, $\mathbf{M}_{j,0}$, or CD extinction coefficient, $\Delta\epsilon_j$, give the calculated or observed rotary strength for the j^{th} electronic transition, R_j, by,[6]

$$R_j = \mathrm{Im}(\mu_{0,j}\mathbf{M}_{j,0}) = 0.248 \int [\Delta\epsilon_j(\lambda)/\beta\lambda]\mathrm{d}\lambda(\mathrm{D-B}) \qquad (2)$$

where D−B is Debye−Bohr Magnetons.

^{1}H NMR. Spectra were recorded at 500 MHz on a Bruker AMX-500 instrument (Billerica, MA). Samples contained 1−2 mM TNP-6p in a DMSO-*d*6/water mixture (99:1 v/v) at 1 °C. The absorption and CD extinction coefficient spectra, observed from peptide samples prepared for the ^{1}H NMR measurements and under the NMR conditions, match those observed under the conditions used for the optical spectroscopic measurements.

Two-dimensional NOESY spectra [7] were collected at 100, 300, and 600 ms mixing times. Phase-sensitive two-dimensional double-quantum filtered correlated spectra (2D DQF−COSY)[8] were obtained with resolution of 2 Hz (1 Hz after zero-filling) and were suitable for extraction of the vicinal spin−spin coupling constants. Short- and long-range totally correlated spectra (TOCSY) [9] were collected at 30 and 100 ms mixing times, respectively. The proton−carbon heteronuclear correlation spectrum was obtained by {^{1}H-^{13}C}-HSQC[10] over one-bond H−C coupling and edited for proton multiplicities (CH, CH_3 vs CH_2) by the double-INEPT editing scheme.[11] All spectra were recorded in phase-sensitive mode by the time-proportional phase incrementation in indirect time domain.[12] All spectra were processed with FELIX 97 (MSI, San Diego, CA) on a Silicon Graphics workstation. Processing involved FFT in the two dimensions after multiplication by cosine squared window functions and zero-filling to double data size. Peaks and peak volumes measured in FELIX were exported to MATLAB (The Math Works, Inc., 1997) for further analysis.

Assignment of the proton resonances was done from the DQF−COSY, TOCSY, and {^{1}H-^{13}C}-HSQC spectra, in combination with NOE data. Chemically equivalent protons of the aromatic rings and geminal protons, except those at β-positions of the residues side chains, were not resolved.

Distance constraints were calculated from the NOESY spectra. Observed spectra were consistent with the spin-diffusion regime of magnetization transfer and required nonlinear analysis of the magnetization buildup. Volumes of cross-peaks were normalized by volumes of the corresponding diagonal peaks. Buildup curves of normalized cross-peak volumes were fit to a quadratic function of the mixing time for determination of initial buildup rates. Eighty-five nontrivial interproton distances that may be affected by conformational changes were determined from cross-relaxation rates and were calculated using geminal methylene protons as the standard. Dihedral angle constraints were determined from vicinal coupling constants measured in the DQF−COSY spectrum.

Structure Determination. The model TNP-6p molecule was built in the program QUANTA (MSI, Quanta 4.0). The peptide structure file of the model was exported to the program XPLOR (MSI) where the annealing procedure with NOE distance constraints was set up. The annealing schedule included three steps: simulated annealing, simulated annealing refinement, and final minimization.

In the first run the annealing procedure started from the extended chemical structure of the model using 85 NOE distance constraints and 10 dihedral angle constraints. One hundred structures were produced and examined for the close contacts not supported by NOE data. Where found, the repelling distance constraint that required the final distance to be >3.5 was added. The procedure was repeated until a full consistency with the NOE spectrum was obtained, generating a total of 110 repelling constraints. In the final run 200 structures were produced, and 47 of them of not more than 10% above the lowest-energy structure were kept. The standard root-mean-squared (RMS) deviation from the average structure in this set was 0.58 Å. We refer to this set of 47 structures as *S*.

