# Determination of Nonligand Amino Acids Critical to $[4Fe\text{-}4S]^{2+/+}$ Assembly in Ferredoxin Maquettes†

Stephen E. Mulholland, Brian R. Gibney, Francesc Rabanal,‡ and P. Leslie Dutton*

*The Johnson Research Foundation, Department of Biochemistry and Biophysics, and School of Medicine, The University of Pennsylvania, Philadelphia, Pennsylvania 19104*

*Received April 15, 1999; Revised Manuscript Received June 16, 1999*

ABSTRACT: The prototype ferredoxin maquette, FdM, is a 16-amino acid peptide which efficiently incorporates a single $[4Fe\text{-}4S]^{2+/+}$ cluster with spectroscopic and electrochemical properties that are typical of natural bacterial ferredoxins. Using this synthetic protein scaffold, we have investigated the role of the nonliganding amino acids in the assembly of the iron−sulfur cluster. In a stepwise fashion, we truncated FdM to a seven-amino acid peptide, FdM-7, which incorporates a cluster spectroscopically identical to FdM but in lower yield, 29% relative to FdM. FdM-7 consists solely of the •CIACGAC• consensus ferredoxin core motif observed in natural protein sequences. Initially, all of the nonliganding amino acids were substituted for either glycine, FdM-7-PolyGly (•CGGCGGC•), or alanine, FdM-7-PolyAla (•CAACAAC•), on the basis of analysis of natural ferredoxin sequences. Both FdM-7-PolyGly and FdM-7-PolyAla incorporated little $[4Fe\text{-}4S]^{2+/+}$ cluster, 6 and 7%, respectively. A systematic study of the incorporation of a single isoleucine into each of the four nonliganding positions indicated that placement either in the second or in the sixth core motif positions, •CIGCGGC• or •CGGCGIC•, restored the iron−sulfur cluster binding capacity of the peptides to the level of FdM-7. Incorporation of an isoleucine into the fifth position, •CGGCIGC•, which in natural ferredoxins is predominately occupied by a glycine, resulted in a loss of [4Fe-4S] affinity. The substitution of leucine, tryptophan, and arginine into the second core motif position illustrated the stabilization of the [4Fe-4S] cluster by bulky hydrophobic amino acids. Furthermore, the incorporation of a single isoleucine into the second core motif position in a 16-amino acid ferredoxin maquette resulted in a 5-fold increase in the level of [4Fe-4S] cluster binding relative to that of the glycine variant. The protein design rules derived from this study are fully consistent with those derived from natural ferredoxin sequence analysis, suggesting they are applicable to both the de novo design and structure-based redesign of natural proteins.

Natural proteins assume complex tertiary and quaternary structures whose determinants are encrypted in their deceptively simple amino acid sequences (*1*). Also encoded in the primary structure are the factors governing the incorporation of cofactors and construction of active sites which complicates analysis of protein structure−function relationships. To better understand the fundamental relationship between primary structure and biological function, we are employing the use of peptide maquettes (*2*), simplified peptide architectures derived from the minimalist hierarchical approach pioneered by DeGrado and co-workers (*3*−*7*). Maquettes are peptide-based synthetic analogues (*8*, *9*) which provide a constructive approach for addressing structural, functional, or biochemical questions that are relevant to more complex natural proteins. Maquettes potentially offer more direct insight into structure−function relationships by minimizing the complications due to issues not intimately involved in the protein aspect being studied.

We have used maquettes to bind a tetranuclear iron−sulfur cluster (*10*), a structure that occurs in many diverse families of proteins found in the Archae, bacteria, and eukaryotes (*11*−*16*). Using insight drawn from natural ferredoxin structures (*17*) and sequences (*18*) (Figure 1), we have designed a prototype 16-amino acid ferredoxin maquette, FdM, which binds a single [4Fe-4S] cluster with spectral and electrochemical characteristics consistent with natural $[4Fe\text{-}4S]^{2+/1+}$ ferredoxins. The minimal sequence of FdM, KLCEGG•CIACGAC•GGW, contains the consensus motif of •CIACGAC• common to bacterial ferredoxins. The modular nature of FdM was illustrated by its incorporation into the design of a complex oxidoreductase maquette containing hemes and [4Fe-4S] clusters (*10*). Recently, we have validated the use of the FdM based on the *Peptococcus aerogenes* sequence as a template for studying [4Fe-4S] assembly within maquettes and utilized FdM to delineate the fundamental ligand requirements for efficient [4Fe-4S] cluster assembly within a maquette scaffold (*19*).

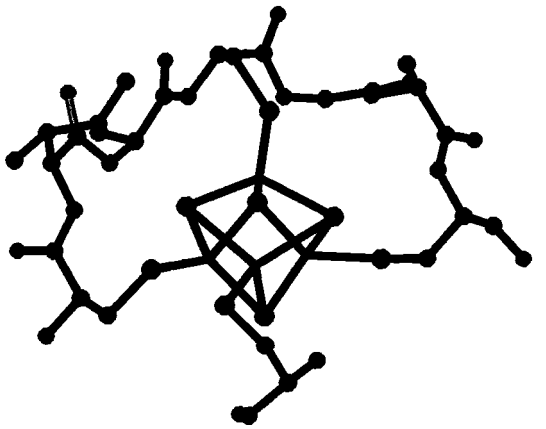
Herein, we address the role of the nonliganding amino acids within the •CXXCXXC• core consensus motif in the assembly of a [4Fe-4S] cluster by altering the intervening

† This work was supported by the National Institutes of Health Grant GM41048, and in part by the NSF MRSEC Program (DMR96-32598).

* To whom correspondence should be addressed: B501 Richards Bldg., Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104. E-mail: dutton@mail.med.upenn.edu. World Wide Web: http://www.med.upenn.edu/biocbiop/faculty/pages/dutton.html.

