Co-operation between the PAI and RED subdomains of Pax-8 in the interaction with the thyroglobulin promoter

Lucia PELLIZZARI¹, Gianluca TELL¹ and Giuseppe DAMANTE²

Dipartimento di Scienze e Tecnologie Biomediche, Università di Udine, Piazzale Kolbe 1, 33100 Udine, Italy

Pax proteins are transcription factors that play an important role in the differentiation of several cell types. These proteins bind to specific DNA sequences through the paired domain. This evolutionarily conserved element is composed of two subdomains (PAI and RED), located at the N- and C-terminals, respectively. Due to the presence of these two subdomains, Pax proteins may recognize DNA in different modes, a possibility that has not been exhaustively explored yet. The C site of the thyroglobulin promoter is bound by the thyroid-specific transcription factor Pax-8. In this study we have characterized the mode by which the Pax-8 paired domain interacts with the C site. Results allow the identification of the respective positions of the PAI and RED subdomains when the full-length protein is bound to the C site. The binding of the isolated PAI and RED subdomains to the C site and to several related mutants was also evaluated. Both subdomains interact with DNA as a monomer and display a

lower binding affinity than the full-length protein. Therefore, the Pax-8 paired domain–C site interaction occurs through a co-operation between the two subdomains. The binding properties of the PAI subdomain suggest that the co-operation between PAI and RED subdomains does not merely consist of the sum of contacts established by the single subdomain: the presence of the RED subdomain is necessary for correct DNA recognition by the PAI subdomain, thus accounting for a sort of chronology of events during DNA binding. Since the RED subdomain is much more variable than the PAI subdomain among Pax proteins, these results could explain how distinct Pax proteins may select different target genes.

Key words: DNA binding, homeodomain, paired domain, transcription factor, TTF-1.

INTRODUCTION

Pax proteins form a family of transcription factors which are essential for a variety of developmental decisions during embryogenesis [1–4]. Pax proteins are defined by the presence of a 128-amino-acid-long DNA-binding domain, the paired domain (Prd domain) [5]. This element is well conserved during evolution [6] and, according to the sequence similarity in this region, Pax proteins can be grouped into distinct subfamilies [7]. Several studies have demonstrated that the Prd domain consists functionally of two distinct subdomains [8,9]. The crystal structure resolution of the Prd domain of the paired protein bound to DNA has also demonstrated the presence of two structurally independent subdomains, each containing a helix–turn–helix motif, joined by a linker region [10]. Recently, the N- and C-terminal subdomains have been named PAI and RED respectively [11].

Prd domains of different Pax proteins recognize similar consensus sequences [10,12]. However, when natural target sequences of Pax proteins are compared with respective consensus sequences, often determined by site-selection experiments *in vitro* [13], significant divergence is observed. This phenomenon suggests a high DNA-binding flexibility of the Prd domain and could be explained by compensatory protein–DNA contacts of a subdomain, when the other is not able to establish an efficient bonding network [8]. According to this view, Prd domains interact with DNA sequences in different ways. The PAI subdomain can bind as a monomer or there can be co-operation between subdomains [11]. Alternative splicing may inactivate the binding function of the PAI subdomain; in this case, DNA

sequences are recognized only through the RED subdomain, with oligomers of the protein assembled on a single DNA molecule [14,15].

The discordance existing between *in vitro*-determined consensus sequences for Pax proteins and natural sequences may also suggest that, in addition to phenomena reproduced *in vitro* by pure proteins and oligonucleotides, other events have shaped sequences of natural Pax protein binding sites. A recent hypothesis states that Pax proteins compete with other proteins to occupy DNA sites [16]. In this way, one of the reasons why Pax protein-bound sequences have diverged could be due to the need to accommodate other DNA-binding proteins.

Pax-8 is expressed in the developing kidney, some areas of the brain and in the follicular thyroid cells; in this cell type Pax-8 is also expressed in adults [17]. Pax-8 activates the promoters of the thyroglobulin (Tg) and thyroperoxidase (TPO) genes in the follicular thyroid cell, by interacting with sites located near the transcriptional start sites [18]. Both of these sites partially overlap with those of a different thyroid-specific transcription factor, TTF-1 (Figure 1). This peculiar arrangement of binding sites is present in Tg and TPO promoters of different species and, therefore, could have a functional relevance [19]. Pax-8 binding sites of Tg and TPO promoters do not completely conform to the *in vitro*-determined consensus sequence [11,12].

In this study, we have investigated how the Pax-8 Prd domain interacts with the C site of the Tg promoter. We demonstrate that the C site is recognized through a co-operation between PAI and RED subdomains. Our data suggest that the RED subdomain–C-sequence interaction is required for correct DNA recognition by the PAI subdomain. Therefore, the co-operation

Abbreviations used: Prd domain, paired domain of Pax-8; Tg, thyroglobulin; TPO, thyroperoxidase; TTF-1, thyroid-specific transcription factor; TTF-1HD, TTF-1 homeodomain; GST, glutathione S-transferase; DTT, dithiothreitol.

