Structural defects of a Pax8 mutant that give rise to congenital hypothyroidism

Gianluca TELL*, Lucia PELLIZZARI*, Gennaro ESPOSITO*, Carlo PUCILLO*, Paolo Emidio MACCHIA†, Roberto DI LAURO† and Giuseppe DAMANTE*¹

*Dipartimento di Scienze e Tecnologie Biomediche, Università di Udine, Piazzale Kolbe 1-33100 Udine, Italy, and †Stazione Zoologica 'A. Dohrn', Villa Comunale 1, 80121 Napoli, Italy

Pax proteins are transcriptional regulators that play important roles during embryogenesis. These proteins recognize specific DNA sequences via a conserved element: the paired domain (Prd domain). The low level of organized secondary structure, in the free state, is a general feature of Prd domains; however, these proteins undergo a dramatic gain in α -helical content upon interaction with DNA ('induced fit'). Pax8 is expressed in the developing thyroid, kidney and several areas of the central nervous system. In humans, mutations of the *Pax8* gene, which are mapped to the coding region of the Prd domain, give rise to congenital hypothyroidism. Here, we have investigated the molecular defects caused by a mutation in which leucine at position 62 is substituted for an arginine. Leu⁶² is conserved among Prd domains, and contributes towards the packing together of helices 1 and 3. The binding affinity of the Leu⁶²Arg

mutant for a specific DNA sequence (the C sequence of thyroglobulin promoter) is decreased 60-fold with respect to the wild-type Pax8 Prd domain. However, the affinities with which the wild-type and the mutant proteins bind to a non-specific DNA sequence are very similar. CD spectra demonstrate that, in the absence of DNA, both wild-type Pax8 and the Leu⁶²Arg mutant possess a low α -helical content; however, in the Leu⁶²Arg mutant, the gain in α -helical content upon interaction with DNA is greatly reduced with respect to the wild-type protein. Thus the molecular defect of the Leu⁶²Arg mutant causes a reduced capability for induced fit upon DNA interaction.

Key words: DNA binding, human mutation, induced fit, Pax genes, protein structure.

INTRODUCTION

Pax proteins form a family of transcription factors that are essential for a variety of developmental decisions during embryogenesis [1–4]. Pax proteins are defined by the presence of a 128-amino-acid DNA-binding domain, the paired domain (Prd domain) [5]. This element has been well conserved during evolution [6] and, according to the sequence similarity in this region, Pax proteins can be grouped into distinct subfamilies [7]. Elucidation of the crystal structure of the Prd domain of the paired protein bound to DNA has revealed the presence of two structurally independent subdomains, PAI and RED [8], each containing a helix–turn–helix motif joined by a linker region (Figure 1) [9]. Functional studies have revealed that both subdomains are able to interact with DNA [10,11]. CD data obtained with Pax6 and Pax8 Prd domains have revealed that these proteins are almost structureless in the free state; however, a dramatic gain in α -helical content occurs upon DNA binding [11,12], suggesting that unstructured-to-structured transitions are necessary for an efficient Prd domain–DNA interaction. These transitions comprise some of the conformational changes that proteins undergo when interacting with DNA, according to a general mechanism referred to as 'induced fit' [13]. The α -helical gain that occurs in the Prd domain during DNA interaction might have a regulatory function, since it is greatly controlled by the redox potential [14].

Pax8 is expressed in the developing kidney, in some areas of the brain and in follicular thyroid cells [15]. In this latter type of

cells, Pax8 is also expressed in adults and activates the thyroglobulin and thyroperoxidase promoters [16]. The developmental role of Pax8 has been demonstrated recently by gene targeting [17]. Pax8 knock-out mice have been shown to possess a complete absence of thyroid follicular structures. A search for *Pax8* mutations in congenital hypothyroidism has been reported recently [18]. One of the hypothyroidism-inducing mutations consists of a Leu → Arg amino acid change at position 62, within the recognition helix of the PAI subdomain (Figure 1, helix α 3). This mutant (Leu⁶²Arg) shows a reduction in DNA-binding activity [18]. Although Leu⁶² is always conserved among Prd domains, it does not directly establish contacts with DNA [9]. This residue contributes to the hydrophobic core of the PAI subdomain, packing together helices α 1 and α 3 [9]. In this study