‡ Present address: Department de Quimica Organica, Universitat de Barcelona, Marti I franques, 1-11, 080028 Barcelona, Spain.

10.1021/bi9908742 CCC: $18.00 © 1999 American Chemical Society
Published on Web 07/22/1999



*Peptococcus aerogenes* S•CIACGAC•K...C
*Bacillus subtilis* T•CIACGAC•G...C
*Clostridium thermocellum* A•CISCGAC•E...C

Ferredoxin Maquette KLCEGG•CIACGAC•GGW

FIGURE 1: Molecular model of the [4Fe-4S] binding site in *P. aerogenes* FdI (PDB file 1fdx) with sequence alignments of several natural ferredoxins and the prototype ferredoxin maquette, FdM. The consensus ferredoxin core motif •CIACGAC• is represented as well as the fourth cysteine ligand that is remote in the primary structure facing the viewer.

amino acid sequence while keeping the number and placement of cysteine residues constant. The 16-amino acid FdM, shown in Figure 1, has been used to derive a minimal sequence competent for [4Fe-4S] cluster assembly, resulting in a truncated seven-amino acid version, FdM-7, containing only the core motif, •CIACGAC•. Amino acid substitutions within the FdM-7 framework demonstrate the preference for bulky amino acids at the second position (Ile) and small amino acids at the fifth position (Gly and Ala). These results are consistent with the observed frequency of amino acids within related natural ferredoxin protein sequences. Furthermore, incorporation of an isoleucine into a 16-amino acid maquette demonstrates information transfer between related maquette scaffolds. The results of this study yield insight into how residues which are not directly involved in cluster ligation play critical roles in maintaining competent cofactor binding sites.

## EXPERIMENTAL PROCEDURES

*Materials and Methods.* The peptides were synthesized on a continuous flow Milligen 9050 or PerSeptive Biosystems Pioneer solid phase synthesizer (Framingham, MA) using the standard Fmoc/$^t$Bu protection strategy (*20*) and purified to homogeneity as previously described (*21*). The identity of each purified protein was confirmed by MALDI-TOF mass spectrometry in the apo state. Solution molecular weight analysis was performed by gel permeation chromatography as previously described (*19*). Quantitative electron paramagnetic resonance (EPR) spectroscopy was performed using a Bruker ESP300E spectrometer operating at X-band frequencies as previously reported (*10*).

*[4Fe-4S]$^{2+/1+}$ Assembly.* The standard FdM reconstitution protocol based on modified biochemical literature methods (*22*, *23*) was utilized to incorporate the [4Fe-4S] cluster into each of the designed ferredoxin maquettes at a peptide concentration of 100 $\mu$M, determined by Trp absorbance or by peptide mass, as previously reported (*19*). Cluster

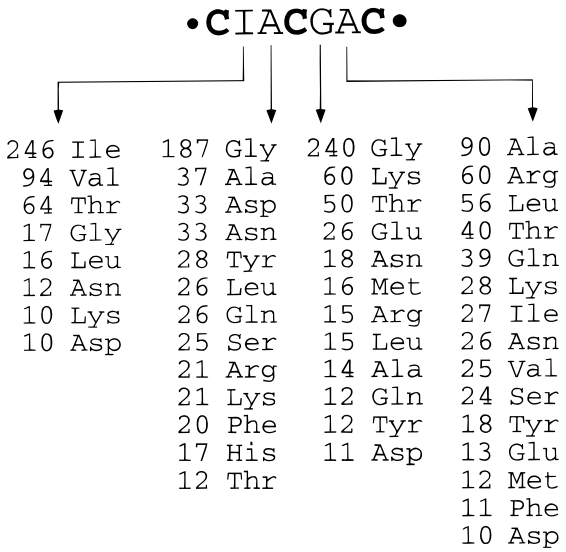



FIGURE 2: Representation of the results of the sequence variability anaylsis of the consensus ferredoxin core motif used in the design of FdM, •CIACGAC•. Shown below each nonliganding position is the number of times each amino acid appeared in the 510 natural [4Fe-4S] proteins that were analyzed.

incorporation yields were determined by spin quantitation of the sodium dithionite-reduced [4Fe-4S] electron paramagnetic resonance (EPR) signal relative to the reduced EPR signal of the prototype FdM reconstituted at the same time as a control. Additionally, concentrated solutions of peptide (>200 $\mu$M) were reconstituted under $FeCl_3$ and $Na_2S$ limiting conditions, i.e., <0.25 equiv of [4Fe-4S] per peptide, affording comparable cluster incorporation yields. Following [4Fe-4S] incorporation, the ferredoxin maquettes were reduced with sodium dithionite and stored frozen in liquid nitrogen prior to EPR spin quantitation. Successive reconstitutions using both methods, peptide or cluster limited, suggest an error of no greater than 10% in the given values.

## RESULTS

*Analysis of the Sequence Variability of Natural Iron–Sulfur Protein Structure.* The sequence of the consensus ferredoxin motif, •CIACGAC•, used for ferredoxin maquette design was compared to natural [4Fe-4S] protein sequences in the Swiss protein database (*24*). Using a locally written Mathematica program, 510 •CXXCXXC• core motifs, from a total of 664 [4Fe-4S] protein sequences, were analyzed for amino acid variation in the four nonliganding amino acid positions, providing insight into natural ferredoxin design. Figure 2 illustrates the results of the sequence variability search for each of the four nonliganding positions of the core motif. The second position, •C**I**ACGAC•, is dominated by the presence of $\beta$-branched amino acids, Ile,Val, or Thr, in 404 of the 510 motifs. The third position, •CI**A**CGAC•, is dominated by glycine (187 of 510) and tolerates all amino acids except proline. Stricter engineering specifications are observed in the fifth position, •CIAC**G**AC•, which is Gly for 240 of the proteins with Lys (60) and Thr (50) also being common with an apparent absence of Phe, Ile, and Pro. The prevalence of Gly at position 5 may reflect the preference for an amino acid capable of adopting a wide range of dihedral angles. The sixth position, •CIACG**A**C•, is rather tolerant with all 20 amino acids observed in a more even

distribution than those of the other three positions. The nonrandom distribution of amino acid content at each position implicates the four nonliganding positions in cluster incorporation and stabilization. Additionally, the absence of amino acids at various positions suggests that some amino acids may be incompatible at certain positions with $[4Fe\text{-}4S]^{2+/+}$ incorporation.

