J|A|C|S COMMUNICATIONS

Published on Web 03/12/2009

Hairpin Structure of a Biarsenical−Tetracysteine Motif Determined by NMR Spectroscopy

Fatemeh Madani,[†] Jesper Lind,[†] Peter Damberg,[†] Stephen R. Adams,[‡] Roger Y. Tsien,*[,‡] and Astrid O. Gräslund*[,†]

Department of Biochemistry and Biophysics, Arrhenius Laboratories, Stockholm University, SE-106 91 Stockholm, Sweden, and Howard Hughes Medical Institute, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0647

Received November 28, 2008; E-mail: rtsien@ucsd.edu; astrid@dbb.su.se

Genetically targeted labeling of proteins with small molecules in intact living cells has become popular in the past decade because such tagging combines the genetic precision of targeting with the versatile properties of small molecules. This strategy requires that the protein of interest be fused genetically to a peptide or protein domain that can bind a small ligand tightly and selectively enough so the ligand labels the chimeric protein nearly stoichiometrically and exclusively in a living cell. Many such peptide domain−small molecule ligand pairs have been introduced, but the first of these, the biarsenical−tetracysteine system, has been the most widely used.[1−4] Here the peptide domain is of the form $X_3C_2X_2C_2X_3$, where C = cysteine and X = a noncysteine amino acid, and the small molecule is a tricyclic fluorophore with As(III) substituents at the 4 and 5 positions. Each As(III) reversibly forms two covalent bonds to cysteine thiols. We have used NMR to determine the solution structure of the 12-mer peptide FLNCCPGCCMEP in complex with the ReAsH fluorophore. This peptide system was previously selected after several generations of optimization, culminating in high-throughput genetic screens and fluorescence-activated cell sorting.[3] The present three-dimensional structure results show a relatively tight peptide hairpin formed around the fluorophore, which interacts closely with the N-terminal phenylalanine side chain.

The advantages of the investigated biarsenical−tetracysteine label system include the small size of the peptide domain (only 12 residues), the modest size of the biarsenical (≤750 Da), the membrane permeability of the latter, and the very tight binding (subpicomolar dissociation constants). Low micromolar concentrations of 1,2-ethanedithiol (EDT) or 2,3-dimercaptopropanol prevent toxicity by diverting the biarsenical from binding to most endogenous molecules containing pairs of thiol groups. Millimolar concentrations of the same *vic*-dithiols can competitively remove the biarsenical from the tetracysteine domains, so the labeling can be reversed if desired. The fluorescence of the biarsenical moiety is strongly quenched until it binds to the peptide domain. However, the biggest limitation remains background labeling of endogenous proteins. The main strategy for reducing such background has been to evolve tetracysteine domains with higher affinities, allowing good labeling to be maintained with lower concentrations of biarsenical, higher concentrations of *vic*-dithiol competitor, or both.[2,3]

Despite many biological applications of this and previous tetracysteine domains, no structure of a biarsenical complex has been reported. Attempts at crystallization either alone or fused to fluorescent proteins have not yet been successful. Here we turned to NMR spectroscopy, which would reveal the three-dimensional structure in aqueous solution. We had a choice between the two most widely used (and commercially available) biarsenicals, "FlAsH" and "ReAsH", which are based on fluorescein and resorufin, respectively. Complexes of FlAsH with tetracysteine peptides usually can be resolved by HPLC into two closely spaced peaks of identical amplitude and mass spectra, which probably correspond to atropisomers with opposite orientations of the carboxyphenyl group relative to the xanthene rings and the peptide. ReAsH lacks the bulky carboxyphenyl substituent and gives single peaks in HPLC; thus, it was considered preferable. The complex of ReAsH with FLNCCPGCCMEP (C-terminus amidated) was not soluble enough for natural-abundance multidimensional NMR spectroscopy, so we acetylated the N-terminus with hydrophilic substituents such as sulfo, succinyl ($^-$OOCCH$_2$CH$_2$CO−), and MeO(CH$_2$CH$_2$O)$_n$CH$_2$CH$_2$CO−. Succinyl was chosen because it gave the greatest solubility increase (to 1 mM).

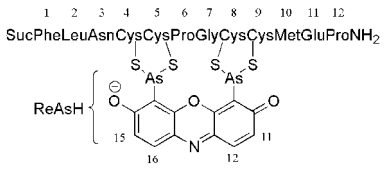

Figure 1. Schematic picture of ReAsH bound to Cys4, Cys5, Cys8, and Cys9 of the peptide. ReAsH protons are numbered 11, 12, 15, and 16. Suc refers to the succinyl group at the N-terminus of the peptide.

