FEBS Letters 458 (1999) 180–184 FEBS 22599

NMR characterization of the NADP(H)-binding domain of *Escherichia coli* transhydrogenase: sequential assignment and global fold

Carina Johansson[a], Anders Bergkvist[a], Ola Fjellström[a], Jan Rydström[a], B. Göran Karlsson[b,*]

[a] *Department of Biochemistry and Biophysics, Göteborg University, Göteborg, Sweden*
[b] *Department of Molecular biotechnology, Chalmers University of Technology, S-413 90 Göteborg, Sweden*

Received 13 July 1999

Abstract The soluble NADP(H)-binding domain of *Escherichia coli* transhydrogenase (186 amino acids, 20.4 kDa, rotational correlation time 14 ns) was characterized using NMR techniques. The global fold is similar to that of a classical dinucleotide-binding fold with six parallel β-strands in a central sheet surrounded by helices and irregular structures, but is lacking both αD and αE. The substrate is bound in an extended conformation at the C-terminal end of the parallel β-sheet and our data support the notion of a redox dependent structural rearrangement.

Key words: NMR; Transhydrogenase; Global fold; *Escherichia coli*

1. Introduction

Nicotinamide nucleotide transhydrogenase (EC 1.6.1.1) is a membrane-bound enzyme that utilizes the energy stored in the electrochemical gradient for the production of NADPH. The reduction of $NADP^+$ by NADH is coupled to proton translocation across the membrane (Fig. 1). Several different transhydrogenases have been cloned and ca. 23% of the amino acids are invariant. The enzyme consists of three domains, domain I and III that bind NAD(H) and NADP(H), respectively, and the membrane spanning domain II, through which protons are translocated. This organization is conserved between species, but the domains are distributed over one (e.g. bovine), two (e.g. *Escherichia coli*) or three (e.g. *Rhodospirillum rubrum*) polypeptide chains and a cyclic permutation is also found (e.g. *Eimeria tenella*). The N-terminal part of the protein harbors the NAD(H)-binding domain, ecI, consisting of residues α1–α404 (*E. coli* terminology), whereas the NADP(H)-binding domain, ecIII, is located in the C-terminal part, residues β286–β462.

Little structural information is known for transhydrogenase. A predicted model of ecI has been presented and it was suggested that NAD(H) binds to a classical dinucleotide-binding domain [1]. The structure of alanine dehydrogenase, a homologue of domain I of transhydrogenase, has recently been determined [2]. The hexameric quarternary structure could be regarded as a trimer of dimers and the monomeric unit consists of two structurally similar domains, one of which binds NAD(H) to a classical dinucleotide-binding fold. Domain III, on the other hand, shows no sequence similarity with known sequences in the databases and the putative dinucleotide-binding site is less obvious [3].

The soluble domains of *E. coli* transhydrogenase were previously characterized using a variety of biophysical techniques [4], including one dimensional NMR spectroscopy of domain I primarily from *R. rubrum*, but also from *E. coli* [5]. Mixtures of domain I and III from the same or different species display catalytic activities, indicating native-like structures [4,6–8].

In later years, the techniques for structure determination of macromolecules have made rapid progress. The combined use of uniform ^{13}C and ^{15}N-labelling and novel pulse sequences have extended the study of proteins up to the size of ca. 20–25 kDa and the use of deuteration to suppress relaxation has extended this range to 30 kDa and above. Recently, we [9] and others [10] used NMR techniques in the characterization of the NADP(H)-binding domain from *E. coli* and *R. rubrum*, respectively, and a secondary structure model of the *R. rubrum* domain was also reported [11].

In this work, we report on the NMR characterization of the NADP(H)-binding domain of *E. coli* transhydrogenase. We have used the above-mentioned techniques, combined with specific labelling experiments, to obtain essentially complete resonance assignments of the protein backbone and secondary structure determination of the 20.4 kDa NADP(H)-binding domain of *E. coli* transhydrogenase. Long-range NOEs identified a central β-sheet and observation of a limited number of methyl-methyl NOEs was used in the definition of the global fold. Our results are not in agreement with the previously published structure model [11], but differ both in core regions of the domain, as well as in the substrate-binding area.