The potential energies of the structures were computed in a vacuum using standard methods. The ∼10 kcal/mol average RMS potential energy fluctuation for structures in *S* is equivalent to that expected from the thermodynamics of the peptide, suggesting that the members of *S* are a reasonable representation for the peptide in solution (see Appendix).

^{1}H NMR detects two isomers of the TNP-6p. The major component (∼80%) produced the structures in *S*. The minor component (∼20%) has a cis-peptide bond joining the acetyl group to the N-terminus residue, while the rest of the structure resembles the major component.

(6) Schellman, J. A. *Chem. Rev.* **1975**, *75*, 323−331.
(7) Macura, S.; Ernst, R. R. *Mol. Phys.* **1980**, *41*, 95−117.
(8) Piantini, U.; Sorensen, O. W.; Ernst, R. R. *J. Am. Chem. Soc.* **1982**, *104*, 6800−6801.
(9) Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521−528.
(10) Bodenhausen, G.; Rubin, D. J. *Chem. Phys. Lett.* **1980**, *69*, 185
(11) Juranić, N.; Zolnai, Z.; Macura, S. *J. Biomol. NMR* **1999**, *9*, 317
(12) Marion, D.; Wutrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *113*, 967

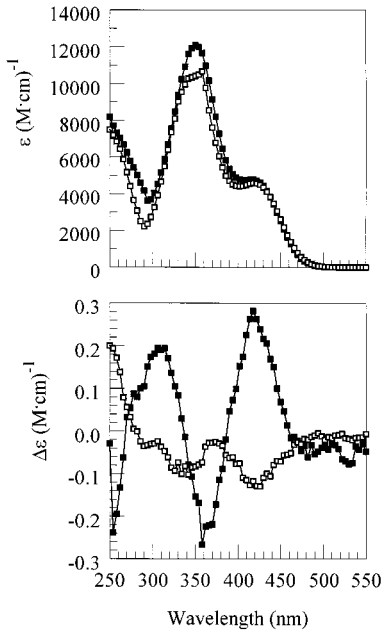

Figure 1. Absorption (top) and CD (bottom) extinction coefficient spectra of TNP-6p in DMSO (■) and HEPES buffer pH 7 (□).

There was no exchange between the two components on the NMR time scale (≲100 s) and no change in the relative amount of the two components over several months.

Calculation of the Optical Signal. The peak absorption energy and dipole and rotary strengths for electronic transitions contributing to the near-UV absorption and induced CD bands of TNP-6p were calculated as described elsewhere[13] by the matrix method.[2] The calculation combines the (n,π*) and (π,π*) transitions of the peptide amide group, with the observable (π,π*) transitions from the side chain of Tyr[4] and TNP modifying Lys[3] in TNP-6p. The Coulomb potential with vacuum dielectric constant couples the transition monopoles of these excited states from interacting groups. The positions and values of the transition monopole charges for the peptide primary, secondary, and tertiary amide groups[14] and for tyrosine[15] were taken from the literature. The positions and values of the transition monopole charges of the TNP were calculated as described elsewhere.[5]

Experiment and calculation are compared for energy, dipole strength, and rotary strength of TNP at its three lowest-energy transitions having unperturbed peak absorption at 418.5, 354.5, and 325.0 nm.

Results

Optical Spectroscopy. Figure 1 shows the absorption and CD extinction coefficient spectra from TNP-6p in DMSO (■) and in HEPES buffer pH 7 (□). Comparison of the absorption spectra from these samples show TNP in TNP-6p to be slightly hyperchromic in DMSO. Comparison of the CD spectra from these samples suggests that local interactions with the folded peptide in DMSO immobilize the TNP. TNP-6p in other organic solvents, including dimethylformamide, acetone, and methanol, had CD spectra similar to that from TNP-6p in DMSO (data not shown). Except for the sign, the CD spectrum from TNP-6p in DMSO resembles that from TNP modifying Lys[84] in skeletal muscle myosin. In the labeled myosin, the protein matrix provides the local interactions that immobilize the probe and induce a large CD signal in the absorption band of the TNP.[5] Table 1 summarizes the dipole and rotary strengths computed from the spectra in Figure 1 using eqs 1 and 2.