Figure 1 shows the schematic biarsenical−tetracysteine system investigated in this study, with the numbering of the peptide residues and the protons of the resorufin moiety.

Circular dichroism spectra of the biarsenical complex showed a minimum at 195 nm (Figure S1 in the Supporting Information), suggesting that a random coil dominates the secondary structure. A certain contribution from β-structure was suggested by a negative shoulder at ∼212 nm.

The peptide−ReAsH complex in aqueous solution was investigated by NMR. All of the peptide backbone proton resonances and essentially all of the side-chain proton resonances were assigned using standard homonuclear NMR procedures[5] (Table S1). ReAsH by itself is symmetrical, and there are two nonequivalent protons in each equivalent ring. A natural-abundance ^{13}C−^{1}H HSQC spectrum gave assignments of proton-bonded carbons in the peptide as well as in resorufin (Table S1) and verified some of the proton assignments. ^{1}H chemical shifts were referenced to the H$_2$O signal at 4.96 ppm,[6] and ^{13}C chemical shifts were referenced indirectly.[7]

There are rather large deviations from random-coil values of chemical shifts[8] for some of the proton resonances (Figure S2). The presence of the thiol−arsenic bonds causes downfield shift of ∼10 ppm in the C$^\beta$ resonances of the cysteines. The tight structure and potential ring-current shifts are confounded with the presence

[†] Stockholm University.
[‡] University of California, San Diego.

10.1021/ja809315x CCC: $40.75

Figure 2. Inter-residue connectivities used for the structure calculation for the peptide bound to ReAsH. The NOE constraints were obtained from a 500 MHz NOESY spectrum recorded with a mixing time of 300 ms at 278 K. Thick and thin bars represent strong and weak intensities of sequential NOEs, respectively, while lines between nonsequential residues indicate observed medium-range NOEs.[5] Data are from CYANA (version 2.0).

Table 1. Structure Calculation Data for the Peptide Bound to ReAsH (30 Structures)

total number of upper distance constraints	233
intraresidual	152
sequential (i, $i + 1$)	64
medium-range [(i, $i + 2$) or (i, $i + 3$)]	11
ReAsH and Phe1	2
ReAsH and S_{Cys}	4
CYANA target function	0.19 ± 0.13
atom rmsd's (Å):	
backbone	1.00 ± 0.48
heavy atoms	1.28 ± 0.42
residues 4−8	0.24 ± 0.16

of the As atoms, making deduction of structural constraints from the secondary chemical shifts uncertain.

An NMR translational diffusion experiment[9] was performed in order to determine the size of the complex and its oligomeric state in solution. The result showed a translational diffusion coefficient of 12.5×10^{-11} m^2 s^{-1}. From this coefficient, a hydrodynamic radius and a molecular weight of 1616 Da was determined[10] for the diffusing species. This molecular weight is close to that of a monomer (1772.7 Da), showing that ReAsH bound to the peptide exists as a monomer in the solution at pH 7.4 and 278 K.

Cross-peaks in a ^{1}H−^{1}H NOESY spectrum recorded with a mixing time of 300 ms were used to calculate distance constraints for the structure calculation by the program CYANA.[11,12] The intensities (volumes) of the crosspeaks were converted into upper-distance constraints. No long-range NOEs were observed, but almost all of the sequential inter-residue NOEs and some nonsequential medium-range NOEs were found (Figure 2). We observed NOEs between the β protons of Cys4 and the α, β, and aromatic protons of Phe1 (Table 1). Two NOEs between ReAsH and the peptide were found: between the δ and ε (ζ) protons of the side chain of Phe1 and the 16 (12) proton of ReAsH (Figure S3).

The β-protons of Pro6 were stereospecifically assigned. No stereospecific assignment was meaningful for the other β-protons, which were treated in the pseudoatom representation. The four distances between the As and S atoms (Table 1) were given the same upper and lower limits of 2.25 Å as reported from a crystal structure.[13] In total, 233 upper-distance and six lower-distance

Figure 3. Structure calculation of the peptide bound to ReAsH, based on the NOESY-derived distance constraints. Arsenic atoms are fixed to the cysteine atoms (colored red) at a distance of 2.25 Å. The peptide backbone and ReAsH are represented by an ensemble of the 30 best structures, which appear as semitransparent, while the average peptide backbone structure from the ensemble, including the Phe1 side chain and the ReAsH moiety, is shown as a solid line. The overlay was made on the backbone atoms from residue 4 to 8 and also included ReAsH.

constraints were used to calculate the structures (Table 1). The upper limits were adjusted iteratively during the structure calculation. Different combinations of binding between the As atoms and the four S atoms in the cysteines were studied by calculating the corresponding structures and comparing the energies and potential constraint violations. Only the isomer where As2 is connected to C4 and C5 and As1 to C8 and C9 results in low-energy structures. Other isomers are incompatible with the experimental distance constraints and result in van der Waals conflicts and high energies. Hence, it is concluded that sequential cysteine residues bind to the same As atom, as previously suggested[2] on the basis of data from mass spectrometry and fluorescence experiments on chemically modified compounds. Altogether, 100 structures were calculated in CYANA. The 30 best structures with respect to violation energies and CYANA target function were selected for further investigation of the structure. Figure 3 shows an ensemble of these 30 best structures of the peptide backbone bound to the ReAsH. The structure calculation data is summarized in Table 1.

