doi:10.1016/j.jmb.2004.07.094 J. Mol. Biol. (2004) 342, 1249–1264

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Diversity in Structure and Function of the Ets Family PNT Domains

Cameron D. Mackereth†, Manuela Schärpf†, Lisa N. Gentile Scott E. MacIntosh, Carolyn M. Slupsky and Lawrence P. McIntosh*

Department of Biochemistry and Molecular Biology Department of Chemistry and The Biotechnology Laboratory, University of British Columbia, Vancouver BC, Canada V6T 1Z3

**Corresponding author*

The PNT (or Pointed) domain, present within a subset of the Ets family of transcription factors, is structurally related to the larger group of SAM domains through a common tertiary arrangement of four α-helices. Previous studies have shown that, in contrast to the PNT domain from Tel, this domain from Ets-1 contains an additional N-terminal helix integral to its folded structure. To further investigate the structural plasticity of the PNT domain, we have used NMR spectroscopy to characterize this domain from two additional Ets proteins, Erg and GABPα. These studies both define the conserved and variable features of the PNT domain, and demonstrate that the additional N-terminal helix is also present in GABPα, but not Erg. In contrast to Tel and Yan, which self-associate to form insoluble polymers, we also show that the isolated PNT domains from Ets-1, Ets-2, Erg, Fli-1, GABPα, and Pnt-P2 are monomeric in solution. Furthermore, these soluble PNT domains do not associate in any pair-wise combination. Thus these latter Ets family PNT domains likely mediate interactions with additional components of the cellular signaling or transcriptional machinery.

Keywords: transcription factor; SAM domain; protein interactions; Erg; GABP

Introduction

Proteins involved in gene regulation and signal transduction are typically composed of a series of modular domains that confer a variety of binding and catalytic functions. As an example, components of signaling cascades use modules such as the SH2, SH3, and PDZ domains to regulate protein–protein association through the binding of specific peptide sequences.[1,2] Transcription factors contain a similar combinatorial assembly of domains, with functions including DNA recognition, transactivation, and auto-regulation. For example, the DNA-binding winged helix-turn-helix ETS domain defines the members of the Ets family of transcription factors.[3–5]

A conserved PNT (or Pointed) domain is also found within a subset of the Ets transcription factors, including mammalian Ets-1, Ets-2, Erg, Fli-1, GABPα, and Tel, as well as *Drosophila* Pnt-P2 and Yan. First identified during cloning of the Pnt-P2 gene product in *Drosophila*[6] and characterized structurally through the NMR analysis of Ets-1,[7] the PNT domain has since been encompassed by the closely related family of SAM domains. These latter domains are found in numerous proteins involved in eukaryotic developmental and signal transduction pathways (Figure 1).[8–10] Following the suggested nomenclature of the SMART[11] and SCOP[12] databases, the PNT domain distinction is restricted to members of the Ets family of transcription factors and represents a specific category within the larger set of SAM domains.

† C.D.M. & M.S. contributed equally to this work

Present address: L. N. Gentile, Department of Chemistry, Western Washington University, Bellingham, WA 98225, USA.

Abbreviations used: CD, circular dichroism; ESI-MS, electrospray ionization mass spectrometry; NOE, nuclear Overhauser effect; HSQC, heteronuclear single quantum correlation; MWCO, molecular weight cut-off; rms, root-mean-square; T_m, midpoint thermal unfolding temperature; protein fragments are denoted in superscripted parentheses by the initial and final residue number from the native sequence, and if present, an asterisk (*) to indicate an N-terminal GSH or GSHM sequence remaining after thrombin cleavage of a His_6 affinity tag.

E-mail address of the corresponding author: mcintosh@otter.biochem.ubc.ca

0022-2836/$ - see front matter

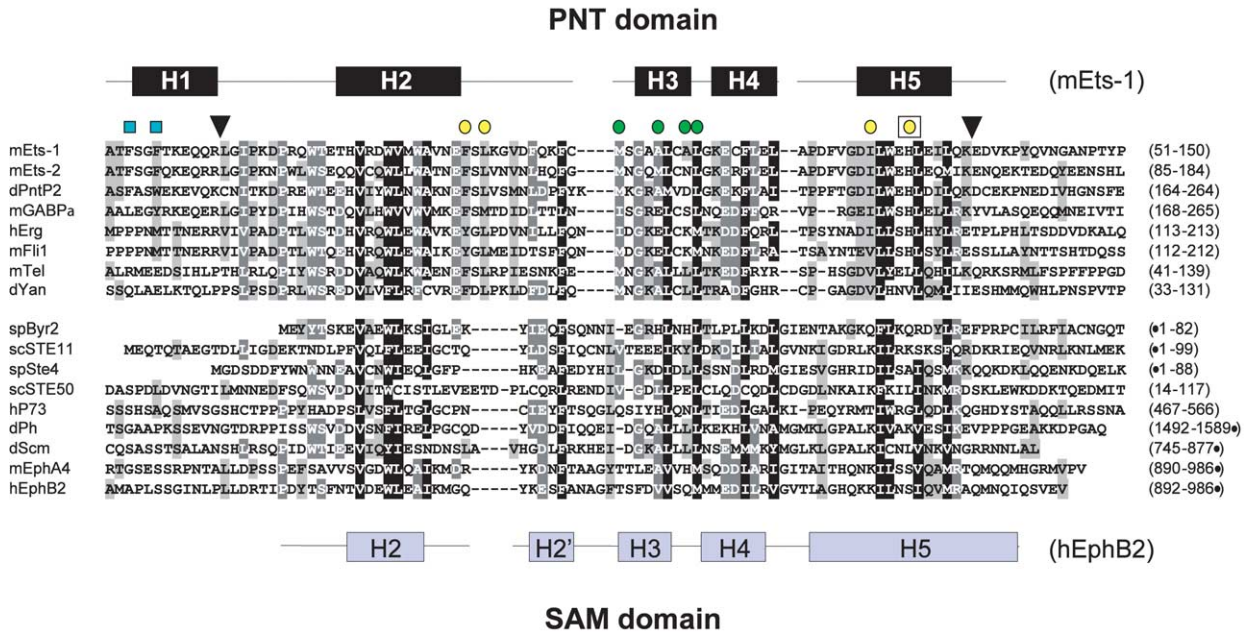

Figure 1. Sequence alignment of PNT (upper) and SAM domains (lower), constructed using CLUSTAL W (http://www.ebi.ac.uk/clustalw),[83] and refined by consideration of available structural information. Black, dark gray, and light gray shading indicates amino acid similarity in 80%, 60%, or 40%, respectively, of all the sequences shown. Also illustrated are the helical secondary structures of an example PNT domain (mEts-1; PDB code 1BQV) and SAM domain (EphB2 receptor; PDB code 1B4F). Note that helix H1 is absent in some PNT domains and in all SAM domains characterized to date, while a small α or 3_{10} helix H2′ is only present in a subset of SAM domains. Green and yellow circles correspond to residues within the EH and ML interfaces involved in Tel PNT domain self-association, respectively.[13] The boxed yellow circle indicates the location of the buried histidine in GABPα and Ets-1. Cyan squares denote residues key to the intramolecular packing of helix H1 in GABPα$^{(*168–254)}$. The approximate boundaries of the minimal-sized structured PNT domain are indicated by inverted arrows. Sequences (numbered in parentheses, with a black dot indicating the N or C terminus of the full-length protein) correspond to PNT domains from murine Ets-1 (mEts-1; SwissProt accession number P27577), murine Ets-2 (mEts-2; P15037), *Drosophila* Pointed-P2 (dPntP2; P51023), murine GABPα (Q00422), human Erg (hErg; P11308, with p55 isoform based on GenBank accession number AAH40168), murine Fli-1 (mFli-1; P26323), murine Tel (P97360), *Drosophila* Yan (dYan; Q01842), and SAM domains from *S. pombe* Byr2 (spByr2; P28829), *S. cerevisae* STE11 (scSTE11; P23561), *S. pombe* Ste4 (spSte4; P36622), *S. cerevisiae* STE50 (scSTE50; P25344), human p73 (hP73; O15350), *Drosophila* Polyhomeotic (dPh; P39769), *Drosophila* Sex combs on midleg (dScm; Q9VHA0), murine EphA4 receptor tyrosine kinase (hEphA4; Q03137), and human EphB2 receptor tyrosine kinase (hEphB2; P29323).

A role in protein–protein association was established for the PNT domain based on the Tel (or ETV6) transcription factor as a model system. Biochemical studies defined the potential of the Tel PNT domain to oligomerize in the context of the native transcriptional repressor,[13] as well as in oncogenic chimeric proteins involving this domain genetically fused with catalytic tyrosine kinase domains including those of PDGFRβ[14] and NTRK3,[15] or with fragments of transcription factors such as AML1.[16] When expressed in isolation, the PNT domain from Tel self-associates to form long insoluble polymers that derive from head-to-tail binding of two hydrophobic surface patches.[13,17] A similar polymeric helical structure is observed for the isolated SAM domain from polyhomeotic, a member of the *Drosophila* Polycomb group of transcriptional repressors.[18] There is also indication of weak dimerization or polymerization of SAM domains from the Ephrin receptor tyrosine kinases EphA4 and EphB2,[19–21] although this interaction appears dispensable when studied by mouse genetic experiments.[22]

In addition to homo-oligomerization, PNT and SAM domains may also form heteromeric complexes. For example, Tel2 can form homotypic polymers, as well as mixed polymers with its paralogue Tel.[23,24] In the case of the Polycomb group proteins, intact SAM domains are required for the interaction of Scm and polyhomeotic,[25,26] and the transduction of the yeast mating signal is mediated by the association of the SAM domains from Ste4 and Byr2 in *Schizosaccharomyces pombe* (STE50 and STE11 in *Saccharomyces cerevisiae*).[27–29] Other functions have also been discovered for this widespread module, including the protein association role of Ets-1 PNT domain as a docking module for the MAP kinase ERK2,[30] the RNA binding activity of the Smaug SAM domain,[31,32] and the interaction of the p73 SAM domain with phospholipids.[33]

Despite this diversity in function, as well as varying levels of sequence identity, the PNT and SAM domains characterized to date share a significant degree of structural similarity based on a core bundle of four α-helices.[10] However,

plasticity in this fold is also apparent with both the significant variation in the length of these core helices, and the inclusion of an additional α helix or 3_{10} helix in specific variants. For example, a fifth helix H1 is present at the N terminus of the PNT domain from Ets-1, whereas several SAM domains have a helix H2′ within the core bundle (Figure 1; for consistency, helices are numbered according to the original description of the Ets-1 PNT domain[7]).

The role of the PNT domain-specific helix H1, present in Ets-1 but absent in Tel, remains to be established. This helix is an integral component of the Ets-1 PNT domain, as evident by NMR-based structural and relaxation studies and its required presence for proper protein expression and folding.[7] However, the sequences corresponding to helix H1 of Ets-1 show limited similarity among Ets family members, and thus it was not apparent whether this is a conserved component of the PNT domain fold. In order to address this issue, and to elucidate the common elements that define the architecture of this domain, the structures of two additional PNT domains from Erg and GABPα were determined by using NMR spectroscopy. Along with Ets-1 and Tel, Erg and GABPα represent the four main sub-families of Ets PNT domains, as determined by a phylogenetic analysis (Supplementary Figure S1).