¹ These authors have contributed equally to the work.

² To whom correspondence should be addressed (e-mail giuseppe.damante@dstb.uniud.it).

Figure 1 Model of the Pax-8 Prd domain–C-sequence interaction and oligonucleotides used in the study

Top: schematic representation of the structure of the paired domain with the PAI and RED subdomains; α -helical regions as described by Xu et al. [10] are indicated. Bottom: all the oligonucleotides used in this study are aligned to the Pax-8 consensus sequence defined previously [11,12]. All the sequences are described in the text. Underlined bases indicate point mutations with respect to the consensus sequence. Numbering of the positions is shown at the bottom of the Figure and is the same as used in [10]. The Pax-8- and TTF-1-binding areas are indicated as defined in [16,20].

between the PAI and RED subdomains is not merely due to the sum of contacts established by single subdomains. Protein–DNA crosslinking experiments indicate that the RED subdomain binds the C site as a monomer. Since the RED subdomain is much more variable than the PAI subdomain among Pax proteins, these results could explain how distinct Pax proteins may select different target genes. Moreover, we demonstrate that the presence of a complete Pax-8 consensus on the C sequence does not allow TTF-1 binding, thus providing one of the reasons why the C sequence does not conform to the Pax-8 consensus sequence completely.

MATERIALS AND METHODS

Oligodeoxynucleotide synthesis and purification

Oligodeoxynucleotides were synthesized on an automated DNA synthesizer (Applied Biosystems, model 380B) according to standard procedures and purified by FPLC on Mono-Q column

(Pharmacia, Milton Keynes, U.K.) utilizing an ammonium bicarbonate gradient. The purity of oligonucleotides was controlled on a 20 % acrylamide/7 M urea electrophoretic gel. The top strand sequence of oligonucleotides used in this study is shown in Figure 1.

Expression vectors

The plasmids pT7.7Pax-8 Prd domain, pT7.7Pax-8 PAI and pT7.7Pax-8 RED were generated by cloning the DNA sequences coding for the Pax-8 Prd domain and the two PAI and RED subdomains into the plasmid pT7.7. The DNA fragment encoding for the Pax-8 Prd domain was amplified by PCR using the oligonucleotides Pb1, 5'-GGCGCGGATCCATCAGATC-CGGCCATGGAGGG-3', and Pb2, 5'-CCGCGGGATCCGG-GGAGGTTGAATGGCTGCTG-3'.

The DNA fragment encoding for the PAI subdomain was obtained by PCR using the oligonucleotides Pb1 and Pb4 (5'-CCGCGGGATCCCACTCCAGGCCGGATGCTGCC-3'), and the DNA encoding for the RED subdomain was obtained by

Figure 2 Primary structure and purification of the Pax-8 Prd domain, and PAI and RED subdomains

(a) Schematic structure of the Pax-8 Prd domain, and PAI and RED subdomains. (b) SDS/PAGE (15%), stained with Coomassie Brilliant Blue, of Pax-8 Prd domain. Lane 1, molecular-mass markers (MW; phosphorylase b, 94 kDa; albumin, 67 kDa; ovalbumin, 43 kDa; carbonic anhydrase, 30 kDa; trypsin inhibitor, 20.1 kDa; α -lactalbumin, 12 kDa); lane 2, Pax-8 Prd (PRD) domain after ion-exchange chromatography purification (Mono-S, Pharmacia), with 5 μ g of protein. (c) Purified PAI and RED subdomains, run under the same experimental conditions as (b).

PCR using the oligonucleotides Pb2 and Pb3 (5'-GGCGCG-GATCCGTGATAGGGGGCTCCAAGCCC-3'). All PCR reactions were performed using the plasmid CMV-Pax8 kindly provided by P. Gruss (Max-Planck Institute of Biophysical Chemistry, Göttingen, Germany) as a template. PCR products were digested with *Bam*HI and cloned into the *Bam*HI site of the pT7.7 polylinker. Due to this cloning strategy, all these proteins contained 9 and 21 extra amino acids at the N- and C-termini, respectively (see Figure 2a). In order to generate the plasmid expressing the GST (glutathione S-transferase)-RED fusion protein, the DNA fragment encoding for the RED subdomain was cloned into the *Bam*HI site of the pGEX-2T vector (Pharmacia). All plasmids were controlled by nucleotide

sequencing. The expression vector of the GST-Prd protein was the plasmid pGEX-Pax8 [12].

