A Consensus Zinc Finger Peptide: Design, High-Affinity Metal Binding, a pH-Dependent Structure, and a His to Cys Sequence Variant

Beth Allyn Krizek, Barbara T. Amann, Valda J. Kilfoil, Denise L. Merkle, and Jeremy M. Berg*

Contribution from the Department of Chemistry, The Johns Hopkins University, Baltimore, Maryland 21218, and Department of Biophysics and Biophysical Chemistry, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205. *Received October 13, 1990*

Abstract: A single zinc finger peptide, ProTyrLysCysProGluCysGlyLysSerPheSerGlnLysSerAspLeuValLys**His**GlnArg-Thr**His**ThrGly, has been designed with the use of a data base of 131 zinc finger sequences. Studies indicated that this peptide binds metal ions such as Zn^{2+} and Co^{2+} and folds in their presence. The affinity of this peptide for metal ions is greater than that demonstrated for any other zinc finger peptide characterized to date. Nuclear magnetic resonance studies revealed that the zinc complex of this peptide adopts a structure similar to that predicted and observed for other zinc finger domains. In addition, these studies led to the discovery that the latter of the histidine residues can be protonated and dissociated from the metal center with only local loss of structure. This histidine residue can also be replaced with a cysteine residue to yield a peptide that has a $(Cys)_3(His)$ rather than a $(Cys)_2(His)_2$ metal binding site.

Introduction

In recent years, a large class of proteins has been discovered that is characterized by the presence of one or more sequences that closely approximate the form (Tyr,Phe)-X-Cys-$X_{2,4}$-Cys-X_3-Phe-X_5-Leu-X_2-His-$X_{3,4}$-His where X represents relatively variable amino acids.[1] Each of these sequences appears to form a small domain (often termed a "zinc finger" domain[2]) organized around a zinc ion tetrahedrally coordinated by the cysteine and histidine residues. Where it has been determined, proteins that contain these domains have been shown to be specific nucleic acid binding proteins. Peptides corresponding to single domains of this type have been used to characterize the folding, metal-binding, and structural properties of such domains although these peptides have been found to be incapable of site-specific interactions with DNA.[3-10] It has been demonstrated that these peptides are largely unfolded in the absence of appropriate metal ions but fold to unique three-dimensional structures in their presence. These peptides have been very useful for structural studies of zinc finger domains by two-dimensional nuclear magnetic resonance (NMR) methods, as they are small (approximately 30 amino acids) yet fold in aqueous solution. Such NMR studies on several different zinc finger peptides have revealed structures that are quite similar to although not identical with one another.[4,5,7,9,10] Moreover, the experimentally determined structures are strikingly similar to a structure predicted for the zinc finger domains on the basis of the observation of recurring metal-chelating substructures in crystallographically characterized metalloproteins.[11] The structure, which consists of two strands and a helix, is shown schematically in Figure 1.

The number of zinc finger sequences that are known is large and growing rapidly.[1] This is true for several reasons. First, as noted above, these sequences often occur as large tandem arrays, with as many as 37 such sequences being found in a single deduced protein sequence. Second, a large number of genes involved in developmental control are being cloned and sequenced and many of these are found to contain zinc finger sequences. Third, it has been possible to intentionally clone zinc finger encoding regions on the basis of hybridization to other zinc finger encoding DNA fragments. Finally, methods based on the polymerase chain reaction are being developed that may make detection and characterization of zinc finger sequences even more facile.[12] When the project described herein was initiated, a total of 131 sequences from 18 proteins,[13-28] tabulated in Table I, were known.

This sequence data base contains a great deal of information. As one illustration of this point, consider the distribution of proline

* To whom correspondence should be addressed at The Johns Hopkins University School of Medicine.

(1) Berg, J. M. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 405 and references therein.
(2) Miller, J.; McLachlan, A. D.; Klug, A. *EMBO J.* **1985**, *4*, 1609.
(3) Frankel, A. D.; Berg, J. M.; Pabo, C. O. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4841.
(4) Párraga, G.; Horvath, S. J.; Eisen, A.; Taylor, W. E.; Hood, L.; Young, E. T.; Klevit, R. E. *Science* **1988**, *241*, 1489.
(5) Lee, M. S.; Gippert, G. P.; Soman, K. V.; Case, D. A.; Wright, P. E. *Science* **1989**, *245*, 635.
(6) Párraga, G.; Horvath, S.; Hood, L.; Young, E. T.; Klevit, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 137.
(7) Klevit, R. E.; Herriott, J. R.; Horvath, S. J. *Proteins: Struct., Funct., Genet.* **1990**, *7*, 215.
(8) Weiss, M. A.; Mason, K. A.; Dahl, C. E.; Keutmann, H. T. *Biochemistry* **1990**, *29*, 5660.
(9) Omichinski, J. G.; Clore, G. M.; Appella, E.; Sakaguchi, K.; Gronenborn, A. M. *Biochemistry* **1990**, *29*, 9324.
(10) Neuhaus, D.; Nakaseko, Y.; Nagai, K.; Klug, A. *FEBS Lett.* **1990**, *262*, 179.
(11) Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 99.
(12) Pellegrino, G. R.; Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 671.
(13) Kadonaga, J. T.; Carner, K. R.; Masiarz, F. R.; Tjian, R. *Cell* **1987**, *51*, 1079.
(14) Vincent, A.; Colot, H. V.; Rosbash, M. *J. Mol. Biol.* **1985**, *186*, 149.
(15) Rosenberg, U. B.; Schröder, C.; Preiss, A.; Kienlin, A.; Côté, S.; Riede, I.; Jäckle, H. *Nature* **1986**, *319*, 336.
(16) Boulay, J. L.; Dennefeld, C.; Alberga, A. *Nature* **1987**, *330*, 395.
(17) Ruis i Altaba, A.; Perry-O'Keefe, H.; Melton, D. A. *EMBO J.* **1987**, *6*, 3065.
(18) Baldarelli, R. M.; Mahoney, P. A.; Salas, F.; Gustavson, E.; Boyer, P. D.; Chang M.-F.; Roark, M.; Lengyel, J. A. *Dev. Biol.* **1988**, *125*, 85.
(19) Stillman, D. J.; Bankier, A. T.; Seddon, A.; Groenhout, E. G.; Nasmyth, K. A. *EMBO J.* **1988**, *7*, 485.
(20) Ginsberg, A. M.; King, B. O.; Roeder, R. G. *Cell* **1984**, *39*, 479.
(21) Hartshorne, T. A.; Blumberg, H.; Young, E. T. *Nature* **1986**, *320*, 283.
(22) Schuh, R.; Aicher, W.; Gaul, U.; Côté, S.; Preiss, A.; Maier, D.; Seifert, U.; Nauber, C.; Schröder, C.; Kemler, R.; Jäckle, H. *Cell* **1986**, *47*, 1025.
(23) Chowdhury, K.; Deutsch, U.; Gruss, P. *Cell* **1987**, *48*, 771.
(24) Tautz, D.; Lehmann, R.; Schnürch, H.; Schuh, R.; Seifert, E.; Kienlin, A.; Jones, K.; Jäckle, H. *Nature* **1987**, *327*, 383.
(25) Pays, E.; Murphy, N. B. *J. Mol. Biol.* **1987**, *197*, 147.
(26) Milbrandt, J. *Science* **1987**, *238*, 797.
(27) Page, D. C.; Mosher, R.; Simpson, E. M.; Fisher, E. M. C.; Mardon, G.; Pollack, J.; McGillivray, B.; de la Chapelle, A.; Brown, L. G. *Cell* **1987**, *51*, 1091.