Figure 1 Structure and purification of Prd domains

Secondary structure of the Pax8 Prd domain. The drawing is on the basis of the data of Xu et al. [9]. Grey boxes indicate α -helical regions; empty ovals indicate β -sheets. The position of the Leu⁶²Arg mutation (L62R) is indicated.

Abbreviations used: DTT, dithiothreitol; Prd domain, paired domain.

¹ To whom correspondence should be addressed (e-mail giuseppe.damante@dstb.uniud.it).

we tested the possibility that the Leu → Arg mutation at position 62 impairs the unstructured-to-structured transition that occurs during DNA recognition. We demonstrate that the Leu⁶²Arg mutation abolishes the specific DNA recognition by hampering the induced fit. Thus the congenital hypothyroidism due to the Leu⁶²Arg mutation provides the first example of a human organogenesis defect caused by a lack of an induced fit during the interaction of protein with DNA.

MATERIALS AND METHODS

Plasmids

The plasmids pT7.7wt Prd and pT7.7Leu⁶²Arg Prd were generated by cloning the DNA sequences encoding for the wild-type and Leu⁶²Arg Pax8 Prd domains into the plasmid pT7.7. The DNA fragments were amplified by PCR using the oligonucleotides 5'hP8 (5'-GCGCGCCATATGGGCCATGGAGGGCT-GAACCAGC-3') and 3'hP8 (5'-GCGCGCGGATCCTCAAG-GGAGGTTGAATGGTTGCTGCAC-3'). All PCR reactions were performed using plasmids described previously [18] as the template. PCR products were digested with *NdeI*/*Bam*HI, and cloned into the corresponding sites of pT7.7 polylinker.

Protein expression, purification and quantification

Plasmids were used to transform the bacterial strain BL21 [19]. Transformed cells were grown at 37 °C to a D_{600} of 0.6–0.7, and then induced by 1 mM isopropyl β -D-thiogalactoside for 3 h. Cells were harvested by centrifugation, and resuspended in lysis buffer [50 mM sodium phosphate buffer (pH 7.0)/0.25 M NaCl/1 mM EDTA/1 mM EGTA/1 mM dithiothreitol (DTT)/2 μ g/ml leupeptin/2 μ g/ml pepstatin/1 mM PMSF] in a vol. of 10 ml/g of bacterial pellet. After cell lysis by sonication, bacterial debris was removed by centrifugation. DNA was removed by addition of protamine sulphate to the supernatant to a final concentration of 0.3 mg/ml, and the precipitate was removed by centrifugation. The supernatant was subsequently loaded on a Mono-S column (Pharmacia, Uppsala, Sweden) pre-equilibrated with lysis buffer. Both the wild-type and the Leu⁶²Arg mutant of the Pax8 Prd domain were purified using a linear gradient of 0.25–0.7 M NaCl in 50 mM phosphate buffer, pH 7.0, containing 1 mM DTT. Protein concentrations were determined using a molar absorption coefficient (ϵ) of 9890 M⁻¹·cm⁻¹, calculated by the method of Wetlaufer [20], and confirmed by the colorimetric assays of

Figure 2 SDS/PAGE analysis of wild-type (WT) and Leu⁶²Arg (L62R) Pax8 Prd domains

Wild-type and L62R Pax-8 Prd domains were analysed using SDS/PAGE (15% gels), and stained with silver nitrate [29]. Molecular mass markers used were: phosphorylase *b*, 94 kDa; albumin, 67 kDa; ovalbumin, 43 kDa; carbonic anhydrase, 30 kDa; trypsin inhibitor, 20 kDa; cytochrome *c*, 12.5 kDa.