Protein expression and purification

Plasmids were used to transform the BL21 bacterial strain [20]. Transformed cells were grown at 37 °C to D_{600} 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 volume of 10 ml/g of bacterial pellets. After cell lysis by sonication, the

bacterial debris was removed by centrifugation. DNA was removed by addition of protamine sulphate to the supernatant to 0.3 mg/ml and the precipitate removed by centrifugation. The supernatant was then loaded on to a Mono-S column (Pharmacia) pre-equilibrated with lysis buffer. The Pax-8 Prd domain and the PAI and RED subdomains 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. The purified proteins gave a single band on an overloaded SDS (15 %)/polyacrylamide gel. Fractions containing purified proteins were dialysed against water and stored at -80°C . The Pax-8 Prd domain, and PAI and RED subdomain concentrations were determined using molar absorption coefficients of 9890, 2800 and 7090 $\text{M}^{-1}\cdot\text{cm}^{-1}$ respectively, calculated according to Wetlauffer [21], and confirmed by the colorimetric assays of Lowry et al. [22] or Bradford [23]. Absorbance measurements were obtained using a Perkin-Elmer Lambda 14 UV-visible spectropolarimeter. GST-Prd and GST-RED fusion proteins were purified according to described procedures [24].

CD

The purified Pax-8 Prd domain and the PAI and RED subdomains were used for CD spectroscopy at a concentration of 1.4 μM using a Jasco J-600 CD/ORD spectropolarimeter interfaced to an Olidata computer for data collection. Standard conditions were 150 mM KCl/10 mM Na_2HPO_4 (pH 7.4), 25°C , 1 cm path-length cuvette. For UV CD, spectra are presented in terms of mean residue molecular ellipticity ($[\theta]$; $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$), based on a mean residue weight of 110.4. In thermal-denaturation studies, the solution was cooled from room temperature to 4°C and allowed to equilibrate for 30 min; then the temperature was raised in 10°C increments up to 90°C . Every time the solution was allowed to equilibrate for 30 min, and its CD spectra were recorded ten times and averaged. The reported results are the means smoothed over ten measurements. Temperature was controlled by a Haake F3 water bath.

Gel-retardation assays and quantification of binding affinity

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, 10 $\mu\text{g}/\text{ml}$ calf thymus DNA and 10 % glycerol for 30 min at room temperature. In the experiments shown in Figures 3, 6 and 8 (see later), calf thymus DNA was omitted. Oligonucleotides used as probes were labelled at the 5' end using polynucleotide kinase and [γ - ^{32}P]ATP and annealed with the complementary strand. At the end of the binding reaction, samples were loaded on to a 7.5 % native polyacrylamide gel and run in $0.5\times\text{TBE}$ (where $1\times\text{TBE} = 45\text{ mM Tris/borate}/1\text{ mM EDTA}$) at 8°C . The relative binding affinity by which the full-length Pax-8 Prd domain, as well as the isolated PAI and RED subdomains, binds to the different oligonucleotides was calculated as follows. Known amounts of proteins (1.5×10^{-8} , 6.5×10^{-8} and 5.0×10^{-8} M for the Pax-8 Prd domain, the PAI and the RED subdomains, respectively) and oligonucleotides ($5\times 10^{-9}\text{M}$) were used in gel-retardation assays. Values of protein-bound and free DNA were quantified by scanning densitometry of autoradiograms by an LKB-Pharmacia (Uppsala, Sweden) laser densitometer. The affinity of each protein–DNA complex was calculated as follows:

$$K_a = \frac{[\text{Protein}]\cdot[\text{DNA}]}{[\text{Protein} - \text{DNA}]}$$

Values were expressed as a fraction of the Pax-8 Prd domain–C-sequence interaction, which was arbitrarily considered to be 1.

Protein–DNA crosslinking

The monomeric binding of the Pax-8 Prd domain and RED proteins on the C sequence was determined by UV crosslink analysis, following the procedure of Molnar et al. [25]. Briefly, proteins and DNA were incubated as described for gel-retardation assay but without glycerol. After 30 min, aliquots were either loaded on to a native polyacrylamide gel to detect the protein–DNA complex or subjected to UV crosslink analysis (2.2 J, in a GS Gene Linker; Bio-Rad, Hemel Hempstead, Herts., U.K.). After UV exposure, samples were loaded on to a 12 % SDS/polyacrylamide gel. Following electrophoresis, the gel was stained by Coomassie Brilliant Blue, to detect molecular-mass markers, and exposed to autoradiography without drying.

RESULTS AND DISCUSSION

In Figure 1, the Pax-8 binding site of the Tg promoter (C sequence) and the other sequences used in this study are aligned with the *in vitro*-determined Pax-8 consensus sequence, derived from previous studies [11,12]. On the C sequence, both TTF-1 and Pax-8 binding areas (mapped by phenanthroline footprinting and methylation interference, see [18,26]) are also shown. Based on the crystal structure of the paired Prd domain–DNA complex and on our previous studies [10,12], the docking of the Pax-8 Prd domain with the C sequence is attempted. In this model, bases 7–15 (approx.) could be contacted by the PAI subdomain, while bases 16–24 could be contacted by the RED subdomain.