0002-7863/91/1513-4518$02.50/0

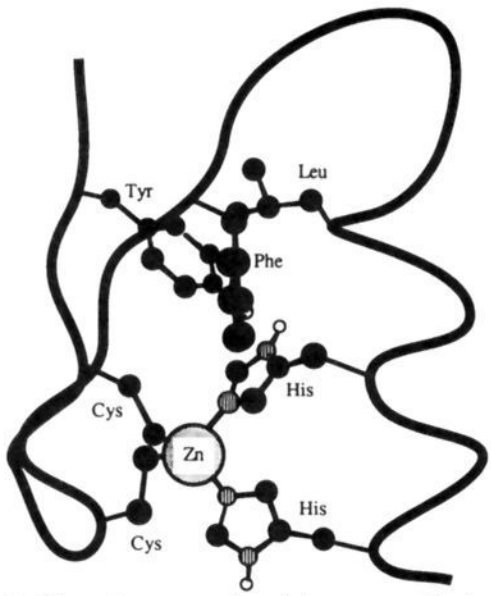

Figure 1. Schematic representation of the structure of a zinc finger domain. The structure consists of two β strandlike structures followed by a helix. The metal-binding and conserved hydrophobic residues are illustrated.

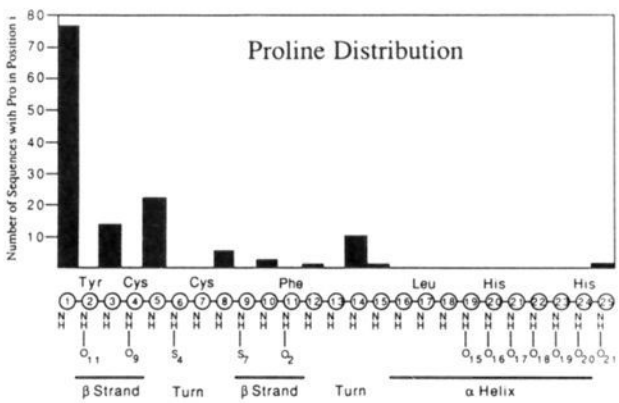

Figure 2. Distribution of proline residues in the 131 sequences from Table I. For comparison, the hydrogen-bonding pattern from the predicted structure of the zinc finger domains is shown below. Note that proline is never found at a position in which the peptide amide hydrogen was proposed to be involved in a hydrogen bond, either to a carbonyl oxygen or to a cysteinyl sulfur.

residues across the zinc finger sequences shown in Figure 2. This proline distribution is remarkably similar to the peptide NH hydrogen-bonding pattern predicted for zinc finger domains in that proline is never present at any residue for which the peptide amide group was proposed to participate in hydrogen bonding, either to a peptide carbonyl oxygen or to a metal-coordinated cysteinyl sulfur atom. Thus, for the two strands, an alternation of proline-tolerant and proline-excluding sites is observed whereas, for the helical region, proline is excluded from nine consecutive positions.

We have elected to use the knowledge of this large data base to design a consensus zinc finger peptide. The sequence of this peptide was determined by selecting the amino acid that occurred in the largest number of zinc finger domains at each position. The rationale behind this approach was that amino acids involved in stabilizing the zinc finger domain should be present with relatively high frequency in the data base and, hence, would be present in the consensus peptide. We hypothesized that the structure of a peptide with the consensus sequence might be unusually stable.

(28) Joho, K. E.; Darby, M. K.; Crawford, E. T.; Brown, D. D. *Cell* **1990**, *61*, 293.

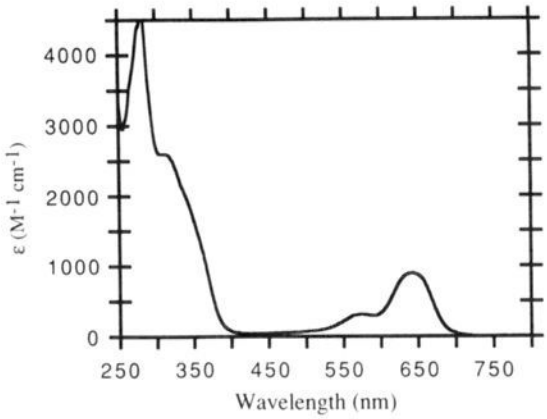

Figure 3. Absorption spectrum of the Co^{2+} complex of the zinc finger consensus peptide, CP-1.

Since the processes of protein folding and metal binding are coupled, this stability should be reflected in the affinity of the consensus peptide for metal ions. The sequence selected for the consensus peptide (hereafter CP-1) is ProTyrLys**Cys**ProGlu**Cys**GlyLysSerPheSerGlnLysSerAspLeu-ValLys**His**GlnArgThr**His**ThrGly. In this report, we describe the synthesis and characterization of CP-1. Our results indicate that CP-1 appears to bind metal ions more tightly than any other zinc finger peptide that has been characterized to date. NMR spectroscopic studies revealed the presence of many features that are consistent with the expected overall structure. In addition, these studies led to the discovery of a form of the peptide–Zn^{2+} complex formed at low pH in which the final histidine residue becomes dissociated from the metal center yet the rest of the structure remains intact. Finally, we are using CP-1 as a framework on which to study the effects of sequence changes on the properties of zinc finger peptides. As one example of this approach, we report the synthesis of a sequence variant in which we have replaced the final histidine with a cysteine. Spectroscopic studies of the Co^{2+} complex reveal that this change leads to the expected differences in the nature of the metal-binding site. The ability to vary the properties of the metal-binding site will be useful for additional studies.

