

^1H , ^{13}C and ^{15}N resonance assignment of a 114-residue fragment of Engrailed 2 homeoprotein, a partially disordered protein

Rafal Augustyniak · Stéphane Balayssac ·
Fabien Ferrage · Geoffrey Bodenhausen ·
Olivier Lequin

Received: 24 January 2011 / Accepted: 14 April 2011 / Published online: 23 April 2011
© Springer Science+Business Media B.V. 2011

Abstract Engrailed 2 is a transcription factor belonging to the class of homeoproteins. These proteins possess a 60-residue DNA binding globular domain and play an important role in the early stages of development. We expressed and purified a 13.4 kDa fragment of Engrailed 2, which comprises a 54-residue N-terminal extension in addition to the homeodomain region. Almost all backbone and side-chain resonances have been assigned. The weak dispersion in the proton dimension of the ^1H – ^{15}N HSQC spectrum indicates the presence of disordered regions that do not belong to the homeodomain. This work is a first step toward the NMR investigation of the structure and

dynamics of Engrailed 2 protein that contains a well-structured globular domain and partially disordered regions.

Keywords Engrailed 2 · Homeodomain · Transcription factor · Disordered proteins · NMR

Biological context

Homeoproteins constitute a large class of transcription factors that are present in a wide range of eukaryotic species, from yeast to humans. They are encoded by homeogenes, characterized by a highly conserved homeobox motif (Gehring et al. 1994) coding for a 60-residue DNA binding domain called homeodomain. Discovered initially in *Drosophila* (McGinnis et al. 1984), homeogenes have since been found in all metazoan organisms (Gehring et al. 1994). Homeoproteins have been shown in particular to play a fundamental role in embryonic development by modulating the spatio-temporal expression of target genes (Gehring 1987).

Engrailed 2 is a key factor in the development of the encephalon at the mesencephalon-metencephalon junction in the embryo (Joyner 1996). This protein is also expressed in adults in dopaminergic neurons of the mesencephalon. The inhibition of the expression of Engrailed 2 in dopaminergic neurons was shown to lead to apoptosis (Simon et al. 2001). This indicates a major physiologic role played by Engrailed 2 (Alberi et al. 2004; Sonnier et al. 2007). The structures of Engrailed and many other homeodomains have been studied in details by X-ray crystallography or NMR spectroscopy (Fraenkel et al. 1998). On the other hand, little information is available about the structure and dynamics of other parts of homeoproteins.

R. Augustyniak · F. Ferrage · G. Bodenhausen · O. Lequin (✉)
UPMC Univ Paris 06, 4 Place Jussieu, 75252 Paris, Cedex 05,
France
e-mail: olivier.lequin@upmc.fr

R. Augustyniak · F. Ferrage · G. Bodenhausen · O. Lequin
Département de chimie, Ecole Normale Supérieure,
24 rue Lhomond, 75231 Paris Cedex 05, France

R. Augustyniak · F. Ferrage · G. Bodenhausen · O. Lequin
UMR 7203 CNRS-UPMC-ENS, Paris Cedex 05, France

S. Balayssac
Groupe de RMN Biomédicale, Laboratoire de Synthèse et
Physico-Chimie de Molécules d'Intérêt Biologique, UMR CNRS
5068 CNRS, Université de Toulouse, 118 route de Narbonne,
31062 Toulouse Cedex 09, France

G. Bodenhausen
Institut des Sciences et Ingénierie Chimiques, BCH,
Ecole Polytechnique Fédérale de Lausanne, 1015
Lausanne, Switzerland

Chicken Engrailed 2 is a 289-residue protein. We are studying a shorter construct encompassing residues 146–259 that comprises the homeodomain (residues 200–259) and an N-terminal extension. This extension contains several known binding sites for other transcription factors (Pbx, FoxA2) which are implicated in the regulation of transcription (Piper et al. 1999; Foucher et al. 2003). It also comprises a serine-rich region whose post-translational phosphorylation controls the secretion of the protein (Maizel et al. 2002). In this article we report the NMR assignment of the 146–259 fragment of Engrailed 2, in view of further investigations of the three-dimensional structure and dynamics of Engrailed 2 protein.