In order to test the model shown in Figure 1, the full-length Pax-8 Prd domain, as well as the isolated PAI and RED subdomains, were expressed in bacteria and purified by ion-exchange column chromatography. In Figure 2, the primary structure of the three proteins are shown (Figure 2a), as well as the high degree of purification obtained (judged by SDS/PAGE, Figures 2b and 2c).

Binding of the Pax-8 Prd domain, and PAI and RED subdomains to deletions of the C sequence

The DNA-binding properties of the isolated subdomains, as well as those of the full-length Pax-8 Prd domain, were evaluated on C_{14} and C_{15} oligonucleotides (Figure 3). As predicted from the model, the RED subdomain efficiently binds to the C_{15} sequence but not to the C_{14} sequence. The PAI subdomain binds to C_{14} oligonucleotides. However, it binds also to C_{15} oligonucleotide, but with lower affinity. The binding properties of the subdomains confirm the model depicted in Figure 1. Moreover, the results indicate that the PAI subdomain possesses only a limited capability to discriminate among different DNA sequences.

Effect of the introduction of Pax-8 consensus sequence in the context of the C sequence

In the C sequence, 5 out of 15 bases do not conform with the consensus sequence (Figure 1). Thus, several mutants of the C sequence were synthesized, each including additional base(s) conforming with the Pax-8 consensus sequence. In the $\text{C}\pi$ oligonucleotide the complete Pax-8 consensus is introduced, whereas in $\text{C}\beta$ only 2 bases are changed, and in $\text{C}\beta 9$ and $\text{C}\beta 11$

Figure 3 Binding activity of the Pax-8 Prd domain, and the PAI and RED proteins on C_{14} and C_{15} oligonucleotides

Gel-retardation assays were performed by using C_{14} and C_{15} oligonucleotides as probes, the sequences of which are shown in Figure 1. The final concentrations of proteins were 1.5×10^{-8} , 6.5×10^{-8} and 5.0×10^{-8} M for the Pax-8 Prd domain, and the PAI and RED subdomains, respectively. The DNA probes were used at the final concentration of 5×10^{-9} M. B, protein-bound DNA; F, free DNA.

only 1 base is changed. If the model shown in Figure 1 is correct, because of the optimization of the PAI subdomain-DNA contacts these oligonucleotides should bind to the Pax-8 Prd domain with higher affinity (proportionally to the number of bases changed) than the wild-type C site. The binding of either the Pax-8 Prd domain or the TTF-1 homoeodomain (TTF-1HD) to these mutants was evaluated by gel-retardation assay. Representative gels are shown in Figure 4 and relative affinities of the

Table 1 Relative K_d of the full-length Pax-8 Prd domain, and PAI and RED subdomains with the C sequence and related mutants

Protein	Oligonucleotide ...	K_d				
		C	$C\pi$	$C\beta$	$C\beta9$	$C\beta11$
Prd domain		1	0.12	0.32	0.40	1.12
PAI		5.4	8.3	9.2	8.8	9.8
RED		14.7	14.0	16.0	15.2	15.5

Pax-8 Prd domain to the different sequences are shown in Table 1. The optimal binding site for the Pax-8 Prd domain is on the $C\pi$ oligonucleotide, which shows a relative affinity about 8-fold higher than the wild-type C sequence. This difference is likely due to optimization of contacts established by helix 3 and the β turn of the PAI subdomain with base pairs 4–6 and 9–11, respectively [10]. Due to optimization of the β turn-DNA contacts, complexes established with the $C\beta$ and $C\beta9$ oligonucleotides also show affinities higher than the Prd domain-C sequence complex. The $C\beta11$ oligonucleotide is bound with the same affinity as the C sequence, indicating that the base pair at position 11 does not play a significant role in the Prd domain-C-sequence interaction. Above all, a good relationship exists between the number of bases conforming with the Pax-8 consensus and the affinity by which the Pax-8 Prd domain binds to sequences. Therefore, these results are highly compatible with the model shown in Figure 1. In contrast to the Pax-8 Prd domain, TTF-1HD binds to the $C\beta9$ and $C\beta11$ oligonucleotides very weakly and it is not able to recognize the $C\pi$ and $C\beta$ sequences at all (Figure 4). Therefore, in the TTF-1 binding area, any attempt to conform the C sequence to the Pax-8 consensus abolishes TTF-1HD binding.

The binding of either the PAI or RED subdomains was also evaluated on the full-length C sequence and relative mutants. Representative autoradiograms are shown in Figure 5. Both PAI and RED subdomains interact with the C sequence at lower affinity than the full-length Pax-8 Prd domain (Table 1). This suggests the occurrence of a DNA-binding co-operation between subdomains during the interaction of the full-length protein. The RED subdomain binds with similar affinity to all tested oligonucleotides. These results are predicted by the model of Figure 1; in fact, none of the mutated oligonucleotides has changes in

Figure 4 Binding activity of the Pax-8 Prd domain and TTF-1HD to mutants of the C sequence

Gel-retardation assay showing the interaction of the sequences described in Figure 1 with the Pax-8 Prd domain and TTF-1HD. For each protein, two final concentrations were used (0.5×10^{-8} and 1.5×10^{-8} M). The DNA probes were used at the final concentration of 5×10^{-9} M. B, protein-bound DNA; F, free DNA.