Experimental Section

All peptides were synthesized on a MilliGen/Biosearch 9050 Pep-Synthesizer using *N*-fluorenylmethoxycarbonyl (Fmoc) amino acid pentafluorophenyl esters. The peptides were cleaved from the resin as suggested by MilliGen/Biosearch protocol. All amino acids and reagents were purchased from MilliGen/Biosearch except for dimethylformamide (Burdick and Jackson). The cleaved peptides were dissolved in water and incubated with an excess of dithiothreitol for 2 h at 50 °C before being purified by reverse-phase high-performance liquid chromatography (HPLC). The single major HPLC peak was collected and dried in an anaerobic chamber (COY) under a 5% hydrogen/95% dinitrogen atmosphere. In all cases, amino acid analysis data were consistent with the sequences synthesized. All samples were stored under anaerobic conditions, and all solvents were degassed with helium prior to use. Metal ion titrations were performed as described previously.[29]

NMR samples were prepared with or without 1.2 equiv of zinc and adjusted to the desired pH with deuterated Tris (MSD Isotopes). ^1H NMR data were obtained with a Varian XL 400, a Bruker AM 600, or a Varian UNITY 400 NMR spectrometer. Two-dimensional NMR experiments were performed by using standard methods.[30] COSY and pH titration experiments were performed on samples in deuterium oxide (Aldrich, 100.00%). COSY,[31] NOESY,[32] and HOHAHA[33] experiments were carried out on 90% H_2O/10% D_2O samples by using either presaturation or jump return[34] methods for water suppression. Mixing times

(29) Berg, J. M.; Merkle, D. L. *J. Am. Chem. Soc.* **1989**, *111*, 3759.
(30) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley-Interscience: New York, 1986.
(31) Bax, A.; Freeman, R. *J. Magn. Reson.* **1981**, *44*, 542.
(32) Jenner, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *J. Chem. Phys.* **1979**, *71*, 4546.
(33) Bax, A. *Methods Enzymol.* **1989**, *176*, 151.

Table I. A Zinc Finger Data Base with 131 Sequences[a]

Human Sp1 (13)

QHICHI QGCGKVYGKTSHLRAHLR WHTG
PFMCTW SYCGKRFTRSDELQRHKR THTG
KFACPE CPKRFMRSDHLSKHIK THQN

Drosophila Serendipity β (14)

EIPCHI CGEMFSSQEVLERHIKADTCQK
QATCNV CGLKVKDDEVLDLHMN LHEG
ELECRY CDKKFSHKRNVLRHME VHWD
KYQCDK CGERFSLSWLMYNHLM RHDA
ALICEV CHQQFKTKRTYLHHLR THQT
-YPCPD CEKSFVDKYTLKVHKR VHQP

Drosophila Serendipity δ (14)

KQECTT CGKVYNSWYQLQKHLS EEHSK
NHICPI CGVIRRDEEYLELHMN LHEG
EKQCRY CPKSFSRPVNTLRHMR SHWD
KYQCEK CGLRFSQDNLLYNHRL RHEA
PIICSI CNVSFKSRKTFNHHTL IHKE
-HYCSV CPKSFTERYTLKMHMK THEG
SGFCLI CNTTFENKKELEHHLQ FHAD

Drosophila Krüppel (15)

SFTCKI CSRSFGYKHVLQNHER THTG
PFECPE CDKRFTRDHHLKTHMR LHTG
PYHCSH CDRQFVQVANLRRHLR VHTG
PYTCEI CDGKFSDSNQLKSHML VHTG
PFECER CHMKFRRRHHLMNHK CGI

Drosophila Snail (16)

RFKCDE CQKMYSTSMGLSKHRQ FHCP
THSCEE CGKLYTTIGALKMHIR HTL
PCKCPI CGKAFSRPWLLQGHIR THTG
PFQCPD CPRSFADRSNLRAHQQ THVD
KYACQV CHKSFSRMSLLNKHSS SNCTI

Xenopus Xfin (17)

SHHCPH CKKSFVQRSDFLKHQR THTG
PYQCVE CQKKFTERSALVNHQR THTG
PYTCLD CQKTFNQRSALTKHRR THTG
PYRCSV CSKSFIQNSDLVKHLR THTG
PYECPL CVKRFAESSALMKHKR THST
PFRCSE CSRSFTHNSDLTAHMR KHTE
PYSCSK CRKTFKRWKSFLNHQQ THSR
PYLCSH CNKGFIQNSDLVKHFR THTG
PYQCAE CHKGFIQKSDLVKHLR THTG
PFKCSH CDKKFTERSALAKHQR THTG
PYKCSD CGKEFTQRSNLILHQR IHTG
PYKCTL CDRTFIQNSDLVKHQK VHAN
PHKCSK CDLTFSHWSTFMKHSK LHSG
KFQCAE CKKGFTQKSDLVKHIR VHTG
PFKCLL CKKSFSQNSDLHKHWR IHTG
PFPCYT CDKSFTERSALIKHHR THTG
PHKCSV CQKGFIQKSALTKHSR THTG
PYPCTQ CGKSFIQNSDLVKHQR IHTG
PYHCTE CNKRFTEGSSLVKHRR THSG
PYRCPQ CEKTFIQSSDLVKHLV VHNG
PYPCTE CGKVFHQRPALLKHLR THKT
RYPCNE CSKEFFQTSDLVKHLR THTG
PYHCPE CNKGFIQNSDLVKHQR THTG
PYTCSQ CDKGFIQRSALTKHMR THTG
PYKCEQ CQKCFIQNSDLVKHQR IHTG
PYHCPD CDKRFTEGSSLIKHQR IHSR
PYPCGV CGKSFSQSSNLLKHLK CHSE
SFKCND CGKCFAHRSVLIKHVR IHTG
PYKCSQ CTRSFIQKSDLVKHYR THTG
PYKCGL CERSFVEKSALSRHQR VHKN
RYSCSE CGKCFTHRSVFLKHWR MHTG
PYTCKE CGKSFSQSSALVKHVR IHTG
PYACST CGKSFIQKSDLAKHQR IHTG
PYTCTV CGKKFIDRSSVVKHSR THTG
PYKCNE CTKGFVQKSDLVKHMR THTG
PYGCNC CDRSFSTHSASVRHQR MCNT