Methods and experiments

The *E. coli* BL21RP expression strain was transformed with a pGEX plasmid (Pharmacia) containing a gene coding for sequence 146–259 of chicken Engrailed 2 fused to Glutathion S-Transferase. The recombinant Engrailed 2 protein was expressed and labelled using the Marley protocol (Marley et al. 2001), commonly used for preparing uniformly labelled proteins for NMR. Cells were grown at 37°C in 4 L of LB medium supplemented with 100 µg/mL of ampicillin and chloramphenicol. When the optical density at 600 nm reached 0.7, the cells were harvested by centrifugation at 4,000g for 10 min, and re-suspended in 1 L of M9 minimal medium containing $^{15}\text{NH}_4\text{Cl}$ (or $^{15}\text{NH}_4\text{Cl}$ and ^{13}C -glucose) as required for ^{15}N (or $^{15}\text{N}/^{13}\text{C}$) labelling. Protein expression was induced by addition of 1 mM isopropyl- β -D-thiogalactoside (IPTG). After 3 h of incubation, the cells were harvested by centrifugation at 4,000g, 4°C for 10 min. The cell pellets were re-suspended in a phosphate buffer (pH 7.4) containing 140 mM NaCl, 1 mM EDTA, 1 µg/mL pepstatin, 1 µg/mL leupeptin, 1.5 mM PMSF, 2 mM TCEP, 0.5% (v/v) Triton X-100 and frozen at -80°C . The cell pellets were thawed and lysed by sonication. After centrifugation at 15,000g, 4°C for 45 min, the supernatant was loaded on GStrap FF columns (GE Healthcare). The GST fusion protein was eluted using a 50 mM Tris-HCl buffer, pH 8, containing 150 mM NaCl, 1 mM EDTA, 1 mM DTT and 30 mM glutathion. The fusion protein was cleaved using PreScission protease (GE Healthcare). Engrailed 2 protein was then loaded on HiTrap Heparin columns (GE Healthcare) and eluted with a phosphate buffer containing 1.5 M NaCl. The collected fractions were concentrated by ultrafiltration using a 3 kDa cutoff membrane (Millipore). Each NMR sample typically contained ~ 0.6 mM of uniformly labelled protein in 40 mM succinate buffer, pH 6.0, supplemented with anti-proteases (leupeptine, pepstatine, EDTA and AEBSF), 0.1 mM DSS and 0.02% (w/v) NaN_3 in $\text{H}_2\text{O}/\text{D}_2\text{O}$ (95/5%

v/v). 10 mM TCEP was added to prevent the oxidation of the C175 residue.

Most NMR experiments were recorded at 30°C on a Bruker Avance III spectrometer equipped with a TCI cryoprobe, operating at a ^1H frequency of 500 MHz. Five millimetre Shigemi tubes were used for all experiments. Proton chemical shifts were referenced to internal DSS. ^{15}N and ^{13}C resonances were referenced indirectly to DSS. The data were processed with NMRpipe (Delaglio et al. 1995) and analyzed with CCPNmr Analysis 2.0 (Vranken et al. 2005).

Backbone resonance assignments were obtained using ^1H - ^{15}N HSQC and a set of 3D spectra HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO and HBHA(CO)NH. Side-chain aliphatic ^1H and ^{13}C resonances were assigned mainly using 2D ^1H - ^{13}C constant time HSQC, 3D ^1H - ^{15}N TOCSY-HSQC, 3D H(CCO)NH-TOCSY and 3D (H)C(CO)NH-TOCSY experiments (Sattler et al. 1999). To overcome problems of overlap, particularly in the case of resonances corresponding to disordered regions, a 4D HC(CC-TOCSY)(CO)NH with random sampling (Kazimierczuk et al. 2009) was recorded on a Varian 700 MHz spectrometer equipped with a Performa XYZ PFG unit. 3D HBCB(CGCD)HD, 3D HBCB(CGCDCE)HE and 3D ^1H - ^{13}C NOESY-HSQC experiments were also recorded in D_2O to assign the aromatic resonances.