Figure 5 Binding activity of the Pax-8 Prd domain, and PAI and RED proteins to mutants of the C sequence

Gel-retardation assay showing the interaction of the sequences described in Figure 1 with the Pax-8 Prd domain and the PAI and RED subdomains. The final concentrations were as for Figure 3. The DNA probes were used at the final concentration of 5×10^{-9} M. B, protein-bound DNA; F, free DNA. Note that the PAI- $C\pi$ and the PAI- $C\beta$ complexes show a slower mobility than the PAI-C complex.

Figure 6 Sequential incubation of PAI and RED subdomains in the same binding mixture

Gel-retardation assay in which the PAI and RED subdomains were incubated sequentially with the C-sequence oligonucleotide. Lane 1, free DNA. In lanes 2 and 3, PAI and RED subdomains, respectively, were used alone. In lane 4 the adding order was: RED, C oligonucleotide, PAI. In lane 5 the adding order was: PAI, C oligonucleotide, RED. After addition of C oligonucleotide, samples were incubated for 10 min at room temperature. Both proteins were used at the final concentration of 2.0×10^{-8} M. C oligonucleotide was used at the final concentration of 5×10^{-9} M.

positions that are contacted by the RED subdomain. Surprisingly, the PAI subdomain binds to mutant oligonucleotides with lower affinity than that displayed with the wild-type C sequence. A possible explanation of this phenomenon could stem from the observation that the PAI- $C\pi$ and PAI- $C\beta$ complexes show slower mobility than the PAI-C complex (Figure 5), suggesting that the PAI subdomain interacts with $C\pi$ and $C\beta$ sequences in a different manner to that used to recognize the C sequence. These results indicate that the binding properties that the PAI subdomain shows in isolation are partially different

from those that exist in the context of the full-length Prd domain. Thus, the presence of the RED subdomain appears to be necessary for a 'normal' PAI-C-sequence interaction. The abnormality of the PAI-C-sequence interaction, when the subdomain is isolated, is confirmed by sequential incubation experiments, adding RED subdomain to DNA and then PAI, or vice versa. PAI and RED were used at equimolar concentrations; the results are shown in Figure 6. When either protein is used in isolation, because of the higher affinity of the PAI-C-sequence interaction the PAI subdomain binds to more than 99 % of the probe, whereas the RED subdomain binds only to 70 % (Figure 6, lanes 2 and 3). When proteins are used in the same binding mixture and RED is added to DNA before PAI (Figure 6, lane 4), a low-mobility band, indicating the presence of a trimeric complex (PAI-RED-C), is evident. A band corresponding to the PAI-C complex and a smeared signal above are also evident. No band corresponding to the RED-C complex is detectable. The smeared signal, going from the low-mobility band to the PAI-C-complex band, suggests that the PAI-C interaction stops the pre-existing RED-C interaction. In fact, the smeared signal would be caused by loss of the RED subdomain from the trimeric complex during gel migration. The fact that the RED-C interaction is weakened by the PAI-C interaction is confirmed when the adding order of the sequential incubation is reversed. In fact, when PAI is added before RED, only a band corresponding to the PAI-C complex is detected (Figure 6, lane 5). The absence of other signals indicates the inability of the RED subdomain to interact efficiently with the C sequence when the isolated PAI subdomain is already bound. Thus, sequential incubation experiments further suggest that, when PAI subdomain is used in isolation, its interaction with C site is different from that in the context of the wild-type Prd domain.

The quantitative evaluation of the DNA-binding affinity by which the isolated PAI and RED subdomains interact with the C sequence and relative mutants (Table 1) indicates that the PAI subdomain binds with higher affinity than the RED subdomain. However, results shown in Figure 3 would indicate that the capability to discriminate subsequences in the context of the C site is higher for the RED than for the PAI subdomain. Altogether, our results allow us to hypothesize that there is a sort of 'chronology of events' in the binding to the C site. The RED subdomain would initially recognize the sequence, setting the specificity of the interaction; subsequently the PAI sub-

domain would be in the correct position to establish DNA contacts, increasing the affinity of the whole Prd domain–DNA complex.

