Domain Architecture of a Bacterial HMGA-like Protein

Domain Architecture of an HMGA-Type Bacterial Transcriptional Factor*

S. Padmanabhan‡¶, M. Elías-Arnanz‡, E. Carpio†#, P. Aparicio†, and F.J. Murillo‡¶

Departamento de ‡Genética y Microbiología and †Area de Inmunología,
Universidad de Murcia, Murcia, Spain.

* Funded by grants PB96-1096 (DGICYT-MEC-Spain) and BMC2000-1006 (DGICYT-MCYT-Spain) to FJM, and PM98-0052 (DGICYT-MEC-Spain) to PA. SP was supported in parts by MEC-Spain and MCYT-Spain.

¶ To whom correspondence should be addressed: Phone: 34-968-364 951; FAX: 34-968 363 963; e_mail: padhu@um.es; araujo@um.es.

Running title. Domain Architecture of a Bacterial HMGA-like Protein
SUMMARY

*Myxococcus xanthus* transcriptional factor CarD participates in carotenogenesis and fruiting-body formation. It is the only reported prokaryotic protein having adjacent “AT-hook” DNA-binding and acidic regions characteristic of eukaryotic high mobility group A (HMGA)\(^1\) proteins. The latter are small, unstructured, non-histone nuclear proteins that function as architectural factors to remodel DNA and chromatin structure and modulate various DNA-binding activities. We find CarD to be predominantly dimeric with two stable domains: (a) an N-terminal domain of defined secondary and tertiary structure which is absent in eukaryotic HMGA proteins; (b) a C-terminal domain formed by the acidic and AT-hook segments and lacking defined structure. CarD, like HMGA proteins, binds specifically to the minor-groove of AT-rich DNA present in two appropriately spaced tracts. As in HMGA proteins, casein kinase II can phosphorylate the CarD acidic region, and this dramatically decreases the DNA-binding affinity of CarD. The acidic region, in addition to modulating DNA-binding, confers structural stability to CarD. We discuss how the structural and functional plasticity arising from domain organization in CarD could be linked to its role as a general transcriptional factor in *M. xanthus*. 
INTRODUCTION

Transcription, replication, recombination and repair are mediated by the assembly of specific nucleoprotein complexes. One essential consequence of this nucleoprotein complex formation is to confer local flexibility to the inherently stiff DNA molecule (1). Various factors guide the specific assembly of these complexes including proteins referred to as architectural factors because they remodel DNA by specific or nonspecific DNA-binding (2). Given their fundamental importance, architectural proteins are found in phages, prokaryotes and eukaryotes. Examples in phages and bacteria include the proteins HU, H-NS, IHF and phage φ29 p6 (see (3) and references therein). In eukaryotes, in addition to histones, the abundant non-histone chromosomal proteins of the HMG family constitute an important group of architectural transcription factors that regulate gene expression and are implicated in a variety of cellular functions (4-6).

The subfamily of HMGA proteins (previously HMGI(Y), (7)) is characterized by the presence of multiple repeats of a conserved RGRP sequence (the "AT-hook" motif) embedded in a less conserved cluster of basic and proline residues (8). The most extensively studied, the mammalian HMGA isoforms, are small proteins (≤ 107 residues; ~12 kDa) with three AT-hooks lying between a highly acidic C-terminal stretch of about fifteen residues and a short N-terminal region of less than twenty-five residues of variable sequence (9; Figure 1A). The AT-hooks bind specifically to the narrow minor groove of AT-rich sequences four to eight base pairs in length occurring in at least two or three appropriately spaced tracts (10,11). The unstructured AT-hooks adopt a defined conformation on binding DNA (12), and their DNA-binding specificity is modulated by the acidic domain (13). Protein conformational changes and decreases in DNA-binding affinity are also brought about by phosphorylation by a number of kinases including CKII, Cdc2 kinase, MAP kinase and...
protein kinase C (14-18). This provokes fluctuations in intracellular protein stability and in DNA-binding thereby fine-tuning their regulatory functions in vivo (16, 18).

The first and, to our knowledge, only prokaryotic protein with multiple AT-hooks and a flanking highly acidic region is the 316 amino acid (34 kDa) product of gene carD in the bacterium *Myxococcus xanthus* (19). Protein CarD is involved in regulating at least two distinct processes in *M. xanthus* (20). It is required for the expression of two different sets of genes that form part of a complex network regulating light-induced carotenogenesis—“the light regulon” (reviewed in 21). In addition, CarD also participates in the starvation-induced formation of fruiting bodies where thousands of cells cluster and subsequently undergo cellular differentiation to myxospores (reviewed in 22). Mutations in *carD* prevent fruiting body formation and block the expression of several developmentally activated genes (20). Thus CarD, like mammalian HMG A proteins, is recruited by distinct gene regulatory circuits in vivo.

In contrast to mammalian HMG A proteins, CarD has one more AT-hook, and its significantly longer acidic region is situated to the N- rather than to the C-terminus of the AT-hook region (Figure 1B). Moreover, the segment containing the acidic and AT-hook portions of CarD is preceded by a considerably larger N-terminal stretch of around 180 amino acids, whose equivalent is absent in eukaryotic HMG A proteins. This N-terminal segment of unknown function includes in its sequence a stretch of heptad repeats of the type found in leucine-zipper coiled-coils (19). Several CKII phosphorylation sites are predicted in the acidic and N-terminal regions of CarD, as also some protein kinase C sites (Fig 1). Equivalent kinases have not thus far been specifically assigned in *M. xanthus*, but it does contain several eukaryotic-like serine/threonine protein kinases (23), and at least one
functionally linked phosphatase which have been proposed to act in concert in the developmental cycle of the bacterium (24).

The focus of the present study is to characterize the structural and functional domain organization in CarD. We have mapped out the domain architecture in CarD by means of biochemical and spectroscopic analyses of the pure protein and several specifically designed fragments, and have probed these for oligomerization, DNA-binding and phosphorylation. The properties thus dissected in CarD are discussed using the reported properties of the considerably smaller mammalian HMGA1a as a benchmark. Based on our findings we hypothesize on how the modular organization in CarD affords the structural stability and malleability that may underlie its regulatory roles in *M. xanthus*.

**EXPERIMENTAL PROCEDURES**

*Cloning of carD and its fragments into expression vectors*

Several constructs were generated for the production of CarD and the following truncated versions: CD(1-104), CD(183-316), CD(225-316), CD(247-316), CD(Δ181-223), and CD(1-215), the numbers referring to amino acid positions in CarD. The expression vector pET11b was used for production of untagged proteins and pET15b for production of His$_6$-tagged ones (25). CarD and the indicated fragments, with the exception of CD(1-215), were obtained by PCR amplification and cloned into the *Nde*I – *Bam*HI sites of the vectors. The *Nde*I site introduces a non-native Met for bacterial expression of the fragments CD(183-316), CD(225-316) and CD(247-316). The procedure to obtain a construct expressing CD(1-215) was as follows. pET11-*carD* was cut with *Bgl*II-*Bam*HI, the fragment containing *carD* was digested with *Hinfl* (which cuts at a position corresponding to the codons for residues 215 and 216), filled with Klenow and cut with *Nde*I. This fragment was cloned into pET11b, first cut with *Bam*HI, filled with Klenow, then cut with *Nde*I. The resulting ligation results in a TGA stop
Codon immediately following the codon for CarD residue 215. pET11b-HMGA1a (human) was constructed from the plasmid pET15b-HMGA1a (generously provided by Prof. T. Maniatis- Harvard University; 26).