The peptide backbone makes a rather well-defined three-dimensional structure in which P6 and G7 contribute to a relatively tight turn. The conformation of the central CPGC fragment resembles a type-II β-turn, where the $i + 1$ and $i + 2$ residues should adopt the dihedral angles (ϕ, ψ) = (−60°, 131°) and (84°, 1°),[14] respectively. For the PG fragment, we found (ϕ, ψ) ≈ (−70°, 163°) for the P residue and (52°, −32°) for the G residue and that the cysteines in the i (C5) and $i + 3$ (C8) positions adopt a β conformation. The resemblance to a type-II turn explains why the peptide with the spacer residues PG displayed much better binding to ReAsH than peptides with other spacer sequences.[2]

The preference of the peptide sequence for a proline in position $i + 1$ in type-II turns is very strong,[15] with proline preferred over all other residue types. The preference for glycine in position $i + 2$ is even stronger.[15] Interestingly, cysteine has the highest preference of all residues in position $i + 3$ and the second highest preference for position i.[15] Hence, the CPGC sequence should fit

Table 2. Absorbance, Fluorescence, and Stability toward EDT of ReAsH Complexes with Tetracysteine Peptides[a]

peptide	λ_{max} (nm)	E_{max} (nm)	ε_{max} (M^{-1} cm^{-1})	ϕ	τ (s)
FLNCCPGCCMEP	597	608	69 000	0.48	476
YLNCCPGCCMEP	594	607	68 000	0.094	256,[b] 1667
WLNCCPGCCMEP	601	607	56 000	0.034	133,[c] 909
HLNCCPGCCMEP	596	607	67 000	0.20	138,[d] 5000
FLNCCPGCCKTF	597	608	65 000	0.47	115

[a] Abbreviations: E_{max}, fluorescence emission maximum; ϕ, fluorescence quantum yield; τ, time constant for displacement of ReAsH from the complex by 0.5 mM EDT in the presence of 5 mM 2-mercaptoethanesulfonate. Spectra and stability toward EDT were measured in 100 mM K·MOPS (pH 7.2). [b] Corresponds to 60% of the total decrease of resorufin fluorescence. [c] Corresponds to 55% of the total decrease. [d] Corresponds to 77% of the total decrease.

almost perfectly into the observed structure when bound to ReAsH, which explains why the apparent dissociation constant K_d is very favorable.[2] However, the ReAsH−peptide complex is not a classical β-hairpin. There is no evidence for hydrogen bonds between Cys5 and Cys8, nor do Cys4 and Cys9 adopt a β conformation or display hydrogen bonds. The cysteine in position 9 adopts an α_L conformation to make the interaction between its sulfur atom and the arsenic atom possible. The cysteine in position 4 has the dihedral angles (ϕ, ψ) $\approx$ (−78°, −72°). The constraints imposed by the covalent S−As bonds to the rigid fluorophore may be the cause of this unusual conformation. If Cys4 and Cys9 would adopt a β conformation, as in a classical β-hairpin, their side chains would point away from the As atoms.

The four S_{Cys}−As−C angles are relatively well-defined: from Cys4, 122 ± 2° standard deviation in the 30 best structures; Cys5, 126 ± 15°; Cys8, 112 ± 15°; Cys9, 113 ± 9°. The position of ReAsH relative to the peptide was investigated by calculating the angle between the plane of the Phe1 side chain and the resorufin plane in the calculated structures. The Phe1 side chain is located almost perpendicular to the resorufin plane (at an angle of 92 ± 21°). This edge−face geometry is not unusual for interacting aromatic systems[16] and allows the partially positive hydrogens of the phenyl ring to contact the highest electron density at the center of the anionic dye. The close positioning of the phenyl ring of Phe1 and ReAsH explains why Phe is strongly preferred at position 1, as already implied by alanine scanning.[3] However, the structure provides no explanation for the slight preference that alanine scanning had revealed[3] for Asn at position 3.