*Analysis of Local Amino Acids in Natural Iron−Sulfur Protein Structure.* Besides the amino acids intervening between the cysteine ligands in the core motif, natural [4Fe-4S] clusters are locally sequestered by amino acids that are distant in the primary structure. Recent analysis of second-shell amino acids (the metal ligands define the first coordination shell) has provided insight into the design of mononuclear proteins as well as complex oxidoreductases by demonstrating that the second shell is often more hydrophilic than previously believed (*25*−*27*). Using a locally written Mathematica program, we evaluated the distribution of amino acids within 5 Å of the metal ions of the [4Fe-4S] clusters in structurally characterized natural ferredoxin and high-potential iron proteins (HiPIPs). The results of each of the nine ferredoxin protein structures (PDB files 1fca, 1fdn, 1fxr, 1vjw, 6fd1, 1aop, 6acn, 1nip, and 1blu) and six HiPIP proteins (PDB files 1hpi, 1hip, 2hip, 1neh, 1isu, and 1hrq) were tabulated. The combined data for each family of [4Fe-4S] proteins, ferredoxin $[4Fe\text{-}4S]^{2+/+}$ or HiPIP $[4Fe\text{-}4S]^{3+/2+}$, were compared to data for a hypothetical natural protein of equivalent size with a statistical distribution of amino acids derived from all the proteins in the Brookhaven Protein Data Bank. The data presented in Figure 3 illustrate a clear differential between ferredoxin and high-potential iron sulfur [4Fe-4S] environments. The ferredoxins exhibit a preponderance of Ala, Gly, Ile, and Val which are all contained in the core motif, consistent with the Swiss protein database analysis above. Additionally, there is a prevalence of proline in ferredoxin [4Fe-4S] second shells due to the propensity of the fourth ligating cysteine, not part of the core motif, to be adjacent to a proline which terminates a helical region. The data for the HiPIP proteins are clearly distinguishable due to the presence of large numbers of aromatic amino acids in the second shell, Phe, Tyr, and Trp, which are known to stabilize the oxidized HiPIP from nucleophilic cluster degradation (*28*−*30*). The differential in the distribution of second-shell amino acids may provide a structural basis for determining the type of tetranuclear iron−sulfur protein, ferredoxin or HiPIP, by tuning the reduction potential of the two redox couples, i.e., $[4Fe\text{-}4S]^{2+/+}$ and $[4Fe\text{-}4S]^{3+/2+}$, by altering the solvent accessibility (*31*, *32*).

*FdM Amino Acid Truncation Defines the Minimal Ferredoxin Maquette.* Previous studies with the 16-amino acid FdM peptide had shown that peptides with less than the full complement of four cysteines could successfully bind a [4Fe-4S] cluster with standard spectroscopic and electrochemical properties, suggesting that FdM could be minimized. The peptide:cluster stoichiometry remained 1:1 after the removal of a single cysteine, but it rose to 2:1 when two cysteines were removed from the peptide sequence. Figure 4 shows that truncation of the FdM peptide sequence to 10 amino acids, FdM-10, with the requisite deletion of one cysteine ligands, yielded an $[4Fe\text{-}4S]^{2+/+}$ protein with a lowered reconstitution efficiency, 32%, relative to FdM, but with spectroscopic (reduced state EPR *g* values, temperature

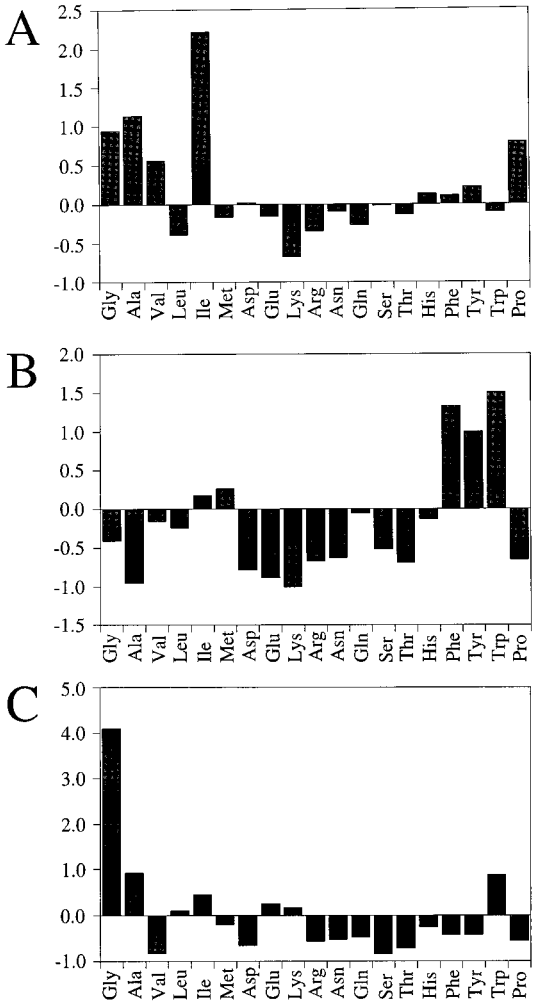



FIGURE 3: Histogram representations of the second-shell amino acid distribuition local to structurally characterized [4Fe-4S] binding sites in natural (A) ferredoxins and (B) high-potential iron proteins with (C) the prototype ferredoxin maquette shown for comparison. The data are presented as the difference between the family of [4Fe-4S] proteins and a hypothetical protein with the identical length and statistical composition. As such, positive values represent a higher content of a particular amino acid in the [4Fe-4S] protein relative to a typical PDB protein.