**Overexpression and purification of CarD and its fragments**

_E. coli_ BL21-(DE3) cells freshly transformed with each construct were cultured overnight in LB-ampicillin medium. After dilution into fresh LB-ampicillin, the cultures were grown at 37 °C to OD_{600} of 0.6-0.8 and induced with 0.4 mM IPTG for five hours at 37 °C or overnight at 25 °C. Protein expression in whole cell extracts was checked by centrifuging 1 ml induced culture (14000 rpm in a microfuge), and the cell pellet was lysed by boiling in SDS-loading buffer for analysis by SDS-PAGE (27) and Western blotting using anti-CarD polyclonal and monoclonal antibodies (see below). To check solubility of the expressed protein, the cell pellet (obtained as above) was suspended in buffer A (50 mM Tris, 2 mM EDTA, 5 mM β-mercaptoethanol, pH 7.5) and 200 mM NaCl, sonicated and centrifuged, and the supernatant and pellet were separately analyzed by SDS-PAGE.

Proteins were purified in ice-cold conditions. Pelleted cells from 1 l induced cultures (4-5 g of cell wet weight) were lysed by grinding with alumina (2g per g cell pellet) in the presence of 1 mM PMSF and benzamidine. Alumina and cell debris were eliminated by centrifugation of the ground cell paste after suspension in 25-30 ml of buffer A containing 1 M NaCl and 1 mM each of PMSF and benzamidine. The resulting supernatant was mixed with PEI to 0.3% final concentration to precipitate out DNA. Expressed proteins were recovered from the PEI-supernatant by ammonium sulphate precipitation at levels determined in pilot experiments: 50% for CarD, His6-CD(1-104) and CD(1-215), and 65% for CD(183-316).

CarD, CD(183-316) and HMGA1a were purified off phosphocellulose and then by HPLC off a MonoS ion-exchange column (AKTA- Amersham-Pharmacia). CD(1-215) was purified off DEAE-Sephadex and then by HPLC off a MonoQ column. CD(1-104) was
expressed as the His\textsubscript{6} metal-affinity tag and purified from inclusion bodies employing TALON metal affinity resin and the accompanying purification protocol (Clontech, Palo Alto, CA). Protein and fragment identities were confirmed by N-terminal amino acid sequencing (Applied Biosystems Procise-494 Protein Sequencer) and by MALDI mass spectrometry. Concentrations of CarD, CD(1-104) and CD(1-215) were determined from the 280 nm absorbance arising from Trp and Tyr present with $\varepsilon_{280}$ (M$^{-1}$cm$^{-1}$)=8480 for CarD and CD(1-215), and 6990 for CD(1-104) (28). The absorbance at 205 nm (29) was used for CD(183-316), and that at 220 nm for HMGA1a ($\varepsilon_{220} =74000$ M$^{-1}$cm$^{-1}$; 30).

**Monoclonal and polyclonal anti-CarD antibodies**

Anti-CarD rabbit polyclonal and mouse monoclonal antibodies were obtained using standard procedures (31). Immunization was performed with CarD excised off SDS-PAGE gels reversibly stained with Zn-imidazole negative staining (32). Epitope mapping of the polyclonal and monoclonal anti-CarD antibodies was done by ELISA and Western blotting (ECL\textsuperscript{TM} kit from Amersham Pharmacia).

**Limited proteolysis**

Subtilisin Carlsberg, proteinase K, papain, chymotrypsin, and trypsin (Sigma) protease stocks were stored at -70 °C as aliquots of 5 µg/µl in 50 mM Tris pH 7.5, 1mM DTT. In pilot experiments at 28, 30 and 37 °C, protease was added to 240 µl purified protein (50 µg/µl) in buffer A containing 0.2 M NaCl at 1:500 or 1:1000 wt:wt protein:protease. 40 µl aliquots were removed at 0, 5, 10, 15, 20, 30, 45, 60 and 90 min intervals, the proteolysis quenched with 1 µl each of 1 M PMSF and benzamidine, and then analyzed by SDS-PAGE. For fragment identification, aliquots of a 15 min or 45 min subtilisin digest at 28 °C were run on separate SDS-PAGE gels for Coomassie Blue staining, and for electrotransfer to nitrocellulose for Western analysis using anti-CarD antibodies or to an Immobilon P\textsuperscript{SQ}...
Domain Architecture of a Bacterial HMGA-like Protein

membrane (Millipore, Bedford, MA, USA) for N-terminal sequencing. Proteolytic fragments for mass spectrometry were purified by reverse-phase HPLC; or by identifying the band in SDS-PAGE gel by reversible Zn-imidazole negative staining (32), excising it and then leaching it off as described by Cohen and Chait (33). Intrinsic subtilisin activities at given salt concentrations were estimated from the hydrolysis of 1 mM TAME by subtilisin at 12.5 µg/ml in 3 ml total reaction volumes (34). Hydrolysis rates were calculated from the initial linear increase in absorbance at 247 nm (ΔA247/min) monitored in a Kontron UVIKON 940 spectrophotometer equipped with a stirrer unit and a constant-temperature circulating waterbath.

Analytical size-exclusion chromatography

Analytical HPLC size-exclusion data were obtained at room temperature using a Superdex-200 column equilibrated with buffer A with 200 mM NaCl, sufficient to minimize nonspecific interactions with the column matrix. Column calibration was done using vitamin B12, (1.355 kDa), cytochrome C (12.4 kDa), carbonic anhydrase (29 kDa), ovalbumin (43 kDa), bovine serum albumin, BSA, (66 kDa), yeast alcohol dehydrogenase (150 kDa), β-amylase (200 kDa) (all from Sigma). 100 µl samples of CarD or each of its fragments at 10-100 µM were injected at 0.4 ml/min, and the elution was tracked by absorbance at 280, 235 and 220 nm. Void (V0) and total (Vt) bed volumes were determined using Blue Dextran (2000 kDa; Sigma) and vitamin B12, respectively. Elution volumes, Vt, were assigned for CarD and each fragment in distinct runs by verifying peak identities by Coomassie-stained SDS-PAGE and Western blotting. Stokes radii, Rs (in nm) for the standards were obtained from Potschka (35). The following calibration curves were generated from the data for the standards employing SigmaPlot (Jandel Scientific) with correlation coefficients ≥ 0.99 in each case: log
Mw =7.91-0.23 $V_e$ and $K_{av} =1.06-1.33\log R_s$, with $K_{av}=(V_e -V_o)/(V_t -V_o)$. These were then used to estimate the apparent Mw and $R_s$ for CarD and each fragment (36).