Lowry et al. [21] or Bradford [22]. Absorbance measurements were obtained using a Perkin–Elmer Lambda 14 UV/visible spectropolarimeter.

Gel-retardation assay

Gel-retardation assays were performed by incubating protein and DNA in a buffer containing 20 mM Tris/HCl, pH 7.6, 75 mM KCl, 0.25 mg/ml BSA, 5 mM DTT and 10% (v/v) glycerol for 30 min at room temperature. Double-stranded oligodeoxynucleotides, labelled at their 5'-end termini with ³²P, were used as probes. The C site is a 24-mer, whose top-strand sequence is 5'-CACTGCCCAGTCAAGTGTCTTGA-3'. The BS2 site is a 20-mer, whose top-strand sequence is 5'-CGAGAAAAAGCCATTAGAGC-3'. At the end of the binding reaction, samples were loaded on to a native 7.5% polyacrylamide gel and run in 0.5 × Tris/borate/EDTA (TBE) buffer (where 1 × TBE = 45 mM Tris/borate/1 mM EDTA) at 8 °C. Both protein-bound and free DNA were quantified by PhosphorImager analysis (Bio-Rad), and the affinity of binding of protein with DNA in protein–DNA complexes was calculated by:

$$K_d = [\text{Protein}] \cdot [\text{DNA}] / [\text{Protein–DNA}]$$

Figure 3 DNA-binding properties of wild-type and Leu⁶²Arg Pax8 Prd domains

Autoradiograms of gel-retardation assays performed with the indicated amounts of protein and 1 nM ³²P-labelled oligonucleotides. Gel-retardation assays were performed as described in the Materials and Methods section, dried and exposed for 12 h to autoradiography. B and F indicate protein-bound and free DNA respectively.

Figure 4 UV CD spectra of the Leu⁶²Arg mutant and the wild-type Pax-8 Prd domain in the absence/presence of the oligonucleotide C

(A) CD spectra of the Leu⁶²Arg Pax8 Prd domain (0.16 μM) in the absence (continuous line) or the presence (broken line) of 0.46 μM oligonucleotide C, after correction for the contribution of the oligonucleotide. (B) CD spectra of the wild-type Pax8 Prd domain (0.33 μM) in the absence (○) or the presence (●) of 0.46 μM oligonucleotide C, after correction for the

CD spectroscopy

The purified Pax8 and Leu⁶²Arg Prd domains were used for CD spectroscopy. A Jasco J-600 CD/optical rotatory dispersion (ORD) spectropolarimeter interfaced with an Olidata computer for data collection was used for measurements. Standard buffer conditions used were 75 mM KCl/10 mM Na₂HPO₄, pH 7.4/0.1 M DTT at 10 °C, and experiments were performed using a 1-cm path-length cuvette. Temperature was controlled using a Haake F3 water bath. UV CD spectra are presented in terms of the mean residue molecular ellipticity, [θ] (in units of deg·cm²·dmol⁻¹), on the basis of a mean residue mass of 110.4 Da. The results shown are the smoothed averages for ten independent measurements.

RESULTS AND DISCUSSION

After expression in *Escherichia coli*, both wild-type and Leu⁶²Arg proteins were purified to homogeneity by ion-exchange column chromatography. The degree of obtained purity was higher than 95 %, as assessed by SDS/PAGE (Figure 2). Gel-retardation assays were performed by using the C site of thyroglobulin [11] and the BS2 oligonucleotide as specific and non-specific DNA sequences respectively. The BS2 oligonucleotide was chosen as a non-specific DNA site, since its sequence does not show significant similarity with the Pax8 consensus sequence [23,24]. Representative autoradiograms of gel-retardation assays are shown in Figure 3. The wild-type Pax8 Prd domain bound to the C site with high affinity (K_d 3.1×10^{-9} M), whereas the Leu⁶²Arg mutation resulted in an $\approx$ 60-fold decrease in binding affinity to the C site (K_d 1.8×10^{-7} M). In contrast, the binding affinities of the wild-type and Leu⁶²Arg proteins to the non-specific sequence (BS2) were very similar (K_d 4.4×10^{-7} M and 4.2×10^{-7} M respectively). These findings demonstrate that the Leu⁶²Arg mutation abolishes specific DNA binding, but has no effect on non-specific binding. In addition, our data demonstrate that the Prd domain exhibits an appreciable affinity for non-specific DNA, suggesting that most Pax8 molecules would be bound *in vivo* to sequences that do not play a role in transcriptional regulation. However, this non-specific binding could be extremely useful, e.g. it decreases dimensionality in the search for high-affinity sites, resulting in an acceleration of target location [25].