dependence, $P_{1/2}$ values, and oxidized state UV−visible $\lambda_{max}$) and electrochemical ($E_{m8}$ of −350 mV) properties identical to those of the prototype (*10*) and the $\alpha_4$-FeS designed by Scott and Biggins (*33*). Additionally, gel permeation chromatography demonstrated that FdM-10 and all the other peptides studied are monomeric in solution; however, during solution molecular weight studies, no UV−vis evidence for cluster incorporation was observed for five peptides, FdM-6, FdM-4, FdM-7-PolyAla, FdM-7-PolyGly, and FdM-7-I3, consistent with their low reconstitution efficiencies or the loss of cluster on the column. The standard spectroscopic properties of the ferredoxin maquettes, shown in Figure 4 for comparison to panels A and B of Figure 5 in ref *10*, allow for direct comparison of reconstitution yields by quantitative reduced state EPR spectroscopy in a standard assay, enumerated in Table 1, and most likely reflect the inability of these minimal peptides to influence the structure of the robust tetranuclear cluster. The FdM-10 yield is significantly lower than the 56% yield observed for FdM-1A, a 16-amino acid FdM variant in which the first cysteine is replaced with an alanine, demonstrating that removal of

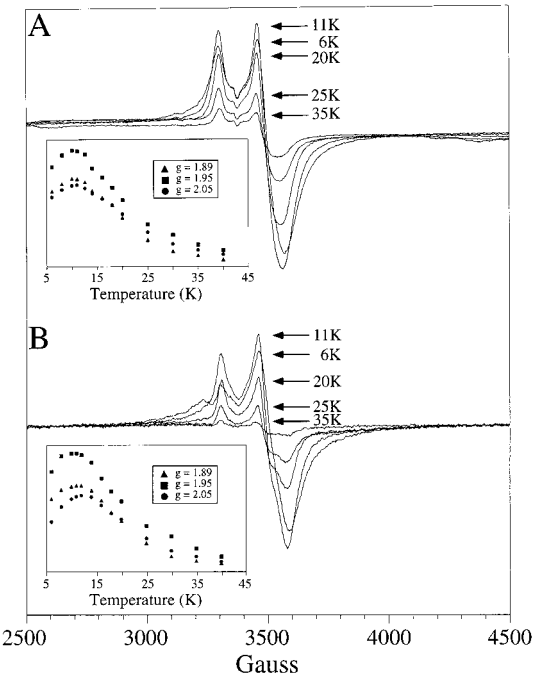



FIGURE 4: Electron paramagnetic resonance (EPR) spectroscopy characterization of (A) the 10-amino acid ferredoxin maquette, FdM-10, and (B) the seven-amino acid ferredoxin maquette, FdM-7. The insets show the temperature dependence of the reduced FdM EPR signal extrema between 6 and 40 K at a 1 mW incident microwave power which are indicative of a tetranuclear iron−sulfur cluster.

six nonligating amino acids results in some loss of reconstitution efficiency.

Reduction of the FdM-10 sequence to the consensus core motif observed in natural ferredoxins, •CIACGAC• FdM-7, did not affect the reconstitution efficiency, 29%, the EPR spectrum, or reduction potential. Removal a single cysteine from FdM-7 yielded FdM-6 which contained only two cysteine residues. FdM-6 exhibited a significantly lower yield, 4%, which was barely competent for reconstitution. On the basis of previous results with ferredoxin maquettes containing only two cysteine residues, FdM-6 is assumed to have a 2:1 peptide:cluster stoichiometry. As noted above, reconstituted FdM-6 eluted as a monomeric peptide devoid of cluster during molecular weight studies of the holo-maquette. The further minimization of FdM-6 to a simple four-amino acid sequence, FdM-4, resulted in a complete loss of cluster assembly. These data indicate that minimum ferredoxin maquette sequence competent for partial [4Fe-4S] cluster incorporation is the heptapeptide core motif from natural ferredoxins, •CIACGAC• FdM-7.

*Intervening Amino Acid Substitutions in FdM-7.* The observation that a glycine rich variant, FdM-PolyGly, of the prototype ferredoxin maquette had a dramatically lowered yield, 11% relative to FdM, indicated that the nonligating residues contained information critical to [4Fe-4S] cluster assembly and stabilization. FdM-7 was chosen as a template for studying the role of nonligating amino acids because it was smaller and monomeric and it has a marginal reconstitution efficiency, 29%, which allows the quantification of both improvements and failures. Additionally, we wanted to use the results of the seven-amino acid ferredoxin maquette, FdM-7, studies to improve the binding efficiency of FdM-PolyGly, a 16-amino acid ferredoxin maquette scaffold. We

Table 1: Sequences and [4Fe-4S]$^{2+/+}$ Cluster Incorporation Efficiencies for the Iron−Sulfur Maquettes in This Study[a]

| Maquette | Sequence | [Spins]/[Peptide] |
|---|---|---|
| FdM | KLCEGG•CIACGAC•GGW | 100 |
| FdM-10 | G•CIACGAC•GW | 32 |
| FdM-7 | •CIACGAC• | 29 |
| FdM-6 | GCIACG | 4 |
| FdM-4 | CIAC | 0 |
| FdM-1A | KLAEGG•CIACGAC•GGW | 56 |
| FdM-Gly | GGCGGG•CGGCGGC•GGW | 11 |
| FdM-Gly-I | GGCGGG•CIGCGGC•GGW | 61 |
| FdM-7-Ala | •CAACAAC• | 7 |
| FdM-7-Gly | •CGGCGGC• | 6 |
| FdM-7-I | •CIGCGGC• | 28 |
| FdM-7-I2 | •CGICGGC• | 15 |
| FdM-7-I3 | •CGGCIGC• | 0 |
| FdM-7-I4 | •CGGCGIC• | 27 |
| FdM-7-L | •CLGCGGC• | 24 |
| FdM-7-W | •CWGCGGC• | 26 |
| FdM-7-R | •CRGCGGC• | 14 |