(13) Burghardt, T. P.; Garamszegi, S. P.; Park, S.; Ajtai, K. *Biochemistry* **1998**, *37*, 8035−8047.
(14) Nielsen, E. B.; Schellman, J. A. *Biopolymers* **1971**, *10*, 1559−1581.
(15) Goux, W. J.; Cooke, D. B.; Rodriguez, R. E.; Hooker, T. M. *Biopolymers* **1974**, *13*, 2315−2329.

Table 1. Dipole and Rotary Strengths of TNP from TNP-6p in DMSO and HEPES Buffer[a]

	peak absorbance (nm)		dipole strength (Debye^2)		rotary strength (D−M)[b]	
transition	DMSO	HEPES	DMSO	HEPES	DMSO	HEPES
1	418.5	418.6	5.5	5.8	+0.008	−0.003
2	354.5	359.8	10.1	5.9	−0.013	<0.001
3	325.0	331.3	8.9	9.8	+0.014	−0.003

[a] Errors in dipole and rotary strengths are ∼5%. [b] D−M is Debye−Bohr Magneton.

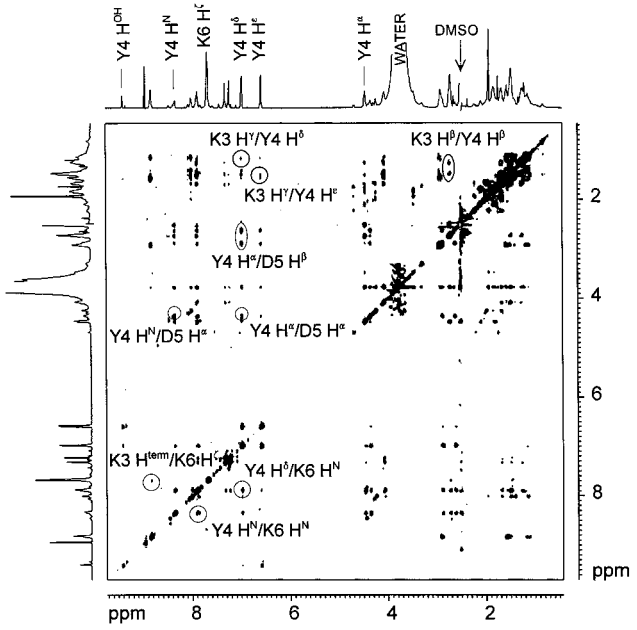

Figure 2. The 500 MHz NOESY spectrum of TNP-6p at 1 °C (mixing time 0.6 s, 1.5 mM in DMSO-*d*6/water mixture: 99:1 v/v). Residue 3 is denoted as K3 although its side chain is modified with TNP. The circled cross-peaks are crucial for the structure determination.

Absorption and CD extinction coefficient spectra from TNP-5p in DMSO and in HEPES pH 7 were also obtained (data not shown). The CD spectra from TNP-5p in either solvent resembles that from the TNP-6p in HEPES, suggesting that no structured environment surrounds TNP in the labeled pentapeptide. These data indicate that Lys[6] in the hexapeptide is necessary for folding.

^{1}H NMR. Figure 2 shows the NOESY spectrum of TNP-6p in DMSO recorded with 0.6 s mixing time. Indicated are the spread of chemical shifts and numerous NOESY cross-peaks from TNP-6p, suggesting that the molecule forms a well-defined structure in solution. Numerous inter-residue cross-peak pairs (e.g., Lys^{3}H^{term}/Lys^{6}H$^\zeta$ at 8.93 ppm/7.68 ppm, Tyr^{4}H^N/Lys^{6}H^N at 7.87 ppm/8.33 ppm) indicate spatial proximity of the respective groups. Several cross-peak pairs between TNP modified Lys[3] and Tyr[4] correspond to the interaction between their side chains. The ^{1}H NMR spectra were also obtained from the hexapeptide without TNP (6p) in DMSO (data not shown). These spectra also indicate a folded structure for 6p resembling that of TNP-6p. Evidently, interactions between TNP and the rest of the hexapeptide are not essential for stabilization of the folded conformation. Structures in *S* characterizing TNP-6p in DMSO were obtained from the NMR data by the protocol outlined in Materials and Methods.