Structural characterization of the Pax-8 Prd domain

CD data obtained with the Pax-8 Prd domain revealed it to be largely structureless in solution, with a dramatic α -helical gain occurring upon DNA binding [9]. In this view, the model that we propose for the Pax-8–C interaction requires a structural folding independence between the PAI and RED subdomains. Such an issue was investigated by CD analysis of the Pax-8 Prd domain. The measure of the α -helical content of the Pax-8 Prd domain free in solution was performed first. The collected spectrum indicates the presence of α -helical regions, as seen from the two relative minima at 208 and 222 nm (Figure 7, top left panel). The helix content is estimated from the mean residue ellipticity at 222 nm, assuming that for a 100% helical peptide this value is $-33000 \text{ deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$ at 0 °C, and at higher temperature is attenuated by 0.3%/°C before unfolding [27]. In this way the calculated α -helical content for Pax-8 Prd domain is 19% at 4 °C. Therefore, a significant amount of α -helix is present in this protein when it is free in solution, indicating that the Prd domain approaches DNA in a partially folded state and not in a random-coil conformation as previously reported [9]. This view could suggest that the α -helical gain induced by DNA binding [28] would occur on a pre-existing structural fold. In this way, the gain in negative entropy during DNA interaction would be minimized, thus the ΔG of the interaction would be more favourable with respect to a situation in which the unbound protein is in a random-coil conformation.

In order to test for the presence of a structural independence in folding between the PAI and RED subdomains within the whole Prd domain, a thermostability study was performed by recording CD spectra at several temperatures. The absence of an isodichroic point (Figure 7, top left panel), together with the profile of the denaturation curve (Figure 7, top right panel) obtained by plotting the $[\theta]_{222}$ values at several temperatures, suggests a lack of co-operation in folding between the two subdomains, which is therefore indicative of a structural independence between the PAI and RED subdomains.

CD analysis of the isolated PAI and RED subdomains reveals that both polypeptides possess a limited amount of α -helical content (Figure 7, middle panel), demonstrating their partial folding in the free state.

CD analysis was also used to study the Prd Pax-8–C interaction (Figure 7, bottom left panel). As indicated above, in the absence of DNA the α -helical content of protein was 19%. The addition of the C sequence ($r = 0.7$) induces an increase (Figure 7, bottom left panel, dotted line) of α -helix (signal at 222 nm). The α -helical content increases from 19 to 33% upon interaction with the C sequence. To verify if structural modifications of DNA could occur as a result of interaction with the Prd domain, CD spectra of the Prd domain–C-sequence complex were recorded and compared with those of the C-sequence oligonucleotide alone. Results are shown in Figure 7 (bottom right panel). Following the interaction with the Prd domain, the long-wavelength positive CD band of DNA decreases with a shift in the 258 nm crossover to a longer wavelength. These modifications are likely due to significant structural alterations of the C-sequence oligonucleotide, since the CD spectrum of the protein alone shows no bands of absorption in the region above 250 nm (results not shown). Similar CD changes have been observed in compaction of the DNA into particles *in vitro*, into phage heads or into nucleosomes [29,30]

The RED subdomain binds to the C sequence as a monomer

By using sequences obtained through selection procedures, it has been demonstrated that the RED subdomain can bind DNA in an oligomeric state [14,15]. However, in the model hypothesized above, the RED subdomain should recognize the C site in a monomeric state; only in this form would the PAI subdomain then be able to interact with DNA and thus increase the affinity of the whole Prd domain–C-sequence complex. In our gel-retardation-assay experiments, the RED–C complex shows a mobility that is intermediate between the PAI–C and the Prd domain–C complexes. These mobility patterns, though suggestive, do not allow us to conclude that the RED–C interaction occurs with the protein in a monomeric form. In order to demonstrate the monomeric state of the RED–C interaction, a protein–DNA crosslinking procedure was utilized. Several different proteins were used: Prd domain, PAI and RED subdomains, and GST–Prd and GST–RED chimaeric proteins, in which the Prd domain and the RED subdomain are fused respectively to the GST protein. These molecules were incubated with the C sequence to allow the interaction. The mixture was subjected to UV crosslinking [25] and samples were run on an SDS/polyacrylamide gel (Figure 8, left panel). The free DNA migrates as a band of about 8 kDa (Figure 8, left panel, lane 1). As expected, due to the monomeric nature of the Prd domain–DNA interaction, the Prd domain–DNA complex migrates between 20 and 30 kDa (Figure 8, left panel, lane 2). PAI–C and RED–C complexes migrate faster than the Prd domain–DNA complex (Figure 8, left panel, lanes 3 and 4), with molecular masses of approx. 18 kDa, the sum of the molecular masses of the oligonucleotide (8 kDa) in addition to a single molecule of either the PAI or RED subdomains (10 kDa each). These results allow us to conclude that the RED subdomain interacts with the C sequence in a monomeric form. This conclusion is strengthened by the observation that the GST–RED protein crosslinked to the C site migrates at approximately 47 kDa (Figure 8, left panel, lane 5), just below the GST–Prd protein, which migrates at about 53 kDa (Figure 8, left panel, lane 6). Since the GST protein has a molecular mass of 29 kDa, the molecular masses reported above are the ones expected in the case of monomeric interaction.