**Chemical cross-linking**

Chemical cross-linking was examined with glutaraldehyde, DSS (Pierce Chemical Co.) or Ni-GGH (37). 2-5 µM pure protein in 200 mM NaCl, 50 mM phosphate buffer (pH 7.5) was treated with a freshly prepared 5 mM glutaraldehyde (in water) or 25 mM DSS (in DMSO) or 2 mM Ni-GGH to final concentrations of 1 mM, 2.5 mM and 1 mM, respectively, and incubated for 1 hr at 30 °C. Total reaction volumes were 100 µl. Cross-linking was quenched with SDS-PAGE gel-loading buffer (150 mM Tris final concentration) and analyzed by SDS-PAGE and Western blotting.

**Analytical ultracentrifugation**

Sedimentation equilibrium measurements were done in a Beckman Optima XL-A analytical ultracentrifuge, a Ti60 rotor and six sector Epon charcoal centerpieces with 12 mm optical pathlength. 70 µl samples in 100 mM phosphate buffer (pH 7.4) containing 200 mM NaCl and 0.1 mM β-mercaptoethanol, were centrifuged to equilibrium at 13000, 15000, 18000, or 25000 rpm at 20 °C. Radial scans were acquired at 2 hr intervals by monitoring at wavelengths between 220 and 280 nm, until successive scans were superimposable indicating equilibrium. 10 µM CarD, 25 µM CD(1-215), and 200-300 µM for CD(183-316) and HMGA1a were used, and a 50 µM CarD sample in buffer containing 1 M NaCl was also examined. The apparent weight-average molecular masses ($M_r$) were determined by fitting data (using the programs EQASSOC- Beckman) to the equation for an ideal solution containing a single species as described elsewhere (38, 39). Partial specific volumes, $\bar{\nu}$, (in ml/g) calculated from the amino acid compositions (40), were set to 0.732 for CarD, 0.720 for CD(183-316), 0.733 for CD(1-215) and 0.718 for HMGA1a.
CD and fluorescence spectroscopy

CD spectra were recorded in a Jasco-810 spectropolarimeter coupled to a Neslab temperature control unit, using 0.2 nm steps at a scan speed of 20 nm/min and a 4 s time constant and averaged over 5 scans. A Hitachi F-4500 spectrofluorimeter equipped with a stirrer unit and a constant-temperature circulating water-bath was used for fluorescence spectra. Sample excitation was at 295 nm or 280 nm for a slit width of 2.5 nm, and the emission spectra, averaged over 2 scans, were recorded between 300 and 400 nm for a slit width of 10 nm at 240 nm/min and a 2 s response time. 10-30 µM protein and a 1 mm pathlength cuvette were employed for far-UV CD, while for fluorescence these were, respectively, 1-2.5 µM and 1 cm. All spectra were recorded at 25 °C in 200 mM NaCl / 100 mM phosphate buffer (pH 7.4). Fluorescence spectra in denatured conditions were obtained in buffer containing 6 M GdmHCl (Ultrapure from ICN Biomedicals, OH, USA).

DNA-binding Assays

EMSA for DNA-binding employed the following synthetic DNA oligonucleotides and their respective complementary strands. (i) IRE: 5’-GAGAAGTGAAAGTGGGAAATT CCTCTGAATAGAGAGGGAC-3’ (HMGA1a binding sites underlined; (26)); (ii) QRS: 5’-AACCCCGTGACTTTCCTAGAGCTTTCCACCGAACC-3’ (proposed CarD binding sites underlined; (19)). One of the probe strands was 32P end-labelled and annealed with cold complementary strand. EMSA were carried out in 20 µl total reaction volumes which contained: 1-3 pM end-labelled double-stranded probe (~13000 cpm); 50-300 nM CD(183-316) or HMGA1a, or 300-1000 nM CarD; and 1 µg double-stranded poly(dA-dT), poly (dG-dC) (Sigma) or poly(dI-dC) (Amersham Pharmacia). The DNA-binding buffer was 50 mM NaCl, 15 mM HEPES-4 mM Tris pH 7.9, 1 mM DTT, 10 % glycerol, 1 mg/ml BSA, and 0.1% Nonidet P-40. After a 30 min equilibration at 4 °C, DNA-binding was analyzed in
nondenaturing 6% polyacrylamide gels (37.5:1 acrylamide:bis-acrylamide) pre-run at 200 V/4 °C for 30 min in 0.5x TBE buffer (45 mM Tris base, 45 mM boric acid, 1 mM EDTA). Samples were electrophoresed for 1-1.5 hr, and the gels dried and analyzed by autoradiography. In competition assays, unlabelled double-stranded IRE or QRS probe was also included at concentrations ranging from 1 to 20 pM. For comparing DNA-binding of pure protein with its CKII-phosphorylated form, CarD, CD(183-316) or HMGA1a at concentrations as above was first incubated at 30 °C for 2 hr with or without 0.375 units of CKII and/or 125 mM ATP in 10 µl of 2X DNA-binding buffer (see below) in separate 1.5 ml tubes. Labelled probe and 1 µg poly (dG-dC) was then added to each tube to a total reaction volume of 20 µl, left to equilibrate for 30 min at 4 °C and analyzed by EMSA as before.

**CKII Phosphorylation Assays**

CKII phosphorylation of purified CarD or fragments was examined using rat liver (Promega) or recombinant human CKII (New England BioLabs). 0.3-1 µM protein was treated with 0.375 units of CKII and 1µCi [γ³²-P]ATP or [γ³²-P]GTP in CKII buffer (200 mM NaCl, 25 mM Tris pH 7.4, 10 mM MgCl₂, 100 µM ATP) or DNA-binding buffer for at least 30 min at 30 °C. Unincorporated [γ³²-P]NTP was removed by passing through a Sephadex G-50, and then examined by SDS-PAGE gel followed by autoradiography.

**RESULTS**

**Limited proteolysis of CarD indicates relatively stable N- and C-terminal domains**

Unstructured or partially structured regions of native proteins are more accessible and so more susceptible to protease action than compact structured domains. The latter are revealed on limited proteolysis of purified protein as the relatively resistant fragments visualized in SDS-PAGE gels (41). Domain organization of CarD was assessed by treatment of the protein with broad specificity proteases such as subtilisin and proteinase K. Figure 2A
shows the proteolytic cleavage pattern from limited proteolysis using subtilisin. (A similar pattern was observed with proteinase K). Three discrete bands (numbered 1-3 in Figure 2A) and a group of bands (4 and 5 in Figure 2A) were detectable 15 min after the initiation of proteolysis. Fragments corresponding to the group 4/5 persisted even after 45 min. Bands 1 to 5 were characterized by N-terminal sequencing, mass spectrometry and Western blot analysis using anti-CarD polyclonal and monoclonal antibodies of known epitope specificities (see below). The analysis, summarized in Figure 2B, indicated the following. Four fragments could be identified from bands 4 and 5 which are the ones most resistant to proteolysis: a fragment corresponding to approximately the first hundred N-terminal residues, a fragment beginning at residue 157 and including the entire acidic region, and fragments which contain all or most of the AT-hook region. Bands 2 and 3 are two discrete, intense bands generated in the early steps of proteolysis. The fragments corresponding to these begin around residues 151 (band 2) or 185 (band 3) and span all of the C-terminus of CarD. Thus the complete acidic-AT-hook segment appears to constitute a relatively stable domain. The segment from residues 100 to 151 is highly susceptible to proteolysis and could constitute a loosely structured part of the protein.