Drosophila Terminus (18)

DLHCRR CRTQFSRRSKLHIHQK LRCGQ

Yeast SWI5 (19)

TFECLF PGCTKTFKRRYNIRSHIQ THLE
PYSCDH PGCDKAFVRNHDLIRHKK SHQE
-YACP CGKKFNREDALVVHRSRMICSG

Xenopus Transcription Factor IIIA (20)

RYICSF ADCGAAYNKNWKLQAHLC KHTG
PFPCKE EGCEKGFTSLHHLTRHSL THTG
NFTCDS DGCDLRFTTKANMKKHFN RFHNI
VYVCHF ENCGKAFKKHNQLKVHQF SHTQ
PYECPH EGCDKRFSLPSRLKRHEK VHAG
-YPCKKDDSCSFVGKTWTLYLKHVA ECHQD
-AVCDV CNRKFRHKDYLRDHQK THEK
VYLCPR DGCDRSYTTAFNLRSHIQ SFHEE
PFVCEH AGCGKCFAMKKSLERHSV VHDP

Yeast ADR1 (21)

SFVCEV CTRAFARQEHLKRHYR SHTN
PYPCGL CNRCFTRRDLLIRHAQ KIHSG

Drosophila Krh (22)

TYRCSE CQREFELLAGLKKHLK THRT
KYQCDI CGQKFVQKINLTHHAR IHSS
PYECPE CQKRFQERSHLQRHQK YHAQ
SYRCEK CGKMYKTERCLKVHNL VHLE
PFACTV CDKSFISNSKLKQHSN IHTG
PFKCNY CPRDFTNFPNWLKHTR RRHKV

Mouse Mkr1 (23)

PFVCNY CDKTFSFKSLLVSHKR IHTG
PYECDV CQKTFSHKANLIKHQR IHTG
PFECPE CGKAFTHQSNLIVHQR AHME
PYGCSE CGKAFTHQSNLIVHQR IHTG
PYECNE CAKTFFKKSNLIIHQK IHTG
RYECSE CGKSFIQNSQLIIHRR THTG
PYECTE CGKTFSQRSTLRLHLR IHTG

Mouse Mkr2 (23)

VYGCDE CGKTFRQSSSLLKHQR IHTG
PYTCNV CDKHFIERSSLTVHQR THTG
PYKCHE CGKAFSQSMNLTVHQR THTG
PYQCKE CGKAFRKNSSLIQHER IHTG
PYKCHD CEKAFSKNSSLTQHRR IHTG
PYECMI CGKHFTGRSSLTVHQV IHTG
PYECTE CGKAFSQSAYLIEHRR IHTG
PYECDQ CGKAFIKNSSLIVHQR IHTG
PYQCNE CGKPFSRSTNLTRHQR THT-

Drosophila Hunchback (24)

NYKCKT CGVVAITKVDFWAHTR THMK
ILQCPK CPFVTEFKHHLEYHIR KHKN
PFQCDK CSYTCVNKSMLNSHRK SHSS
QYRCAD CDYATKYCHSFKLHLRHYGHKP
IYECKY CDIFFKDAVLYTIHMG YHSC
VFKCNM CGEKCDGPVGLFVHMARNAHS-

Trypanosome TRS-1 (25)

PTKCTE CDATYQCRSSAVTHMV NKHGF
VLHCTI CASKFAVPGRLLHHLR TIHGI
PFQCDL CEASFGTHSSLSLHKK LKHKS
EVQCGV CQKVLSCRDSLIRHCK AFHKG
MLVCPT CGRQCASKTGLTLHQK KMHGM

Mouse NGFI-A (26)

PYACPV ESCDRRFSRSDELTRHIR IHTG
PFQCRI CMRNFSRSDHLTTHIR THTG
PFACDI CGRKFARSDERKRHTK IHLR

Human ZFY (27)

VYPCMI CGKKFKSRGFLKRHMK NHPE
KYHCTD CDYTTNKKISLHNHLE SHKL
AIECDE CGKHFSHAGALFTHKM VHKE
MHKCKF CEYETAEQGLLNRHLL AVHSK
PHICVE CGKGFRHPSELRKHMR IHTG
PYQCQY CEYRSADSSNLKTHIK TKHSK
PFKCDI CLLTFSDTKEVQQHTL VHQE
THQCLH CDHKSSNSSDLKRHVI SVHTK
PHKCEM CEKGFHRPSELKKHVA VHKG
MHQCRH CDFKIADPFVLSRHIL SVHTK
PFRCKR CRKGFRQQNELKKHMK THSG
VYQCEY CEYSTTDASGFKRHVI SIHTK
PHRCEY CKKGFRRPSEKNQHIM RHHK

Xenopus p43 5S RNA Binding Protein (28)

LLRCPA AGCKAFYRKEGKLQDHMA GHSE
PWKCGI KDCDKVFARKRQILKHVK RHLA
KLSCPT AGCKMTFSTKKSLSRHKL YKHGE
PLKCFV PGCKRSFRKKRALRRHLS VHSN
LSVCDV PGCSWKSSSVAKLVAHQK RHRG
-YRCSY EGCQTVSPTWTALQTHVK KHPL
-LQCAA CKKPFKKASALRRHKA THAK
QLPCPR QDCDKTFSSVFNLTHHVARKLHLC
THRCPH SGCTRSFAMRESLLRHLV VHDP

[a] Numbers in parentheses correspond to references in the text. One-letter codes for amino acids are used throughout (A = Alanine, C = Cysteine, D = Aspartic Acid, E = Glutamic Acid, F = Phenylalanine, G = Glycine, H = Histidine, I = Isoleucine, K = Lysine, L = Leucine, M = Methionine, N = Asparigine, P = Proline, Q = Glutamine, R = Arginine, S = Serine, T = Threonine, V = Valine, W = Tryptophan, Y = Tyrosine).

of 350–400 ms were used in the NOESY experiments shown.