The monomeric nature of the RED–C interaction has also been evaluated by a different approach. RED and GST–RED proteins were incubated alone or together and then used in a gel-retardation assay. Results are shown in Figure 8 (right panel). When mixed together, RED and GST–RED proteins produce complexes with mobility identical to those observed when proteins are used alone. No additional bands are observed. Signals with intermediate mobility were instead expected in the case of dimeric–oligomeric interaction with DNA, due to hybrid complexes containing RED and GST–RED proteins. Note that autoradiograms shown in Figure 8 (right panel) have been greatly overexposed to detect even faint signals.

Conclusions

The C site of the Tg promoter was chosen as an example of a natural binding sequence for two reasons: (i) this sequence has been extensively studied and characterized in great detail for both Pax-8 and TTF-1 binding [12,31]; and (ii) Pax-8 plays a critical role in Tg gene expression [32].

From our data, the occurrence of a DNA-binding co-operation between the PAI and RED subdomains during the interaction of the Prd domain with the C sequence can be inferred. However, the co-operation between the two subdomains during DNA binding does not merely consist of the sum of the contacts established by the two elements; the presence of the RED

Figure 7 For legend see opposite page.

Figure 8 The RED subdomain interacts with the C sequence as a monomer

(Left panel) Protein–DNA-crosslink analysis was performed as described by Molnar et al. [25]. After UV crosslinking (2.2 J in a GS Gene Linker, Bio-Rad) samples were run on 12% SDS/PAGE. The gel was then stained and exposed to autoradiography without drying. (Right panel) Gel-retardation assay of RED and GST–RED proteins used alone or in combination. When in combination, proteins were mixed together in the gel-retardation mixture and incubated for 10 min at room temperature. Then the C oligonucleotide was added at a final concentration of 5×10^{-9} M and incubation continued for an additional 30 min before loading on to a native polyacrylamide gel.

subdomain is necessary to allow a correct DNA recognition by the PAI subdomain. Of note is that the C sequence of the Tg promoter is the first natural sequence recognized by an isolated RED subdomain. Moreover, the RED subdomain binds to the C sequence as a monomer. The only RED–DNA interaction reported so far consists of a complex between an artificial sequence and RED subdomain oligomers [14,15]. Therefore, the Pax-8–C-sequence interaction provides a new way by which paired domains interact with DNA, further underscoring the flexibility of these structures in DNA recognition.

As previously noted [7,12,33], the high degree of sequence similarity of the PAI subdomain among different Prd domains would lead to the recognition of similar sequences by Pax proteins, raising the question of how distinct DNA-binding specificity is achieved *in vivo*. Among Pax proteins the RED subdomain is much more variable than the PAI subdomain. The model that we propose, since indicating the RED subdomain–DNA interaction as the initial event during the Prd domain–DNA docking, may suggest that the more divergent RED subdomain might play a relevant role in the discrimination of target genes.

From a structural point of view, our CD data indicate that reduced levels, but not an absence, of secondary structure in the free state is a characteristic of Prd domains. Both isolated PAI

and RED subdomains show partial structural folding in the free state. Therefore, the α -helical content that the Pax-8 Prd domain shows in the free state (19 % at 4 °C) is contributed to by both subdomains. Spectra recorded in the presence of DNA indicate that both the Prd domain and the C sequence undergo conformational changes upon binding, further revealing the complexity of this protein–DNA interaction.

In a previous work, by using *in vitro* site-selection experiments, we and other groups have determined the consensus sequence for the Pax-8 Prd domain [11,12]. Similarly to other Pax proteins, natural Pax-8 binding sequences significantly diverge from the *in vitro*-determined consensus. These differences suggest that during evolution, in addition to the high-affinity binding (reproduced by the *in vitro* selection), other events have shaped the sequence of Pax-protein natural binding sites. Our data demonstrate that one of the reasons by which the C sequence does not completely conform to the Pax-8 consensus obtained by *in vitro* selection (a procedure in which the Pax-binding preference is the only driving force of evolution) is the necessity for TTF-1 binding.

We thank P. Gruss for providing the CMV-Pax-8 plasmid and E. Tell for improvement to the English. This study is supported by grants to G. D. by the Associazione Italiana Ricerca Cancro (A.I.R.C.) and Consiglio Nazionale delle Ricerche (C.N.R., Target