Expression of CarD fragments suggests a protein-stabilizing role for the acidic region through interactions with the AT-hooks

Structural stability is an important determinant of proteolytic susceptibility and so of intracellular degradation in E.coli (42). Consequently, fragments corresponding to stably folded domains are usually expressed to higher levels than those with little or no defined structure. This provides a means to assess whether a given protein segment constitutes an independently folded domain, and can be used to corroborate limited proteolysis data. We have done so for CarD by checking the expression of the following fragments: (i) the N-terminal region spanning residues 1-104, CD(1-104); (ii) the acidic and basic AT-hook
regions alone, CD(183-316); (iii) the segment containing all four AT-hooks, CD(225-316) or just the last three C-terminal ones, CD(247-316). These fragments were chosen taking into account the results of the limited proteolysis experiments described above. All these fragments were expressed off plasmids constructed as described in Experimental Procedures, and none have as the penultimate N-terminal amino acid one that would confer a short half-life in bacteria (43).

Figure 3 shows the protein expression patterns for total extracts from cells in which expression of CarD (lane 3) or one of its fragments (lanes 4-9) was induced under equivalent conditions of growth and induction times. Like CarD, CD(1-104) and CD(183-316) were expressed at levels visually detectable by Coomassie-Blue staining (shown boxed in lanes 4 and 7, respectively, in Figure 3A), but not CD(225-316) or CD(247-316) (lanes 5 and 6, respectively). An unambiguous identification of the above overexpressed bands was obtained from Western blots using polyclonal and monoclonal antibodies generated against purified CarD. The results obtained with polyclonal anti-CarD antibodies and with one of the two monoclonal antibodies generated are shown in Figures 3B and 3C, respectively. Polyclonal anti-CarD antibodies detected every one of the above fragments, including CD(225-316) or CD(247-316) which were not apparent in Coomassie-stained gels. This was also the case with the anti-CarD monoclonal antibody shown in Figure 3C, except for the N-terminal fragment which is not detected by this antibody (epitope specificities are summarized in Figure 3D). The relative intensities of the bands in Western blots paralleled the expression levels inferred from Figure 3A: the observed bands for CD(1-104) and CD(183-316) in the Western blots were quite intense, but were barely perceptible for CD(225-316) or CD(247-316).

CarD has an apparent molecular weight, Mw, of 41 kDa in SDS-PAGE, (Figure 2 or 3), higher than the value of 33.9 kDa calculated from sequence or determined by mass
spectrometry. Similar anomalous mobilities in SDS-PAGE are observed for mammalian and insect HMGA proteins (13, 44), as also for some other highly charged proteins. CD(225-316), CD(247-316) and CD(183-316), all of which contain the AT-hook region, also run anomalously in SDS-PAGE with apparent Mw 8-10 kDa higher than their true values of 9.6, 7.3 and 14.3 kDa, respectively, but CD(1-104) (true Mw=11.7 kDa) does not exhibit this anomalous behaviour.

Consistent with the expression levels observed, sufficient amounts of CD(1-104) and CD(183-316) could be purified, but not CD(225-316) or CD(247-316). CD(183-316), like CarD, was purified off cation-exchange columns, whereas CD(1-104) was expressed as a His6-tag fusion protein for affinity purification, the tag subsequently cleaved off by thrombin (see Experimental Procedures). N-terminal sequencing of the purified proteins indicated results expected for proteins expressed in E. coli (45): the initiator N-terminal Met, N-Met, in CarD is processed out leaving an N-terminal Pro, the non-native N-Met introduced for bacterial expression of CD(183-316) is retained, and thrombin-cleaved His6-CD(1-104) has the expected N-terminal GSH. The molecular masses of 14.25 kDa for CD(183-316), and 11.7 kDa for CD(1-104) determined by mass spectrometry match the calculated values, confirming that both were purified as the full-length proteins.

The results for limited proteolysis of the whole protein are thus in accord with the stable expression observed for fragments CD(1-104) and CD(183-316), but in apparent contrast to the low expression of CD(225-316) or CD(247-316). If the latter two fragments containing only the AT-hooks are devoid of defined structure as has been reported for human HMGA1a (12), they would be expected to be more susceptible to intracellular proteolytic degradation and so poorly expressed (42), as is actually observed. Consequently, the stable expression of CD(183-316) implies that the acidic region, when simultaneously present, is sufficient to stabilize the basic AT-hooks, which by themselves are not stable. In the context
of limited proteolysis of the whole protein, the coexisting acidic region could help to stabilize, intra- or intermolecularly, the AT-hook regions thereby accounting for their apparent proteolytic resistance. Any interactions between the basic AT-hook region and the adjacent acidic region are most likely electrostatic in origin. In accord with this, we have observed in SDS-PAGE that bands corresponding to the acidic-AT hook fragments generated by limited subtilisin proteolysis of CarD are weaker in intensity at higher salt (0.9 M NaCl) than at lower salt (0.08 M NaCl) where the intrinsic activity of subtilisin is only ~5 % lower (data not shown). Added support for the stabilizing role of the acidic region also comes from our observation that the fragment CD(Δ181-223), which lacks only the acidic segment of CarD, was poorly expressed (lane 8, Figure 3A-C). The construct CD(1-215) containing most of the acidic region but lacking the entire AT-hook region, was more stably expressed (lane 9, Figure 3A-C). Thus the observed expression for CD(Δ181-223) and CD(1-215) suggests that the presence of the acidic region is required for CarD stability, whereas the protein can be stably expressed in the absence of the AT-hook segment. We purified CD(1-215), and verified that its N-terminal sequence matched that in CarD, and that its molecular mass from mass spectrometry corresponded to that calculated from the sequence (= 23.39 kDa).

**Tests for CarD oligomerization**

Distinct segments of CarD may interact with one another as discussed above, and sequence analysis of CarD suggested leucine-zipper-type heptad repeats between residues 120 to 141 (19). Moreover, the considerably smaller mammalian HMGA1a has been reported to oligomerize in solution, and to interact with other proteins (6,46,47). Consequently, we tested CarD for oligomerization.

The oligomeric nature of proteins, their shapes and sizes can be assessed by analytical gel-filtration HPLC (36). Figure 4A summarizes the results of analyzing CarD, CD(1-104), CD(183-316) and CD(1-215) as well as human HMGA1a in a Superdex-200 HPLC gel-
filtration column equilibrated with buffer at 0.2 M NaCl. Each of these proteins eluted as a single symmetrical peak of sharpness comparable to the standards, and no additional peaks were detected. This either indicates a homogeneously populated conformation, or different conformations exchanging rapidly relative to their mobilities in the column (48). The apparent Mw estimated from this data are indicated in Figure 4A. CD(1-104) appears to be a globular monomer with apparent Mw close to the expected monomer value. On the other hand, CarD, CD(183-316), CD(1-215), as also HMGA1a, elute with apparent Mw considerably higher than their expected monomer values. (The highly charged CarD, CD(183-316) and HMGA1a when examined in 1 M NaCl exhibited essentially identical elution behaviour as at 0.2 M NaCl- data not shown). Apparent Mw (in kDa) of 47 for CD(1-215) is that expected for a dimer, whereas values of 129, 70 and 37 for CarD, CD(183-316) and HMGA1a, respectively, suggest even higher order oligomers. Alternatively, the slower mobilities could reflect extended molecular shapes (49). To distinguish between these possibilities, CarD and its fragments were further examined by chemical cross-linking and analytical ultracentrifugation.