On the basis of these binding data, it may be estimated that, if the binding reaction is performed with protein and DNA concentrations in the micromolar range, a significant amount of Leu⁶²Arg protein would be in the bound state. This allows us to examine whether or not the Leu⁶²Arg protein undergoes conformational changes upon interaction with DNA, and to compare its structural properties with those of the wild-type protein. Protein secondary structure (in terms of α -helical content) was determined by CD analysis. Proteins were assayed at final concentrations of 0.33 and 0.16 μM for the wild-type and Leu⁶²Arg mutant respectively; DNA (the C site) was present at a concentration of 0.46 μM. Under these conditions, most of the Leu⁶²Arg protein was bound to the C site. The CD spectrum of the Leu⁶²Arg mutant in the absence of DNA (Figure 4A, continuous line) exhibits the typical minima at 208 and 222 nm characteristic of an α -helix conformation, but is slightly less structured in comparison with the wild-type protein (as shown by

contribution of the oligonucleotide. (C) Gel-retardation assays were performed with samples used for the collection of CD spectra, as described in the Materials and Methods section but, owing to the use of unlabelled C sequence, the gel, after completion of the run, has been subjected to silver nitrate staining. F, free DNA; B, protein-bound DNA.

Figure 5 Position of Leu⁵⁴ and Ile²⁶ side chains

Schematic structure of the PAI subdomain of the *Drosophila* paired protein, achieved by using the software MOLMOL [30]. Helices ($\alpha 1$, $\alpha 2$ and $\alpha 3$) are shown as cylinders, and Ile²⁶ and Leu⁵⁴ residues are shown emboldened in black.

the spectrum indicated with open circles in Figure 4B). The α -helix content, estimated from the mean residue ellipticity at 222 nm, was 17% and 19% for the Leu⁶²Arg and wild-type proteins respectively. Thus, in accordance with previous results [11,12], the α -helical content of the Prd domain is low in the absence of DNA. Addition of a stoichiometric excess of C oligonucleotide to the wild-type protein induces a dramatic conformational change towards a more structured state (34% α -helix; Figure 4B, filled circles). On the other hand, the gain in helicity upon addition of the C oligonucleotide was much less pronounced in the Leu⁶²Arg mutant, where only 21% of the residues adopt the α -helix conformation (Figure 4A, broken line). Gel-retardation analysis of the samples used for CD spectra collection (Figure 3C) clearly demonstrates a similar level of protein–DNA complex formation both for the wild-type and the Leu⁶²Arg forms. This allows us to conclude that, in the experimental conditions used for collection of CD spectra, the lack of α -helix induction for the Leu⁶²Arg protein during DNA binding is due to neither a different equilibrium relative to the wild-type form nor a lack of DNA-binding activity, but may be attributed to an intrinsic structural deficiency of this mutant. It is noteworthy that, when CD experiments were attempted at higher protein or DNA concentrations, precipitation occurred, making it impossible to record CD spectra; indeed, a similar problem has been reported by Epstein et al. [12] in their studies of the Pax6 Prd domain interacting with a DNA sequence. Taken together, these data demonstrate that the Leu⁶²Arg mutation of the Pax8 Prd domain does not allow the structural transitions essential for proper, specific DNA recognition.