[a] The core •CIACGAC• motif taken from *P. aerogenes* is between the dots. All of the peptides are C-terminally amidated.

chose to substitute the nonligating amino acids with either glycine or alanine, since these were the most prominent in the natural sequence analysis. Initially, we replaced all of the nonligating residues of FdM-7 with alanine, FdM-7-PolyAla, or glycine, FdM-7-PolyGly, to evaluate their role in conferring cluster binding competency to the peptide. Despite maintaining the number and relative placement of the ligating cysteine residues, substitution of the intervening positions with alanine or glycine dramatically lowered the cluster binding efficiencies which are fully consistent with the observation with the 16-amino acid ferredoxin maquettes. These results clearly demonstrate that the nonliganding residues are critical to the assembly of the [4Fe-4S] cluster.

*Role of Isoleucine Incorporation in FdM-7.* Starting from the FdM-7-PolyGly sequence, in which all the nonligating residues had been replaced by glycines, a single isoleucine residue was placed in each of the four intervening positions to delineate the potential role of isoleucine on cluster assembly. Incorporation into the second position, FdM-7-I2, fully restored the cluster binding efficiency to the 28% that was evident with the parent FdM-7 design. Placing an isoleucine in the third position, FdM-7-I3, resulted in a peptide that bound a [4Fe-4S] about half as well as FdM-7 or FdM-7-I2. When the isoleucine was placed in the fifth nonligating position, FdM-I5, the peptide lost all ability to stabilize the iron−sulfur cluster, consistent with the sequence variability results that show glycine strictly predominates in this position with isoleucine absent. Last, substitution of an isoleucine at the sixth position restored cluster binding ability, 27%, to a level similar to that measured for FdM-7 and FdM-7-I2. These data demonstrate clearly the positive influence of an Ile residue at position 2 or 6 in the construction of ferredoxin maquettes. Additionally, these data illustrate the intricacies of the fundamental engineering specifications for [4Fe-4S] protein design since all four peptides that were

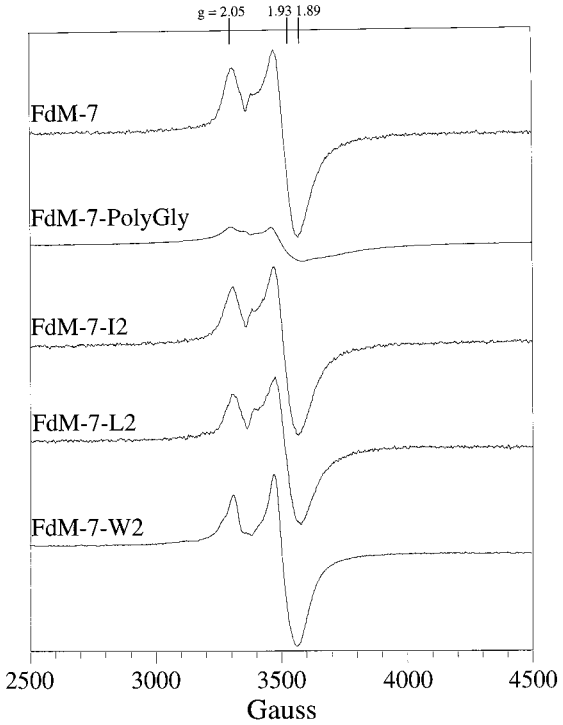



FIGURE 5: EPR characterization of several seven-amino acid ferredoxin maquettes. Incorporation of a single bulky amino acid, Ile, Leu, or Trp, into the second position of •CGGCGGC•, FdM-7-PolyGly, results in cluster reconstitution similar to that of FdM-7. EPR spectral acquisition parameters are given in Experimental Procedures. Peptide sequences are given in Table 1.

studied have identical amino acid composition, ligand quantity, type, and placement but differ only in the relative placement of a single bulky $\beta$-branched amino acid.

*Side Chain Steric Bulk at Position 2.* The difference in the reconstitution efficiencies of FdM-7-I2 and FdM-7-PolyGly demonstrates that single, conservative amino acid changes can have a significant influence on cluster incorporation. Figure 5 shows that addition of a bulky side chain, leucine, tryptophan, or arginine, at position 2 affects the cluster binding efficiency of the peptide. Substitution of the bulky hydrophobic residues, leucine or tryptophan, produced peptides whose cluster binding efficiencies were roughly equivalent to those of FdM-7, 24% for FdM-7-L2 and 26% for FdM-7-W2. Alternatively, Ile replacement with the large polar side chain of arginine only partially restored some of the cluster binding capacity, 14% for FdM-7-R2, half of that of the original FdM-7 design. These ferredoxin maquettes underscore the necessity for a bulky hydrophobic amino acid in the second position of the natural core consensus motif.

*Redesign of the Full-Size 16-Amino Acid FdM-PolyGly.* The results of the nonligand amino acid substitutions of FdM-7 illustrate a preference for an isoleucine at position 2 and a glycine at position 5 of the core motif. These results were incorporated into the 16-amino acid FdM-PolyGly to test the transferability of the results from one maquette scaffold to another related scaffold. Figure 6 shows that the resulting FdM-PolyGly-I2 peptide has a 5-fold higher reconstitution efficiency, 61% compared to 11% for FdM-PolyGly, clearly demonstrating that the isoleucine in position 2 of the core motif is critical to [4Fe-4S] assembly regardless of the maquette scaffold.