2. Materials and methods

The C-terminal 177 residues of the protein (286–462) were subcloned in a T7 expression system [12] following a N-terminal Met-$(His)_6$-Ser-Ser-tag, expressed in BL21(DE3) cells and purified as previously described [4]. In some cases, 2 mM $NADP^+$ was present in buffers used in size exclusion chromatography. Absorbance spectroscopy was used to characterize substrate-binding. The theoretical ε_{280} of the protein is 10 800 M^{-1} cm^{-1}, whereas ε_{260} of the $NADP^+$ substrate is 18 000 M^{-1} cm^{-1}. Thus, λ_{max} is an indicator of the fraction of apo-protein. The absorbance at 340 nm was used to determine the oxidation state of the protein-bound substrate. Following purification, the integrity of the protein and level of labelling were checked with MALDI-TOF mass spectrometry.

Labelled protein was prepared by growing cells in M9 minimal medium [13], using [^{15}N]ammonium chloride (Martek) and/or [^{13}C]glucose (Martek) as nitrogen and carbon source, respectively. Deuterated protein was obtained by growing cells in 99% D_2O (Cambridge Isotope Laboratories), using either protonated glucose or deuterated glycerol (Martek) as carbon source. Labelling of specific residues was performed by adding ^{15}N-labelled amino acids (Cambridge Isotope Laboratories) to the minimal medium, in the presence or absence of unlabelled amino acids. The concentrations of amino acids in the medium were as described in [14].

*Corresponding author. Fax: (46) (31) 773 39 10.
E-mail: goran@bcbp.chalmers.se

PII: S0014-5793(99)01156-4

NMR samples contained 0.9–1.5 mM protein in 10 mM phosphate, 100 mM NaCl, pH 7.0, at 25°C. NMR experiments were performed at the Swedish NMR Center on Varian Inova spectrometers. A sensitivity-enhanced version of ^{15}N HSQC [15] was used. HNCA, HNCO, HN(CO)CA, HNCACB and HN(CO)CACB experiments for assignment purposes were performed at 600 MHz, essentially as described by Kay and co-workers [16–18]. ^{15}N-filtered NOESY-HSQC spectra were recorded at 600 and 800 MHz, with NOESY mixing times between 50 and 200 ms. In ^{15}N-filtered TOCSY-HSQC experiments at 600 MHz, mixing times of 50 ms were used and in HCCH-TOCSY and ^{13}C NOESY-HSQC experiments at 800 MHz, mixing times of 21 and 50 ms were used, respectively. The ^{1}H chemical shifts were referenced by using DSS as internal reference. ^{15}N and ^{13}C chemical shifts were referenced indirectly by multiplying the ^{1}H reference with appropriate conversion factors [19]. Spectra were analyzed using ANSIG [20].

T_1 and T_2 relaxation experiments were performed as described [21]. Experiments using T_1 relaxation delays of 0.0111 (×2), 0.111, 0.2775, 0.555 and 1.221 s and T_2 relaxation delays of 0.014 (×2), 0.028, 0.042, 0.070 and 0.140 s were collected at 600 MHz and spectra were collected as 2048×256 complex points with eight transients per increment. The T_1 and T_2 relaxation times were determined and the modelfree software package [22] was used to estimate τ_c.

Exchange of amide protons was studied by lyophilizing the NMR sample and dissolving it in D_2O. H^N exchange was observed in two dimensional (2D) HSQC experiments and the first experiment was recorded within 10 min after transfer to D_2O. NMR spectra were recorded with increasing intervals during a total period of 3 weeks. Between experiments, the sample was kept at 25°C.

Bound $NADP^+$ was exchanged by the addition of stoichiometric amounts of NADPH. The exchange was completed within a few minutes. Residues that were affected were analyzed in 2D HSQC experiments.