Our results allow two major conclusions to be drawn. First, the unstructured-to-structured transition occurring during the Pax8 Prd domain–DNA interaction is necessary for high-affinity binding, but is non-essential for low-affinity binding. Therefore our data could suggest that the Leu⁶²Arg mutation does not allow the structural transitions to occur that are necessary for

specific DNA recognition by the Pax8 Prd domain. Alternatively, the molecular effect of the Leu⁶²Arg mutation may be explained by the fact that we have recently demonstrated that C site recognition by the Pax8 Prd domain is due to functional co-operativity between the PAI and RED subdomains [11]. This functional co-operativity does not merely result from an algebraic sum of the contacts established by the two elements, but, rather, the RED subdomain is necessary for correct DNA recognition by the PAI subdomain, therefore leading to an ordering of events during DNA binding. This model predicts a primary role for the RED subdomain in determining specificity and, consequently, a role for the PAI subdomain in determining stable binding. Since the Leu⁶²Arg mutation affects the correct folding (upon DNA interaction) of the PAI domain only, it might also be concluded that the mutation prevents the conformational change that is a consequence of sequence-specific DNA interaction, thus preventing the formation of stable binding. Leu⁶² is conserved in all Prd domains. It should be noted that, upon sequence alignment, Leu⁶² of Pax8 is found to correspond to Leu⁵⁴ in the Prd domain of the *Drosophila* paired protein [9], and therefore this conclusion should hold true for the Prd domains of Pax proteins in general. Since these proteins have appreciable affinity for non-specific DNA sequences, most of the Pax molecules will be bound to transcriptionally irrelevant DNA sites, while searching for the high affinity-sites by sliding along the DNA [25]. Our findings, together with those reported by Epstein et al. [12], suggest that Pax proteins would scan the genome with a poorly folded or almost totally unfolded Prd domain. Only upon reaching high-affinity sites should the Prd domain adopt a highly folded conformation. Available evidence indicates that DNA-binding domains can also contact transcriptional cofactors [26,27]. Reduced levels of Prd domain folding might abolish interactions with cofactors during non-specific DNA binding, leaving productive protein–protein interactions to be formed only upon reaching the high-affinity sequences.

Secondly, a fraction of the human diseases caused by mis-sense mutations of DNA-binding domains (which result in structural modifications that affect the specific DNA-binding capability) could be due to an absence or a reduction of the induced-fit capability. For example, a human *PAX3* mutation has been described recently that gives rise to Waardenburg syndrome [28], in which the isoleucine at position 59 in helix 1 is substituted by phenylalanine. Ile⁵⁹ of Pax3 corresponds to Ile²⁶ of the Prd domain of the *Drosophila* paired protein, and is conserved among Prd domains. It forms part of the hydrophobic core that packs together helices 1 and 3. In the *Drosophila* paired protein, Ile²⁶ faces Leu⁵⁴ (Leu⁶² in Pax8) with a closest packing distance of 3.9 Å (0.39 nm) between heavy atoms of respective side chains, i.e. ≈ 2 Å when hydrogens are also taken into consideration (Figure 5). Such a close contact could not be accommodated within the hydrophobic core of the molecule if Ile²⁶ was replaced by a phenylalanine residue (or if an arginine replaced Leu⁵⁴, as in the Leu⁶²Arg mutant of Pax8). The overall packing pattern between helices 1 and 3 would be affected, which would prevent the structural rearrangement into a correctly folded PAI sub-domain, and therefore impair the specific protein–DNA interaction. On the basis of these considerations, it would be expected that both the structural and the DNA-binding properties of the Ile⁵⁹Phe mutant of Pax3 should be similar to those observed with the Pax8 Leu⁶²Arg mutant.

This study was supported by grants to G.D. from Consiglio Nazionale delle Ricerche (C.N.R., Target Project on Biotechnology). L.P. was supported by the Ph.D. program 'Diagnosi e Terapie Cellulari e Molecolari' of Udine University.

REFERENCES

- 1 Stuart, E. T., Kiussi, C. and Gruss, P. (1993) *Annu. Rev. Genet.* **27**, 219–236
- 2 Noll, M. (1993) *Curr. Opin. Genet. Dev.* **3**, 595–605
- 3 Gruss, P. and Walther, C. (1992) *Cell* **69**, 719–722
- 4 Mansouri, A., Hallonet, M. and Gruss, P. (1996) *Curr. Opin. Cell. Biol.* **8**, 851–857
- 5 Treisman, J., Harris, E. and Desplan, C. (1991) *Genes Dev.* **5**, 594–604
- 6 Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. and Noll, M. (1989) *Cell* **47**, 1033–1040
- 7 Walther, C., Guenet, J. L., Simon, D., Deutsch, U., Jostes, B., Goulding, M. D., Plachov, D., Balling, R. and Gruss, P. (1991) *Genomics* **11**, 424–434
- 8 Jun, S. and Desplan, C. (1996) *Development* **122**, 2639–2650
- 9 Xu, W., Rould, M. A., Jun, S., Desplan, C. and Pabo, C. O. (1995) *Cell* **80**, 639–650
- 10 Czerny, T., Schaffner, G. and Busslinger, M. (1993) *Genes Dev.* **7**, 2048–2061
- 11 Pellizzari, L., Tell, G. and Damante, G. (1998) *Biochem. J.* **337**, 253–262
- 12 Epstein, J., Cai, J., Glaser, T., Jepeal, L. and Maas, R. (1994) *J. Biol. Chem.* **269**, 8355–8361
- 13 Spolar, R. S. and Record, Jr., M. T. (1994) *Science* **263**, 777–784
- 14 Tell, G., Scaloni, A., Pellizzari, L., Formisano, S., Pucillo, C. and Damante, G. (1998) *J. Biol. Chem.* **273**, 25062–25072
- 15 Plachov, D., Chowdhury, K., Walther, C., Simon, D., Guenet, J. L. and Gruss, P. (1990) *Development* **110**, 643–651
- 16 Zannini, M., Francis-Lang, H., Plachov, D. and Di Lauro, R. (1992) *Mol. Cell. Biol.* **12**, 4230–4241
- 17 Mansouri, A., Chowdhury, K. and Gruss, P. (1998) *Nat. Genetics* **19**, 87–90
- 18 Macchia, P. E., Lapi, P., Krude, H., Pirro, M. T., Missero, C., Chiovato, L., Souabni, A., Baserga, M., Tassi, V., Pinchera, A. et al. (1998) *Nat. Genetics* **19**, 83–85
- 19 Studier, F. W. and Moffat, B. A. (1986) *J. Mol. Biol.* **189**, 113–130
- 20 Wetlaufer, D. B. (1962) *Adv. Protein Chem.* **17**, 303–309
- 21 Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275
- 22 Bradford, M. (1976) *Anal. Biochem.* **72**, 248–254
- 23 Jun, S. and Desplan, C. (1996) *Development* **122**, 2639–2650
- 24 Pellizzari, L., Fabbro, D., Lonigro, R., Di Lauro, R. and Damante, G. (1996) *Biochem. J.* **315**, 363–367
- 25 Berg, O. G., Winter, R. B. and von Hippel, P. H. (1981) *Biochemistry* **20**, 6929–6948
- 26 Babb, R., Cleary, M. A. and Herr, W. (1997) *Mol. Cell. Biol.* **17**, 7295–7305
- 27 Wolberger, C. (1996) *Curr. Opin. Struct. Biol.* **6**, 62–68
- 28 Soejima, H., Fujimoto, M., Tsukamoto, K., Matsumoto, N., Yoshiura, K.-I., Fukushima, Y., Jinno, Y. and Niikawa, N. (1997) *Hum. Mutat.* **9**, 177–180
- 29 Oakley, B. R., Kirsch, D. R. and Morris, N. R. (1980) *Anal. Biochem.* **105**, 361–367
- 30 Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graphics* **14**, 51–55

Received 15 February 1999/12 March 1999; accepted 20 April 1999