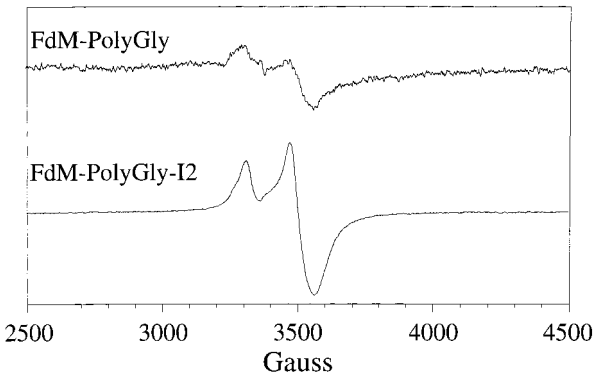



FIGURE 6: Effect of introduction of a single isoleucine at motif position 2 in FdM-Gly as evidenced by reduced state EPR spectroscopy. Placing an Ile into the second core motif position of the 16-amino acid FdM-PolyGly maquette increases the reconstitution yield from 11 to 61%. The spectrum of FdM-PolyGly is expanded 5-fold for clarity.

## DISCUSSION

In this study, we have used a series of peptide maquettes to study the fundamental nonligand amino acid requirements for $[4Fe\text{-}4S]^{2+/+}$ binding sites. Starting from a 16-amino acid peptide that quantitatively binds a single [4Fe-4S] cluster via four cysteine ligands and possesses spectroscopic and redox characteristics consistent with natural ferredoxins, progressive reduction of the peptide size yielded the smallest possible peptide that could bind a characterizable iron−sulfur cluster. From the resultant seven-amino acid FdM-7, systematic alteration the four remaining nonligating residues has allowed us to focus on the structural status of the consensus sequence for maintaining the binding competency of the site. The understanding of binding site engineering specification and tolerances is critical not only for the design of complex oxidoreductase maquettes which utilize FdM as a design module but also for a fundamental understanding of the factors which control natural [4Fe-4S] protein structure and function.

The small size of our original FdM design, a 16-amino acid maquette that binds a [4Fe-4S] cluster, is critical for modular metalloprotein design purposes (*34*, *35*). The FdM peptide was intended to be an independent module for integration into larger peptide structures for constructing complex oxidoreductase maquettes. To this end, it is somewhat important to establish the minimal structural unit that can efficiently incorporate the $[4Fe\text{-}4S]^{2+/+}$ cluster. These and previous results clearly illustrate that the prototype FdM efficiently binds a single cluster and is small enough to be incorporated into larger maquette structures without large changes in the [4Fe-4S] cluster properties. While the shorter ferredoxin maquettes studied do incorporate variable amounts of [4Fe-4S] clusters, their low yields obviate their general utility as design modules. However, the consistent cluster nuclearity, EPR spectral characteristics, and redox properties of the peptide-bound $[4Fe\text{-}4S]^{2+/+}$ clusters clearly illustrate the inherent predictability of some key properties of the cofactor in these minimal maquettes. Thus, the prototype ferredoxin maquette, FdM, is validated as a transferable protein design module which provides a low-potential [4Fe-4S] binding site with predictable spectroscopic and electrochemical properties.

Comparison of the ferredoxin maquette sequences with natural ferredoxins and high-potential iron proteins (HiPIPs) suggests that nonliganding amino acids are involved in cofactor redox modulation. The second shell of the prototype ferredoxin maquette, FdM, correlates well with the natural ferredoxins, which is not surprising since FdM was based on a consensus motif present in *P. aerogenes* FdI. However, the second shell of HiPIPs is significantly different from that of ferredoxins containing a high proportion of aromatic amino acids. The distribution of local amino acids potentially provides a structural basis for the rational design of future HiPIP maquettes such as that designed by Hellinga and Caradonna (*36*) by alteration of the second-shell amino acid content as a high-potential module electron acceptor for photoinduced electron transfer maquette design (*37*).

Producing peptides where the cysteines were spaced appropriately using simply alanines or glycines as spacers produced peptides that bound the iron−sulfur cluster very poorly due to the loss of the information contained in the consensus motif. The observation that both •CGGCGGC• and •CAACAAC• peptides bound [4Fe-4S] poorly prompted us to investigate the natural proteins whose sequences were homologous to the sequences of these peptide. The glycine peptide, •CGGCGGC•, is similar in sequence to a variety of metallothionen proteins (*38*) as well as MST87P which contains multiple •CGP• repeats (*39*). The low reconstitution efficiency of the glycine rich peptides with cysteines employed by us (*10*) and others (*40*) as iron−sulfur ligands is consistent with the Fasta3 sequence search (*41*) results which demonstrate that these sequences are most commonly associated with Zn(II) binding sites in biochemical systems. The alanine peptide, •CAACAAC•, is close in sequence with only seven proteins, including the methyl-CPG binding protein MBD1 (*42*). The only iron−sulfur protein with which FdM-7-PolyAla exhibits sequence similarity is *Azotobacter vinelandii* FixP which contains a putative •CTACAAC• core motif (Swiss protein database file Q44508). These results indicate that the nonintervening amino acids may play a role in metal ion selectivity perhaps by imposing steric restrictions on the size of the metal chelate by altering the ligand geometries.

Placing bulky hydrophobic amino acids in the second position in the core motif improved the reconstitution efficiency of the resultant peptides. While the natural proteins predominately used Ile, Val, and Thr for this core motif position, we observed that Leu and Trp were also competent for aiding cluster assembly and stabilization. As shown in Figure 2, Leu is occasionally found in position 2, and a Fasta3 search found a recent report of an iron−sulfur protein with a Trp at position 2, the •CWTCGAC• motif of indolepyruvate ferredoxin oxidoreductase (*43*). Additionally, arginine in the second position had a positive effect on cluster efficiency, but less that of the hydrophobes that were studied. The influence of single bulky hydrophobic residues on [4Fe-4S] binding efficiency is not limited to the seven-amino acid maquettes as it can be transferred to the 16-amino acid ferredoxin maquette scaffold. Incorporation of a single isoleucine into the 16-amino acid FdM-PolyGly design increased the yield 5-fold to 61%. Since $\beta$-branched amino acids have been shown to introduce rigidity into peptides (*44*, *45*) which aided the NMR structural determination of an apo four-helix bundle maquette scaffold (*46*, *47*), the presence of Ile or Val may serve to restrict the beginning of the core motif in a favorable geometry for [4Fe-4S] binding.

Consistent with the strict engineering specifications observed in the consensus sequence variability results, substitution of the glycine at core motif position 5 with an isoleucine resulted in a complete loss of [4Fe-4S] reconstitution yield. The two analogous glycines in *P. aerogenes*, Gly 12 and Gly 41, have $\phi$ angles of −81.9 and −78.1° and $\psi$ angles of −0.6 and 0.1° (*17*), respectively, which places them outside of the restricted Ramachandran range typically observed for amino acids with side chains. Clearly, the preference for a specific amino acid conformation determines the strict engineering specification at this motif position. This contrasts with the variability of the sixth motif position which can accommodate the majority of the natural amino acids and represents a site of protein plasticity.

Overall, the experimental results of the FdM-7 variants correlate well with the observed variability in the sequences of natural ferredoxins, illustrating the applicability of maquettes in deriving protein design rules in a constructive manner. This implies that the local sequence context of the core motif in natural ferredoxins, which is absent in the ferredoxin maquettes, is not the sole factor dictating the nuclearity of the iron−sulfur cluster but, as discussed previously, may modulate the cluster electrochemistry. Additionally, the results illustrate information transfer between a maquette scaffold and a natural protein. The protein design rules delineated from the seven-amino acid maquettes can be applied to the 16-amino acid ferredoxin maquettes; e.g., incorporation of the isoleucine into FdM-PolyGly increased the reconstitution efficiency by 50% (>5-fold), and are consistent with a sequence analysis of natural ferredoxins. The observed information transfer between different ferredoxin maquette scaffolds as well as natural proteins aids in the development of modular oxidoreductase maquettes since results with one maquette scaffold can in some cases be transplanted into another related maquette scaffold with predictable consequences.

In the process of validating the prototype ferredoxin maquette as a design module, we have delineated several ferredoxin protein design concepts. First, thiolate ligation is highly preferred over nitrogen and oxygen ligation (*48*, *49*), consistent with the relative rarity of nonthiolate ligation in [4Fe-4S] proteins (*50*−*53*). Second, only three cysteine residues are required for monomeric cluster ligation, in which case the fourth exogenous ligand may be hydroxide or $\beta$-mercaptoethanol. Third, only the consensus core motif, •CIACGAC•, is required for stabilization of the [4Fe-4S] cluster. Fourth, the spacing of the cysteine residues, especially in the •CXXCXXC• motif, is crucial for cluster ligation. Fifth, *the identities of the nonligating amino acids in FdM are as important as each of the metal ligands in assembling and stabilizing the* $[4Fe\text{-}4S]^{2+/+}$ *cluster*. These simple iron−sulfur protein design concepts derived from the constructive, biochemically relevant maquette approach correlate with natural ferredoxins, illustrating that maquettes provide an alternative route to understanding natural proteins just as natural proteins provide critical insight into the design of maquettes.

## ACKNOWLEDGMENT

We thank Julie A. Baker, Chris C. Page, and Professor Michael J. Therien for assistance in the assembly of the Mathematica program for analyzing second-shell amino acid content in [4Fe-4S] proteins. Mass spectroscopic analyses were performed by the Protein Chemistry Laboratory of the University of Pennsylvania.

## REFERENCES

1. Anfinsen, C. B. (1973) *Science 181*, 223−230.
2. Robertson, D. E., Farid, R. S., Moser, C. C., Urbauer, J. L., Mulholland, S. E., Pidikiti, R., Lear, J. D., Wand, A. J., DeGrado, W. F., and Dutton, P. L. (1994) *Nature 368*, 425−431.
3. Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. J., Zhou, H. X., O'Neil, K. T., and DeGrado, W. F. (1995) *Science 270*, 935−941.
4. DeGrado, W. F., Wasserman, Z. R., and Lear, J. D. (1989) *Science 243*, 622−628.
5. Dieckmann, G. R., McRorie, D. K., Tierney, D. L., Utschig, L. M., Singer, C. P., O'Halloran, T. V., Penner-Hahn, J. E., DeGrado, W. F., and Pecoraro, V. L. (1997) *J. Am. Chem. Soc. 119*, 6195−6196.
6. Dieckmann, G. R., McRorie, D. K., Lear, J. D., Sharp, K. A., DeGrado, W. F., and Pecoraro, V. L. (1998) *J. Mol. Biol. 280*, 897−912.
7. Choma, C. T., Lear, J. T., Nelson, M. J., Dutton, P. L., Robertson, D. E., and DeGrado, W. F. (1994) *J. Am. Chem. Soc. 116*, 856−865.
8. Ibers, J. A., and Holm, R. H. (1980) *Science 209*, 233−235.
9. Karlin, K. D. (1993) *Science 261*, 701−708.
10. Gibney, B. R., Mulholland, S. E., Rabanal, F., and Dutton, P. L. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 15041−15046.
11. Beinert, H., Holm, R. H., and Münck, E. (1997) *Science 277*, 653−659.
12. Beinert, H., Kennedy, M. C., and Stout, C. D. (1996) *Chem. Rev. 96*, 2335−2373.
13. Golinelli, M.-P., Chmiel, N. H., and David, S. S. (1999) *Biochemistry 38*, 6997−7007.
14. Ohnishi, T., and Salerno, J. C. (1982) in *Iron−Sulfur Proteins* (Spiro, T. G., Ed.) pp 285−328, John Wiley and Sons Inc., New York.
15. Hidalgo, E., and Demple, B. (1997) *EMBO J. 22*, 207−210.
16. Flint, D. H., and Allen, R. M. (1996) *Chem. Rev. 96*, 2315−2334.
17. Adman, E. T., Sieker, L. C., and Jensen, L. H. (1976) *J. Biol. Chem. 251*, 3801−3806.
18. Cammack, R. (1992) *Adv. Inorg. Chem. 38*, 281−322.
19. Mulholland, S. E., Gibney, B. R., Rabanal, F., and Dutton, P. L. (1998) *J. Am. Chem. Soc. 120*, 10296−10302.
20. Bodanszky, M. (1993) *Peptide Chemistry: A Practical Approach*, 2nd ed., Springer-Verlag, New York.
21. Gibney, B. R., Johansson, J. S., Rabanal, F., Skalicky, J. J., Wand, A. J., and Dutton, P. L. (1997) *Biochemistry 36*, 2798−2806.
22. Lovenberg, W., Buchanan, B. B., and Rabinowitz, J. C. (1963) *J. Biol. Chem. 238*, 3899.
23. Zhao, J., Li, N., Warren, P. V., Golbeck, J. H., and Bryant, D. A. (1992) *Biochemistry 31*, 5093.
24. Bairoch, A., and Apweiler, R. (1999) *Nucleic Acids Res. 27*, 49−54.
25. Karlin, S., Zhu, Z. Y., and Karlin, K. D. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 14225−14230.
26. Karlin, S., and Zhu, Z. Y. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 14231−14236.
27. Karlin, S., Zhu, Z. Y., and Karlin, K. D. (1998) *Biochemistry 37*, 17726−17734.
28. Bian, S. M., Hemann, C. F., Hille, R., and Cowan, J. A. (1996) *Biochemistry 35*, 14544−14552.
29. Agarwal, A., Li, D., and Cowan, J. A. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 9440−9444.
30. Cowan, J. A., and Lui, S. M. (1998) *Adv. Inorg. Chem. 45*, 313−350.
31. Jensen, G. M., Warshel, A., and Stephens, P. J. (1994) *Biochemistry 33*, 10911−10924.
32. Stephens, P. J., Jollie, D. R., and Warshel, A. (1996) *Chem. Rev. 96*, 2491−2513.
33. Scott, M. P., and Biggins, J. (1997) *Protein Sci. 6*, 340−346.
34. Gibney, B. R., Rabanal, F., and Dutton, P. L. (1997) *Curr. Opin. Chem. Biol. 1*, 537−542.
35. Lu, Y., and Valentine, J. S. (1997) *Curr. Opin. Chem. Biol. 7*, 495−500.
36. Coldren, C. D., Hellinga, H. W., and Caradonna, J. P. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 6634−6640.
37. Sharp, R. E., Rabanal, F., Moser, C. C., and Dutton, P. L. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 10465−10470.
38. Cousins, R. J. (1998) *Proc. Nutr. Soc. 57*, 307−311.
39. Kuhn, R., Schaefer, U., and Shaefer, M. (1988) *EMBO J. 7*, 447−454.
40. Que, L., Jr., Anglin, J. R., Bobrik, M. A., Davison, A., and Holm, R. H. (1974) *J. Am. Chem. Soc. 96*, 6042−6048.
41. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2444−2448.
42. Cross, S. H., Meehan, R. R., Nan, X., and Bird, A. (1997) *Nat. Genet. 16*, 256−259.
43. Siddiqui, M. A., Fujiwara, S., and Imanaka, T. (1997) *Mol. Gen. Genet. 254*, 433−439.
44. Creamer, T. P., and Rose, G. D. (1992) *Proc. Natl. Acad. Sci. U.S.A. 92*, 5937−5941.
45. Furukawa, K., Oda, M., and Nakamura, H. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 13583−13588.
46. Skalicky, J. J., Gibney, B. R., Rabanal, F., Bieber-Urbauer, R. J., Dutton, P. L., and Wand, A. J. (1999) *J. Am. Chem. Soc. 121*, 4941−4951.
47. Gibney, B. R., Rabanal, F., Skalicky, J. J., Wand, A. J., and Dutton, P. L. (1999) *J. Am. Chem. Soc. 121*, 4952−4960.
48. Shen, B., Jollie, D. R., Diller, T. C., Stout, C. D., Stephens, P. J., and Burgess, B. K. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 10064−10068.
49. Bobrik, M. A., Que, L., Jr., and Holm, R. H. (1974) *J. Am. Chem. Soc. 96*, 285−287.
50. Peters, J. W., Lanzilotta, W. N., Lemon, B. J., and Seefeldt, L. C. (1998) *Science 282*, 1853−1858.
51. Nicolet, Y., Piras, C., Legrand, P., Hatchikian, C. E., and Fontecilla-Camps, J. C. (1999) *Structure 7*, 13−23.
52. Volbeda, A., Charon, M. H., Piras, C., Hatchikiam, E. C., Frey, M., and Fontecilla-Camps, J. C. (1995) *Nature 373*, 580−587.
53. Calzolai, L., Gorst, C. M., Bren, K. L., Zhao, Z. H., Adams, M. W. W., and LaMar, G. N. (1997) *J. Am. Chem. Soc. 119*, 9341−9350.

BI9908742