Calculation of CD Signals. The ensemble of TNP-6p structures in *S* has subclasses containing 27 and 20 structures, s_+ and s_-, that differ by their peptide conformation in the

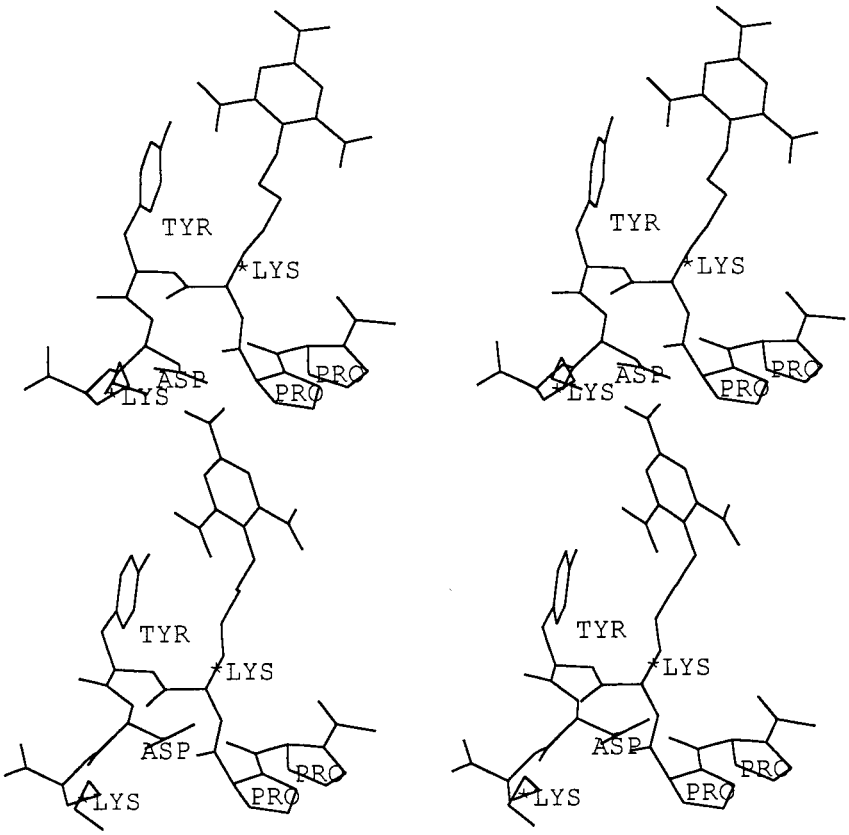

Figure 3. Uncrossed stereoview of the two types of TNP-6p structures present in solution. The upper panel corresponds to the L-shaped (s_+), the lower to the stacked (s_-), structure of the TNP-6p in DMSO. The asterisk indicates modification at Lys^3 (side chain modified by TNP) and Lys^6 (C-term modified by NH_2).

Table 2. Computed and Observed Rotary Strengths of TNP from TNP-6p in DMSO and the Total Potential Energy of the Gas-Phase Peptide[a]

	computed averages				
	S (47)[b]	s_- (20)	s_+ (27)	weighted (47)[c]	observed
rotary strength					
$\langle R_1 \rangle$	-0.017 ± 0.055[d]	-0.075 ± 0.027	$+0.025 \pm 0.017$	$+0.013 \pm 0.048$	$+0.008$
$\langle R_2 \rangle$	-0.001 ± 0.045	$+0.033 \pm 0.024$	-0.025 ± 0.015	-0.012 ± 0.037	-0.013
$\langle R2 \rangle$	$+0.002 \pm 0.039$	-0.035 ± 0.023	$+0.030 \pm 0.019$	$+0.024 \pm 0.037$	$+0.014$
potential energy					
$\langle E \rangle$	215.5 ± 10	217.9 ± 10	213.8 ± 10	204.3 ± 2	