Chemical cross-linking of CarD and its fragments was examined using two reagents that cross-link primary amino groups- glutaraldehyde and DSS. Oxidative Ni-GGH cross-linking which has been proposed to involve aromatic amino acids (37) was also investigated. With all three cross-linking agents, dimers and to a lower extent higher order oligomers were observed for CarD and CD(1-215). For CD(183-316), as for HMGA1a, dimers were observed with glutaraldehyde and DSS but not with Ni-GGH possibly because both polypeptides lack aromatic amino acids. Defined cross-linked products could not be observed with CD(1-104), consistent with its monomeric nature as suggested by analytical gel filtration. Figure 4 B shows representative cross-linking data obtained for the fragments CD(1-215) and CD(183-316). In all cases, the non-crosslinked form appeared as an intense
band even at the highest protein or cross-linker concentrations that we employed. Therefore, CarD and all its fragments except CD(1-104) are present as an equilibrium distribution of monomers, dimers and possibly higher order oligomers according to cross-linking data.

As a final diagnostic of oligomerization, we performed sedimentation equilibrium experiments of CarD, CD(1-215), CD(183-316) and HMGA1a by analytical ultracentrifugation (38, 39). For each of these, the observed sedimentation equilibrium gradients were fit to the equation that describes an ideal single component situation to obtain the apparent weight-average molecular mass, \( M_r \). Figure 4C shows such an analysis for CarD in 200 mM NaCl buffer, which yields a best fit \( M_r = 56 \pm 3 \) kDa at 15000 rpm, with small though not randomly scattered residuals around the fit. The latter is indicative of the presence of several species (38, 39). For comparison, Figure 4C also shows deviations from experimental data when the ideal-single component fits were performed with \( M_r \) fixed to the calculated monomer or dimer value. Similar results were obtained with CarD at 13000 rpm (best-fit \( M_r = 63 \pm 3 \) kDa), and also for a five-fold higher protein concentration in 1 M NaCl buffer (best-fit \( M_r = 54 \pm 2 \) kDa). The best-fit \( M_r \) values for CarD are close to the dimer value, thus suggesting that the protein exists as a largely dimeric form in equilibrium with monomeric and possibly some higher order forms. The same conclusion appears to hold for CD(1-215): \( M_r \) of 48\( \pm \)1 kDa at 13000 rpm or 45\( \pm \)2 kDa at 15000 rpm close to the calculated dimer value, but with small and non-randomly scattered residuals around the fit. By contrast, \( M_r \) was 19\( \pm \)2 kDa for CD(183-316) and 14\( \pm \)1 kDa for HMGA1a at 25000 rpm, and the residuals around the fit were small and randomly scattered even at the highest protein concentrations examined (200-300 \( \mu \)M), consistent with the predominant species being the monomer.

\( M_r \) and the Stokes radii, \( R_s \) (in m) values determined from gel filtration data for each protein may be used to estimate its approximate frictional ratio (\( f/f_0 \)), which is an indicator of
particle shape. $f/f_0 = R_s/\left[3 \bar{v} M_r/(4\pi N)\right]^{1/3}$, where $\bar{v}$ is the partial specific volume of the protein (listed in Experimental Procedures) and $N$ is Avogadro’s number (50, 51). Gel filtration data for CarD, CD(183-316), CD(1-215) and HMGA1a, provided $R_s$ (in nm) of 3.63, 3.15, 2.89 and 2.72, respectively. Estimates for $f/f_0$ obtained were $\sim 1.4$ for CarD and $\sim 1.2$ for CD(1-215) both within or close to the range observed for the standards cytochrome C, BSA and yeast alcohol dehydrogenase: $f/f_0$=1.09, 1.30 and 1.28, respectively (50, 51). By contrast, CD(183-316) and HMGA1a with $f/f_0$ of 1.8 and 1.7, respectively, appear to deviate significantly from a spherical shape and are probably elongated molecules, in accord with their slower gel filtration mobilities.

In summary, gel filtration, chemical cross-linking and analytical ultracentrifugation data together suggest that CarD and its fragment lacking the AT-hook, CD(1-215), are predominantly dimers, its N-terminal fragment CD(1-104) forms a compact monomeric domain, and the C-terminal acidic-AT hook segment, CD(183-316), is largely monomeric and elongated as is HMGA1a. The compact nature of the CD(1-215) dimer suggests that the acidic region in this molecule (and possibly in CarD) may not be in an extended conformation, but that it may be so in CD(183-316) or HMGA1a. The above data also hint that residues within the 104-215 segment may be involved in dimerization.

**CarD is structurally well-defined at its N-terminus but random at its C-terminus**

The presence of protein secondary and tertiary structure is readily assessed by CD and intrinsic fluorescence of the purified protein (52). In far-UV CD spectra, $\alpha$-helices are characterized by two minima at 222 nm and 208 nm and a maximum at 192 nm, $\beta$-sheets by a weaker and broader minimum around 215 nm and a maximum at 198 nm, and random coils by an intense minimum at 198 nm (53). The far-UV CD spectra of CarD and its N-terminal fragments CD(1-215) and CD(1-104) exhibit the characteristic minima for $\alpha$-helical and $\beta$-
sheet conformations. CD(183-316), however, has a far-UV CD spectrum expected for random coils as does HMGA1a (Figure 5). Thus, defined α-helical and/or β-sheet secondary structural elements in CarD are confined to its N-terminal region, and the C-terminus is randomly structured. These observations are in qualitative accord with the secondary structure predictions using the PHD algorithm (54). The far-UV CD spectrum of CD(1-215) resembles that of whole CarD, and, in terms of their –[Θ] values at 222 nm, the two have similar helix contents. –[Θ]222 is smaller for CD(1-104). The region between residues 104 and 215 may therefore be intrinsically more helical, or it may be that oligomerization in CarD and CD(1-215) drives additional folding in these molecules relative to CD(1-104) (55).

A Trp (Trp92) and one or two Tyr (Tyr18 and Tyr135) are present in CarD, CD(1-104) and CD(1-215) but not in CD(183-316) or HMGA. The first three can therefore be examined by intrinsic Trp fluorescence measurements. Trp emission maxima for native CarD and CD(1-215) are both around 343 nm while that for CD(1-104) is about 349 nm (Figure 5B). In all three, the maximum is red-shifted to 354 nm in denaturing 6 M GdmHCl solutions with about a 20% loss in intensity for CarD and CD(1-215), and a slight increase in intensity for CD(1-104). Therefore the environment of Trp in these proteins in the native state differs from that in the denatured, solvent-exposed forms and points to the existence of defined tertiary structure (52). As inferred from their similar intrinsic Trp fluorescence behaviour, the tertiary fold is probably similar in native CarD and CD(1-215) but varies somewhat in CD(1-104).