GABP (GA-binding protein), also known as NRF-2 (nuclear respiratory factor 2) and E4TF1 (adenovirus E4 transcription factor 1), consists of an α and β subunit that form a dimeric complex in solution[34] and bind with high affinity as a hetero-tetramer to tandem pairs of 5′-GGA-3′ motifs within GA-rich promoter elements.[35] The GABPα subunit is a member of the Ets family, characterized by a central PNT domain and a C-terminal DNA-binding ETS domain (Supplementary Figure S2).[36] The non-Ets family GAPBβ subunit homodimerizes *via* a coiled-coil motif, and contains ankyrin repeats that bind to the C-terminal sequences of two separate α subunits, which include residues from their ETS domains, to form a high affinity GAPB($\alpha\beta_2\alpha$)–DNA complex on tandem promoter sites.[35,36] Through this binding, GABP is involved in the regulation of a wide variety of genes, such as those required for T-cell activation, post-synaptic neuromuscular differentiation, and expression of the herpes simplex virus immediate-early genes.[37–42]

Erg was identified in a screen for Ets family members in a colon cancer cell line.[43,44] Transcription of the human *erg* gene leads to at least five possible splicing isoforms (Supplementary Figure S2), each of which functions as a transcriptional activator. However, they differ in the length of an N-terminal region and in the alternative splicing of exons between the central PNT domain and the C-terminal DNA-binding ETS domain.[43,45–48] Within frogs and mammals, Erg has been associated with the gene regulation of a variety of proteins implicated in endothelial cell differentiation and neural crest development.[49–52] In chicken, different Erg isoforms play opposing roles in bone morphogenesis.[53]

As demonstrated here, the additional N-terminal helix H1 is present within the PNT domains of Ets-1 and GABPα, but is absent in the domains from Tel and Erg. This difference in architecture results in surface variations that likely contribute to a range of biological functions specific to each of these transcription factors. The potential role of this structural variation in homo- and hetero-oligomerization was investigated using a set of isolated PNT domains. It was found that only the PNT domains from Yan and Tel form polymers, whereas those from Ets-1, Ets-2, Pnt-P2, Erg, Fli-1, and GABPα are monomeric in solution and fail to form hetero-oligomers with one another. The roles of these latter domains thus remain largely elusive, but may include interaction with any of the growing types of macromolecules thus far identified as binding partners for PNT and SAM domains.[10]

Results

Determination of the boundaries of the Erg and GABPα PNT domains

The boundaries of the Erg and GABPα PNT domains were defined using a similar strategy as that followed with Ets-1 (Supplementary Figure S2).[7] Briefly, partial proteolysis of an initial construct, $Erg^{(*1–201)}$, revealed that fragments from Arg124 or Val125 through Pro201 were resistant to trypsin cleavage (Figure 2(a)). In parallel, NMR spectroscopy demonstrated that the first ~120 amino acid residues of $Erg^{(*1–201)}$ displayed random coil chemical shifts and were dispensable in terms of stability and structure of the PNT domain (Figure 2(b)). These results strongly suggested that a fifth N-terminal helix is not present in the Erg PNT domain. However, limited susceptibility to trypsin cleavage was also observed for the well-structured helix H1 in Ets-1.[7] Therefore, to prevent inadvertent deletion of a potentially unstable helix H1 in Erg, we chose to study a fragment containing residues 108–201. The final Erg construct, $Erg^{(*108–201)}$, corresponds to the full length of the Ets-1 PNT domain and contains sequences invariant between all five Erg isoforms. The purified protein formed a stable structure that yielded a circular dichroism (CD) signal indicative of a significant helical content and that exhibited a reversible two-state folding with a T_m of 64 °C at pH 7.0. However, further studies, shown below, demonstrated that the first ~20 residues of $Erg^{(*108–201)}$ are indeed conformationally disordered and do not adopt a helical structure in solution.

In the case of GABPα, an initial 117 residue fragment, $GABP\alpha^{(138–254)}$, corresponding to $Ets\text{-}1^{(29–138)}$, was cloned and characterized. Consistent with the presence of a stable and folded PNT domain, CD spectroscopic studies revealed that the protein had significant helical content and a $T_m >$ 80 °C at pH 7.2. However, the first ~30 residues of

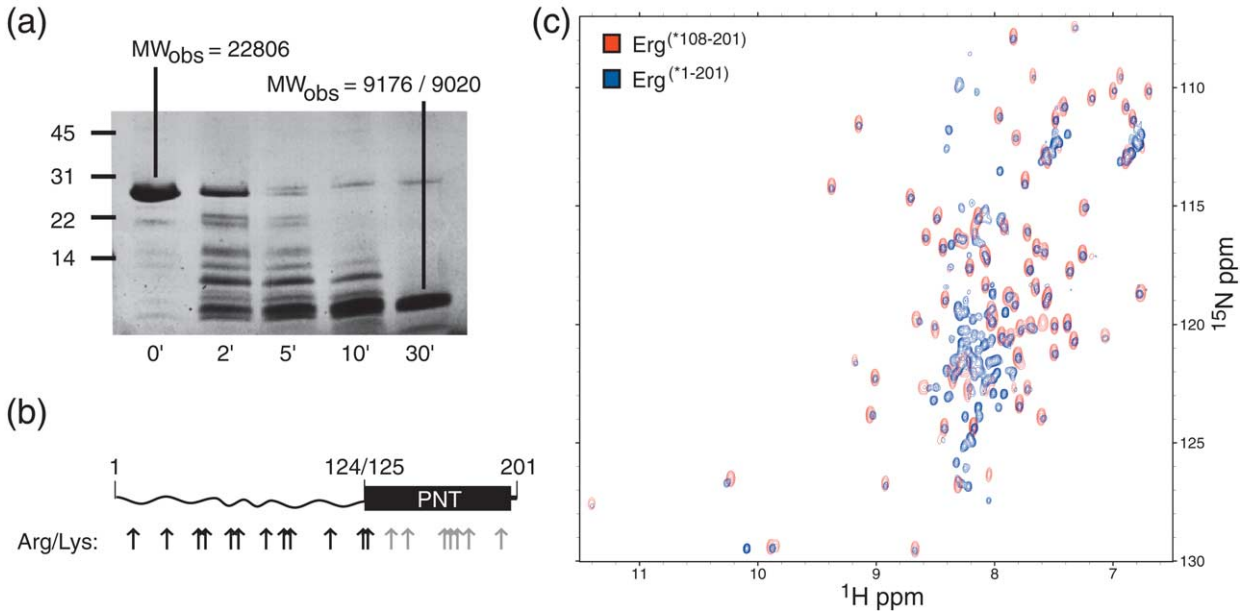

Figure 2. Determination of the Erg PNT domain boundaries. (a) The time-course of the limited trypsin digestion of Erg$^{(*1–201)}$, monitored by a 20% SDS-PAGE PhastGel (Amersham Biosciences) and visualized by Coomassie staining, revealed that the structured PNT domain is encompassed within residues 124/125 to 201. ESI-MS analysis of the ^{15}N-labeled protein before digestion yielded a molecular mass of 22,806 Da (corresponding to residues 1–201 with a GSH tag), whereas after 30 minutes of digestion, the peptides were 9176 Da (residues 124–201) and 9020 Da (residues 125–201). (b) Schematic of all potential trypsin cleavage sites in Erg$^{(*1–201)}$ (arrows). Following the 30 minutes incubation, residues susceptible to cleavage are shown with black arrows, and those resistant to trypsin proteolysis with gray. Also indicated are the location of the Erg PNT domain (black box), as well as the conformationally flexible N-terminal region of Erg$^{(*1–201)}$ (wavy line). (c) The exact superimposition of the corresponding peaks in the ^{1}H–^{15}N HSQC spectra of ^{15}N-labeled Erg$^{(*1–201)}$ (blue) and Erg$^{(*108–201)}$ (red) demonstrates that the PNT domain structure is independent of the presence or absence of residues 1–107. Note that the residues 1–107 in Erg$^{(*1–201)}$ yield resonances with ^{1}H chemical shifts clustered in the region of the spectra diagnostic of a flexible, unfolded polypeptide. The single downfield peak at 10.1 ppm specific to Erg$^{(*1–201)}$ arises from the indole N$^{\varepsilon 1}$H of Trp56.

GABP$\alpha^{(138–254)}$, and the last ~70 residues of a second construct, GABP$\alpha^{(168–319)}$ that also contained the MAP kinase phosphoacceptor site Thr280,[42] were found to be conformationally disordered as evident by random coil amide chemical shifts, characteristic ^{15}N relaxation properties, and partial proteolysis (data not shown). Accordingly, the minimal-sized construct GABP$\alpha^{(*168–254)}$ was utilized for detailed structural studies of its PNT domain.

Erg and GABPα PNT domains are monomeric and do not form Tel-like polymers

Since the function of the Tel PNT domain as a self-association module is important in several oncogenic fusion proteins, and likely in the full-native species,[13–16] the oligomeric state of the Erg and GABPα PNT domains were investigated. As demonstrated by the following evidence, the PNT domains from GABP$\alpha^{(168–254)}$ and Erg$^{(*108–201)}$ are similar to that of Ets-1, being monomeric in isolation. In the case of GABP$\alpha^{(168–254)}$, equilibrium ultracentrifugation studies (70 μM protein in 20 mM sodium phosphate (pH 7.2), 100 mM NaCl) yielded an apparent molecular mass of 11.7 kDa, in agreement with that calculated for the single polypeptide chain of this species (10,320 Da). Similarly, ^{15}N relaxation measurements, recorded for GABP$\alpha^{(168–254)}$ at 30 °C, revealed a global rotational correlation time of 5.3 ns. This value is consistent with that expected for a compact protein of ~10 kDa.[54] The ^{1}H–^{15}N heteronuclear single quantum coherence (HSQC) spectra of GABP$\alpha^{(168–254)}$ were also invariant between 0.1 and 3.7 mM, disfavoring the possibility of weak self-association even at elevated protein concentrations. Finally, GABP$\alpha^{(168–254)}$ behaved as a monomer during analysis by gel-filtration (apparent molecular mass of 10(±2) kDa) and native gel electrophoresis, and did not exhibit cross-linking when treated with glutaraldehyde (see below and Figure 5).

In the case of Erg$^{(*108–201)}$, gel-filtration experiments yielded an apparent molecular mass of 14(±3) kDa, consistent with a calculated molecular mass of 11,381 Da for the single polypeptide chain. Native gel electrophoresis and glutaraldehyde cross-linking also failed to detect the presence of dimeric or oligomeric species (Figure 5). Finally, the NMR spectroscopic linewidths exhibited by Erg$^{(*108–201)}$ are in keeping with a monomeric protein with the above molecular mass.

Determination of the Erg and GABPα PNT domain structures by NMR spectroscopy

Similar to the previously characterized Ets-1$^{(29–138)}$,[7] both Erg$^{(*108–201)}$ and GABPα$^{(*168–254)}$ yield well dispersed ^{1}H–^{15}N HSQC spectra with line-widths indicative of stably folded, low molecular mass proteins (Supplementary Figure S3). The resonances from the ^{1}H, ^{13}C, and ^{15}N nuclei in the backbones and side-chains of these two PNT domain-containing fragments were essentially completely assigned using a standard suite of multidimensional heteronuclear NMR correlation experiments.[55,56] Based on these assignments, extensive distance, dihedral angle, and orientational restraints were derived for Erg$^{(*108–201)}$ and GABPα$^{(*168–254)}$ from ^{13}C and ^{15}N-resolved nuclear Overhauser effect spectroscopy (NOESY) spectra, chemical shift and *J* coupling information, and residual dipolar coupling measurements (Table 1).

The ensembles of structures calculated for Erg$^{(*108–201)}$ and GABPα$^{(*168–254)}$ using ARIA/CNS[57–60] are presented in Figure 3. Both proteins have a well-defined core of four α-helices, whereas GABPα$^{(*168–254)}$ has an additional N-terminal helix (Erg$^{(*108–201)}$: H2, Thr136–Glu149; H3, Gly165–Cys169; H4, Lys173–Gln177; and H5, Asn184–Arg196; GABPα$^{(*168–254)}$: H1, Ala168–Leu180; H2, Thr191–Glu204; H3, Gly219–Leu222; H4, Gln227–Arg233; and H5, Glu238–Lys249). In contrast to this fifth helix in GABPα$^{(*168–254)}$, the N-terminal ~20 residues of Erg$^{(*108–201)}$ exhibit high rms deviations within the structural ensemble, indicative of conformational disorder.