Figure 7 Far-UV CD analysis of the Pax-8 Prd domain, and PAI and RED subdomains

(Top left panel) CD spectrum of Pax-8 Prd domain. Ten spectra were acquired in 10 mM Na_2HPO_4 /150 mM KCl (pH 7.4) at 25 °C. The protein concentration was 1.4 μM and the path length was 1 cm. (Top right panel) Thermal-denaturation curve of the Pax-8 Prd domain obtained from the ellipticity at 222 nm ($[\theta]_{222}$) versus temperature. The protein concentration and path length were as above. (Middle panel) CD spectra of the PAI (○) and RED (●) subdomains (1.4 μM) recorded in 10 mM Na_2HPO_4 /150 mM KCl (pH 7.4) at 25 °C. The inset shows the helical content estimated from the mean residue ellipticity at 222 nm assuming that for a 100% helical peptide this value is $-33\,000 \text{ deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$. (Bottom left panel) CD spectra of the Prd domain of Pax-8 (0.33 μM) in the absence (unbroken line) and the presence (dotted line) of 0.4 μM C-sequence oligonucleotide after correction for the contribution of the oligonucleotide. Spectra were recorded in 10 mM Na_2HPO_4 /75 mM KCl (pH 7.4) at 10 °C. (Bottom right panel) Near-UV CD spectrum of 0.4 μM C-sequence oligonucleotide in the presence (dotted line) or absence (unbroken line) of 0.33 μM Pax-8 Prd domain ($r = 0.7$), with correction for the contribution of the protein. The ellipticity values are given in terms of the molar concentration of oligonucleotide. Spectra were recorded as for the bottom left panel.

Project on Biotechnology). L.P. is supported by the Ph.D. programme '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 Czerny, T., Schaffner, G. and Busslinger, M. (1993) *Genes Dev.* **7**, 2048–2061
- 9 Epstein, J., Cai, J., Glaser, T., Jepeal, L. and Maas, R. (1994) *J. Biol. Chem.* **269**, 8355–8361
- 10 Xu, W., Rould, M. A., Jun, S., Desplan, C. and Pabo, C. O. (1995) *Cell* **80**, 639–650
- 11 Jun, S. and Desplan, C. (1996) *Development* **122**, 2639–2650
- 12 Pellizzari, L., Fabbro, D., Lonigro, R., Di Lauro, R. and Damante, G. (1996) *Biochem. J.* **315**, 363–367
- 13 Thiesen, H. J. and Bach, C. (1990) *Nucleic Acid Res.* **18**, 3203–3209
- 14 Epstein, J. A., Glaser, T., Cai, J., Jepeal, L., Walton, D. S. and Maas, R. L. (1994) *Genes Dev.* **8**, 2022–2034
- 15 Kozmik, Z., Czerny, T. and Busslinger, M. (1997) *EMBO J.* **16**, 6793–6803
- 16 Read, A. P. (1995) *Nat. Genet.* **9**, 333–334
- 17 Plachov, D., Chowdhury, K., Walther, C., Simon, D., Guenet, J. L. and Gruss, P. (1990) *Development* **110**, 643–651
- 18 Zannini, M., Francis-Lang, H., Plachov, D. and Di Lauro, R. (1992) *Mol. Cell. Biol.* **12**, 4230–4241
- 19 Damante, G. and Di Lauro, R. (1994) *Biochim. Biophys. Acta* **1218**, 255–266
- 20 Studier, F. W. and Moffat, B. A. (1986) *J. Mol. Biol.* **189**, 113–130
- 21 Wetlaufer, D. B. (1962) *Adv. Protein Chem.* **17**, 303–309
- 22 Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275
- 23 Bradford, M. (1976) *Anal. Biochem.* **72**, 248–254
- 24 Smith, D. B. and Johnson, K. S. (1988) *Gene* **67**, 31–40
- 25 Molnar, G., O'Leary, N., Pardee, A. B. and Bradley, D. W. (1995) *Nucleic Acid Res.* **23**, 3318–3326
- 26 Damante, G., Fabbro, D., Pellizzari, L., Civitavecchia, D., Guazzi, S., Polycarpou-Schwartz, M., Cauci, S., Quadrioglio, F., Formisano, S. and Di Lauro, R. (1994) *Nucleic Acid Res.* **22**, 3075–3083
- 27 Chen, Y. H., Yang, J. T. and Chau, K. H. (1974) *Biochemistry* **13**, 3350–3359
- 28 Spolar, R. S. and Record, Jr., M. T. (1994) *Science* **263**, 777–784
- 29 Dorman, B. P. and Maestre, M. F. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 255–266
- 30 Mandel, R. and Fasman, G. D. (1976) *Nucleic Acids Res.* **3**, 1839–1848
- 31 Damante, G., Pellizzari, L., Esposito, R., Fogolari, F., Viglino, P., Fabbro, D., Tell, G., Formisano, S. and Di Lauro, R. (1996) *EMBO J.* **15**, 4992–5000
- 32 van der Kallen, C. J., Spierings, D. C., Thijssen, J. H., Blankenstein, M. A. and de Bruin, T. W. (1996) *J. Endocrinol.* **150**, 377–382
- 33 Vogan, K. J. and Gross, P. (1997) *J. Biol. Chem.* **272**, 28289–28295

Received 16 June 1998/8 October 1998; accepted 2 November 1998