Results

As an initial probe of its metal-binding properties, the interaction between CP-1 and Co^{2+} was examined. The peptide does bind Co^{2+} to yield a complex that has an absorption spectrum, shown in Figure 3, that is consistent with a tetrahedral metal-binding site similar to that seen in other zinc finger peptides.[3,4,6,8] The affinity of the peptide for metal ions was probed in several ways. First, a metal ion titration[29] with Co^{2+} produced a curve that could be fit with a dissociation constant $< 1 \times 10^{-7}$ M, as shown in Figure 4a. The curvature observed is sufficiently small that it was not possible to determine a well-defined lower bound for the dissociation constant from this experiment alone. Second, a pH titration of the peptide–Co^{2+} complex (shown in Figure 4b) revealed a sharp drop in the concentration of the complex near pH = 5.0. Hill plot analysis revealed that this process is a two-proton transition, consistent with an NMR-monitored pH titration

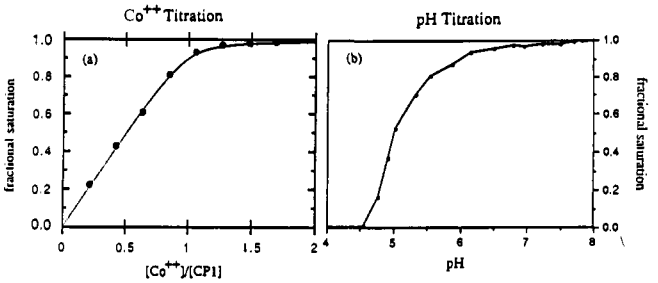

Figure 4. Determination of the affinity of CP-1 for metal ions. (a) 20.3 nmol of CP-1 was titrated with $CoCl_2$ in 100 mM HEPES, 50 mM NaCl, pH 7.0 buffer. The absorption spectrum of the sample was monitored and used to calculate the degree of saturation of the peptide with Co^{2+}. The curve could be fit with any dissociation constant less that 10^{-7} M. (b) 22.2 nmol of CP-1 plus 61.8 nmol of $CoCl_2$ in 50 mM NaCl was titrated with HCl. The absorption spectrum was monitored and the extent of saturation determined. The results are consistent with a two-proton process being responsible for the loss of the characteristic absorption spectrum and with a dissociation constant near 10^{-7} M for the Co^{2+} complex, given a pK_a value of 6.5 for the two histidine residues.

(34) Bax, A.; Sklenar, V.; Clore, G. M.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1987**, *109*, 6511.

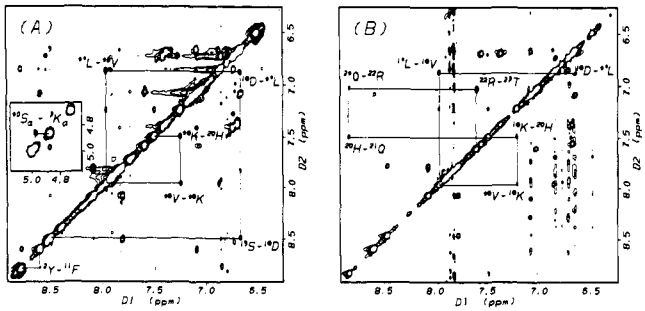

Figure 5. The NH–NH portion of the contour plot of NOESY spectra of CP-1 in the presence of Zn^{2+}. The spectra were recoreded at 600 MHz in 90% H_2O/10% D_2O at 25 °C. (A) Data were collected at pH 4.65. Cross peaks ^{2}Tyr(NH)–^{11}Phe(NH) and ^{3}Lys(C_αH)–^{10}Ser(C_αH) (inset) from the β sheet region and NH(i)–NH(i + 1) cross peaks for i = 15–19 from the helical region are indicated. (B) Data were collected at pH 6.45. NH(i)–NH(i + 1) cross peaks for i = 16–22 from the helical region are indicated.

of another zinc finger peptide.[6] Given the pK_a value of 6.5 for the two histidine residues in the unbound peptide (determined by NMR spectroscopy), the observations are consistent with an apparent K_d near 1×10^{-7} M at pH 7.0 calculated from eq 1,

$$K_d = [M]((1 + 10^{-7}/K_a)^2/(1 + ([H^+]_{1/2}/K_a)^2)) \quad (1)$$

where [M] is the free metal concentration at half-saturation, $[H^+]_{1/2}$ is the proton concentration when the peptide–metal complex is half-saturated, and K_a is the equilibrium constant for protonation of the histidine residues (found to be equal for CP-1 and assumed to be equal in the derivation of eq 1).[35] Finally, competition experiments with other metal-binding peptides are consistent with a dissociation constant of 5×10^{-8} M.[36] The dissociation for the CP-1–Zn^{2+} complex was determined to be approximately 2×10^{-12} M via Zn^{2+}/Co^{2+} competition experiments.[29]

NMR studies of CP-1 were performed. The addition of Zn^{2+} dramatically changes the one-dimensional NMR spectrum of CP-1, causing both upfield and downfield shifts in the aromatic region and increased dispersion of the aliphatic proton signals (not shown). To examine the structure of the peptide–Zn^{2+} complex, nuclear Overhauser exchange spectroscopy (NOESY) studies were undertaken. A portion of the NOESY spectrum recorded at pH 4.65 is shown in Figure 5A. Cross peaks between the NH protons of ^{2}Tyr and ^{11}Phe and between the α protons of ^{3}Lys and ^{10}Ser as well as a chain of sequential NH to NH cross peaks from ^{15}Ser to ^{20}His were evident. Furthermore, numerous NOE connections between protons from ^{2}Tyr, ^{4}Cys, ^{11}Phe, ^{17}Leu, and ^{20}His were also observed. NOESY studies at pH 6.45 (Figure 5B) revealed that the chain of sequential NH–NH connectivities extended through ^{23}Thr at this higher pH.