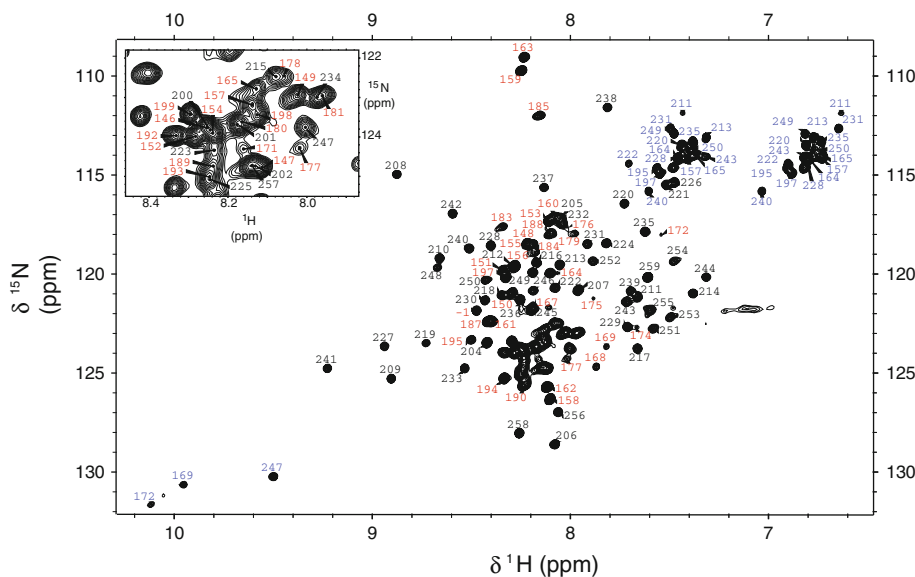
Assignments and data deposition

The assigned 2D ^1H - ^{15}N HSQC is shown in Fig. 1. Most backbone NH groups were assigned, with the exception of the N-terminal residue and V173. The absence of correlation for V173 may be due to line broadening. Indeed, all NH resonances of residues in the W169–Y178 segment show lower signal intensities. Moreover, additional weak resonances are observed in the ^1H - ^{15}N HSQC spectrum corresponding to the aromatic group of W169. These observations suggest some particular dynamic effects in this region that might be due to *cis-trans* isomerism of the P170 residue. The NH resonances corresponding to homeodomain residues are well dispersed, while those belonging to the N-terminal extension are all in a narrow range between 7.6 and 8.6 ppm (see Fig. 1) which indicates intrinsic disorder in regions outside the homeodomain.

A few side-chain ^{15}N resonances could not be assigned (^{15}N of Pro, $^{15}\text{N}^\zeta$ of Lys, $^{15}\text{N}^\eta$ and $^{15}\text{N}^\epsilon$ of Arg). The assignment is complete for all $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ and ^{13}CO and reaches 98% for $^1\text{H}^\alpha$. The majority of side-chain resonances have been assigned (91% for ^{13}C , 75% for ^{15}N and 96% for ^1H). The chemical shifts have been deposited in the BioMagResBank (<http://www.bmrb.wisc.edu>) under accession number 17325.

Fig. 1 Assigned 2D ^1H - ^{15}N HSQC spectrum of the ^{15}N -labelled 146–259 fragment of Engrailed 2 protein acquired at 500 MHz. A zoom of the central crowded region of the spectrum is shown in *upper-left*.

Backbone resonances corresponding to residues in the homeodomain (200–259) and in the N-terminal extension (146–199) are marked in *black* and *red*, respectively. The side-chain resonances of Trp, Asn and Gln residues are indicated in *blue*



Acknowledgments Financial support by the Access to Research Infrastructures activity of the 7th Framework Programme of the EU (Contract 228461, EAST-NMR) for conducting research in Warsaw is gratefully acknowledged. Wiktor Kozminski and Jan Stanek (Department of Chemistry, University of Warsaw) are deeply acknowledged for their help in NMR data collection and fruitful discussions. We thank Alain Prochiantz (UMR 7233 CNRS, Collège de France) for kindly providing the expression plasmid.

References

- Alberi L, Sgado P, Simon HH (2004) Engrailed genes are cell-autonomously required to prevent apoptosis in mesencephalic dopaminergic neurons. *Development* 131:3229–3236
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293
- Foucher I, Montesinos ML, Volovitch M, Prochiantz A, Trembleau A (2003) Joint regulation of the MAP1B promoter by HNF3beta/Foxa2 and Engrailed is the result of a highly conserved mechanism for direct interaction of homeoproteins and Fox transcription factors. *Development* 130:1867–1876
- Fraenkel E, Rould MA, Chambers KA, Pabo CO (1998) Engrailed homeodomain-DNA complex at 2.2 Å resolution: a detailed view of the interface and comparison with other engrailed structures. *J Mol Biol* 284:351–361
- Gehring WJ (1987) Homeo boxes in the study of development. *Science* 236:1245–1252
- Gehring WJ, Affolter M, Buerklin T (1994) Homeodomain proteins. *Ann Rev Biochem* 63:487–526
- Joyner AL (1996) Engrailed, Wnt and Pax genes regulate midbrain-hindbrain development. *Trends Genet* 12:15–20
- Kazmierczuk K, Zawadzka A, Kozminski W (2009) Narrow peaks and high dimensionalities: exploiting the advantages of random sampling. *J Magn Reson* 197:210–228
- Maizel A, Tassetto M, Filhol O, Cochet C, Prochiantz A, Joliot A (2002) Engrailed homeoprotein secretion is a regulated process. *Development* 129:3545–3553
- Marley J, Lu M, Bracken C (2001) A method for efficient isotopic labelling of recombinant proteins. *J Biomol NMR* 20:71–75
- McGinnis W, Levine MS, Hafen E, Kuroiwa A, Gehring WJ (1984) A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and bithorax complex. *Nature* 308:428–433
- Piper DE, Batchelor AH, Chang CP, Cleary ML, Wolberger C (1999) Structure of the HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and the fourth homeodomain helix in complex formation. *Cell* 96:587–597
- Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc* 34:93–158
- Simon HH, Saueressig H, Wurst W, Goulding MD, O’Leary DD (2001) Fate of midbrain dopaminergic neurons controlled by the engrailed genes. *J Neurosci* 21:3126–3134
- Sonnier L, Le Pen G, Hartmann A, Bizot JC, Trovero F, Krebs MO, Prochiantz A (2007) Progressive loss of dopaminergic neurons in the ventral midbrain of adult mice heterozygote for Engrailed1. *J Neurosci* 27:1063–1071
- Vranken WF, Boucher W, Stevens TJ et al (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59:687–696