The average distance between the N-terminus (amide N atom) and the C-terminus (carboxy C atom) of the peptide in the 30 best structures was determined to be 12.5 ± 2.2 Å. This distance is relatively small compared to, for example, the distance of 17.8 Å in an α-helix with the same number of residues, confirming the hairpin nature of the construct. The hairpin geometry suggests that this biarsenical−tetracysteine label system may also be internally inserted within a protein structure without too much structural disturbance. Previous insertions were limited to minimal CCXXCC sequences.[17] The present dodecapeptide sequence should enable insertions with increased affinity and fluorescence brightness.

To test whether Phe1 is uniquely favorable for the fluorescence properties and complex stability, complexes were prepared with Phe1 replaced by His, Tyr, or Trp. The results (Table 2) showed that the fluorescence of all three variants was quenched compared to the Phe1 complex, probably by electron transfer from the electron-rich side chains to the fluorophore. The complex stability as assessed by time constants for ReAsH displacement by EDT was decreased by factors of ∼2 or more (Table 2). The NMR structure indicated one surprisingly accessible face of the resorufin that could be available for additional interaction with the C-terminal flanking residues. Interchanging this sequence with the corresponding three residues from another optimized peptide, HRWCCPGCCKTF,[3] which also contains a potentially interacting F, to give FLNCCPGCCKTF did not change the fluorescence properties but decreased the stability toward EDT 4-fold.

The NMR results and solution structure calculations of the optimized ReAsH peptide complex suggest that the N-terminus of the 12-mer peptide is relatively tightly constrained. The residues on the N-terminal side of the binding site interact strongly with the resorufin. A hairpin structure with some resemblance to a type-II β-turn is formed. The C-terminus is not as well constrained by the NMR results. This could mean that the C-terminus is more dynamic and does not strongly interact with the fluorophore. The N- and C-terminal tripeptides may interact with each other while the fluorophore is bound. Further optimization of the tetracysteine motif will require screening of the full 12-mer or even longer sequences.

Acknowledgment. This study was supported by the Swedish Research Council (to A.O.G.), the Knut and Alice Wallenberg Foundation (to A.O.G.), NIH GM072033 (to R.Y.T.), and HHMI (to R.Y.T.).

Supporting Information Available: Experimental methods, Figures S1−S3, and Table S1. This material is available free of charge via the Internet at http://pubs.acs.org. The structure coordinates of the FLNCCPGCCMEP−ReAsH motif and the NMR chemical shift list have been deposited with BioMagResBank (ID code 16041).

References

(1) Griffin, B. A.; Adams, S. R.; Tsien, R. Y. *Science* **1998**, *281*, 269–272.
(2) Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y. *J. Am. Chem. Soc.* **2002**, *124*, 6063–6076.
(3) Martin, B. R.; Giepmans, B. N.; Adams, S. R.; Tsien, R. Y. *Nat. Biotechnol.* **2005**, *23*, 1308–1314.
(4) Luedtke, N. W.; Dexter, R. J.; Fried, D. B.; Schepartz, A. *Nat. Chem. Biol.* **2007**, *3*, 779–784.
(5) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley & Sons: New York, 1986.
(6) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G.; Rance, M.; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*, 2nd ed.; Elsevier Academic Press: Burlington, MA, 2007; pp 236−237.
(7) Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wüthrich, K. *J. Mol. Biol.* **1998**, *280*, 933–952.
(8) Wishart, D. S.; Sykes, B. D. *Methods Enzymol.* **1994**, *239*, 363–392.
(9) Damberg, P.; Jarvet, J.; Gräslund, A. *J. Magn. Reson.* **2001**, *148*, 343–348.
(10) Danielsson, J.; Jarvet, J.; Damberg, P.; Gräslund, A. *Magn. Reson. Chem.* **2002**, *40*, 89–97.
(11) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283–298.
(12) Güntert, P. *Methods Mol. Biol.* **2004**, *278*, 353–378.
(13) Cruse, W. B. T.; James, M. N. G. *Acta Crystallogr., Sect. B* **1972**, *28*, 1325–1331.
(14) Hutchinson, E.; Thornton, J. *Protein Sci.* **1994**, *3*, 2207–2221.
(15) Guruprasad, K.; Rajkumar, S. *J. Biosci.* **2000**, *25*, 143–156.
(16) Waters, M. L. *Curr. Opin. Chem. Biol.* **2002**, *6*, 736–741.
(17) See, for example: Ignatova, Z.; Gierasch, L. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 523–528. Hoffmann, C.; Gaietta, G.; Bünemann, M.; Adams, S. R.; Oberdorff-Maass, S.; Behr, B.; Vilardaga, J. P.; Tsien, R. Y.; Ellisman, M. H.; Lohse, M. J. *Nat. Methods* **2005**, *2*, 171–176.

JA809315X