One-dimensional NMR spectra of CP-1–Zn^{2+} were examined as a function of pH. Figure 6 shows the aromatic region of the NMR spectra of CP-1–Zn^{2+} at three different pH values. As the pH was lowered, three distinct forms of the peptide were observed. At high pH (pH = 7.54), the spectrum of the fully folded form of CP-1–Zn^{2+} was observed, with the ^{20}His and ^{24}His 4-H ring protons appearing at 6.82 and 6.53 ppm and the 2-H ring protons at 7.82 and 7.91 ppm, respectively. As the pH was lowered, the ring protons of ^{24}His broadened and shifted slightly downfield, with little change observed for any other signals. When the pH was decreased to 2.69, an unfolded form of the peptide was observed, with both of the histidine 4-H protons appearing at 7.15 ppm and the 2-H protons at 8.44 and 8.49 ppm. Studies at a more extensive series of pH values revealed that the first process, characterized by broadening of the ^{24}His resonances, occurred

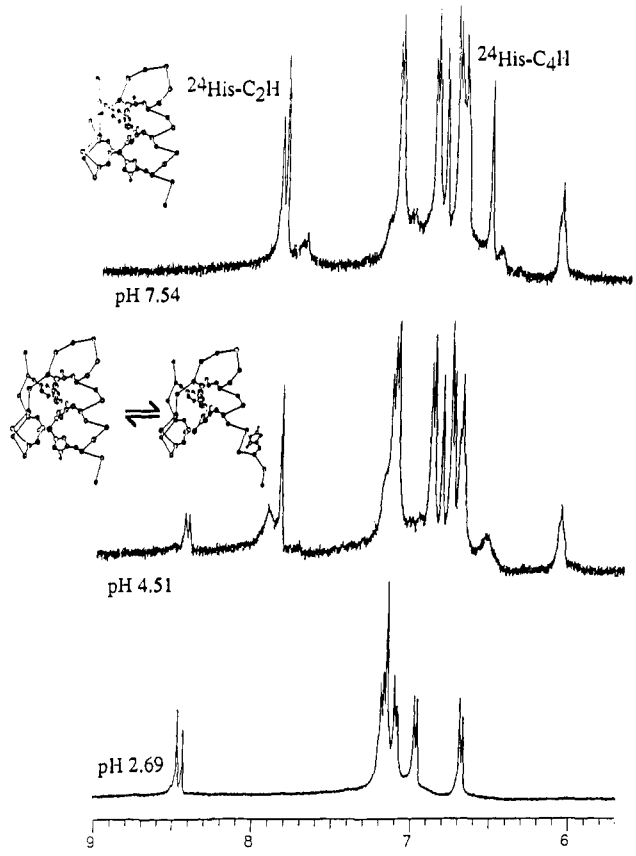

Figure 6. One-dimensional NMR spectra of CP-1 in the presence of Zn^{2+} in D_2O as a function of pH. At pH 7.54, only the spectrum due to the fully folded form is observed. At pH 4.51, the resonances due to ^{24}His have broadened whereas the remainder of the signals are essentially unchanged. A small amount of unfolded peptide is also present at this pH. At pH 2.69, only the unfolded form of the peptide is observed. Schematic structures with only α-carbon positions for all but the most highly conserved residues are shown.

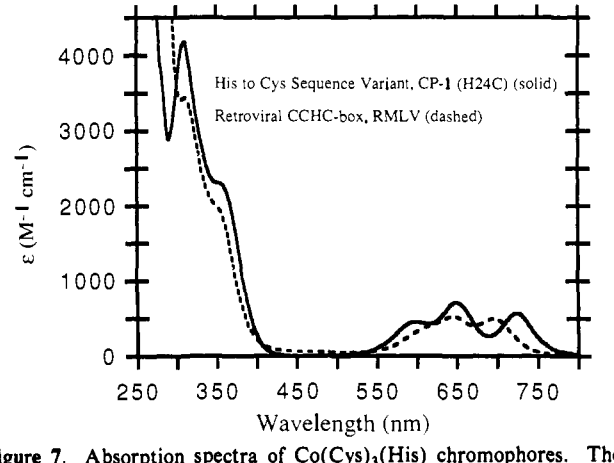

Figure 7. Absorption spectra of $Co(Cys)_3(His)$ chromophores. The spectra of the Co^{2+} complexes of CP-1-H24C and of a peptide of the form X_2-Cys-X_2-Cys-X_4-His-X_4-Cys-X_2[41] derived from the nucleocapsid protein of Rauscher murine leukemia virus are shown.

with a midpoint at pH 5.3 whereas the second process, involving the more dramatic changes, showed a midpoint near pH 3.5.

Examination of the sequence data base revealed that six of the sequences lack the final histidine residue but have a cysteine residue within five amino acids of the conserved histidine. This observation suggested that cysteine could replace histidine as a metal-binding ligand from this position. A sequence variant peptide in which the final histidine in the consensus sequence was replaced by cysteine was prepared. This peptide (CP-1-H24C) binds metal ions such as Co^{2+} and Zn^{2+} as shown by metal ion titration experiments. The absorption spectrum of the CP-1-

(35) This equation was derived by assuming coupled equilibria involving histidine protonation and binding. No correction for the formation of protonated metal-bound species was applied, as this appeared not to occur under the conditions used with Co^{2+}.

(36) Krizek, B. A.; Merkle, D. L.; Berg, J. M. Manuscript in preparation.

H24C–Co^{2+} complex, shown in Figure 7, is clearly distinct from that of the parent complex. A titration experiment with Co^{2+} yielded a dissociation constant of 1×10^{-7} M.

Discussion

A 26 amino acid peptide corresponding to the consensus sequence derived from a data base of 131 zinc finger sequences has been synthesized. Examination of the absorption spectrum of the Co^{2+} complex of this peptide revealed it to be quite similar to those of zinc finger peptides corresponding to naturally occurring sequences. These spectra are indicative of tetrahedral coordination by two cysteinyl sulfur ligands and two histidyl nitrogen ligands. The sensitivity of the absorption spectra of such Co^{2+} complexes to the detailed nature of the ligating atoms is directly demonstrated by studies of a histidine to cysteine sequence variant, as can be seen by comparing Figures 3 and 7.