3. Results and discussion

Expression levels using M9 minimal medium were in the order of 6–8 mg purified protein per l culture. These levels were not affected when producing ^{13}C- and/or ^{15}N-labelled protein, but decreased by a factor two when growing cells in 99% D_2O. The composition of the medium used during cultivation affected both the redox state of the bound substrate and the proportion of apo-protein. ecIII purified from minimal medium in H_2O bound ca. 80% $NADP^+$ and 20% NADPH as determined from absorbance spectroscopy and NMR experiments. When growing cells in minimal medium supplemented with all amino acids, this ratio was changed to 50% $NADP^+$ and 50% NADPH. Cultivation in D_2O affected substrate-binding since only 80% of the deuterated ecIII was purified with bound substrate. In size exclusion chromatography, which was used as second purification step, deuterated

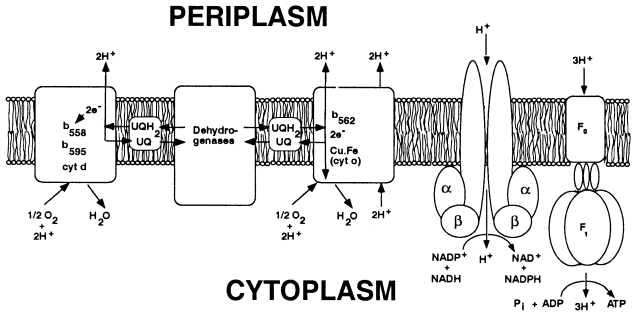

Fig. 1. The major pumps of the plasma membrane of *E. coli*. Transhydrogenase is believed to be important for the redox balance of NADP(H) in the cell. Transhydrogenase is depicted in its active dimeric form.

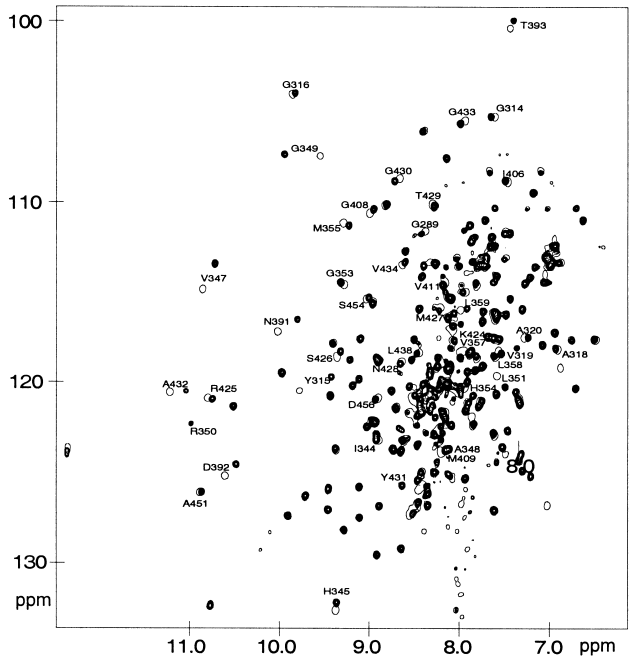

Fig. 2. 2D ^{15}N HSQC spectra of ecIII with bound $NADP^+$ (10 contour levels) and NADPH (one contour level). Backbone amide groups that shift on substrate reduction are labelled.

ecIII eluted from the column as two peaks corresponding to molecular masses of 24 and 19 kDa, respectively. The first peak contained no bound substrate as judged from its maximum absorbance at 280 nm. This assumption was substantiated by the observation that the enzymatic activity increased upon addition of $NADP^+$. The second peak showed maximum absorbance at 267 nm and consequently contained bound substrate. The change in elution volumes reflects a change in the hydrodynamic radius of the molecule. This indicates that the apo-form of ecIII existed in a somewhat less compact state than when substrate is bound. The apo-protein was not stable and to protect it from degradation, $NADP^+$ was added in the final step of this preparation.