The conserved architecture of the core four-helix bundles of Erg$^{(*108–201)}$ and GABPα$^{(*168–254)}$ is clearly evident in Figure 3. This common fold reflects the conservation of amino acids contributing to the hydrophobic interior of each of these PNT domains. In particular, their interiors are formed by the van der Waals packing of non-polar atoms from residues in the amphipathic helix H2 (Val139/Val194, Arg140/Leu195, Trp142/Trp197, Leu143/Val198, Ala146/Val201, and Val147/Met202 from Erg/GABPα), in the loop between helices H2 and H3 (Val155/Ile210, Ile157/Leu212, and Phe160/Leu215), in helix H3 (Leu168/Leu222), in or following helix H4 (Phe176/Phe230 and Leu179/Arg223), and in the amphipathic helix H5 (Leu188/Leu240, Leu189/Trp241, Leu192/Leu244, and Leu195/Leu247). Not surprisingly, as seen in the sequence alignment of Figure 1, this pattern of

Table 1. Statistics for the NMR-derived structures of Erg$^{(*108–201)}$ and GABPα$^{(*168–254)}$

	Erg$^{(*108–201)}$	GABPα$^{(*168–254)}$
A. *Summary of restraints*		
NOEs[a]		
Intraresidue	687 (160)	1004 (69)
Sequential	390 (201)	544 (104)
Medium-range ($1<\|i-j\|<5$)	298 (209)	452 (118)
Long-range ($\|i-j\|\geq 5$)	474 (385)	660 (236)
All	1849 (955)	2660 (527)
Dihedral angles		
ψ	58	67
ϕ	47	67
χ_1	8	0
Hydrogen bonds	2×14	2×14
Residual dipolar couplings	259	276
B. Deviation from restraints		
NOE restraints (Å)	0.022±0.001	0.016±0.002
Dihedral restraints (deg.)	0.90±0.11	0.35±0.05
Residual dipolar coupling restraints (Hz)	1.34±0.04	1.10±0.02
Deviation from idealized geometry		
Bond lengths (Å)	0.003±0.000	0.004±0.000
Bond angles (deg.)	0.68±0.01	0.76±0.02
Improper angles (deg.)	0.73±0.02	1.91±0.11
Mean energies[b] (kcal mol^{-1})		
E_{vdw}	−267±27	−326±15
E_{bonds}	39±3	25±2
E_{angle}	289±18	241±11
E_{NOE}	104±13	45±10
E_{cdih}	3±1	1.0±0.3
E_{SANI}	467±17	274±7
rmsd from average structure (Å)		
All heavy atoms	0.72±0.06[c]	0.62±0.09[d]
Backbone (C′, C$^{\alpha}$, N)	0.26±0.04[c]	0.20±0.03[d]

[a] Number of unambiguous restraints, with ambiguous restraints in parentheses.
[b] Final ARIA/CNS energies for van der Waals (vdw), bonds, angles, NOE restraints (NOE), dihedral restraints (cdih) and residual dipolar coupling restraints (SANI).
[c] Structured region spanning residues 127–197 (i.e. ^{1}H–{^{15}N} NOE >0.5).
[d] Structured region spanning residues 168–254.

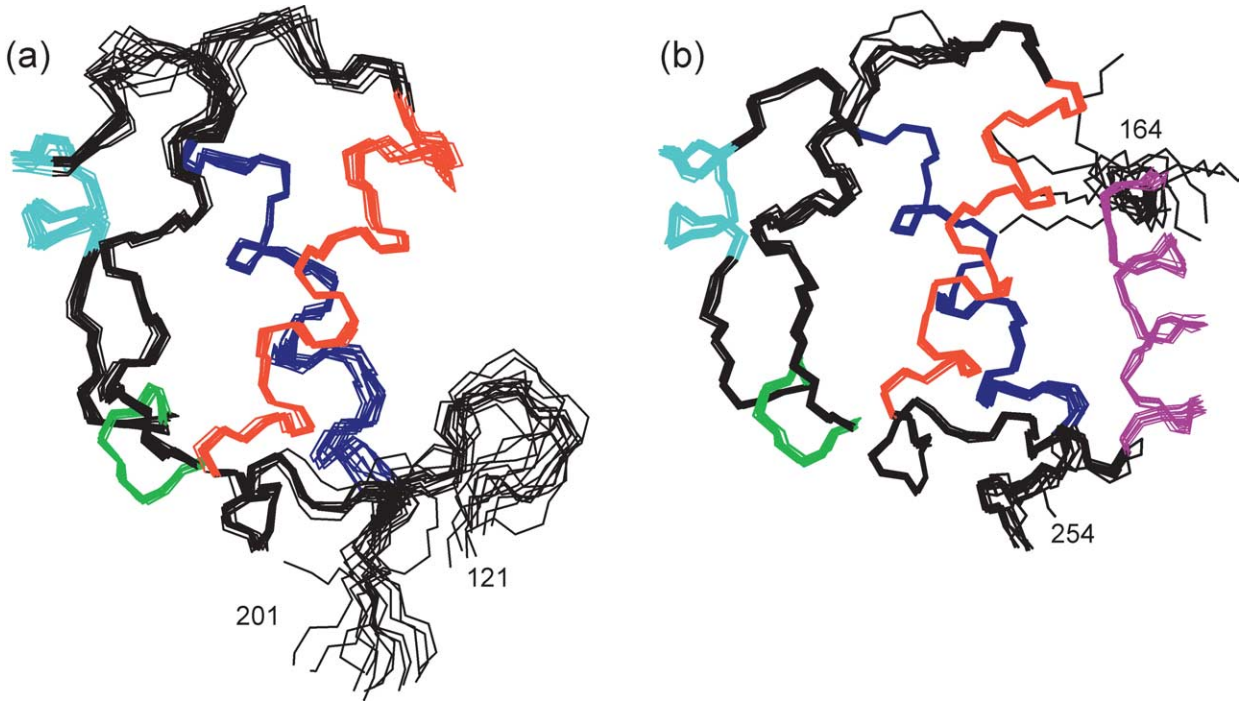

Figure 3. NMR-derived structural ensembles of the Erg and GABPα PNT domains. Superimpositions of the 15 best water-refined structures of (a) Erg$^{(*108–201)}$ and (b) GABPα$^{(*168–254)}$ are shown. The core four α-helices H2 (red), H3 (green), H4 (cyan) and H5 (blue) common to both PNT domains are indicated, along with the additional N-terminal helix H1 (purple) present in GABPα$^{(*168–254)}$. The high rms deviations for the N-terminal 22 residues of Erg$^{(*108–201)}$ (not shown fully for clarity) reflect conformational flexibility as indicated by susceptibility of these residues towards tryptic digestion, as well as by random coil chemical shifts[55] (Figure 2) and $\{^1H\}^{15}N$-NOE values <0.5 (Figure 5). In addition, mobility is observed in the loop between helices H4 and H5 of Erg$^{(*108–201)}$ based on reduced $\{^1H\}^{15}N$-NOE values, weak amide peaks for these residues in the $^1H–^{15}N$ HSQC spectrum, likely due to exchange broadening with the solvent water protons, and a limited number of interproton NOE restraints observed for Thr76, Pro77, Ser78, Tyr79 and Asn80. The corresponding H4-H5 loop is also dynamic in Ets-1$^{(29–138)}$. For GABPα$^{(168–254)}$, conformational flexibility on the sub-nanosecond time-scale is observed for residues 208–210, located in the loop between helices H2 and H3, evident by longer ^{15}N T_1 and T_2 lifetimes (data not shown) and lower $\{^1H\}^{15}N$-NOE values.

hydrophobic residues is conserved across the entire family of Ets PNT domains and is reflected in the tertiary structures of the Ets-1 and Tel PNT domains.[7,13]

The structural ensemble of GABPα$^{(*168–254)}$ also includes a well-defined N-terminal helix H1 (Figure 3(b)). Although predominantly polar in nature, this helix is appended to the core four-helix bundle through van der Waals contacts between Leu170 and Tyr173 in helix H1 with Leu246 in helix H5 and with Trp197 and Trp200 in helix H2, respectively. Interestingly, this helical packing leads to the burial of His243. Indeed, during the chemical shift assignments of GABPα$^{(*168–254)}$, it was noted that the neutral imidazole ring of this histidine was protected from hydrogen exchange with water as evident by the observation of a signal at 13.8 ppm and 167.5 ppm from its labile $^{15}N^{\varepsilon 2}H$. The corresponding His128 in Ets-1$^{(29–138)}$ also yields a detectable $^{15}N^{\varepsilon 2}H$ signal, indicative of protection from exchange due to its burial by helix H1.[7] In contrast, due to the absence of helix H1 in Erg$^{(*108–201)}$, the corresponding His191 is solvent-exposed and undergoes rapid hydrogen exchange under the conditions used for this NMR study.

^{15}N relaxation analyses confirm that the PNT domains from Ets-1 and GABPα, but not Erg, contain helix H1

As summarized in a brief communication on the resonance assignments of the Ets family PNT domains,[55] a striking similarity exists between the secondary ($\delta_{obs} - \delta_{random\ coil}$) $^{13}C^{\alpha}$, $^1H^{\alpha}$, $^{13}C'$, $^1H^N$, and ^{15}N chemical shifts of residues in Erg$^{(*108–201)}$ and GABPα$^{(168–254)}$ corresponding to helices H2–H5 of Ets-1$^{(29–138)}$ plus ~ten residues encompassing a turn preceding these helices. This suggested that the three Ets proteins share a conserved four-helix bundle core. However, only GABPα$^{(168–254)}$ exhibited similar secondary shifts as found for residues in the N-terminal helix H1 of Ets-1$^{(29–138)}$, whereas those of Erg$^{(*108–201)}$ were indicative of a random coil conformation. The NMR-derived structures presented in Figure 3 verify the presence of five α-helices in the Ets-1 and GABPα PNT domains, whereas Erg contains only the four core helices exhibited by SAM domains.

The divergence of the structures of the Ets family PNT domains is supported further by $\{^1H\}^{15}N$-NOE measurements (Figure 4). The $^1H–^{15}N$

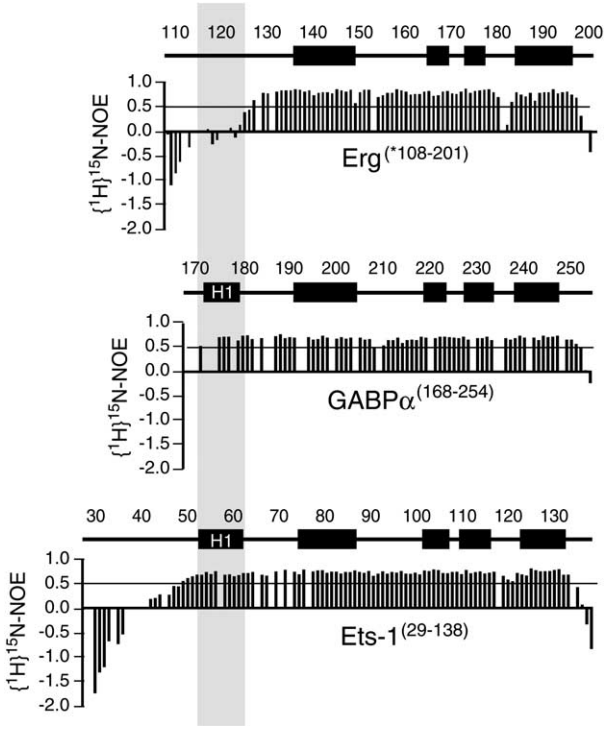

Figure 4. $\{^1H\}^{15}N$-NOE relaxation studies of ^{15}N-labeled Erg$^{(*108–201)}$, Ets-1$^{(29–138)}$, and GABP$\alpha^{(168–254)}$ distinguish regions exhibiting conformational mobility on a sub-nanosecond time-scale ($\{^1H\}^{15}N$-NOE <0.5) from those with restricted motions ($\{^1H\}^{15}N$-NOE ≥ 0.5). These data demonstrated that the Erg PNT domain lacks the N-terminal helix H1 found in Ets-1 and GABPα, and that this helix is an integral component of the folded structures of these latter two proteins. Residues for which no data is shown are either proline or have weak or overlapping ^1H–^{15}N signals.

heteronuclear NOE is a highly sensitive measure of motions on the sub-nanosecond time-scale, with values <0.5 indicative of significant backbone conformational mobility. For Erg$^{(*108–201)}$, only amides of residues 127–197, encompassing the core helices H2–H5 and a turn preceding helix H2, display restricted mobility ($\{^1H\}^{15}N$-NOE ≥ 0.5), with clear conformational flexibility evident at the N and C-terminal regions ($\{^1H\}^{15}N$-NOE <0.5), including that corresponding to helix H1. This is consistent with the elevated rms deviations of these latter regions in the structural ensemble of Erg$^{(*108–201)}$ (Figure 3(a)), as well as their proteolytic sensitivity (Figure 2). In contrast, amides forming helix H1 in Ets-1$^{(29–138)}$ and GABP$\alpha^{(168–254)}$ have heteronuclear NOE values comparable to those from the remainder of their PNT domains. Therefore, the Erg PNT domain clearly lacks a structured helix H1, whereas this helix is an integral component of the PNT domains of Ets-1 and GABPα.

Screen for *in vitro* association of Ets family PNT domains

The hypothesis of the PNT domain as a protein–protein oligomerization module has its origins in the clear association properties of Tel, both in the context of wild-type and oncogenic fusion proteins.[13,14] In addition, several groups have reported the association of other PNT domains in either a homotypic or heterotypic manner. Examples include putative interactions between Erg and Ets-1, Fli-1, Ets-2 and between Tel and Fli-1.[61–63] We have tested these potential associations *in vitro* using a matrix of PNT domains from Ets-1, Ets-2, Pnt-P2, GABPα, Fli-1, Tel, and Yan.

As discussed above, extensive structural and biophysical characterization prove that isolated PNT domains of Ets-1, Erg, and GABPα are monomeric. In contrast, the domain of Tel self-associates to form insoluble polymers.[13] The PNT domain of Yan was also found to be insoluble when expressed in *E. coli* (data not shown). In keeping with its sequence similarity to Tel, this behavior likely results from self-association *via* a head-to-tail interaction of two hydrophobic surfaces.

The remaining PNT domains from Ets-2, Pnt-P2, and Fli-1 also produced soluble proteins that exclusively displayed monomeric behavior in gel filtration, native gel electrophoresis, and glutaraldehyde cross-linking studies (Figure 5, and data not shown). This conclusion is reinforced by the narrow line-widths observed in the ^1H-NMR spectra of these proteins, indicative of well-folded species within the 10 kDa range (Supplementary Figure S4).

The soluble PNT domains were also tested in a pair-wise manner for heterotypic interactions using native gel electrophoresis and glutaraldehyde cross-linking. No evidence of association was observed by either method with any combination of the Ets-1, Erg, GABPα, Ets-2, Pnt-P2, or Fli-1 proteins (data not shown). Since Tel has been implicated in complex formation with some of these PNT domains, two soluble variants of Tel$^{(38–127)}$ were also examined. These variants contain single point mutations (A94D or V113E) that inhibit polymerization due to incorporation of a negative charge into one of the two hydrophobic association interfaces identified through X-ray crystallographic analysis of this protein.[13] However, since one native interface is retained in each mutant Tel PNT domain, the two form a high-affinity soluble dimer upon mixing. Glutaraldehyde cross-linking confirmed that Tel$^{(*36–127)}$-A94D and Tel$^{(*36–127)}$-V113E interact in solution (Figure 5, lane 3). However, we did not observe any crosslinking between either of these soluble Tel variants and any of the Ets family PNT domains tested, including that of Fli-1. Furthermore, no changes in the ^1H–^{15}N HSQC spectra of either ^{15}N-labelled Tel$^{(*36–127)}$-A94D or Tel$^{(*36–127)}$-V113E were observed when unlabelled Fli-1$^{(*106–200)}$ or Erg$^{(*108–201)}$ were added to two molar equivalents ($\sim$1 mM protein; data not shown).

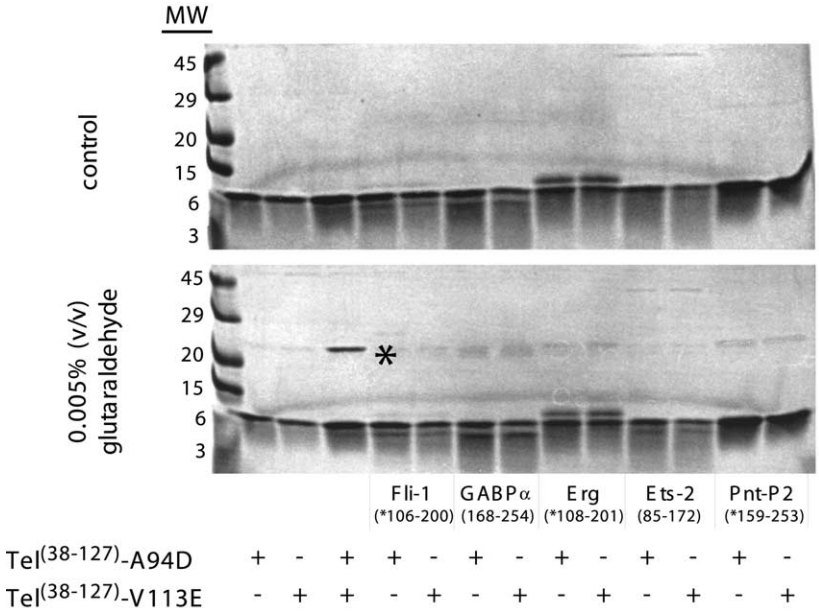

Figure 5. Test for *in vitro* homo- and heteromeric complex formation by isolated Ets family PNT domains. SDS-PAGE gels are shown for 20 μM protein samples before (top) and after (bottom) treatment with 0.005% (v/v) glutaraldehyde for 30 minutes at 22 °C. Consistent with the report by Kim *et al.*,[13] the Tel variants Tel$^{(*38–127)}$-A94D and Tel$^{(*38–127)}$-V113E do not self-associate, yet form a heterodimer that can be detected by chemical cross-linking (marked by *). In contrast, no dimeric or oligomeric species are detected for Fli-1$^{(*106–201)}$, GABPα$^{(168–254)}$, Erg$^{(*108–201)}$, Ets-2$^{(85–172)}$, or Pnt-P2$^{(*159–253)}$, either without or with the addition of Tel$^{(*38–127)}$A94D or Tel$^{(*38–127)}$V113E. Similar results were found when the five PNT domains were tested in pairwise combinations with each other, as well as with Ets-1$^{(29–138)}$ (not shown). The purified PNT domains have approximately the same molecular mass, and thus the monomeric species that appear at the bottom of the gel are not well resolved.

Discussion

PNT domain architecture

The structures of the Erg and GABPα PNT domains, added to the previously characterized domains from Ets-1 and Tel, complete our view of the four phylogenetic sub-families of the Ets family PNT domain (Supplementary Figure S1). Ribbon models of these proteins reveal a similar arrangement of a bundle of four α-helices, augmented with an additional N-terminal α-helix in GABPα and Ets-1 (Figure 6(a)). This fold similarity extends to the molecular level, as the C$^{\alpha}$ atoms in helices H2–H5 of the average structures of Erg$^{(*108–201)}$ and GABPα$^{(*168–254)}$ superimpose with an rmsd value of 1.1 Å. Comparison to sequence-related C$^{\alpha}$ atoms in the previously determined Tel crystal structure reveals comparable similarities, with rmsd values of only 0.9 and 1.3 Å, for Erg$^{(*108–201)}$ and GABPα$^{(*168–254)}$, respectively.

Inspection of ribbon models for representative SAM domains of STE50,[29] polyhomeotic,[18] EphB2 receptor,[19] and p73,[64] illustrates the common and variable attributes that exist amongst the larger group of both PNT and SAM domains (Figure 6(b)). Although the core bundle of four helices is clearly preserved, variations exist for specific members. For example, helix H2 is lengthened in the PNT domain relative to the SAM domain, whereas the reverse tends to hold for helix H5. This is particularly noticeable with the Ephrin receptor SAM domains, for which helix H5 may play a role in self-association.[19,21] A variable helix, denoted herein as H2′, is also present in all SAM domains (Figures 1 and 6(b)). Furthermore, helices H2′ and H3 in these SAM domains may adopt either an α or 3_{10} helical conformation as defined by VADAR[65] and Promotif.[66] Interestingly, residues corresponding to helix H2′ of the SAM domain form a helical-like turn in the PNT domains (Figure 6(a)), allowing a similar contribution of moderately conserved non-polar side-chains to their hydrophobic cores (Figure 1).

Variable presence of the N-terminal helix H1

The most prominent structural difference exhibited by the four characterized PNT domains is the presence of an N-terminal helix in the domains from GABPα and Ets-1, but not Erg and Tel. NMR relaxation measurements (Figure 4), chemical shift analyses,[55] and structure calculations (Figures 3(b) and 6(a)) all confirm that residues Ala168–Leu180 of GABPα$^{(*168–254)}$ are an integral component of its PNT domain fold and adopt a stable helical conformation, with limited mobility on the sub-nanosecond time-scale. Similar techniques were used to characterize helix H1 in Ets-1$^{(29–138)}$.[7] In contrast, this element of secondary structure is clearly absent in Erg$^{(*108–201)}$, as also revealed by multiple lines of evidence. The residues corresponding to helix H1 are sensitive to limited

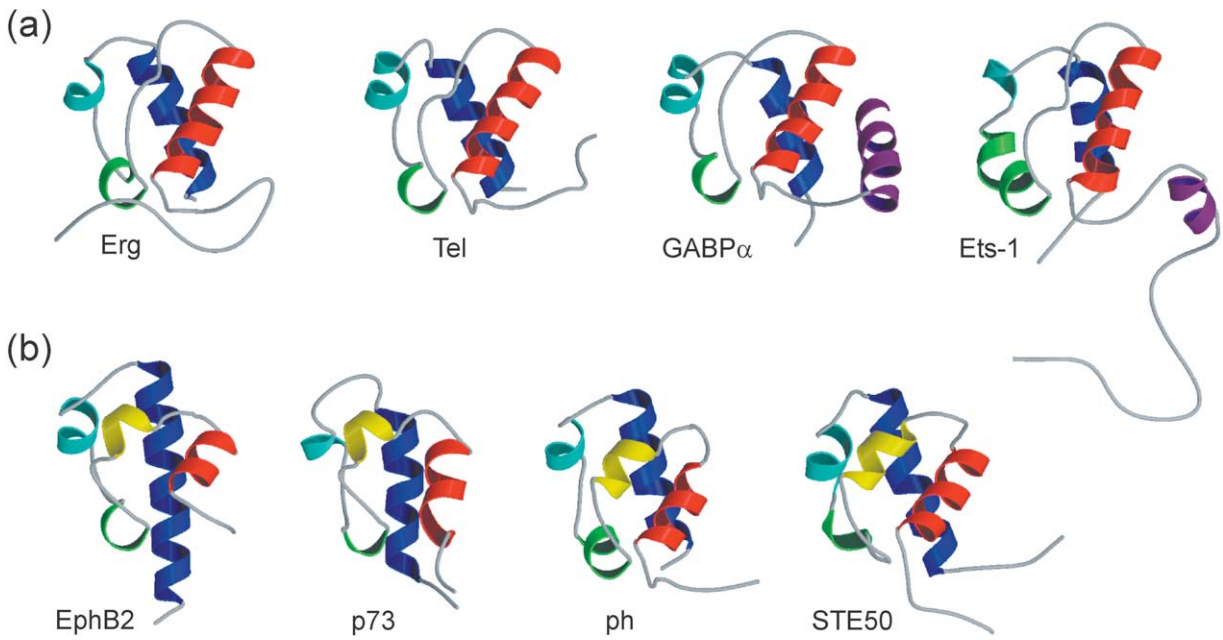

Figure 6. Ribbon diagrams of the tertiary structures of PNT and SAM domains. (a) PNT domains from Erg, GABPα, Tel (PDB code 1LKY), Ets-1 (PDB code 1BQV). (b) SAM domains from EphB2 receptor tyrosine kinase (PDB code 1B4F), p73 (PDB code 1COK), polyhomeotic (PDB code 1KW4) and STE50 (PDB code 1UQV). The proteins share an architecture of a core four α-helix bundle (H2, red; H3, green; H4, cyan; H5, blue). An N-terminal α-helix (H1, purple) is present in the Ets-1 and GABPα PNT domains, while a 3_{10} or α helix (H2′, yellow) is often found in the SAM domains. The positions of helices were determined using a consensus of VADAR[65] and Promotif[66] analyses. The Figure was made using MOLSCRIPT v2.1.2[84] and Raster3D v2.4b.[85]

proteolysis (Figure 2), exhibit random coil secondary chemical shifts,[55] are conformationally disordered by ^{15}N relaxation measurements (Figure 4), and display minimal interproton NOE interactions leading to high rms deviations within the structural ensemble calculated for this protein (Figures 3(a) and 6(a)). In addition, a conserved histidine that yields a distinctive downfield NMR signal due to its complete burial in the GABPα$^{(*168–254)}$ and Ets-1$^{(29–138)}$ structures by helix H1 is not protected from hydrogen exchange in Erg$^{(*108–201)}$.

Based on the sequence alignment (Figure 1) and phylogenetic clustering (Supplementary Figure S1), we predict that helix H1 is present in the PNT domains from all members the Ets-1 and GABPα sub-families, including Ets-2 and Pnt-P2. In particular, residues corresponding to Leu170 and Tyr173, which lead to the packing of helix H1 against helices H5 and H2, respectively, in GABPα$^{(*168–254)}$, are conserved only within these two sub-families. In contrast, along with an overall divergence in sequence of this N-terminal region, these residues are replaced by proline or charged amino acids in members of the Tel and Erg sub-families (Figure 1). Both substitutions would disrupt the formation and packing of a helix in these latter proteins. As also shown in Figure 6(a), the position of helix H1 appears to vary between the structures calculated for GABPα$^{(*168–254)}$ and Ets-1$^{(29–138)}$. However, as discussed previously,[7] the orientation of helix H1 could not be well defined for Ets-1$^{(29–138)}$ using NOE data collected with a 500 MHz spectrometer, and thus the significance of this difference is uncertain.

Even in the absence of a helix H1, it should be noted that the backbone amides of Val127–Ser135 preceding helix H2 in Erg$^{(*108–201)}$ are well-structured, based on heteronuclear {^{1}H}^{15}N-NOE relaxation values greater than 0.5 (Figure 4). In addition, these residues exhibit relatively low rms deviations within the Erg$^{(*108–201)}$ ensemble (Figure 3(a)), as do the corresponding residues in the NMR-derived structures of GABPα$^{(*168–254)}$ and Ets-1$^{(29–138)}$. Similarly, while helix H1 is absent in the Tel PNT domain, 16 amino acid residues N-terminal to helix H2 are also ordered within the crystal lattice formed by this protein.[13] Closer inspection of the structure of Erg$^{(*108–201)}$ reveals hydrophobic contacts between residues throughout this region (Val125, Val127, Ala129, and Pro131) with three conserved tryptophan side-chains (Trp134, Trp142, and Trp145) before and within helix H2. Furthermore, this proline is part of a distinct reverse turn involving $_{130}$DPTL$_{133}$. These features are completely conserved in the sequences (Figure 1), structures (Figures 3 and 6) and, where determined, main-chain chemical shifts[55] and ^{15}N relaxation properties (Figure 4), of the PNT domains from Ets-1, GABPα, and Tel. Therefore, the ~ten residues preceding helix H2 must be also included in the definition of the minimal-sized PNT domain due to their contribution towards its folded structure.

Although well characterized conformationally, the role of the variable helix H1 is less certain. As discussed below, the presence of this additional

helix may relate to the monomeric nature of the GABPα and Ets-1 PNT domains. Another potential role for helix H1 is linked to the involvement of the Ets-1 PNT domain as an ERK2 docking module.[30] The presence of this helix may be required for proper positioning of the phosphoacceptor Thr38, located in the flexible N-terminal region of Ets-1, upon docking of its PNT domain with this kinase. Finally, helix H1 may be required for specific interactions of Ets-1 and GABPα with yet unidentified partners involved in cellular signaling or transcriptional regulation.

Surface elements of Erg, GABPα, and Ets-1 conflict with a role in PNT–PNT association

Based on several *in vitro* assays for homo- and heterotypic interactions, it is evident that the PNT domains from Tel and Yan show the potential to self-associate. In contrast, the isolated domains from Ets-1, Ets-2, Pnt-P2, GABPα, Erg, and Fli-1 are both monomeric in solution and do not measurably associate in any pairwise combination tested. This lack of self-association has been previously described for full-length Ets-2, and the PNT domains of Ets-1, GABPα, Erg, and Fli-1,[7,13,14,61] and can be understood at the molecular level through comparison to the oligomerizing interfaces elegantly defined for the Tel PNT domain by a mutational solubility screen. Subsequent X-ray crystallographic and biophysical analyses demonstrated that two hydrophobic patches, designated ML and EH, mediate the head-to-tail association of the Tel PNT domain to form open-ended polymers (Figures 1 and 7).[13,17] The introduction of an acidic residue within either of these interfaces prevents the polymerization of the soluble mutants Tel$^{(*36–127)}$-A94D and Tel$^{(*36–127)}$-V113E. In the case of Erg$^{(*108–201)}$, the EH region retains a predominantly hydrophobic character, albeit with the inclusion of His191, whereas the ML surface is distinctly hydrophilic, with the incorporation of Glu167 and Lys170 (Figures 1 and 7).[13,17] The presence of these latter charged residues likely mimics the solubilizing mutation A94D generated for Tel, thereby preventing the head-to-tail association of the two interfaces in Erg$^{(*108–201)}$ as required for polymer formation. For GABPα$^{(*168–254)}$ there is a similar incorporation of a charged residue, Glu221, in the ML interface, which again likely mimics the Tel A94D solubilizing mutation (Figures 1 and 7). Furthermore, it is also possible that the presence of the N-terminal helix H1 directly prevents self-association of GABPα$^{(*168–254)}$ through the partial steric blockage of the EH interface. As noted previously, the imidazole ring of His243 is protected from solvent exchange due to its burial by helix H1; this histidine corresponds in primary and tertiary structure to Leu117 on the EH surface of Tel. A similar scenario may exist for Ets-1$^{(29–138)}$, for which His128 in the EH interface is also protected from solvent exchange, likely due to occlusion by helix H1 (Figure 7).

An explanation for the lack of pair-wise binding of Erg$^{(*108–201)}$ or Ets$^{(29–138)}$ to other PNT domains, such as those of the solubilized Tel mutants, is less obvious. These proteins still retain at least one solvent-exposed hydrophobic patch (the EH region for Erg, albeit with His191, and the ML region for Ets-1; Figure 7) that could participate in heterotypic protein interactions. Therefore, additional factors such as incompatible juxtapositioning of charges flanking these interfaces or the lack of precise van der Waals surface complementarity must inhibit complex formation, and allow for a separate, non-oligomerizing role for these monomeric PNT domains.

The observations by others of the self- or cross-association of various Ets PNT domain-containing constructs, such as those of Erg and Ets-2 by two-hybrid analysis,[61,62] may indicate a very weak binding affinity (>1 mM) that is not observed in our experiments. Alternatively, regions flanking the minimal PNT domain may be important for association, or unknown bridging partners and post-translation modifications could mediate the interaction in the yeast system. However, it is noteworthy that *in vitro*, several PNT domains have the potential for forming disulfide bonds if a reducing environment is not maintained. Indeed, a high propensity of cystine formation was observed for GABPα$^{(*168–254)}$ and Erg$^{(*108–201)}$, as well as Ets-1$^{(29–138)}$,[7] during their structural characterization.

PNT domain function

The functions of Ets family PNT domains established to date include the self-association of Tel,[14,15], the docking of Ets-1 to ERK2,[30] and the interaction of Ets-2 with the CREB-binding protein.[67] The MAP kinase phosphorylation of Yan and Pnt-P2 has also been reported to depend upon their interactions with Mae, a small protein containing a PNT domain yet lacking the ETS domain.[68] Unfortunately, the biological roles of the Erg and GABPα PNT domains have not yet been established. For example, we were unable to identify verifiable binding partners of the GABPα PNT domain using a yeast two-hybrid screen against a HeLa cDNA library (S.E.M. & I. Pot, unpublished results). Further investigation of Erg and GABPα proteins is therefore necessary to gain insight into the role of their respective PNT domains, which may include binding to a specific protein component of a signaling or transcription process, or indeed to one of the additional classes of biomolecules such as RNA and lipids that are connected with SAM domain function. It is also possible that a form of post-translation modification is required for the biological activity of the Erg and GABPα PNT domains. Such a finding would further increase the functional diversity displayed by this conserved protein module.

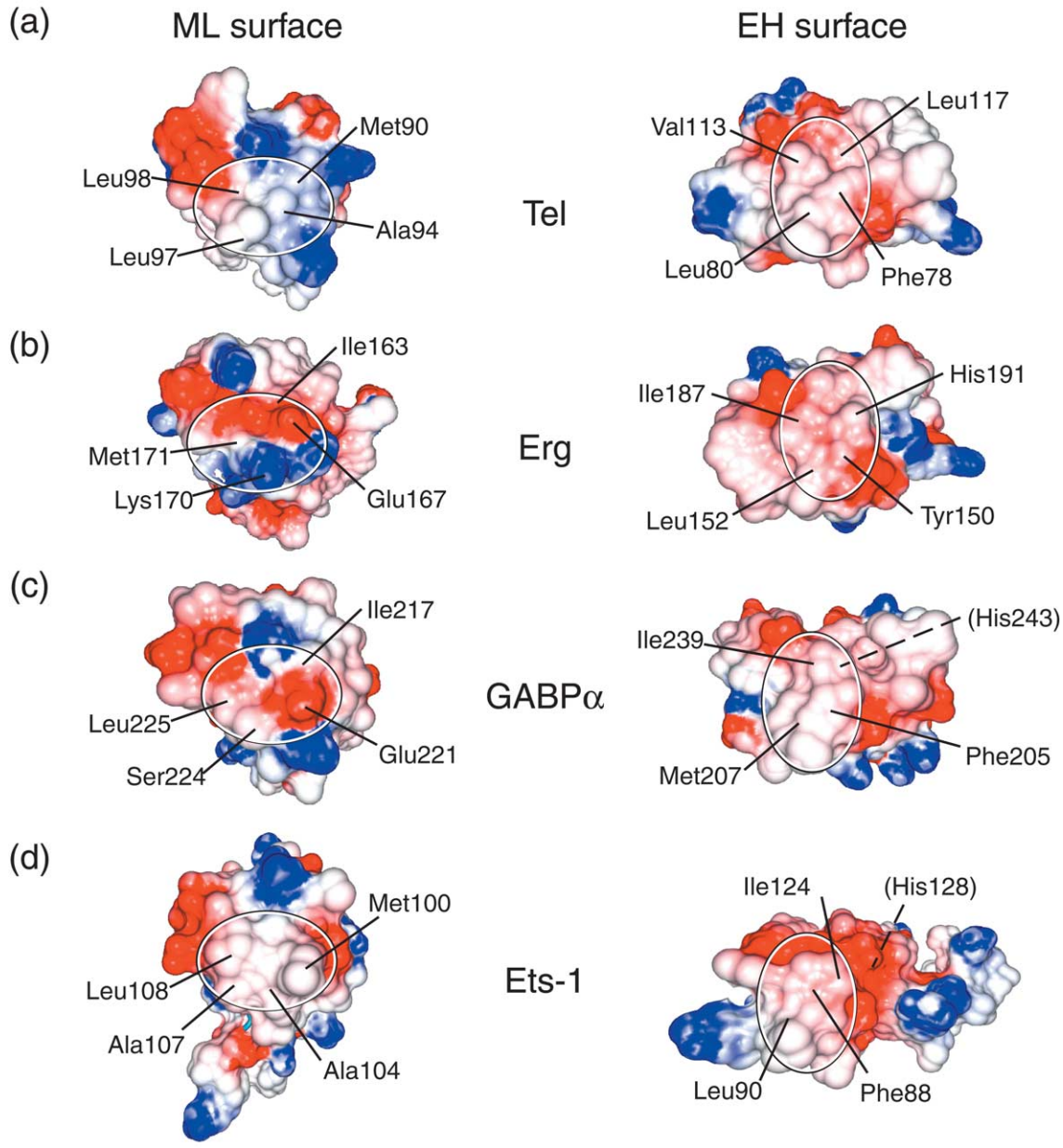

Figure 7. Surface representations of the PNT domains of (a) Tel, (b) Erg, (c) GABPα, and (d) Ets-1 corresponding to the ML (left) and EH (right) dimerization interfaces (ellipses) identified by mutational and crystallographic analyses of Tel.[14] The surfaces are color coded by electrostatic potential (blue, positive; red, negative), calculated using GRASP.[86] Consistent with experimental studies of Erg, GABPα and Ets-1, histidine residues are depicted as neutral at pH 7. The introduction of charged residues within the ML surface of Erg$^{(*108–201)}$ and GABPα$^{(*168–254)}$ and/or the partial steric blockage of the EH surface by helix H1 in GABPα$^{(*168–254)}$ and Ets-1$^{(29–138)}$ may prevent the self-association of these PNT domains. Shown in parentheses are His243 and His128 from GABPα$^{(*168–254)}$ and Ets-1$^{(29–138)}$, respectively, that are occluded on the EH interface by helix H1.

Materials and Methods

Protein expression and purification

A summary of the PNT domain-containing fragments used here is given in Supplementary Figure S2. Cloning of the Ets-1 constructs, Erg$^{(*108–201)}$ and GABPα$^{(168–254)}$ have been described.[7,55] Erg$^{(*1–201)}$ was cloned from the p55 isoform of human Erg (cDNA generously provided by M. Duterque-Coquillard, Institut de Biologie de Lille, France) into the pET28a vector (Novagen) using PCR-generated restriction enzyme sites. GABPα$^{(138–254)}$, GABPα$^{(*168–254)}$ and GABPα$^{(*168–319)}$, as well as Ets-2$^{(85–172)}$, Fli-1$^{(*106–200)}$ and Tel$^{(*38–127)}$ were PCR-amplified from a 10.5 day mouse embryo cDNA library. Pnt-P2$^{(*159–253)}$ and Yan$^{(*22–129)}$ were PCR-amplified from their full-length coding sequences in plasmids kindly provided by H. Ruohola-Baker (University of Washington, USA). The clones were inserted into pET28a vectors (Novagen) using PCR-introduced NdeI and HindIII or BamHI restriction sites, to allow for both exclusion and inclusion, respectively, of an N-terminal His_6 purification tag. Due to the presence of an internal NdeI site, Ets-2 was cloned

solely with NcoI and BamHI restriction sites into pET28a, resulting in exclusion of the N-terminal His_6-tag. In order to produce soluble $Tel^{(*38–127)}$, the A94D and V113E mutations were introduced into the pET28a:$Tel^{(*38–127)}$ plasmid using the QuikChange kit (Stratagene). All clones were verified by DNA sequence analysis.

Plasmids were transformed into *E. coli* BL21(λDE3) and grown in Luria broth (LB) or minimal M9 medium supplemented with the following: 1 g/l (^{15}N, 99%)-NH_4Cl for uniform ^{15}N labeling; 1 g/l (^{15}N, 99%)-NH_4Cl and 3 g/l ($^{13}C_6$, 99%)-glucose for uniform ^{13}C–^{15}N labeling; and 0.3 g/l ($^{13}C_6$, 99%)-glucose and 2.7 g/l ($^{12}C_6$)-glucose for non-random 10% ^{13}C-labelling. Samples of $GABP\alpha^{(168–254)}$, specifically deuterated with $^2H_5$-$\delta^{1,2}$,$\varepsilon^{1,2}$,ζ-Phe/$^2H_4$-$\delta^{1,2}$,$\varepsilon^{1,2}$-Tyr or $^2H_5$-$\delta^{1,2}$,$\varepsilon^{1,2}$,ζ-Phe/$^2H_5$-δ^1,ζ^2,$\zeta^3$$\varepsilon^3$,$\eta^2$-Trp, were prepared as described.[56] Protein expression was induced at $A_{600\ nm} \sim 0.6$ using 1 mM IPTG, followed by growth at 30 °C for three hours. Cells were collected by centrifugation at 3000*g*, and resuspended in binding buffer (50 mM Hepes (pH 7.5), 500 mM NaCl, 5% (v/v) glycerol, 5 mM imidazole) for His_6-tagged proteins, 50 mM Tris (pH 8.5) for pET22b-encoded $GABP\alpha^{(168–254)}$ and $Ets\text{-}1^{(29–138)}$ and pET28a-encoded $Ets\text{-}2^{(85–172)}$. Cells were disrupted by two cycles through a French press at 10,000 psi, followed by one minute of sonication with cooling on ice. Lysates were cleared with centrifugation at 10,000*g* for 30 minutes, followed by filtration through a 0.8 μm cut-off membrane.

All His_6-tagged proteins (including $Erg^{(*1–201)}$, $Erg^{(*108–201)}$, $GABP\alpha^{(*168–254)}$, $GABP\alpha^{(*1–319)}$, $Ets\text{-}1^{(*1–138)}$, $Fli\text{-}1^{(*106–200)}$, $Pnt\text{-}P2^{(*159–253)}$, $Tel^{(*38–127)}$, $Tel^{(*38–127)}$-A94D, $Tel^{(*38–127)}$-V113E, and $Yan^{(*22–129)}$) were purified by using Ni^{2+}-affinity chromatography. Following loading onto a 5 ml Hi-Trap column (Amersham Biosciences), non-specific bound proteins were removed by a 100–150 ml of wash buffer (50 mM Hepes (pH 7.5), 500 mM NaCl, 5% (v/v) glycerol, 60 mM imidazole). The His_6-tagged proteins were eluted at 1 ml/minute into 1.5 ml Eppendorf tubes with elution buffer (50 mM Hepes (pH 7.5), 500 mM NaCl, 5% (v/v) glycerol, 250 mM imidazole). Fractions were tested qualitatively for the presence of protein using a standard Bradford assay (BioRad), and protein-containing fractions were pooled and placed in 2K MWCO dialysis tubing. The His_6-tag was removed by incubation with thrombin (~0.5 mg) during overnight dialysis at 4 °C into NMR buffer containing 20 mM sodium phosphate (pH 7.0), 50 mM NaCl, and 2 mM β-mercaptoethanol. Proteolysis was monitored by a reduction in apparent molecular mass using SDS-PAGE, and upon completion, was terminated by incubation with 200 μl of *p*-aminobenzamidine beads (Sigma) for 15 minutes. Five hundred microlitres of water-rinsed Talon (Clontech) was added to the samples to remove the cleaved His_6-tag and any uncleaved full-length protein. The supernatant contained purified protein with a residual N-terminal Gly-Ser-His sequence. The samples were cleared with centrifugation at 1000*g* for two minutes, and the supernatant concentrated using a 1K MWCO filter (Pall Filtron) in an Amicon stir-cell concentrator. The proteins were stored at 4 °C, and kept in the reduced state with addition of 5 mM fresh dithiothreitol.

$GABP\alpha^{(168–254)}$, $Ets\text{-}1^{(29–138)}$, and $Ets\text{-}2^{(85–172)}$, all lacking an N-terminal tag, were purified by FPLC using a Q-Sepharose column and 50 mM Tris (pH 8.5) with a gradient of 0 to 300 mM NaCl over 80 minutes, at a flow rate of 4 ml/minute. For $Ets\text{-}1^{(29–138)}$ and $Ets\text{-}2^{(85–172)}$, the protein eluted with 120 mM NaCl. Non-His-tagged $GABP\alpha^{(168–254)}$ eluted in two peaks (80 mM NaCl and 150 mM NaCl), with only the first retained to avoid an irreversible conformational change that was not accompanied by a measurable change in the mass of the protein by ESI-MS, nor in its pI by isoelectric focusing, and did not arise from the oxidation of its sole cysteine residue.

Erg, Ets-2, Pnt-P2, Fli-1, and Tel protein samples were dialyzed into 20 mM potassium phosphate (pH 7.0), 50 mM NaCl, then stored at 4 °C with fresh dithiothreitol to maintain a reduced state. For Ets-1 and GABPα samples, dialysis utilized 10 mM potassium phosphate (pH 6.5) and 10 mM KCl, and 20 mM sodium/potassium phosphate (pH 7.2) and 20 mM KCl (or NaCl), respectively.

The purity and identity of each sample was confirmed by using SDS-PAGE and ESI-MS. Protein concentrations were determined using a theoretical value of ε_{280} (see Gill & von Hippel[69]) calculated with the ProtParam program†.

Limited trypsin digestion

Samples were digested with type XIII bovine pancreatic trypsin (Sigma) at 25 °C in 25 mM Tris (pH 7.9), 50 mM KCl, 0.1 mM EDTA, 1 mM dithiothreitol at a ratio of 1 : 250 (w/w) of enzyme to protein. Proteolysis was stopped at defined intervals with 5 mM phenylmethanesulfonyl fluoride, coupled with heat inactivation at 95 °C for five minutes. Samples were analyzed both by SDS-PAGE and by ESI-MS.

Circular dichroism spectroscopy

CD spectra were recorded using either a Jasco-J720 spectropolarimeter with an RTE-111 accessory, or with a Jasco-810 spectropolarimeter equipped with a PFD-452S Pelletier thermal control. Thermal denaturation studies were measured by monitoring the CD signal at 222 nm as a function of increasing temperature. $Erg^{(*108–201)}$ was in 20 mM potassium phosphate (pH 7.0), 50 mM NaCl, while $GABP\alpha^{(138–254)}$ was in 20 mM potassium phosphate (pH 7.2), 20 mM NaCl.

Native gel electrophoresis

Protein samples at a concentration of 20 μM were prepared in loading buffer (125 mM Tris (pH 7.0), 10% glycerol) and separated using native gels consisting of 125 mM Tris (pH 7.0), 25% (w/v) polyacrylamide (acrylamide/bis-acrylamide, 29 : 1, w/w), 0.05% (v/v) ammonium persulfate and 0.005% (v/v) *N*,*N*,*N*′,*N*′-tetramethylethylenediamine. Separation occurred at a constant 30 mA in a Tris–glycine buffer (25 mM Tris, 19 mM glycine, pH 7.5) at 4 °C, and resulting gels were visualized by using standard Coomassie staining.

Gel-filtration

Using a Dionex HPLC system and a Waters 125 Protein-Pak column, protein samples (50 μl at a concentration of 50 μM) were separated using 20 mM potassium phosphate (pH 7.0), 50 mM NaCl, 5 mM dithiothreitol at a flow rate of 0.5 ml/minute. Additional size determination was achieved using an FPLC Sephadex S-100 column (Amersham Biosciences) with the same buffer as above. Molecular mass calibration of the columns utilized a mixture of blue dextran, albumin, ovalbumin,

† http://www.expasy.org/tools/protparam.html

chymotrypsinogen A and ribonuclease A (Amersham Biosciences).

Glutaraldehyde cross-linking

Protein samples at a concentration of 20 μM, alone or in pair-wise combination, were prepared in 20 mM potassium phosphate (pH 7.0), 50 mM NaCl, 10 mM dithiothreitol. Cross-linking utilized 0.005% (v/v) glutaraldehyde for 30 minutes at 22 °C. Reactions were quenched by using standard SDS-PAGE loading buffer and analyzed with SDS/15% PAGE and Coomassie staining.

NMR spectral assignments

All samples contained 0.5 to 2.5 mM protein in 20 mM potassium phosphate (pH 7.0), 50 mM NaCl, 2 mM dithiothreitol with ~10% or 99% ^2H_2O added for the lock. GABPα samples were prepared in a similar manner, except using a buffer of 20 mM potassium phosphate (pH 7.2), 20 mM potassium chloride. All spectra were recorded at 30 °C using a 500 MHz Varian Unity or 600 MHz Varian Inova NMR spectrometer equipped with a triple resonance gradient probe. ^1H and ^{13}C chemical shifts were referenced to an external sample of DSS (sodium 2,2-dimethyl-2-silapentace-5-sulfonate) and ^{15}N referenced *via* indirect chemical shift ratios. Spectra were processed and analyzed using Felix 2000 (Accelrys, Inc.), NMRPipe/Draw,[70] and Sparky 3 (T. D. Goddard & D. G. Kneller, University of California, San Francisco). The assignment of the resonances from main-chain and side-chain ^1H, ^{13}C, and ^{15}N nuclei, including prochiral Hβ,β′ and methyl groups, were obtained as outlined previously.[7,55] The assignment of the resonances from aromatic rings was carried out as described.[56] In the case of GABPα$^{(*168–254)}$, this was facilitated using selective ring deuteration.

NMR relaxation measurements

Amide ^{15}N relaxation parameters were acquired at 500 or 600 MHz and 30 °C using the water-selective pulse sequences described by Farrow *et al*.[71] Steady-state heteronuclear $\{^1H\}^{15}N$-NOE spectra were recorded with and without three seconds of ^1H saturation and a total recycle delay of 5.017 seconds. Relaxation rates and Lipari–Szabo model-free relaxation parameters, including global tumbling times, were calculated using software provided by Niel Farrow,[71] as well as Curvefit (A. Palmer) and Tensor2.[72]

Residual dipolar coupling measurements

Phage (Pf1) were generated in *Pseudomonas aeruginosa* by using the protocol outlined by Hansen *et al*.[73] Addition of phage was monitored *via* measured splitting of the $^1HO^2H$ signal using 1D ^2H NMR spectroscopy. The final phage concentration was ~17 mg/ml, corresponding to a deuterium splitting of ~18 Hz. $^1H^N$–^{15}N, $^1H^\alpha$–$^{13}C^\alpha$, $^1H^N$–$^{13}C'$, $^{15}N^H$–$^{13}C'$ and $^{13}C'$–$^{13}C^\alpha$ residual dipolar couplings (RDCs) of $^{13}C/^{15}N$-labelled Erg$^{(*108–201)}$ and GABPα$^{(*168–254)}$ were measured using a combination of ^1H–^{15}N IPAP-HSQC,[74] and 2D/3D HNCO-based coupling experiments.[75,76] Values for Da and *R* were initially derived from a histogram analysis of normalized RDC data and later refined *via* a grid search using a preliminary generation of structures derived from dihedral, hydrogen-bond, assigned methyl–methyl and methyl–amide NOE restraints. The lowest ARIA energies and minimal SANI violations resulted in Da values of 13 and 12 Hz, and *R* values of 0.65 and −0.4 for Erg$^{(*108–201)}$ and GABPα$^{(*168–254)}$, respectively.

NMR-based structure calculations for Erg$^{(*108–201)}$ and GABPα$^{(*168–254)}$

Structure calculations were carried out using ARIA/CNS v1.2,[57–60] with a starting extended chain and an extensive set of distance, dihedral angle and RDC restraints. Distance restraints were acquired from a combination of 3D ^{15}N-NOESY-HSQC (125 ms mixing time), 3D simultaneous ^{13}C- and ^{15}N-NOESY-HSQC (120 ms mixing time), 3D aromatic ^{13}C-NOESY-HSQC (120 ms mixing time), and a simultaneous constant time methyl–methyl and amide–methyl NOESY spectra (140 ms mixing time). The methyl–methyl and amide–methyl NOESY spectra were manually assigned, providing unambiguous 51 methyl–methyl and 54 amide–methyl proton distance restraints for Erg$^{(*108–201)}$, which were constrained to a generous range of 1–6 Å. For GABPα$^{(*168–254)}$, 64 methyl–methyl and 156 amide–methyl restraints were used, with distances calibrated using ARIA. The remaining NOE peaks were manually picked but left for assignment by ARIA. Backbone dihedral angles were based on TALOS[77] with errors at the greater of 30° or twice the TALOS prediction; data were also included for residues for which φ angles were confirmed by measured $J_{HNH\alpha}$ coupling constants,[78] and for which helical secondary structure was predicted by CSI analysis.[79] χ_1 angles were included for Thr, Ile and Val based on the pattern of $^3J_{NC\gamma}$ and $^3J_{C'C\gamma}$ coupling constants, determined from long-range C^γ–N^H and C^γ–C' spectra, according to a staggered rotamer model.[80,81] Two restraints per hydrogen bond were included for amide protons displaying protection from hydrogen–deuterium exchange, as measured by ^1H–^{15}N-HSQC spectroscopy following transfer to ^2H_2O buffer. The hydrogen-bond partners were assumed to be to the $i-4$ C' residue for α-helical regions only. All His imidazole side-chains were determined to be in the neutral ε2 tautomeric state by using long-range ^1H–^{15}N correlation measurements. All Xaa-Pro amides were constrained to the *trans* conformation based on a chemical shift analysis using the program POP.[82] Starting at iteration 4, RDC restraints were included using the SANI protocol. In the final set of ten water-refined structures, there were no violations greater than 0.5 Å and 5° for distance and angle restraints, respectively.

Data Bank accession codes

The NMR chemical shifts of Erg$^{(*108–201)}$ (accession number 5399) and GABPα$^{(*168–254)}$ (accession number 6287) have been deposited in the BioMagResBank†.[55] NOE distance restraints and structure ensembles have been deposited in the Protein Data Bank‡ with accession codes 1SXE and 1SXD for Erg$^{(*108–201)}$ and GABPα$^{(*168–254)}$, respectively.

† http://www.bmrb.wisc.edu/
‡ http://www.rcsb.org/pdb

Acknowledgements

We thank James Bowie, Logan Donaldson, Niel Farrow, Barbara Graves, Shou-ming He, Lewis Kay, Gregory Lee, Ranjith Muhandiram, and Isabelle Pot for their assistance with this project, Les Hicks and Cyril Kay for ultracentrifugation analyses, and Barbara Graves, Hannele Ruohola-Baker, and Martine Duterque-Coquillard for providing cDNA. This research was supported by grants from the National Cancer Institute of Canada with funds from the Canadian Cancer Society (L.P.M.), the Natural Sciences and Engineering Research Council of Canada (C.D.M.), the Deutsche Forschungsgemeinschaft (M.S.), the Jane Coffin Childs Memorial Fund for Medical Research (L.N.G.), and the Leukemia Research Fund of Canada (C.M.S.). Instrument support was provided by the Government of Canada's Network of Centres of Excellence Program supported by the Canadian Institutes of Health Research (CIHR) and the Natural Sciences and Engineering Research Council (NSERC) through the Protein Engineering Network of Centres of Excellence (PENCE, Inc.). L.P.M. is a CIHR Scientist.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2004.07.094.

The Supplementary data comprises four Figures illustrating the phylogenetic alignment of Ets family PNT domains, schematic representations of PNT domain constructs studied, annotated $^{1}H–^{15}N$ HSQC spectra of the Erg and GABPα PNT domains, and ^{1}H-NMR spectra of the Ets-2, Pnt-P2, Fli-1, and Tel PNT domains.

References

1. Pawson, T., Gish, G. D. & Nash, P. (2001). SH2 domains, interaction modules and cellular wiring. *Trends Cell. Biol.* **11**, 504–511.
2. Hung, A. T. & Sheng, M. (2002). PDZ domains: structural modules for protein complex assembly. *J. Biol. Chem.* **277**, 5699–5702.
3. Graves, B. J. & Petersen, J. M. (1998). Specificity within the Ets family of transcription factors. *Advan. Cancer Res.* **75**, 1–55.
4. Sharrocks, A. D. (2001). The ETS-domain transcription factor family. *Nature Rev. Mol. Cell Biol.* **2**, 827–837.
5. Oikawa, T. & Yamada, T. (2003). Molecular biology of the Ets family of transcription factors. *Gene*, **303**, 11–34.
6. Klambt, C. (1993). The *Drosophila* gene *pointed* encodes two ETS-like proteins which are involved in the development of the midline glial cells. *Development*, **117**, 163–176.
7. Slupsky, C. M., Gentile, L. N., Donaldson, L. W., Mackereth, C. D., Seidel, J. J., Graves, B. J. & McIntosh, L. P. (1998). Structure of the Ets-1 pointed domain and mitogen-activated protein kinase phosphorylation site. *Proc. Natl Acad. Sci. USA*, **95**, 12129–12134.
8. Ponting, C. P. (1995). SAM: a novel motif in yeast sterile and Drosophila polyhomeotic proteins. *Protein Sci.* **4**, 1928–1930.
9. Schultz, J., Ponting, C. P., Hofmann, K. & Bork, P. (1997). SAM as a protein interaction domain involved in developmental regulation. *Protein Sci.* **6**, 249–253.
10. Kim, C. A. & Bowie, J. U. (2003). SAM domains: uniform structure, diversity of function. *Trends Biochem. Sci.* **28**, 625–628.
11. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998). SMART, a simple modular architecture tool: identification of signalling domains. *Proc. Natl Acad. Sci. USA*, **95**, 5857–5864.
12. Murzin, A. G., Brenner, S. E., Hubbard, T. & Clothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
13. Kim, C. A., Phillips, M. A., Kim, W., Gingery, M., Tran, H. H., Robinson, M. A. *et al*. (2001). Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *EMBO J.* **20**, 4173–4182.
14. Jousset, C., Carron, C., Boureux, A., Quang, C. T., Oury, C., Dusanter-Fourt, I. *et al*. (1997). A domain of TEL conserved in a subset of ETS proteins defines a specific oligomerization interface essential to the mitogenic properties of the TEL-PDGFRβ oncoprotein. *EMBO J.* **16**, 69–82.
15. Tognon, C. E., Mackereth, C. D., Somasiri, A. M., McIntosh, L. P. & Sorensen, P. H. (2004). Mutations in the SAM domain of the ETV6-NTRK3 chimeric tyrosine kinase block polymerization and transformation activity. *Mol. Cell. Biol.* **24**, 4636–4650.
16. McLean, T. W., Ringold, S., Neuberg, D., Stegmaier, K., Tantravahi, R., Ritz, J. *et al*. (1996). TEL-AML-1 dimerizes and is associated with a favorable outcome in childhood acute lymphoblastic leukemia. *Blood*, **88**, 4252–4258.
17. Tran, H. H., Kim, C. A., Faham, S., Siddall, M.-C. & Bowie, J. U. (2002). Native interface of the SAM domain polymer of TEL. *BMC Struct. Biol.* **2**, 5.
18. Kim, C. A., Gingery, M., Pilpa, R. M. & Bowie, J. U. (2002). The SAM domain of polyhomeotic forms a helical polymer. *Nature Struct. Biol.* **9**, 453–457.
19. Thanos, C. D., Goodwill, K. E. & Bowie, J. U. (1999). Oligomeric structure of the human EphB2 receptor SAM domain. *Science*, **283**, 833–836.
20. Smalla, M., Schmieder, P., Kelly, M., ter Laak, A., Krause, G., Ball, L. *et al*. (1999). Solution structure of the receptor tyrosine kinase EphB2 SAM domain and identification of two distinct homotypic interaction sites. *Protein Sci.* **8**, 1954–1961.
21. Stapleton, D., Balan, I., Pawson, T. & Sicheri, F. (1999). The crystal structure of an Eph receptor SAM domain reveals a mechanism for modular dimerization. *Nature Struct. Biol.* **6**, 44–49.
22. Kullander, K., Mather, N. K., Diella, F., Dottori, M., Boyd, A. W. & Klein, R. (2001). Kinase-dependent and kinase-independent functions of EphA4 receptors in major axon tract formation *in vivo*. *Neuron*, **29**, 73–84.
23. Poirel, H., Lopez, R. G., Lacronique, V., Della Valle, V., Mauchauffe, M., Berger, R. *et al*. (2000). Characterization of a novel ETS gene, TELB, encoding a protein structurally and functionally related to TEL. *Oncogene*, **19**, 4802–4806.
24. Potter, M. D., Buijs, A., Kreider, B., van Rompaey, L. & Grosveld, G. C. (2000). Identification and

characterization of a new human ETS-family transcription factor, *TEL2*, that is expressed in hematopoietic tissues and can associate with *TEL1/ETV6*. *Blood*, **95**, 3341–3348.
25. Peterson, A. J., Kyba, M., Bornemann, D., Morgan, K., Brock, H. W. & Simon, J. (1997). A domain shared by the *Polycomb* group proteins Scm and ph mediates heterotypic and homotypic interactions. *Mol. Cell. Biol.* **17**, 6683–6692.
26. Kyba, M. & Brock, H. W. (1998). The SAM domain of polyhomeotic, RAE28, and Scm mediates specific interactions through conserved residues. *Dev. Genet.* **22**, 74–84.
27. Tu, H., Barr, M., Dong, D. L. & Wigler, M. (1997). Multiple regulatory domains on the Byr2 protein kinase. *Mol. Cell. Biol.* **17**, 5876–5887.
28. Ramachander, R., Kim, C. A., Phillips, M. L., Mackereth, C. D., Thanos, C. D., McIntosh, L. P. & Bowie, J. U. (2002). Oligomerization-dependent association of the SAM domains from *Schizosaccharomyces pombe* Byr2 and Ste4. *J. Biol. Chem.* **277**, 39585–39593.
29. Grimshaw, S. J., Mott, H. R., Scott, K. M., Nielsen, P. R., Evetts, K. A., Hopkins, L. J. *et al*. (2004). Structure of the sterile α motif (SAM) domain of the *Saccharomyces cerevisiae* mitogen-activated protein kinase pathway-modulating protein STE50 and analysis of its interaction with the STE11 SAM. *J. Biol. Chem.* **279**, 2192–2201.
30. Seidel, J. J. & Graves, B. J. (2002). An ERK2 docking site in the Pointed domain distinguishes a subset of ETS transcription factors. *Genes Dev.* **16**, 127–137.
31. Green, J. B., Gardner, C. D., Wharton, R. P. & Aggarwal, A. K. (2003). RNA recognition *via* the SAM domain of Smaug. *Mol. Cell*, **11**, 1537–1548.
32. Aviv, T., Lin, Z., Lau, S., Rendl, L. M., Sicheri, F. & Smibert, C. A. (2003). The RNA-binding SAM domain of Smaug defines a new family of post-transcriptional regulators. *Nature Struct. Biol.* **10**, 614–621.
33. Barrera, F. N., Poveda, J. A., González-Ros, J. M. & Neira, J. L. (2003). Binding of the C-terminal sterile α motif (SAM) domain of human p73 to lipid membranes. *J. Biol. Chem.* **278**, 46878–46885.
34. Chinenov, Y., Henzl, M. & Martin, M. E. (2000). The α and β subunits of the GA-binding protein form a stable heterodimer in solution. *J. Biol. Chem.* **275**, 7749–7756.
35. Thompson, C. C., Brown, T. A. & McKnight, S. L. (1991). Convergence of Ets- and notch-related structural motifs in a heteromeric DNA binding complex. *Science*, **253**, 762–768.
36. Batchelor, A. H., Piper, D. E., de la Brousse, F. C., McKnight, S. L. & Wolberger, C. (1998). The structure of GABPα/β: an ETS-domain-ankyrin repeat heterodimer. *Science*, **279**, 1037–1041.
37. Triezenberg, S. J., LaMarco, K. L. & McKnight, S. L. (1988). Evidence of DNA:protein interactions that mediate HSV-1 immediate early gene activation by VP16. *Genes Dev.* **2**, 730–742.
38. Avots, A., Hoffmeyer, A., Flory, E., Cimanis, A., Rapp, U. R. & Serfling, E. (1997). GABP factors bind to a distal interleukin 2 (IL-2) enhancer and contribute to c-Raf-mediated increase in IL-2 induction. *Mol. Cell. Biol.* **17**, 4381–4389.
39. Fromm, L. & Burden, S. J. (1998). Synapse-specific and neuregulin-induced transcription require an ETS site that binds GABPα/GABPβ. *Genes Dev.* **12**, 3074–3083.
40. Hoffmeyer, A., Avots, A., Flory, E., Weber, C. K., Serfling, E. & Rapp, U. R. (1998). The GABP-responsive element of the interleukin-2 enhancer is regulated by JNK/SAPK-activating pathways in T lymphocytes. *J. Biol. Chem.* **273**, 10112–10119.
41. Schaeffer, L., Duclert, N., Huchet-Dymanus, M. & Changeux, J. P. (1998). Implication of a multi-subunit Ets-related factor in synaptic expression of the nicotinic acetylcholine receptor. *EMBO J.* **17**, 3078–3090.
42. Fromm, L. & Burden, S. J. (2001). Neuregulin-1-stimulated phosphorylation of GABP in skeletal muscle cells. *Biochemistry*, **40**, 5306–5312.
43. Rao, V. N., Papas, T. S. & Reddy, E. S. P. (1987). *erg*, a human ets-related gene on chromosome 21: alternative splicing, polyadenylation, and translation. *Science*, **237**, 635–639.
44. Reddy, E. S. P., Rao, V. N. & Papas, T. S. (1987). The *erg* gene: a human gene related to the *ets* oncogene. *Proc. Natl Acad. Sci. USA*, **84**, 6131–6135.
45. Reddy, E. S. P. & Rao, V. N. (1991). *erg*, an *ets*-related gene, codes for sequence-specific transcriptional activators. *Oncogene*, **6**, 2285–2289.
46. Duterque-Coquillard, M., Niel, C., Plaza, S. & Stehelin, D. (1993). New human *erg* isoforms generated by alternative splicing are transcriptional activators. *Oncogene*, **8**, 1865–1873.
47. Prasad, D. D., Rao, V. N., Lee, L. & Reddy, E. S. P. (1994). Differentially spliced erg-3 product functions as a transcriptional activator. *Oncogene*, **9**, 669–673.
48. Owczarek, C. M., Portbury, K. J., Hardy, M. P., O'Leary, D. A., Kudoh, J., Shibuya, K. *et al*. (2004). Detailed mapping of the *ERG-ETS2* interval of human chromosome 21 and comparison with the region of conserved synteny on mouse chromosome 16. *Gene*, **324**, 65–77.
49. Baltzinger, M., Mager-Heckel, A. M. & Remy, P. (1999). Xl erg: expression pattern and overexpression during development plead for a role in endothelial cell differentiation. *Dev. Dyn.* **216**, 420–433.
50. Vlaeminck-Guillem, V., Carrere, S., Dewitte, F., Stehelin, D., Desbiens, N. & Duterque-Coquillard, M. (2000). The Ets family member Erg gene is expressed in mesodermal tissues and neural crests at fundamental steps during mouse embryogenesis. *Mech. Dev.* **91**, 331–335.
51. Hewett, P. W., Nishi, K., Daft, E. L. & Murray, J. C. (2001). Selective expression of erg isoforms in human endothelial cells. *Int. J. Biochem. Cell. Biol.* **33**, 347–355.
52. McLaughlin, F., Ludbrook, V. J., Cox, J., von Carlowitz, I., Brown, S. & Randi, A. M. (2001). Combined genomic and antisense analysis reveals that the transcription factor Erg is implicated in endothelial cell differentiation. *Blood*, **98**, 3332–3339.
53. Iwamoto, M., Higushi, Y., Koyama, E., Enomoto-Iwamoto, M., Kurisu, K., Yeh, H. *et al*. (2000). Transcription factor ERG variants and functional diversification of chondrocytes during limb long bone development. *J. Cell Biol.* **150**, 27–39.
54. Dayie, K. T., Wagner, G. & Lefèvre, J.-F. (1996). Theory and practice of nuclear spin relaxation in proteins. *Annu. Rev. Phys. Chem.* **47**, 243–282.
55. Mackereth, C. D., Schärpf, M., Gentile, L. N. & McIntosh, L. P. (2002). Chemical shift and secondary structure conservation of the PNT/SAM domains from the Ets family of transcription factors. *J. Biomol. NMR*, **24**, 71–72.
56. Slupsky, C. M., Gentile, L. N. & McIntosh, L. P. (1998). Assigning the NMR spectra of aromatic amino acids in proteins: analysis of two Ets pointed domains. *Biochem. Cell Biol.* **76**, 379–390.

57. Nilges, M. (1995). Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245**, 645–660.
58. Nilges, M. & O'Donoghue, S. (1998). Ambiguous NOEs and automated NOE assignment. *Prog. NMR Spectrosc.* **32**, 107–139.
59. Linge, J. P. & Nilges, M. (1999). Influence of non-bonded parameters on the quality of NMR structures: a new force-field for NMR calculation. *J. Biomol. NMR*, **13**, 51–59.
60. Linge, J. P., O'Donoghue, S. & Nilges, M. (2001). Assigning ambiguous NOEs with ARIA. *Methods Enzymol.* **339**, 71–90.
61. Basuyaux, J. P., Ferreira, E., Stehelin, D. & Buttice, G. (1997). The Ets transcription factors interact with each other and with the c-Fos/c-Jun complex *via* distinct protein domains in a DNA-dependent and -independent manner. *J. Biol. Chem.* **272**, 26188–26195.
62. Carrere, S., Verger, A., Flourens, A., Stehelin, D. & Duterque-Coquillard, M. (1998). Erg proteins, transcription factors of the Ets family, form homo, heterodimers and ternary complexes *via* two distinct domains. *Oncogene*, **16**, 3261–3268.
63. Kwiatkowski, B. A., Bastian, L. S., Bauer, T. R., Jr, Tsai, S., Zielinska-Kwiatkowska, A. G. & Hickstein, D. D. (1998). The *ets* family member Tel binds to Fli-1 oncoprotein and inhibits its transcriptional activity. *J. Biol. Chem.* **273**, 17525–17530.
64. Wang, W. K., Bycroft, M., Foster, N. W., Buckle, A. M., Fersht, A. R. & Chen, Y. W. (2001). Structure of the C-terminal sterile α-motif (SAM) domain of human p73α. *Acta Crystallog. sect. D*, **57**, 545–551.
65. Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R. F., Sykes, B. D. & Wishart, D. S. (2003). VADAR: a web server for quantitative evaluation of protein structure quality. *Nucl. Acids Res.* **31**, 3316–3319.
66. Hutchinson, E. G. & Thornton, J. M. (1996). PROMOTIF–a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220.
67. Matsuda, S., Harries, J. C., Viskaduraki, M., Troke, P. J. F., Kindle, K. B., Ryan, C. & Heery, D. M. (2004). A conserved α-helical motif mediates the binding of diverse nuclear proteins to the SRC1 interaction domain of CBP. *J. Biol. Chem.* **279**, 14055–14064.
68. Baker, D. A., Mille-Baker, B., Wainwright, S. M., Ish-Horowicz, D. & Dibb, N. J. (2001). Mae mediates MAP kinase phosphorylation of Ets transcription factors in Drosophila. *Nature*, **411**, 330–334.
69. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326.
70. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeiffer, J. & Bax, A. (1995). NMRPipe: a multi-dimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
71. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G. *et al.* (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry*, **19**, 5984–6003.
72. Dosset, P., Hus, J.-C., Blackledge, M. & Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR*, **16**, 23–28.
73. Hansen, M. R., Mueller, L. & Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Struct. Biol.* **5**, 1065–1074.
74. Ottiger, M., Delaglio, F. & Bax, A. (1998). Measurement of *J* and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373–378.
75. Yang, D., Tolman, J. R., Goto, N. K. & Kay, L. E. (1998). An HNCO-based pulse scheme for the measurement of the $^{13}C^{\alpha}$–$^{1}H^{\alpha}$ one-bond dipolar couplings in ^{15}N, ^{13}C labeled proteins. *J. Biomol. NMR*, **12**, 325–332.
76. Yang, D., Venters, R. A., Mueller, G. A., Choy, W. Y. & Kay, L. E. (1999). TROSY-based HNCO pulse sequences for the measurement of ^{1}HN–^{15}N, ^{15}N–^{13}CO, ^{1}HN–^{13}CO, $^{13}CO^{\alpha}$–$^{13}C^{\alpha}$ and ^{1}HN–$^{13}C^{\alpha}$ dipolar couplings in ^{15}N, ^{13}C, ^{2}H-labeled proteins. *J. Biomol. NMR*, **14**, 333–343.
77. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
78. Kuboniwa, H., Grzesiek, S., Delaglio, F. & Bax, A. (1994). Measurement of HN–H^{α} *J* couplings in calcium-free calmodulin using 2D and 3D water-flip-back methods. *J. Biomol. NMR*, **4**, 871–878.
79. Wishart, D. S. & Sykes, B. D. (1994). The ^{13}C Chemical Shift Index: a simple method for the identification of protein secondary structure using ^{13}C chemical shift data. *J. Biomol. NMR*, **4**, 171–180.
80. Grzesiek, S., Anglister, J. & Bax, A. (1993). Correlation of backbone amide and aliphatic side-chain resonances in ^{13}C/^{15}N-enriched proteins by isotropic mixing of ^{13}C magnetization. *J. Magn. Reson. ser. B*, **101**, 114–119.
81. Vuister, G. W., Wang, A. C. & Bax, A. (1993). Measurement of three-bond nitrogen carbon J-couplings in proteins uniformly enriched in ^{15}N and ^{13}C. *J. Am. Chem. Soc.* **115**, 5334–5335.
82. Schubert, M., Labudde, D., Oschkinat, H. & Schmieder, P. (2002). A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^{13}C chemical shift statistics. *J. Biomol. NMR*, **24**, 149–154.
83. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
84. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 945–949.
85. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.
86. Nicholls, A., Sharp, K. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.

Edited by A. G. Palmer III

(Received 12 April 2004; received in revised form 23 June 2004; accepted 16 July 2004)