**CarD and its C-terminal fragment share the DNA-binding specificity of human HMGA1a**

HMGA1a binds specifically to two appropriately spaced AT-rich tracts present in the -77 to -37 region just upstream of the interferon-β promoter- the interferon response element or IRE (26). EMSA analysis shows that CarD and its fragment CD(183-316) containing the acidic-AT-hook segment bind specifically to this 40 bp IRE fragment, and that the binding is
competed away by poly (dA-dT) or poly(dI-dC) but not poly(dG-dC) (Figure 6A). This is identical to the behaviour reported for human HMGA1a (10), and also shown in Figure 6A. The minor grooves of G-C, A-T and I-C base pairs differ solely in the presence of a 2-amino group in G but which is a hydrogen in A or I. As a consequence, I-C resembles G-C in the major groove and A-T in the minor groove. The ability of poly(dI-dC) to compete as well as poly (dA-dT) and far more effectively than poly(dG-dC) has therefore been used as evidence for the binding of HMGA1a to the minor-groove of AT-rich DNA (10). Thus CarD and CD(183-316), like HMGA1a, exhibit minor-groove DNA-binding specificity. For the same solution conditions HMGA1a binds to IRE with a $K_D \approx 40$ nM (56). As judged by the concentrations of CarD and CD(183-316) required for EMSA analysis, the specific DNA-binding affinity for the acidic-AT hook fragment of CarD is slightly lower than for HMGA1a, while it may be as much as an order of magnitude weaker for CarD (Figure 6A).

CarD is essential for the expression of the light-inducible carQRS operon, a key gene cluster in M. xanthus carotenogenesis (19). Two 5'-GGAAA-3' repeats 5 bp apart in the -88 to -55 promoter upstream region of this operon have been suggested to constitute a CarD binding site (19). A double-stranded oligonucleotide probe containing this site (QRS) is bound by CarD, CD(183-316) and HMGA1a, and the binding was competed away by poly (dA-dT) or poly(dI-dC) but not by poly(dG-dC) (data not shown), as was observed with probe IRE. This reiterates the minor-groove binding specificity of CarD. In EMSA analysis, specific binding to probe QRS required three- to five-fold higher protein concentrations relative to those used with probe IRE. This indicates that the binding affinity for probe QRS is less than for probe IRE. Consistent with this, binding to probe QRS is competed away far more effectively by IRE than by QRS in competition binding assays (Figure 6B). It has been reported that optimal site-specific binding of AT-hooks requires a minimum of two appropriately spaced tracts of AT-rich sequences at least four base pairs in length (10,11).
The lower affinity for probe QRS relative to probe IRE may therefore be related to its two AT-rich stretches being only three base pairs long. Thus the proposed CarD binding site in *M. xanthus* is not optimized for maximal binding affinity, and the possible consequences of this will be examined in the Discussion.

**CK II phosphorylates CarD leading to decreased DNA-binding affinity**

Mammalian HMGA proteins are phosphorylated by CKII kinase (14-18). Protein sequence analysis predicts the presence of such sites in the acidic and N-terminal regions of CarD (Figure 1). We tested this by examining CKII phosphorylation of CarD *in vitro*. Figure 7A shows that CarD, CD(183-316) and CD(1-215) but not CD(1-104) are phosphorylated *in vitro* by CKII. The CKII-phosphorylation sites thus map to the acidic region of CarD as in the case of HMGA proteins. Typical substrates for CKII are short, unstructured peptides (~10 residues long) containing the required phosphorylation sites. CKII phosphorylation of the acidic region is therefore consistent with its relatively open structure, whereas in the more compactly structured CD(1-104) the putative CKII sites appear to be inaccessible.

Phosphorylation of HMGA proteins causes marked decreases in their DNA-binding affinities (16-18). Figure 7B shows that CKII phosphorylation of CarD or CD(183-316) also dramatically decreases DNA-binding affinity. HMGA-DNA binding is highly dependent on ionic conditions (16), which implies large coulombic contributions to DNA-binding (55). CKII phosphorylation of the acidic region would increase its negative charge. This, it may be argued, would boost coulombic repulsions from the DNA backbone and thereby diminish DNA-binding affinity. The simultaneous enhancement expected for any favorable electrostatic interactions between the acidic region and the basic AT-hooks would also lead to the latter being further sequestered from DNA-binding.
DISCUSSION

Our analysis reveals that CarD consists of two relatively stable domains: (i) a C-terminal HMGA-like region that consists of all of the acidic and AT-hook regions and spans residues 183 to 316 at the C-terminal end of the molecule; (ii) an N-terminal domain of about a hundred residues that is absent in eukaryotic HMGA proteins. The two domains are linked by a segment whose stretch between residues 105 and 155 is quite susceptible to proteolysis and is thus likely to be a flexible linker region.

The CarD HMGA-like domain

CD(183-316) corresponds to the HMGA-like domain in CarD. It shares all the attributes of its eukaryotic counterparts. It consists of adjacent highly acidic and basic regions though juxtaposed differently. CD(183-316) lacks defined structure as does HMGA1a (12). Both appear to be largely monomeric, although there may be intermolecular interactions that occur in a highly transient fashion and explain the experimentally observed chemical cross-linking. CD(183-316) is the seat of DNA-binding in CarD and is akin to HMGA1a in its minor-groove binding specificity. Consistent with large electrostatic contributions to DNA-binding, the affinity is lowered dramatically by CKII-phosphorylation that is localized to the acidic portion of CD(183-316) (and CarD) as in HMGA proteins.

The lack of intrinsic structure in CD(183-316) or HMGA1a is not unexpected since both are composed of mostly Pro and highly charged residues and have few hydrophobic residues. These characteristics are unfavorable for the formation of defined structural elements like α-helices and β-sheets or a stable compact core (57-59). Absence of defined structure usually correlates with low intracellular stability because of greater susceptibility to intracellular proteases (42). Sequences such as PEST if present also predispose a protein to intracellular proteolysis (60,61), but these are absent in CarD or HMGA proteins. Our results
show that the presence of the acidic region is required for stable expression of whole CarD or its basic AT-hook region. This leads us to propose that an important role for the acidic region in the protein architecture is that of stabilizing the randomly structured AT-hooks. The resultant acidic-AT hook interactions would necessarily affect DNA-binding by the AT-hook both by sequestering the latter from DNA, as well as by charge repulsions between the acidic region and the DNA backbone as reasoned earlier. These inferences on the roles of the acidic region in protein stability and DNA-binding very likely carry over to the members of the eukaryotic HMGA protein family, all of which have the acidic region (9, 13).