The 2D HSQC spectrum of ^{15}N-labelled ecIII with bound $NADP^+$ is shown in Fig. 2. The signals are well-dispersed although some signals overlap. On addition of stoichiometric amounts of NADPH, the substrate exchanges readily. The exchange is conveniently followed in the 2D ^{15}N HSQC spectrum, where some signals shift considerably (Fig. 2).

In the analysis of ^{15}N relaxation data at 25°C and pH 7, 66 isolated peaks distributed over the ecIII sequence were used. The average and S.D. of the T_1/T_2 ratio were calculated and used as input into the tmest program from the modelfree software package. The resulting calculated value of τ_c was 14.23 ± 2.2 ns ($T_1/T_2 = 20.4 \pm 5.7$). The 66 peaks used in the relaxation analysis were grouped according to their secondary structure type. Of these, 16 residues were located in α-helices, 26 were located in β-strands and 24 residues were found in loop regions (this also included what we believe to be substrate-binding regions). The τ_c value was calculated to 13.45 ± 1.3 ns for residues in helical regions ($T_1/T_2 = 18.3 \pm 3.2$) and 14.73 ± 1.5 ns for residues from β-strands ($T_1/T_2 = 21.9 \pm 3.9$), suggesting (as expected) a somewhat more rigid structure for the β-strands compared to the α-helices. Since the S.D.s for the loop region peaks were very high

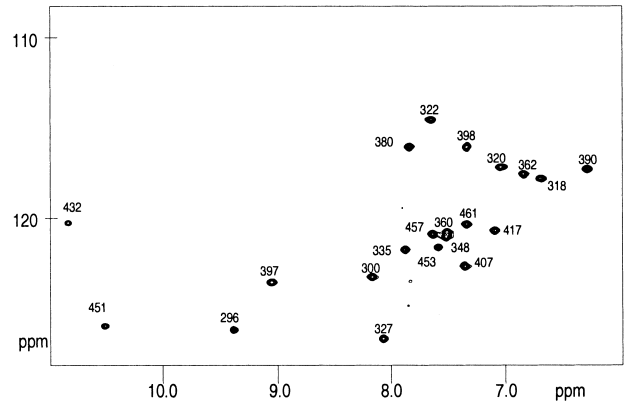

Fig. 3. 2D ^{15}N HSQC spectra of ecIII specifically labelled with [^{15}N]alanine.

($T_1/T_2 = 20.1 \pm 8.0$), the tmest program was unable to yield a reliable error estimate for these regions. Even so, the average value of τ_c was also around 14 ns.

The value of τ_c, 14 ns, is larger than expected for a globular monomeric protein of 20 kDa at room temperature. The large value could be due to the shape of the protein or because of dimerization or aggregation. The observed τ_c value, 14 ns, corresponds well with what would be expected for a dimer of ecIII. The intact enzyme is purified in a dimeric state in membrane preparations, but in size exclusion chromatography, ecIII behaves as if it was a monomer, also at high protein concentrations ($\sim$10 mg/ml). Similarly, dynamic light scattering experiments at concentrations around 1 mM (Johansson et al., unpublished) indicated that the solution molecular weight of ecIII is 20 kDa and, thus, that ecIII exists in a monomeric form. The relaxation behavior of ecIII suggested the use of a deuterated sample in order to perform the sequential assignment.

The sequential assignment was based on U(^{13}C, ^{15}N, 85% ^{2}H) labelling with backbone experiments, in combination with specific [^{15}N]Ala, Val and Lys labelling (Fig. 3). ^{15}N-filtered TOCSY and NOESY experiments were used to assign H^{α} resonances and a U(^{13}C, ^{15}N) sample was used in HCCH-TOCSY and ^{13}C NOESY-HSQC experiments to assign the side chains of all amino acids that contain methyl groups. Resonance assignments of H^N, N, C^{α}, C′ and H^{α} were deposited in the BioMagResBank (accession number 4329). The secondary structure elements were identified using chemical shift data (Fig. 4), short-range NOEs in α-helices and β-strands and amide exchange data (Fig. 5). Long-range NOEs across β-strands (Fig. 6) defined a six-stranded parallel β-sheet with a strand order 3, 2, 1, 4, 5, 6. The first β-strand is preceded by an α-helix, α1, and α-helices also connect β-strands 1–4 (α2, α3 and α4). β-strand 4 is followed by a short α-helix (α5) but the connection to β5 and β6 does not involve regular secondary structure elements. Finally, β-strand 6 is followed by the last α-helix, α6.

The chemical shift changes that are observed on exchange of $NADP^+$ for NADPH define the substrate-binding region in ecIII. Affected residues are primarily located C-terminal to β-strands (with the exception of β3) and extend several residues away in these loops. In the sequence that follows β1, residues in α2 are affected until the kink-inducing proline. Similarly, following β2, both the loop and the greater part of α3 are affected. The sequence that follows β4 and β5 has a little ordered secondary structure (except for the short α5) and here, the chemical shift changes can be observed in regions that extend more than 20 residues away, primarily residues 406–412 and 425–434 following β4 and β5, respectively. The observed chemical shift changes would primarily be expected to be located in one region of the substrate-binding area, reflecting the change in the oxidation state of the nicotinamide ring. However, the extension of chemical shift changes in essentially the entire width of the the dinucleotide-binding region suggests that additional structural changes are coupled to the substrate redox reaction.

The long-range H^N-H^N NOEs that were used to define the secondary structure cannot be used to unambiguously determine the topology of the β-sheet nor the location of α-helices relative to the sheet. For this reason, the side chains of amino acids containing methyl groups were assigned and NOEs involving methyl groups were used to determine the global fold. The observed side chain NOE between methyl groups of residues Val-309 (β1) and Thr-384 (β4) defines the front of the β-sheet and the NOE between Ile-310 (β1) and Val-385 (β4) originates from side chains that protrude from the back of the β-sheet. Similar to Thr-384, the side chain of Leu-386 is also in front of the sheet and the observed NOE to Ile-421 (β5) and the NOE from Ile-421 to Leu-447 (β6) define the side chain topology for strands β1, β4, β5 and β6. The observed NOE between Ile-310 (β1) and Ile-375 (α4) places α4 behind the sheet and for topology reasons, α2 and α3 must not be

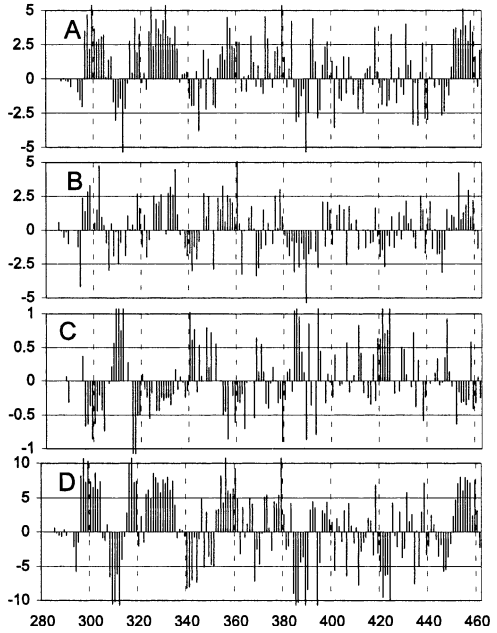

Fig. 4. Chemical shift differences for C^{α} (A), C′ (B) and H^{α} (C) were calculated by subtracting an average chemical shift value for each amino acid from the actual chemical shift [24]. In the case of C^{α} and C′, consecutive values above zero indicate the presence of an α-helix and consecutive values below zero indicate the presence of a β-strand. H^{α} has the opposite relationship. The chemical shift differences are of the same order of magnitude in C^{α} and C′, but the chemical shift differences of H^{α} are ca. five times lower and of opposite sign. A summary of the chemical shift information is presented in (D), where the chemical shift differences have been added according to the formula: diff (C^{α})+diff (C')−5×diff (H^{α}). In this graph, positive values are indicative of α-helices and negative values define β-strands.

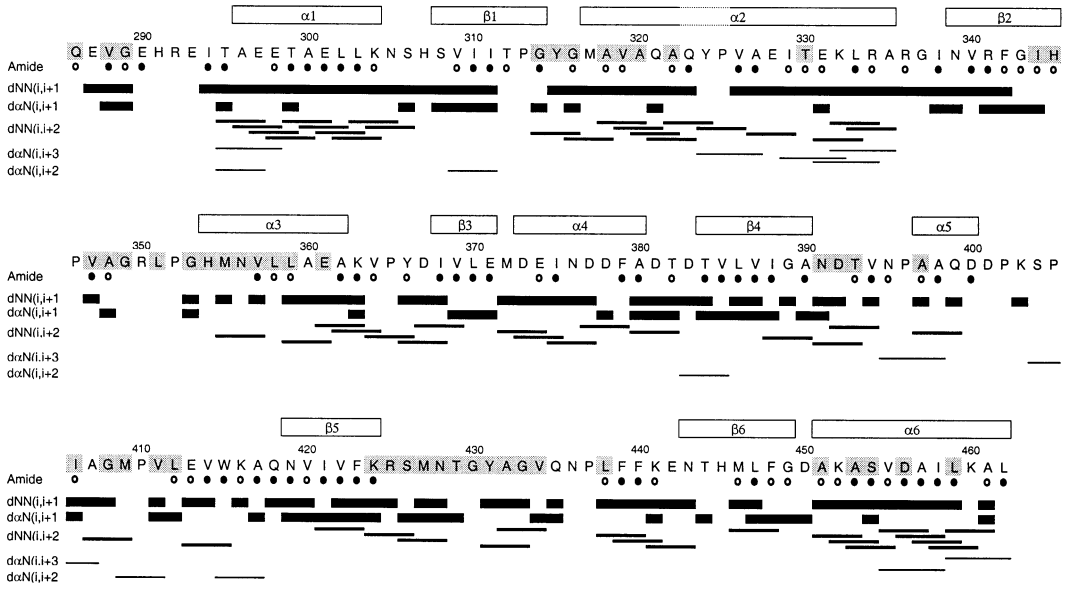

Fig. 5. Summary of amide proton exchange rates and sequential and medium-range NOE connectivities for the ecIII domain of *E. coli* transhydrogenase. Residues for which the amide proton signal was still observed in the 2D ^{15}N HSQC spectrum after the lyophilized protein has been dissolved in D_2O are marked with ○ (after 20 min) or ● (after 32 h). NOE connectivities between residue *i* and *i*+1 are indicated with a bar under residue *i*. Medium-range NOEs are represented by continuous lines connecting the residues. The secondary structures based on NOEs, exchange data and the chemical shift index are indicated as bars above the sequence. Residues for which the amide group experiences a chemical shift change upon titration of NADPH are shaded.

located on the same side as α4. NOEs between Ala-318 (α2) and Leu-358 (α3) and Ala-327 (α2) and Ala-362 (α3) confirm that the helices are located on the same side and pack against each other. Furthermore, Leu-359 (α3) makes contact with Val-369 (β3), which in turn has a NOE to the β-proton of Phe-342 (β2), and the side chain topology of the entire β-sheet is determined.

Finally, our data can now define the position of α-helices 1 and 6, relative to the β-sheet. The two observed NOEs between α1 and β4 (Leu-302–Leu-386) and α1 and β6(Thr-299–Leu-447), respectively, position α1 in front of the sheet. Similarly, the observed NOE between Val-340 (β2) and the penultimate Ala-461 (α6) together with the two NOEs between β4 and α6 (Ile-388–Ala-451), (Ile-388–Val-455) define the position of α6 in front of the sheet. Together, the observed NOEs define the global fold in Fig. 7. Interestingly, we also observe a few NOEs that are located in the loop regions following the β-strands. Especially, the NOEs in the loop between β4 and β5 (Val-394–Val-411) and (Ala-397–Ile-406), respectively, are separated by the short α5 and define an extended loop, the top of which (residues 406–412) is involved in substrate interaction.

Both the secondary structure and the model differ from the model recently published for a homologous domain III from *R. rubrum* [11]. In this model, β-strand 3 is not observed in the sequence and an extra β-strand is suggested to exist in the sequence between β-strands 4 and 5. Furthermore, α5 is not observed but the presence of an α-helix in the sequence between β5 and β6 is suggested. Considering that the domains are of equal length and $>50\%$ identical, the discrepancies are substantial. Despite the differences in the location of the secondary structure elements, a parallel six-stranded β-sheet is suggested with the same strand order and as a consequence, the suggested substrate-binding site is severely distorted.

The global fold of ecIII shares many of the features com-

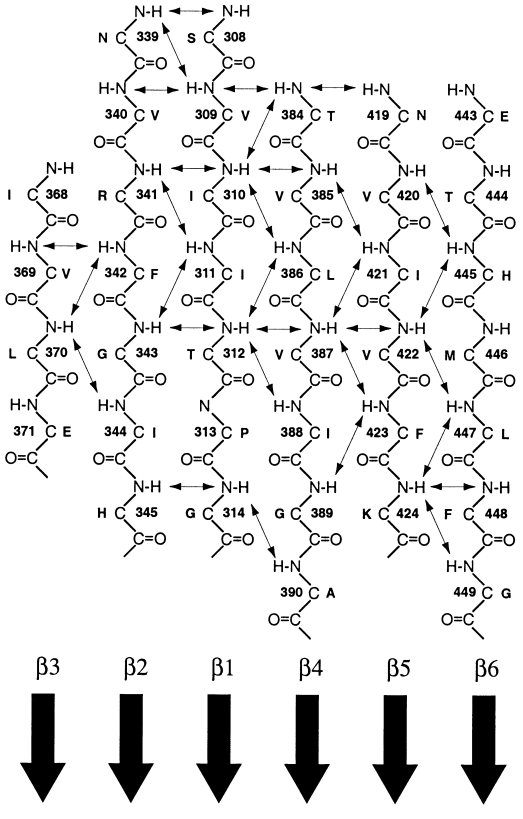

Fig. 6. The central β-sheet in ecIII was defined by NOEs observed between β-strands in ^{15}N NOESY-HSQC spectra of perdeuterated protein with NOESY mixing times between 50 and 200 ms.

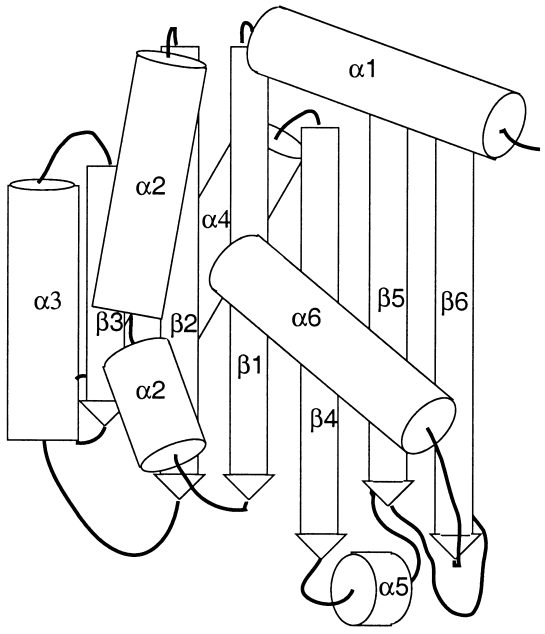

Fig. 7. Global fold based on NMR constraints. The substrate-binding site is located C-terminal to the β-strands in the β-sheet. The fold is similar to that of a classical dinucleotide-binding domain, but αD and αE are missing and instead, the C-terminal α6 is located on the same side as α2 and α3.

monly found in dinucleotide-binding domains. The domain is dominated by the central β-sheet and both helices α2 and α3, as well as α4, involved in the cross over between β1 and β4, are found. The classical dinucleotide-binding domain is often associated with a 2-fold symmetry, caused by two additional α-helices, αE and αF, connecting the C-terminal β-strands [23]. In ecIII, however, the 2-fold symmetry does not exist, since neither of the two α-helices is present. Instead, the C-terminal α6 is found on the same side as α2 and α3, relative to the β-sheet. At present, though, we can only speculate how the observed changes in domain architecture will be reflected in substrate-binding and domain interactions within the protein.

Acknowledgements: This work was supported by grants from the Swedish Natural Science Research Council. NMR experiments were performed at the Swedish NMR Centre. C.J. acknowledges a grant from the Sven and Lilly Lawski Foundation.

References

[1] Fjellström, O., Olausson, T., Hu, X., Källebring, B., Ahmad, S., Bragg, P.D. and Rydström, J. (1995) Proteins 21, 91–104.
[2] Baker, P.J., Sawa, Y., Shibata, H., Sedelnikova, S.E. and Rice, D.W. (1998) Nat. Struct. Biol. 5, 561–567.
[3] Olausson, T., Fjellström, O., Meuller, J. and Rydström, J. (1995) Biochim. Biophys. Acta 1231, 1–19.
[4] Fjellström, O., Johansson, C. and Rydström, J. (1997) Biochemistry 36, 11331–11341.
[5] Diggle, C., Cotton, N.P., Grimley, R.L., Quirk, P.G., Thomas, C.M. and Jackson, J.B. (1995) Eur. J. Biochem. 232, 315–326.
[6] Yamaguchi, M. and Hatefi, Y. (1995) J. Biol. Chem. 270, 28165–28168.
[7] Diggle, C., Bizouarn, T., Cotton, N.P. and Jackson, J.B. (1996) Eur. J. Biochem. 241, 162–170.
[8] Yamaguchi, M. and Hatefi, Y. (1997) Biochim. Biophys. Acta 1318, 225–234.
[9] Johansson, C., Bergkvist, A., Fjellström, O., Rydström, J. and Karlsson, B.G. (1999) J. Biomol. NMR.
[10] Jeeves, M., smith, K.J., Quirk, P.G., Cotton, N.P.J. and Jackson, J.B. (1999) J. Biomol. NMR 13, 305–306.
[11] Quirk, P., Jeeves, M., Cotton, N.P.J., Smith, J.K. and Jacksson, B.J. (1999) FEBS Lett. 446, 127–132.
[12] Studier, F.W. and Moffatt, B.A. (1986) J. Mol. Biol. 189, 113–130.
[13] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Cold Spring Harbor Press, Cold Spring Harbor, NY.
[14] Wanner, B.L., Kodaira, R. and Neidhardt, F.C. (1977) J. Bacteriol. 130, 212–218.
[15] Stonehouse, J., Shaw, G.L., Keeler, J. and Laue, E.D. (1994) J. Magn. Res. Ser. A 107, 178–184.
[16] Yamazaki, T., Lee, W., Revington, M., Mattiello, D.L., Dahlquist, F.W., Arrowsmith, C.H. and Kay, L.E. (1994) J. Am. Chem. Soc. 116, 6464–6465.
[17] Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R. and Kay, L.E. (1994) J. Am. Chem. Soc. 116, 11655–11666.
[18] Shan, X., Gardner, K.H., Muhandiram, D.R., Rao, N.S., Arrowsmith, C.H. and Kay, L.E. (1996) J. Am. Chem. Soc. 118, 6570–6579.
[19] Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E. and Wuthrich, K. (1998) J. Mol. Biol. 280, 933–952.
[20] Kraulis, P.J. (1989) J. Magn. Res. 84, 627–633.
[21] Farrow, N.A. et al. (1994) Biochemistry 33, 5984–6003.
[22] Mandel, A.M., Akke, M. and Palmer, A.G. (1995) J. Mol. Biol. 246, 144–163.
[23] Brändén, C. and Tooze, J. (1991) Introduction to Protein Structure, Garland, New York.
[24] Wishart, D.S. and Sykes, B.D. (1994) Methods Enzymol. 239, 363–392.