[a] Rotary strength in Debye−Bohr Magnetons and potential energy in kilocalories per mole. [b] Parentheses contain the number of structures in the ensemble. [c] Modified Boltzmann averaging, see Appendix. [d] Standard deviation.

vicinity of TNP. These subclasses, represented in Figure 3, differ optically by producing about equal intensity but oppositely signed CD signals in the absorption bands of the TNP. If the CD spectrum in Figure 1 from TNP-6p in DMSO is represented as an ordered triplet (+,−,+) corresponding to the sign of the optical activity from the three lowest-energy transitions, then s_+ and s_- have optical activities characterized by (+,−,+) and (−,+,−), respectively. The magnitude and sign of the optical activity from s_+ and s_- are principally the effect of the Kirkwood−Kuhn−Moffitt mechanism coupling the (π,π^*) transition electric dipoles from TNP and the side chain of Tyr^4.[2] Qualitatively, the two types of structures differ by the TNP and Tyr^4 side chain forming either an L-shaped or a stacked complex.

The ensemble-averaged rotary strengths for the three lowest-energy electronic transitions of TNP in TNP-6p, $\langle R_1 \rangle$, $\langle R_2 \rangle$, and $\langle R_3 \rangle$, were computed by averaging over all of the structures in S or those in subclasses s_+ or s_-. Table 2 lists these ensemble averages and compares them to the pertinent experimental data taken from Table 1. The standard deviation associated with $\langle R_1 \rangle$, $\langle R_2 \rangle$, and $\langle R_3 \rangle$ shows that structures within each subclass are homogeneous with respect to their optical activity. The average rotary strengths for the structures in S and s_- disagree, while those from s_+ agree, with the experimentally observed results.

Alternatively, $\langle R_1 \rangle$, $\langle R_2 \rangle$, and $\langle R_3 \rangle$ were computed using eq A8 and the structures in S as described in the Appendix. Structures were weighted by a modified Boltzmann factor using the peptides' computed potential energy. Table 2 lists these ensemble averages that were obtained with 18 structures (14 from s_+) contributing significantly to the average by having a normalized weight of more than 0.02 (the weighting when every structure is equally probable). With this method, both the s_+ and s_- enantiomers contribute to the ensemble-averaged rotary strengths; however, structures in s_+ dominate due to their lower potential energy (see Table 2). The modified Boltzmann weighting effectively selects the s_+ structures as the solution set.

TNP-6p in solution contains major (80%) and minor (20%) isomer components detected independently by NMR. The minor

component structure resembles that of the major component except for a cis-peptide bond joining the acetyl group to the N-terminus residue. Set *S* contains only the major isomer form. The CD spectrum does not apparently distinguish between these isomer forms so that both of them contribute to the spectrum. We investigated the impact of the presence of the two isomer forms of TNP-6p on the CD spectrum by assuming that the minor isomer structures are identical to those in *S* except for the above-mentioned cis-peptide bond. Then, we constructed a new ensemble of structures that combine the assumed conformations for the minor form with the conformations in *S*. Structures were weighted by concentration during ensemble-averaging. Under these circumstances, $\langle R_1 \rangle$, $\langle R_2 \rangle$, and $\langle R_3 \rangle$ did not differ significantly from values reported in Table 2, suggesting that we could ignore the minor isomer component of the TNP-6p.

Discussion

The hexapeptide *N*-Acetyl-Pro-Pro-Lys-Tyr-Asp-Lys-NH_2, with and without TNP modifying Lys^3, folds in solution with well-defined secondary and tertiary structure. This hexapeptide, with a sequence imitating that found in chicken skeletal muscle myosin surrounding the reactive lysyl residue, requires the presence of the Lys^6 residue to fold. The TNP group interacts with the folded peptide such that a strong CD signal is induced in the three lowest-energy electronic transitions of its absorption band (Table 1). This CD signal senses the 3-dimensional structure of the folded peptide in a manner that can be modeled with the matrix method for calculating rotary strengths.[2]

The 3-dimensional structure of the TNP-6p was solved using ^{1}H NMR. The solution set of structures, *S*, satisfying the NMR constraints contains two homogeneous subsets, s_+ and s_-, that are readily distinguished by their oppositely signed CD signatures, (+,−,+) and (−,+,−). When ensemble-averaged rotary strengths were computed using structures in *S*, or s_-, they did not agree with experimental observations. However, ensemble-averaged rotary strengths computed using structures in s_+ gave good qualitative agreement with experiment (Table 2). The coordination of the aromatic groups of TNP and Tyr^4 in subsets s_+ and s_- is L-shaped (s_+) or stacked (s_-) (Figure 3). These findings show that the matrix method for computing CD signals from peptides works under conditions most closely simulating that of a peptide in solution (i.e., using ensemble-averaging of NMR solution structures rather than a single average solid-state structure from crystallography) while accounting for all of the interactions among the residues in the system.

Our findings demonstrate the use of CD in the final refinement of structures for the solution set satisfying the NMR distance constraints. The CD data supplies one constraint for each observable electronic transition for which the corresponding average rotary strength can be calculated. As the CD distinguishes the enantiomers of an optically active molecule while distance constraints do not, just the sign of the CD provides an independent constraint on structure. The quantitative value of the observed rotary strength has less impact, as we are able to reproduce signals observed here only to within a factor of 1−3. The least tested element of our computational method is the estimate of the monopole charges for the lowest-energy electronic transitions in TNP.[5] Refinement of these charges will likely narrow the quantitative differences between the observed and calculated CD signals.

Conclusions

We have calculated the near-UV optical activity for a folded trinitrophenylated peptide from an ensemble of peptide structures consistent with NMR constraints. The optical activities show that the set of NMR determined structures is made up of two homogeneous subsets differing in the coordination, either L-shaped or stacked, of the aromatic groups in the peptide. Only the L-shaped complex gives good qualitative agreement between calculated and observed optical activities. This demonstrates the usefulness of optical activity for structure refinement of chiral molecules such as small peptides, where a limited number of available experimental NMR constraints often preclude unambiguous structure determination. Additionally, the agreement between calculated and observed optical activities for the peptide builds confidence in our ability to correctly compute these optical parameters with existing approaches and suggests that the hexapeptide might be useful in other optical signal calibration applications with tyrosine as the signal donor or when involving more sophisticated computational approaches that can only be applied to the smallest peptides.

Acknowledgment. We thank Dr. Daniel J. McCormick from the Mayo Protein Core Facility for synthesizing the peptides used in this project. This research was supported by the National Institutes of Health Grant R01 AR39288 and the Mayo Foundation.

Appendix

Energy Fluctuations in TNP-6p and an Alternative Method for Calculating Ensemble Averages. Consider a gas of TNP-6p molecules at thermal equilibrium. The molecules consist of *L* atoms with 3*L* generalized coordinates (q_1, ..., q_{3L}) and momenta (p_1, ..., p_{3L}). In the classical regime the energy of the molecule, *E*, is,

$$E = \frac{\mathbf{p}_c \cdot \mathbf{p}_c}{2M} + \sum_{k=1}^{3L} \frac{p_k^2}{2m_k} + V(q_1, ..., q_{3L}) \qquad \text{(A1)}$$

where $\mathbf{p}_c$ is the molecular center of mass momentum, *M* the molecular mass, m_k the mass of the k^{th} atom, and *V* the potential energy. The partition function for this system, *Z*, is,

$$Z = \int \mathrm{d}\mathbf{p}_c \int \mathrm{d}p_1 ... \mathrm{d}p_{3L} \int \mathrm{d}q_1 ... \mathrm{d}q_{3L}\, \mathrm{e}^{-E/kT} \qquad \text{(A2)}$$

where *kT* is Boltzmann's constant times temperature. Replacing the integral over the generalized coordinates in eq A2 with a sum over the *N* (=47) structures in *S* gives,

$$Z = \int \mathrm{d}\mathbf{p}_c \int \mathrm{d}p_1 ... \mathrm{d}p_{3L} \sum_{i=1}^{N} g_i \exp(-E_i/kT) \qquad \text{(A3)}$$

where g_i is the unknown degeneracy factor accounting for structures of energy E_i not enumerated by the structure annealing procedure described in Materials and Methods and,

$$E_i = \frac{\mathbf{p}_c \cdot \mathbf{p}_c}{2M} + \sum_{k=1}^{3L} \frac{p_k^2}{2m_k} + V_i \qquad \text{(A4)}$$

From the partition function we may compute the average energy, $\langle E \rangle$, and the specific heat at constant volume, C_V, using,

$$\langle E \rangle = kT^2\, \partial \log Z / \partial T \qquad \text{(A5)}$$

and

$$C_V = \partial\langle E\rangle/\partial T = (kT^2)^{-1}\,(\langle E^2\rangle - \langle E\rangle^{\,2}) = (^3/_2)k(L+1) + (kT^2)^{-1}(\langle V^2\rangle - \langle V\rangle^{\,2}) \quad \text{(A6)}$$

where

$$\langle V^l\rangle = \frac{\sum_{i=1}^{N} V_i^l g_i \exp(-f_E V_i/kT)}{\sum_{i=1}^{N} g_i \exp(-f_E V_i/kT)} = \sum_{i=1}^{N} \xi_i V_i^l \quad \text{(A7)}$$

for ξ_i the equilibrium probability for the i^{th} structure. The energy scaling constant, f_E (≈ 0.2), in eq A7 modifies the Boltzmann factor to help remedy various inaccuracies in the computed potential energies.[16] The potential energy, V_i, for the i^{th} structure in S is computed in a vacuum with the molecular mechanics force field MM+ implemented in the commercial computational chemistry program HyperChem (Hypercube Inc., Gainesville, FL).

The CD signal, like potential energy, depends on conformation (not momentum) such that $\langle R_j\rangle$, the ensemble averaged rotary strength for the j^{th} electronic transition, is,

$$\langle R_j\rangle = \sum_{i=1}^{N} \xi_i R_{j,i} \quad \text{(A8)}$$

where $R_{j,i}$ is the calculated rotary strength for the j^{th} electronic transition of the i^{th} structure.

We computed C_V from eqs A6 and A7 assuming g_i in eq A7 is the same for each structure, so they do not contribute to averages. We found that the term $(kT^2)^{-1}$ $(\langle V^2\rangle - \langle V\rangle^2)$ contributes negligibly to C_V for $0 \le f_E \le 1$, giving $C_V \approx (^3/_2)k(L+1) \approx 0.4$ kcal mol^{-1} K^{-1} for TNP-6p. The gas-phase C_V for TNP-6p, estimated by summing up contributions from the constituent groups,[17,18] gives $C_V \approx 0.3$ kcal mol^{-1} K^{-1} (Data taken from Table 5, Makhatadze & Privalov,[17] assuming $C_V = C_p - \gamma T$, where C_p is the specific heat at constant pressure and γ is a constant independent of temperature. The constant γ was computed from the linear temperature dependence of the peptide bond C_p in the 5−125 °C range, also from Table 5). These two estimates for C_V give the RMS energy fluctuation $\langle \Delta E^2\rangle^{1/2} \approx$ 7−8 kcal/mol for TNP-6p in a vacuum at room temperature.

JA991445U

(16) Fennen, J.; Torda, A. E.; van Gunsteren, W. F. *J. Biomol. NMR* **1995**, *6*, 163−170.

(17) Makhatadze, G. I.; Privalov, P. L. *J. Mol. Biol.* **1990**, *213*, 375−384.

(18) Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.* **1990**, *213*, 385−391.