An intrinsic lack of structure has been argued to confer on a protein the inherent flexibility and structural plasticity required in fine-tuning its regulatory or signaling functions (61, 62). Indeed, the ability to tweak both protein stability and conformation of the intrinsically unstructured HMGA proteins by covalent modifications such as phosphorylation or acetylation underlies their participation in diverse biological processes from transcription to recombination (18, 62). The C-terminal HMGA-like part of CarD would be similarly malleable to such conformational and so functional alterations. Moreover, like HMGA proteins, CarD is also a multifunctional regulator being involved in the distinct processes of carotenogenesis and multicellular development in *M. xanthus* (20). Work currently in progress in this laboratory has also underscored the involvement of CarD in processes other than carotenogenesis and fruiting-body development\(^2\). An array of signaling processes are known to be involved in *M. xanthus* development including a number of eukaryotic-like serine/threonine protein kinases and functionally linked phosphatases. However, the identification of specific phosphorylatable substrates has remained elusive (23, 24). CarD would therefore be an attractive candidate given its involvement in *M. xanthus* development, as are HMGA proteins in the eukaryotic cell cycle and development, and based on the analogies between the two proteins that we have enumerated.
According to hydrodynamic experiments, CarD and its fragment lacking the AT-hooks exist largely as dimers in equilibrium with monomers, and these dimers can be chemically cross-linked with a variety of agents. Hydrodynamic data suggest that CD(183-316) and HMGA1a are elongated monomeric species, but they seem capable of being cross-linked by bifunctional amino-reactive agents. Although this may be a cross-linking artefact caused by the presence of a large number of lysines, it may also reflect transient association. The latter explanation would be compatible with our earlier inference of intra- or intermolecular interactions of the acidic region with the AT-hooks that make them resistant to proteolysis. Although we have been unable to provide additional experimental evidence for this, protein-protein interactions between different HMGA molecules or of HMGA with other protein molecules have been invoked to explain the highly cooperative enhanceosome assembly (13, 26, 46, 47, 56, 62). The inherent difficulty in pinpointing the mechanism of HMGA cooperativity has been noted before and attributed to, among others, its lack of defined structure (56). The latter imposes technical challenges which would also apply to CD(183-316), and to this segment in CarD.

**The CarD N-terminal domain**

CD(1-215), which lacks the AT-hooks, appears to be both compact and dimeric. By contrast, CD(1-104) appears to be a compact monomer. This would suggest that dimerization in CarD is mediated by a stretch or stretches between the N-terminal and acidic domains, and this would include the heptad leucine-zipper-type repeats present between residues 120 to 141. This region is quite susceptible to proteolysis, is predicted to have low coiled-coil forming probability by available programs and contains two Pro residues that are not common to leucine-zippers. As a consequence, attempts to obtain purified CarD fragments containing this region and lacking segments N- or C-terminus to it in the CarD sequence have not been successful thus far. Hence the actual dimerization segment remains to be specified.
The CarD N-terminal region as in fragments CD(1-104) and CD(1-215) constitutes a stable, compact domain. It appears to be the part that contains most of the secondary and tertiary structural elements of the protein based on our spectroscopic data. A specific function has yet to be attributed to this domain unique to CarD that is absent in its eukaryotic HMGA counterparts. Sequence analysis indicates that the region corresponding to CD(1-104) shares significant homology with an approximately 200-residue segment of a family of bacterial proteins referred to as transcription repair coupling factors or TRCFs where they constitute the RNA polymerase interacting module (63, 64). TRCFs stimulate the repair of lesions in the transcribed strand by interacting with RNA polymerase (63, 65). The sequences of \textit{E.coli} and \textit{B. subtilis} TRCFs involved in RNA polymerase-binding share a 32% identity, and this appears to be sufficient for interaction with heterologous RNA polymerases (66). The CarD N-terminal domain is, respectively, 25% and 26% identical in sequence to the \textit{E.coli} and \textit{B.subtilis} TRCF RNA polymerase-binding modules, and could conceivably be involved in interactions with RNA polymerase.

TRCFs appear to bind to both the holo and apo forms of RNA polymerase indicating that the $\sigma$ subunit does not interfere with the binding (63). In the case of the CarD-dependent activation of the carQRS operon in \textit{M. xanthus}, genetic analyses have revealed that the carQ gene product is also essential (67, 68). The amino acid sequence of CarQ revealed that it may be a member of the extracytoplasmic function (ECF) subfamily of RNA-polymerase sigma factors (69). Whether CarD interacts with CarQ or with the RNA polymerase needs to be experimentally determined and is beyond the scope of the present study. Nevertheless, drawing from parallels with HMGA proteins, where specificity and affinity are both enhanced by the highly cooperative assembly of transcriptional complexes (62), it is tempting to speculate that CarD may interact with any of CarQ, RNA polymerase or other to-be-discovered factors in assembling an enhancesosome-like complex in the vicinity of the
carQRS promoter region. Our experimental data demonstrate that CarD does exhibit HMGA-like minor-groove binding to a specific AT-rich sequence upstream of the −35 region of the carQRS promoter, albeit with a lower binding affinity relative to CD(183-316) or HMGA, or to the HMGA IRE-binding site in eukaryotic DNA. Both the specificity and affinity of CarD could be enhanced by interactions with additional factors, and by the fact that the search for its specific AT-rich binding sites would be facilitated by the highly GC-rich nature (67.5% GC) of M. xanthus DNA (70). Moreover, interactions of CarD with itself and with other proteins could also serve to maintain and modulate the intracellular levels of this protein which has regions with considerable lack of defined structure (61). Based on the structural and functional information generated in this study, we are currently examining the existence of other DNA-binding sites and protein factors that interact with CarD, using a battery of techniques including two-hybrid analysis, coimmunoprecipitation, and the effects in vivo of specifically truncated fragments.

ACKNOWLEDGEMENTS

We are indebted to Dr. J. M. Sanz (Universidad Miguel Hernandez-Elche) for use of the CD spectropolarimeter and Dr. G. Rivas (CIB-Madrid) for the analytical ultracentrifuge data. We acknowledge the instrumental facilities at CIB (Madrid) for DNA-sequencing (Dr. A. Díaz-Carrasco), N-terminal amino acid sequencing (Dr. J. Varela), and mass spectrometry (Dr. A. Prieto). Our thanks to Dr. S. Streitenberger for help with intrinsic protease activity measurements, and to him and A. F. Martínez in polyclonal anti-CarD preparation, to J. M. Lazaro (CBM-Madrid) and to Drs. F. Solano and M. L. Cayuela for helpful discussions, and to A. Loba, J. A. Madrid and A. C. García for assistance.
REFERENCES


**FOOTNOTES:**

1 Abbreviations used: BSA, bovine serum albumin; CD, circular dichroism; Cdc2, cyclin dependent cell-cycle; CKII, casein kinsase II; DMSO, dimethyl sulfoxide; DSS, disuccinimidyl suberate; DTT, dithiothreitol; EMSA, electrophoretic mobility shift assay; far-UV, far ultra-violet; GdmHCl, guanidinium hydrochloride; HMG, high mobility group; HPLC, high performance liquid chromatography; IPTG, isopropyl-D-thiogalactoside; LB, Luria broth; MALDI, matrix assisted laser desorption ionization; MAP, mitogen activated protein; Ni-GGH, nickel(II)-Gly-Gly-His; PCR, polymerase chain reaction; PEI, polyethyleneimine; PMSF, phenylmethyl sulphonyl fluoride; SDS, sodium dodecyl sulphate; PAGE, polyacrylamide gel electrophoresis; TAME, p-tosyl-L-arginine-methyl ester; TE, Tris-EDTA.

N.B. Nomenclature of high mobility group proteins was recently revised. See reference 7 and webpage: http://www.informatics.jax.org/mgihome/nomen/genefamilies/hmgfamily.shtml. Accordingly, HMGI is referred to as HMGA1a.