The initial rationale for designing and synthesizing CP-1 was that residues involved in stabilizing the three-dimensional structure of the metal-bound forms of zinc finger domains should be present in such a consensus sequence, leading to enhanced affinity for metal ions. The metal ion affinities of several other zinc finger peptides have been studied previously. Metal ion titrations under conditions identical with those reported here for CP-1 were performed with a peptide corresponding to the second domain from *Xenopus* transcription factor IIIA, revealing dissociation constants of 3.8×10^{-6} M for the Co^{2+} complex and 2.8×10^{-9} M for the Zn^{2+} complex.[29] Optically monitored pH titrations were reported for the Co^{2+} complexes of a series of peptides corresponding to domains of the human ZFY.[8] While insufficient information is given to evaluate metal dissociation constants from eq 1, in all cases the pH values at which the peptide–metal complexes are half-diassociated are greater than 5.5. Finally, an NMR-monitored pH titration of the Zn^{2+} complex of a peptide from yeast ADR1 has been reported.[6] A cooperative transition involving protonation of both histidine residues was observed with the halfway point at pH 5.2. By use of the conditions reported (peptide concentration 1.0 mM, total zinc concentration 1.1 mM, which corresponds to a free metal concentration of 0.6 mM at half-saturation, and pK_a's of histidines 6.2), an apparent dissociation constant at pH 7.0 of 7×10^{-6} M for the zinc complex could be estimated from eq 1. This may be an underestimate of the affinity of this peptide for metal ions, given that it was shown to bind Co^{2+} with appreciable affinity. All of these results may be compared with those from analogous experiments with CP-1. In every case CP-1 behaves in a manner consistent with higher affinity metal binding. The dissociation constants for the complexes of CP-1 were determined to be $K_d^{Co} = 5 \times 10^{-8}$ M and $K_d^{Zn} = 2 \times 10^{-12}$ M, estimated to be accurate to within 30%. These values confirm that complexes of CP-1 are more stable than the corresponding complexes of any of the other zinc finger peptides characterized to date. The apparent increases in affinity range from a factor of 76 compared with the affinity of the Co^{2+} complex of the transcription factor IIIA derived peptide to a factor perhaps as large as 10^6 compared with the affinity of the Zn^{2+} complex of the ADR1 peptide, corresponding to free energy differences from 2.5 to 8 kcal/mol.

Initial NMR studies of CP-1 and its metal complexes have been performed as a probe of three-dimensional structure. A complete structure determination is in progress and will be reported in a subsequent publication. The first suggestions concerning the structure of this class of zinc finger domain came from theoretical analysis. Brown and Argos noted that secondary structure predictions applied to a small data base of zinc finger sequences predicted an α helix involving the conserved leucine and the first of the histidine residues.[37] Berg proposed the first detailed model for the complete structure.[11] The availability of this predicted structure (as well as experimental structures of other zinc finger peptides subsequently determined[4,5,7,9,10]) allows some structural inferences to be drawn prior to completion of a full three-dimensional structure determination. The major features of the structure are a helix beginning prior to the conserved leucine residue and extending through the metal-ligating histidine residues, an antiparallel sheetlike structure aligning the two conserved aromatic residues, and a hydrophobic core involving all three conserved hydrophobic residues as well as the metal-binding residues. Inspection of the NOESY data from the CP-1–Zn^{2+} complex revealed that all of these features are present. As shown in Figure 5B, NH(i) to NH($i + 1$) cross peaks were seen starting at $i = {}^{16}$Asp and extending to ^{23}Thr for spectra recorded at an apparent pH of 6.45. Further support for the helical conformation of this region came from the presence of numerous cross peaks involving $C_\alpha(i)$ and NH($i + 3$) protons for $i = 15$–20 and $C_\alpha(i)$ and $C_\beta(i + 3)$ protons for $i = 15$–18. Evidence for the β sheetlike structure was seen in the form of cross peaks involving the amide protons of ^{2}Tyr and ^{11}Phe and the α-hydrogens of ^{3}Lys and ^{10}Ser, as shown in Figure 5A. Finally, the presence of the hydrophobic core was evident from the presence of numerous cross peaks involving the side chains of the conserved hydrophobic and metal-binding residues. Taken together, these data strongly support the claim that the CP-1–Zn^{2+} has a structure that closely resembles the canonical structure. More detailed comparison with the other experimentally determined structures will have to await further analysis.

In the course of these NMR studies, we discovered a localized structural transition that occurs for the CP-1–Zn^{2+} complex as a function of pH. Figure 6 shows the aromatic region of the NMR spectra of this complex at three different pH values. As the pH was lowered, three distinct forms of the peptide were observed. At high pH, a spectrum consistent with the fully folded form of CP-1–Zn^{2+} was observed. As the pH was lowered, the ring protons of ^{24}His specifically broadened, indicating the presence of an equilibrium between the fully folded form and a form in which ^{24}His had become protonated and dissociated from zinc. The NOESY spectrum at pH 4.65 indicates the loss of structure from ^{21}Gln through the carboxyl terminus. Upon further decrease of the pH, a new spectrum appeared that was identical with that observed for CP-1 in the absence of zinc and at the same pH, indicating that the complex loses its remaining structure. As noted above, a similar pH study has been reported for a single zinc finger peptide from yeast ADR1 (ADR1a).[6] The results differ significantly from what we have observed for CP-1. As the pH was lowered below 5.5, the spectra of ADR1a–Zn^{2+} showed two distinct forms of the peptide, folded and unfolded, in slow exchange. Unlike the case of CP-1–Zn^{2+}, no intermediate form was observed. The apparent pK_a of the interconversion of the two forms of ADR1a was 5.2. Comparison of the results from these two NMR studies provides further evidence for the enhanced structural stability of the metal complexes of CP-1 over naturally occurring zinc finger peptides. This enhanced stability is apparently required for the stability of the intermediately folded form.

The observed local unravelling of CP-1 as a function of pH is not unprecedented. The crystal structure analysis of reduced poplar plastocyanin at six different pH values showed similar results.[38] In the high-pH (pH = 7.8) crystal structure, the Cu(I) atom shows distorted tetrahedral coordination by two histidines, a cysteinate and a methionine. At low pH (pH = 3.8), the copper atom becomes trigonally coordinated with one of the histidine–Cu(I) bonds being severed and the histidine protonated. The coordination number at the Zn^{2+} site in the intermediate-pH form of CP-1–Zn^{2+} is not clear. It may be three-coordinate, a rare but occasionally observed structural type for Zn^{2+},[39] or four-coordinate, with a water replacing the protonated imidazole. Studies with Co^{2+}-substituted CP-1 are inconclusive with regard to this point. Due to the weaker binding of Co^{2+} compared with Zn^{2+} to CP-1, Co^{2+} binding appears to be completely lost as the pH is lowered. However, a truncated CP-1 variant in which the final four residues (including the final histidine) were deleted has been synthesized

(37) Brown, R. S.; Argos, P. *Nature* **1986**, *324*, 215.

(38) Guss, J. M.; Harrowell, P. R.; Murata, M.; Norris, V. A.; Freeman, H. A. *J. Mol. Biol.* **1986**, *192*, 361.

(39) Gruff, E. S.; Koch, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8762.

and characterized.[40] This peptide appears to bind Co^{2+} in a tetrahedral site, with water as the fourth ligand, supporting the plausibility of this sort of structure.

Finally, examination of the sequence data base revealed that three of the metal-binding residues appeared to be completely invariant but the last histidine was occasionally absent with a cysteine appearing in a nearby position. Each of the domains that lack the final histidine is the last one of a tandem array, and the spacings between the histidine and the cysteine are variable. No data have been reported to date concerning whether these domains bind metal ions. To test the hypothesis that a cysteine residue in this region is capable of acting as a ligand, a peptide in which the final histidine in the consensus sequence was replaced by cysteine was synthesized. This peptide (CP-1-H24C) binds metal ions such as Co^{2+} and Zn^{2+}. The absorption spectrum of the Co^{2+} complex is shown in Figure 7. The assignment of the chromophore as $Co(Cys)_3(His)$ is supported by examination of the spectrum of the Co^{2+} complex of a peptide derived from a retroviral nucleocapsid protein that contains an otherwise unrelated metal-binding sequence of the form Cys-X_2-Cys-X_4-His-X_4-Cys, also shown in Figure 7.[41] The spectra are very similar in both the ligand-field and charge-transfer regions. For CP-1-H24C-Co^{2+}, the $^4A_2$ to $^4T_1(P)$ ligand field transition in the visible region is clearly resolved into its three components in a manner similar to that observed for certain synthetic tetrahedral Co^{2+} complexes.[42] Furthermore, this band is red-shifted relative to the parent complex, as would be expected given the positions of thiolate and imidazole in the spectrochemical series.[42,43] The metal ion affinity of CP-1-H24C is comparable to that of the parent peptide, indicating that ligand substitution in this position occurs without a large change in overall metal-binding plus folding energy. In contrast, a His to Cys sequence change resulted in a 10-fold increase in dissociation constant for a retroviral nucleocapsid protein-derived peptide.[44] The availability of zinc finger domains with chemically and spectroscopically differentiated metal-binding sites will be useful for testing models for metal ion binding specificity and for studies of peptides and proteins containing multiple zinc finger domains.

Acknowledgment. Financial support for this work has been provided by the National Institutes of Health (Grant GM-38230), the National Science Foundation (Grant DMB-8850069), and the Chicago Community Trust/Searle Scholars Program. We thank Dr. Joseph Vaughn for assistance in recording some of the NMR spectra. NMR spectrometers were purchased with assistance from grants from the National Institutes of Health (RR-01934, RR-03518) and the National Science Foundation (PCM 83-03176). B.T.A. is the recipient of a fellowship from the Institute for Biophysical Studies of Macromolecular Assemblies supported by the National Science Foundation and V.J.K. is the recipient of a scholarship from the Fulbright Foundation.

Registry No. CP-1, 133551-05-0; [Cys^{24}]-CP-1, 133551-06-1; Co, 7440-48-4; Zn, 7440-66-6.

(40) Merkle, D. L.; Schmidt, M. H.; Berg, J. M. *J. Am. Chem. Soc.*, in press.
(41) Green, L. M.; Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 4047.
(42) Lane, R. W.; Ibers, J. A.; Frankel, R. B.; Papaefthymiou, G. C.; Holm, R. H. *J. Am. Chem. Soc.* **1977**, *99*, 84.
(43) Davis, W. J.; Smith, J. *J. Chem. Soc. A* **1971**, 317.
(44) Green, L. M. Ph.D. Thesis, The Johns Hopkins University, 1990.

Low-Temperature Crystal Structure of Superoxocobalamin Obtained by Solid-State Oxygenation of the B_{12} Derivative Cob(II)alamin†

Erhard Hohenester,‡ Christoph Kratky,*,‡ and Bernhard Kräutler§

Contribution from the Institut für Physikalische Chemie, Universität Graz, A-8010 Graz, Austria, and Laboratorium für Organische Chemie, Eidgenössische Technische Hochschule (ETH), CH-8092 Zürich, Switzerland. *Received September 17, 1990*

Abstract: Exposure of a large (0.3 × 0.3 × 1.0 mm^3) "wet" single crystal of cob(II)alamin (B_{12r}, **1**) to gaseous dioxygen (10 atm, 4 °C, 2 days) leads to the formation of the dioxygen adduct $B_{12r}O_2$ (**2**) within the crystal. This solid-state reaction proceeds without affecting the crystalline order, permitting a low-temperature single-crystal structure analysis of $B_{12r}O_2$ (**2**) after rapid cooling of the oxygenated crystal to 96 K. Crystals of $B_{12r}O_2$ (**2**) at 96 K are orthorhombic: space group $P2_12_12_1$, a = 15.749 (13) Å, b = 21.673 (19) Å, c = 25.916 (20) Å, with 4 formula units ($C_{62}H_{88}N_{13}O_{16}PCo$) plus 68 water and 8 acetone molecules per unit cell. The crystallographic refinement (4070 significant observations, 534 parameters) converged at R = 0.095 and R_w = 0.078 and yielded a degree of oxygenation of 0.7. The $B_{12r}O_2$ (**2**) crystal structure is isomorphous to the one of B_{12r} (**1**), but oxygenation leads to structural changes (binding of O_2, upward shift of the cobalt center and of the benzimidazole base) in the cobalamin molecule and to extensive rearrangements in the solvent region. The dioxygen molecule is attached to the metal center in a bent end-on fashion at the β-face of the cobalamin molecule, and it is observed in a single conformation pointing approximately toward C(10) of the corrin ring. All structural evidence is consistent with a formulation of $B_{12r}O_2$ (**2**) as superoxocob(III)alamin. The cobalt-coordinated dioxygen forms hydrogen bonds to two water molecules. The whole solvent region is observed to be fully ordered at 96 K.

Introduction

At low temperature in methanolic solution, the vitamin B_{12} derivative cob(II)alamin (B_{12r}, **1**) is reversibly oxygenated according to Bayston et al.[1] to form what has been termed "superoxocobalamin" ($B_{12r}O_2$, **2**). This species, which is also obtained by the reaction of aquocob(III)alamin (**3**) with superoxide,[2] is assumed to be the first intermediate in the autoxidation

† Dedicated to Prof. A. Eschenmoser on the occasion of his 65th birthday.
‡ Universität Graz.
§ ETH.

(1) Bayston, J. H.; King, N. K.; Looney, F. D.; Winfield, M. E. *J. Am. Chem. Soc.* **1969**, *91*, 2755.
(2) Ellis, J.; Pratt, M. J.; Green, M. *J. Chem. Soc., Chem. Commun.* **1973**, 781.

0002-7863/91/1513-4523$02.50/0