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. Leu<sup>62</sup> is conserved among Prd domains, and contributes towards the packing together of helices 1 and 3. The binding affinity of the Leu<sup>62</sup>Arg 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 Leu<sup>62</sup>Arg mutant possess a low α-helical content; however, in the Leu<sup>62</sup>Arg 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 Leu<sup>62</sup>Arg 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 z3). This mutant (Leu<sup>62</sup>Arg) shows a reduction in DNA-binding activity [18]. Although Leu<sup>62</sup> 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 z1 and z3 [9]. In this study...
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 Leu^{62}Arg mutation abolishes the specific DNA recognition by hampering the induced fit. Thus the congenital hypothyroidism due to the Leu^{62}Arg 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 pT7.7Leu^{62}Arg Prd were generated by cloning the DNA sequences encoding for the wild-type and Leu^{62}Arg Pax8 Prd domains into the plasmid pT7.7. The DNA fragments were amplified by PCR using the oligonucleotides 5'GGAGGTTGAATGGTTGCTGCAC-3' and 3'GGAGGTGTGAATGGTTGCTGCAC-3'. All PCR reactions were performed using plasmids described previously [18] as the template. PCR products were digested with NdeI/BamHI, and cloned into the corresponding sites of pT7.7 poly linker.

**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_{590} of 0.6-0.7, and then induced by 1mM isopropyl-β-D-thiogalactoside for 3 h. Cells were harvested by centrifugation, and resuspended in lysis buffer [50mM sodium phosphate buffer (pH 7.0)/0.25 M NaCl/1mM EDTA/1mM EGTA/1mM dithiothreitol (DTT)] 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.

**Gel-retardation assay**

Gel-retardation assays were performed by incubating protein and DNA in a buffer containing 20 mM Tris/Cl, pH 7.6, 75 mM KCl, 0.25 mg/ml BSA, 5mM 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'-CAGCGCCAGCTCAAGTGTTCTTGA-3'. The BS2 site is a 20-mer, whose top-strand sequence is 5'-GGAGGTATGGTTGCTGCAC-3'. The BS2 site is a 20-mer, whose top-strand sequence is 5'-GGAGGTATGGTTGCTGCAC-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 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:

\[ K_d = \frac{[\text{Protein}][\text{DNA}]}{[\text{Protein-DNA}]} \]
CD spectroscopy

The purified Pax8 and Leu<sup>62</sup>Arg 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 Na<sub>2</sub>HPO<sub>4</sub>, 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 cm<sup>2</sup>·dmol<sup>−1</sup>·l<sup>−1</sup>), 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 Leu<sup>62</sup>Arg 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<sub>d</sub> 3.1 x 10<sup>−7</sup>M), whereas the Leu<sup>62</sup>Arg mutation resulted in an ~60-fold decrease in binding affinity to the C site (K<sub>d</sub> 1.8 x 10<sup>−8</sup>M). In contrast, the binding affinities of the wild-type and Leu<sup>62</sup>Arg proteins to the non-specific sequence (BS2) were very similar (K<sub>d</sub> 4.4 x 10<sup>−7</sup>M and 4.2 x 10<sup>−7</sup>M respectively). These findings demonstrate that the Leu<sup>62</sup>Arg 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 Leu<sup>62</sup>Arg protein would be in the bound state. This allows us to examine whether or not the Leu<sup>62</sup>Arg 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 Leu<sup>62</sup>Arg mutant respectively; DNA (the C site) was present at a concentration of 0.46 μM. Under these conditions, most of the Leu<sup>62</sup>Arg protein was bound to the C site. The CD spectrum of the Leu<sup>62</sup>Arg 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

**Figure 4** UV CD spectra of the Leu<sup>62</sup>Arg mutant and the wild-type Pax-8 Prd domain in the absence/presence of the oligonucleotide C

(A) CD spectra of the Leu<sup>62</sup>Arg 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 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 Leu\textsuperscript{14} and Ile\textsuperscript{26} side chains

Schematic structure of the PAI subdomain of the \textit{Drosophila} paired protein, achieved by using the software MOLMOL [30]. Helices (\(\alpha_1\), \(\alpha_2\) and \(\alpha_3\)) are shown as cylinders, and Ile\textsuperscript{26} and Leu\textsuperscript{54} residues are shown emboldened in black.

Figure 5 Position of Leu\textsuperscript{14} and Ile\textsuperscript{26} side chains

Schematic structure of the PAI subdomain of the \textit{Drosophila} paired protein, achieved by using the software MOLMOL [30]. Helices (\(\alpha_1\), \(\alpha_2\) and \(\alpha_3\)) are shown as cylinders, and Ile\textsuperscript{26} and Leu\textsuperscript{54} residues are shown emboldened in black.
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. Ile39 of Pax3 corresponds to Ile36 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, Ile39 faces Leu34 (Leu32 in Pax8) with a closest packing distance of 3.9 Å (0.39 nm) between heavy atoms of respective side chains, i.e. \( \approx 2 \) 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 Ile39 was replaced by a phenylalanine residue (or if an arginine replaced Leu34, as in the Leu34Arg 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 Ile39Phe mutant of Pax3 should be similar to those observed with the Pax8 Leu34Arg 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
13 Spolar, R. S. and Record, Jr., M. T. (1994) Science 263, 777–784