2 M. Galbis, M. Fontes, F. J. Murillo- unpublished data.
FIGURE LEGENDS

FIG. 1. Schematic representation and sequences of mammalian HMGA1a protein and M. xanthus protein CarD. A. HMGA1a. B. CarD. The stippled boxes represent the AT-hooks (three in HMGA1a and four in CarD), and the horizontally striped boxes indicate the acidic regions. The open box in CarD is a putative leucine-zipper coiled-coil motif. Dotted and continuous lines at the bottom of the sequences indicate the acidic and the basic AT-hook regions. The RGRP repeats and the proposed leucine-zipper are underlined and in boldface. Arrows indicate CKII phosphorylation sites predicted by PROSITE in CarD, and experimentally identified in HMGA1a.

FIG. 2. Analysis of stable domains in CarD by limited proteolysis. A. Coomassie-stained SDS-PAGE (15%) of 0, 15 min and 45 min digestion of CarD by subtilisin Carlsberg, with molecular weight markers as indicated. B. Fragments corresponding to the numbered bands in A identified by N-terminal sequencing and mass spectrometry. Arrows indicate sites susceptible to subtilisin cleavage in CarD (band 1).

FIG. 3. Expression of CarD and its fragments. A. Cell extracts analyzed by SDS-PAGE and Coomassie Blue-staining as described in Experimental Procedures. Lane 1: molecular weight markers (in kDa); lane 2: control (pET11b with no insert); lanes 3-9 are CarD, CD(1-104), CD(225-316), CD(247-316), CD(183-316), CD(Δ181-223) and CD(1-215) in that order. Boxes indicate overexpressed protein when visually detectable. B. Western blots for the samples in A using anti-CarD polyclonal antibodies. C. Western blots for the samples in A using one of the anti-CarD monoclonal antibodies. In B and C filled arrowheads point out bands for CarD or its fragments. D. Summary of CarD fragments used in panels A-C and analysis of their expression by Coomassie-staining (column C), anti-CarD polyclonal
(column P) or the two monoclonal (columns M1 and M2) antibodies. “+” indicates detected and “-” not detected.

**FIG. 4.** Tests for CarD oligomerization. **A.** Size-exclusion analysis of CarD and fragments. The straight line is the calibration curve obtained for 0.2 M NaCl in buffer A. The dots are data obtained for CarD and its fragments with the apparent molecular weight, Mw, in parentheses. **B.** Chemical cross-linking of CD(183-316) (top) and CD(1-215) (bottom) using glutaraldehyde (G), DSS (D) or Ni-GGH (N) analyzed in Western blots using anti-CarD polyclonal antibody as described in the text. Non-crosslinked monomers (open arrowheads), and cross-linked dimers and higher order oligomers (filled arrowheads) are indicated. **C.** Sedimentation equilibrium data for CarD. The data were fit to the equation for a single ideal species with $M_r$ as a variable parameter, and the residuals for this fitting are shown in the top panel. Fits with $M_r$ fixed to the expected monomeric (M) or dimeric value (D) are shown by the dotted or dashed line, respectively.

**FIG. 5.** Secondary and tertiary structure of CarD and its fragments by far-UV CD and fluorescence spectroscopy. **A.** Far-UV CD spectra of CarD, its fragments and HMGA1a. **B.** Fluorescence emission spectra for CarD, CD(1-215) and CD(1-104) with excitation at 295 nm. In B the continuous lines are data obtained in native solution conditions while the dashed lines correspond to denaturing conditions (6 M GdmHCl) as described in the text.

**FIG. 6.** Analysis of the DNA-binding of CarD and its fragments by EMSA. **A.** Binding of CarD (295 nM), CD(183-316) (65 nM) and HMGA (60 nM) to probe $[^{32}\text{P}]$-IRE at ~2 pM (13000 cpm), in the presence of 1 µg poly(dA-dT), “A”, 1 µg poly(dI-dC), “I”, or 1 µg poly(dG-dC), “G”. **B.** Binding of CarD (740 nM; bottom) and CD(183-316) (160 nM; top) to probe $[^{32}\text{P}]$-QRS in the presence of increasing amounts of cold unlabelled IRE or QRS as
specific competitor and 1 μg poly(dG-dC) as nonspecific competitor. Solution conditions are otherwise identical in A and B and are as described in the text.

**Fig. 7.** Phosphorylation of CarD and fragments by CKII and its effect on DNA-binding.

A. Phosphorylation of CarD and fragments *in vitro* in the presence of CKII and \( \gamma^{32}\text{P}\)ATP.

B. CarD (1.5 μM) or CD(183-316) (330 nM) was treated with (+) or without (-) 250 mM ATP and/or 0.375 units of CKII, and then examined for binding to probe \( ^{32}\text{P}\)-QRS by EMSA (see Experimental Procedures).
A

MSESSSKSSQ PLASKQEKD G TEKRGRGRPR KOPPVSPGTA LVGSQKEPSE VPTPKRPRGRP
KGSKNGAA KTRKTTTTPG RKPRGRPKKL EKEEEEGISQ ESSEEEQ

B

MEPGBASLQL AVGDRVYPN QGVCRVSAID VKEVAGQKLT FVTMRREEDG AVVMVPEGKV
LAIGVRKVAS AEDVEAITFT LRSDDSDKDL DWKQRARTNL DRMTQGGIMG LAEVVKGLQV
LSELRPPTK ERELYDNARH LLVTEVAAAL GTAENAYEDA IDIVLPFFPGR ERPKRTAAEF
QREEGDLDLD SDLGLDSSDL DLPSDEESE PSEEESSEEG ESSEAAASSEE SAPKPRGRPP
AKKTEAPEGA EPPAPKRGR PPKPKPEATA EGAEPAPKK RGRPPKPKPVE EVEGAAPAAP
KKGKPRPVPKV PPEGES

Fig. 1, Padmanabhan et al.
A

![Image of a gel electrophoresis result with marker bands and subtilisin digestion at 15' and 45'.]

B

<table>
<thead>
<tr>
<th>Band</th>
<th>N-term. Sequence</th>
<th>Fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PEGSA</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>GTAEV, AEDAI</td>
<td>151/157</td>
</tr>
<tr>
<td>3</td>
<td>DLDDL, GLDSD, DSDLD</td>
<td>185/195/197</td>
</tr>
<tr>
<td>4</td>
<td>GAEPP</td>
<td>249</td>
</tr>
<tr>
<td>5</td>
<td>GAEPP, EESAP, AEDAI, PEGSA</td>
<td>249/271, 229, 157, 2</td>
</tr>
</tbody>
</table>

Fig. 2, Padmanabhan et al.
Fig. 3, Padmanabhan et al.
**Fig. 4.** Padmanabhan et al.
Fig. 5, Padmanabhan et al.
Fig. 7, Padmanabhan et al.
Domain architecture of an HMGA-type bacterial transcriptional factor
S. Padmanabhan, Montserrat Elías-Arnanz, Emilio Carpio, Pedro Aparicio and Francisco Jose Murillo

J. Biol. Chem. published online August 31, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M106352200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts