Sequence and 3D structural relationships between mammalian Ras- and Rho-specific GTPase-activating proteins (GAPs): the cradle fold

Thierry P.G. Calmels[a], Isabelle Callebaut[b], Isabelle Léger[a], Patrick Durand[b], Antoine Bril[a], Jean-Paul Mornon[b,*], Michel Souchet[a,*]

[a]*Smithkline-Beecham Laboratoires Pharmaceutiques, 4 rue Chesnay-Beauregard, 35760 Saint-Grégoire, France*
[b]*Systèmes moléculaires et Biologie structurale, LMCP, CNRS UMR C7590, UP6, Case 115, 4 Place Jussieu, 75252 Paris Cedex 05, France*

Received 9 March 1998

Abstract An extensive study of both sequence and recent 3D structural data concerning GTPase interacting domains of Ras- and Rho-specific GTPase-activating proteins (GAPs) shows that these two subfamilies share a same 3D scaffold and are thus related to each other. This relationship has heretofore remained undetected although these domains of similar size are both totally α-helical and activate nearly structurally identical targets (Ras and Rho proteins). In this report, sequence similarities correlated to 3D structures of p120rasGAP and p50rhoGAP were detected using the sensitive two-dimensional method hydrophobic cluster analysis (HCA). These patterns were further extended to other members in each subfamily and the geometry orientation of crucial arginines R789 in p120 and R282 in p50 and of important stabilizing residues like p120R903 and p50N391 was confirmed. This overall structural relationship is centered on an invariant motif of three consecutive helices that we suggest to name the 'cradle fold'. This observation opens new perspectives to understand how small GTPases are specifically regulated.

Key words: Ras; Rho; GAP; HCA; 3D structure

1. Introduction

Ras proteins and their relatives play critical roles in the control of normal and transformed cell growth. Based on sequence similarities, Ras-related proteins can be grouped into five subfamilies: Ras, Rho, Rab, ARF and Ran [1]. Ras is a ubiquitously expressed GTP-binding protein that is a key regulator of eukaryotic cell growth and differentiation [2]. Recent breakthroughs demonstrating a role for Rho-related proteins in controlling morphogenesis and cytoskeleton organization have increased the interest in these proteins [3,4]. Ras superfamily proteins function as regulated molecular switches that alternate between active GTP-bound and inactive GDP-bound states [5,6]. Thus far, all known GTPases are activated by nucleotide exchange and inactivated through an intrinsic GTPase activity. The conversion of the inactive, GDP-bound form to the active GTP-bound form is dependent on guanine nucleotide exchange factors (GEFs) [7]. Since GTP is present in excess over GDP in the cell, once the small GTPase is nucleotide-free, GTP will bind preferentially. Conversion back to the inactive state is achieved by hydrolysis of GTP, an event mediated by GTPase-activating proteins (GAPs).

Understanding how these regulatory proteins control the function of their respective substrates, and how each GAP is itself regulated, should provide important clues about how the different Ras-related proteins mediate their respective roles.

In contrast to the rather simple organization of the small GTPase of the Ras superfamily, GAPs are larger proteins which, in addition to their GAP activity region, contain several domains such as SH2, SH3, proline-rich regions, Ser/Thr kinase domains and pleckstrin homology domains [8]. These domains are known to regulate protein-protein interactions and mediate subcellular localization, which is presumably related to regulatory processing of GTPases.

Recent structure determinations of the GTPase-activating domains of p120 (Ras-specific GAP) [9,10] and p50 (Rho-specific GAP) [11–13], isolated or in interaction with H-Ras and Cdc42Hs or RhoA, respectively, have shown that both GAPs are predominantly composed of helical secondary structures. Indeed, p120 contains eight α-helices in the proposed GTPase interacting domain [9] whereas p50 possesses nine α-helices including a classical four helix bundle [11]. Despite these recent and rapid advances in the structural characterization of Ras- and Rho-specific GAPs, no detectable sequence or three-dimensional similarities have been reported so far between members of the two GAP families (e.g. [14]).

In this report, we have used hydrophobic cluster analysis (HCA) [15], a sensitive two-dimensional method of sequence analysis and comparison able to detect structural similarities between protein sequences sharing low levels of sequence identity (typically ∼15%, for a review see [16]). Its sensitivity, particularly below the so-called 'twilight zone' (below 25–30% sequence identity), stems from its ability to detect and visualize secondary structure elements [17]. By representing protein sequence alignments in a 2D structuration, HCA makes it possible to overcome the limitations of lexicographic 1D analysis. Using this methodology, GTPase interacting domains of Ras- and Rho-specific GAPs have been analyzed and consensus 2D signatures identified.

Furthermore, these structural similarities are consistent with the presence of important residues involved in the catalytic activity and are fully supported by the comparison of the recently available 3D structures of p120 and p50 GAP interacting domains. They led to the identification of a common scaffold for these two subfamilies, built around three consecutive invariant α-helices, for which we suggest the name 'cradle fold'.

*Corresponding authors. J.-P. Mornon: Fax: +33 144 273 785, E-mail: mornon@lmcp.jussieu.fr; M. Souchet: Fax: +33 299 280 444, E-mail: michel_souchet@sbphrd.com

PII S0014-5793(98)00331-7

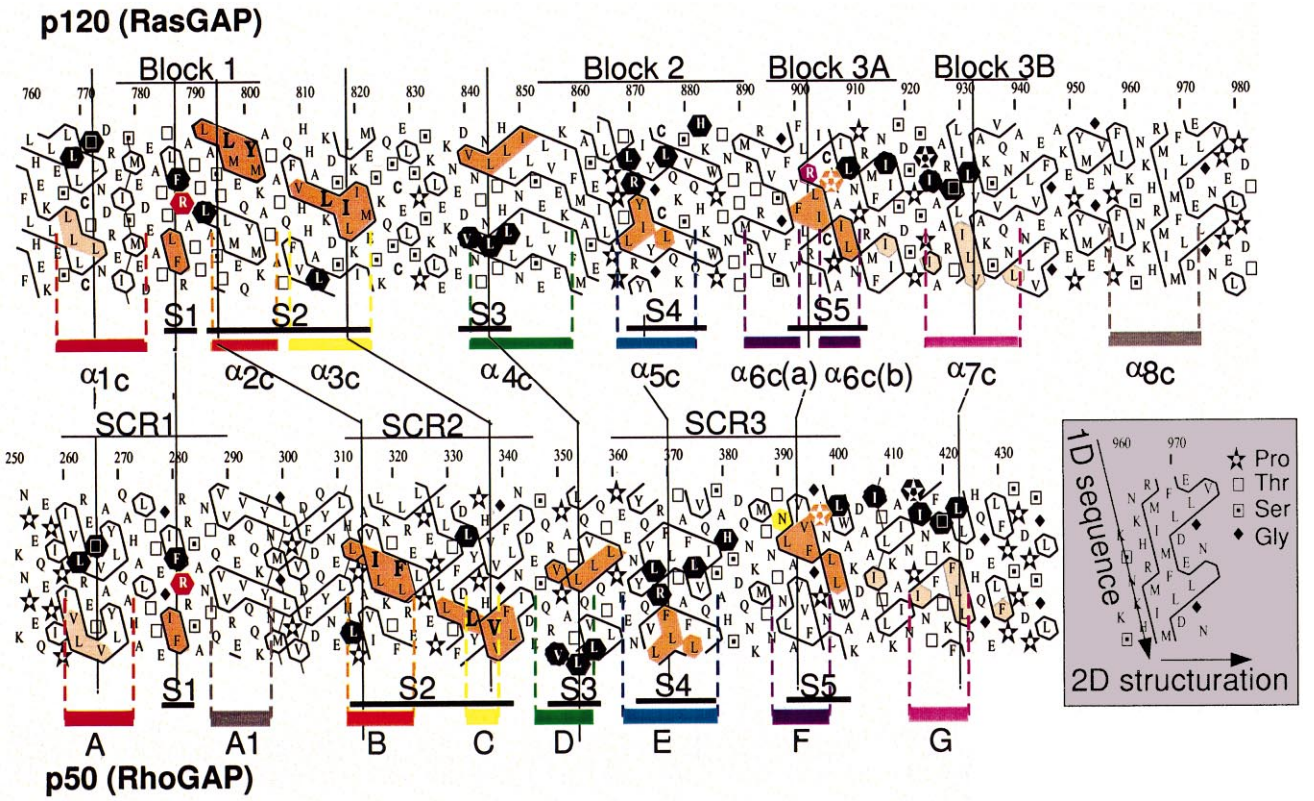


Fig. 1. HCA sequence comparison of p120 and p50 GAP domains. Experimentally observed helices α1c to α8c (p120) and A to G (p50) are indicated below the 2D sequence plots. The more conserved segments previously delineated for the RasGAP (Block1 to Block3 [9]) and RhoGAP (SCR1 to SCR3 [6]) families are indicated above the HCA plot. Vertical lines indicate proposed correspondence between the two sequences. Identities are shown with white letters on a black or colored background. Hydrophobic similar positions are shown in light orange and those belonging to the strikingly conserved segments S1 to S5 are highlighted in dark orange. As an example, within the S2 segment, the amino acids shown in detail in Fig. 4A are in bold, illustrating there the perfect correspondence between HCA 2D sequence information and 3D structures.

2. Material and methods

Guidelines for the use of HCA have been published previously [15,16,18]. Protein sequences are displayed on a duplicated helical net using one-letter code for amino acids except for proline (star), glycine (diamond), threonine (square) and serine (dotted square); hydrophobic residues are automatically contoured. The pitch of the α-helical net has been shown to offer the best correspondence between the positions of hydrophobic clusters and regular secondary structures for all classes of proteins (α, β, α/β) [17].

The statistical significances of alignments are assessed through the calculation of Z-scores [16]. Values between 3.0 and 6.0 can be considered to represent strong support for relationships at low levels of sequence identity. Three Z-score indices are currently computed, relative to the sequence identity, the sequence similarity (using the Blosum62 matrix) and the HCA score (highlighting the conservation of the hydrophobic character in similar positions in the HCA-deduced alignment). The product of these three Z-scores (Z3) is also used in comparison to the best observed Z3 random score.

Three-dimensional visualization as well as manipulations of coor-

→

Fig. 2. HCA-deduced 1D alignments of representative RasGAP (A) and RhoGAP (B) interacting domains. α-Helices experimentally observed for p120 and p50 are boxed. Identities are white on a black background and positions which are mainly hydrophobic are shaded in gray. The N- and C-termini of the domains are indicated in bold. Segments S1 to S5, defined in Fig. 1, are boxed below the multiple alignment on a consensus line (with 1 indicating positions where at least four strong hydrophobic amino acids (VILFMYW) or A, C, T are present; 3: positions mainly occupied by group III amino acids [16] (PGDNS) or A, C, T; ϕ: aromatic positions; o: small amino acids, +: basic residues, −: acidic residues).

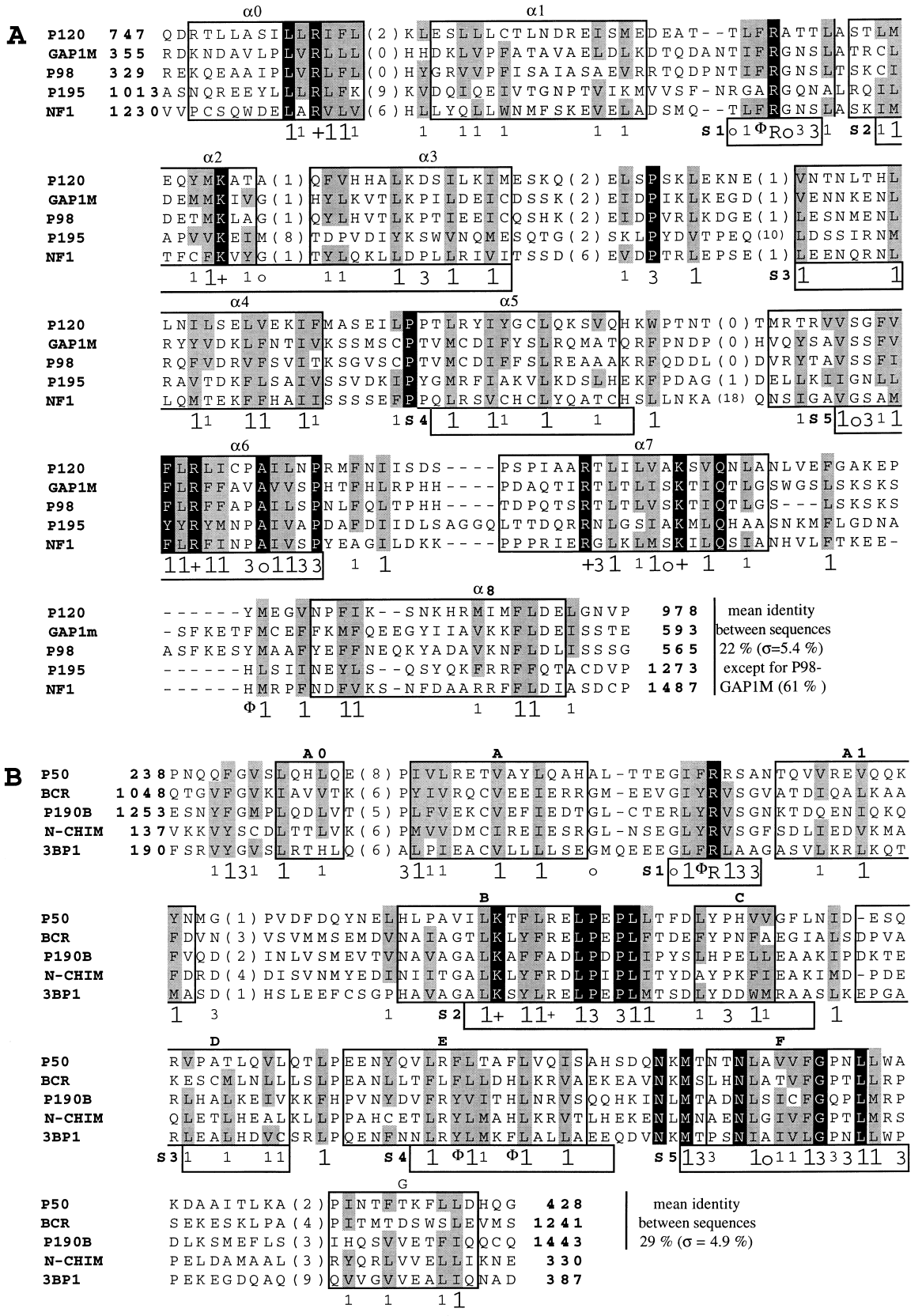
```
S1    RasGAP o1ϕRo33
      RhoGAP o1ϕR133,
S2    RasGAP 11xx11+x1oxx11xxx1x3x1xx1x
      RhoGAP x1+x11+x13x311xxx1x3x11xxx,
S3    RasGAP 1xxxxxx1
      RhoGAP 1xx1xx11,
S4    RasGAP x1(+)x11xx1xxx1xx
      RhoGAP x1+ϕ11xx1xx1xxx,
S5    RasGAP 1031111+11x301133
      RhoGAP 133x31011133311x3.
```

Protein identifiers in the SwissProt (sw) and GenBank (gb) databases are: RasGAP: sw:gtpa_human (P120), gb:d78155 (GAP1M-human), gb:u30857 (P98-bovin), sw:iqga_human (P195), sw:nf1_human (NF1); RhoGAP: gb:z23024 (P50-human), sw:bcr_human (BCR), gb:u17032 (P190B-human) sw:chin_human (N-CHIM), sw:3bp1_mouse (3BP1).

dinates were performed using INSIGHT II (Molecular Simulation Inc.). P50 numbering is given according to its sequence reference (add 197 relative to the numbering used for the structure description [11]).

3. Results and discussion

To determine whether there is a structural relationship be-



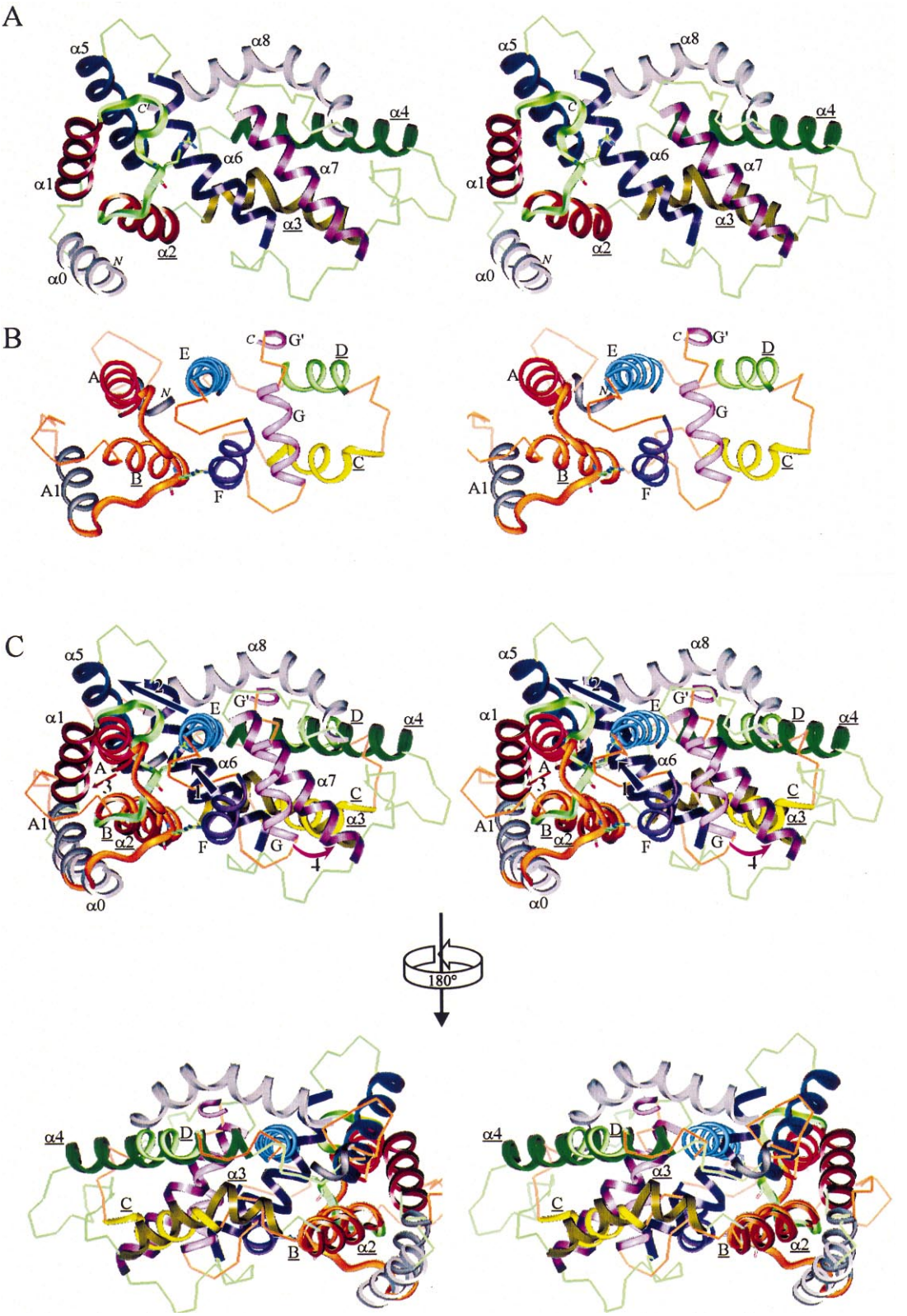


Fig. 3. Simplified stereographic representations of the 3D structures of GAP interacting domains of p120 (A: PBD identifier: 1WER) and of p50 (B: PBD identifier: 1RGP). The four helix (A-B-E-F) bundle of p50 is shown perpendicular to its axis. The two important loops α1c-α2c (p120) and A-A1 (p50) displayed as green and orange solid ribbons, respectively, present the essential arginine residues (R789 and R282, respectively) depicted in a stick rendering. The ribbons are identically colored for equivalent helices, rectangular and darker for p120 and oval and lighter for p50. Only the extra helices (p120 α0c and α8c, p50 A0 and A1) are colored in gray. C: Superimposition of the two structures as diplayed in A and B panels (top), based on the fit between B-C-D and α2c-α3c-α4c scaffolds (underlined letters). The rear view is obtained after a 180° rotation. Arrows 1 to 4 suggest the deformation of the canonical p50 A/B/E/F four helix bundle and the associated induced movement which lead to the p120 structure from the p50 one, relative to the common invariant α2c-α3c-α4c/B-C-D scaffold (see text).

tween p120 (Ras-specific GAP) and p50 (Rho-specific GAP), the sequences of their interacting domains were compared through the sensitive hydrophobic cluster analysis (HCA) method ([16] for a review). This comparison shows that several motifs are conserved between the two sequences, in particular the five motifs labelled S1 to S5 (Fig. 1). The p120 and p50 sequences considered here from aa 760 to 980 and from aa 260 to 430, respectively, share a low but significant identity level (15.3%) with a clearly good hydrophobic matching (65%) as illustrated by colored shading. The corresponding statistical Z-scores (see Section 2) are consistent with a genuine sequence and structural relationship between the two domains and are similar to many well-characterized relationships previously reported (e.g. [16]). Indeed, Z-scores for sequence identity, sequence similarity and hydrophobic scores are 4.2, 5.4 and 6.0 σ, respectively. The ratio between the product of these three scores and the best randomly observed equivalent product is 4.9 (10 000 shuffled sequences). HCA similarities detected between other members of the two GAP families support this putative relationship (data not shown).

To further assess this likely correspondence between the two GAP domains, accurate 2D HCA alignments of five well-documented sequences for both GAP families were performed. The corresponding 1D multiple alignments are shown in Fig. 2A and B. The similarity between the above described regions S1 to S5 is clearly confirmed by the multiple alignments. Moreover, these regions are associated with a fair level of chemical conservation for all the considered domains, indicating that they would be important for the structure and/or functions of these proteins.

The above sequence comparisons suggest that helices α1c to α7c of the Ras-specific GAP family may be structurally equivalent to helices A to G of the Rho-specific GAP family (Fig. 1) since hydrophobic structuration of clusters composing these helical structures is similar. Indeed and importantly, it can be seen that the connectivities between the above two sets of seven helices are sequentially identical. In such a situation, the A1 helix of p50, as well as the C-terminal α8c helix of p120, would constitute extra elements of a common fold.

The direct comparison of the 3D structures of p120 and p50 GAPs (Fig. 3) provides definite proof of this likely relationship. Indeed, the superimposition of the p120 α2c-α3c-α4c helices on the p50 B-C-D helices leads to a good overall match, including the loops between these elements (Fig. 3C). Taking these three successive structural elements as a reference, the transposition of the canonical four helix bundle (A-B-E-F) of the p50 structure on the p120 structure can be described as follows. The main driving effector of this transconformation appears to be the backward tilt of the p120 α6 helix relative to its p50 F counterpart (arrow 1 of Fig. 3C), which gives an anticlockwise push of the p120 α5c and α1c from the places occupied in the common reference by p50 E and A helices, relative to the invariant P120 α2c and p50 B helices (orange in Fig. 3C). Meanwhile, the p120 α7c helix partly follows the movement of α6c (arrow 4).

The invariant and common scaffold p120 α2c-α3c-α4c/p50 B-C-D helices (Fig. 4A) constitutes a cradle-like structure for the secondary elements in direct interaction with the small GTPase. The entire GAP interacting domain would therefore be built from this common core.

Interestingly, the extra p50 A1 helix occupies the p120 α0c N-terminal helix (Fig. 3C) and the extra p120 α8c helix partly occupies the space filled by E and G′ in p50.

Moreover, the p120 and p50 GAP interacting domains can strikingly be superimposed using only two modules, thus confirming their structural relationship (Fig. 4B). Module 1 is composed of the p50 four helix bundle and its p120 distorted counterpart, whose superimposition through the following segments results in an overall root mean square distance (rms) of 3.0 Å (33 Cα positions, p50 A 265–272, B 314–323, E 370–377, F 390–396 and p120 α1c 772–779, α2c 793–802, α5c 876–883, α6c 893–899). Module 2 is built with the remaining helices added to the helix α2c or B which is in common with module 1. Superimposition of these two modules 2 led to a rms of 3.1 Å (27 Cα positions, p50 B 318–323, C 334–340, D 349–354, G 417–423 and p120 α2c 798–803, α3c 814–820, α4c 847–852, α7c 927–933). Each entire domain results from the concatenation of these two modules rotated each other around the common junction between helices B and C or α2c and α3c within the cradle scaffold.

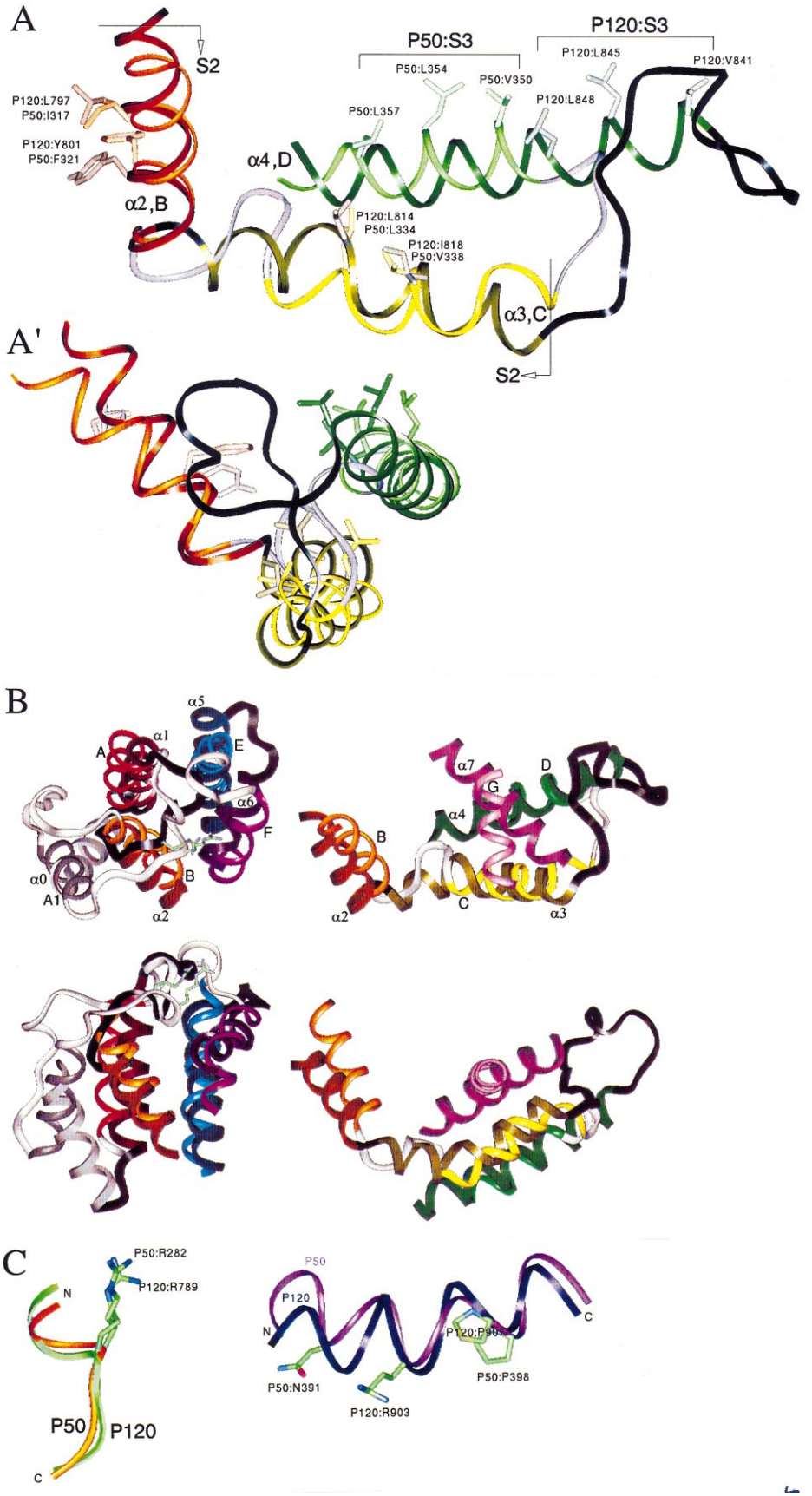
However, with respect to evolutionary considerations and to accurate correspondence between sequence and 3D data (as illustrated in Figs. 1, 2 and 4), the GAP interacting domains should be considered as the result of concerted displacements of several helices on both sides of the common α2c-α3c-α4c/B-C-D cradle scaffold described above (e.g. the angles [19] between the four helices of the two bundles are p50 A/B +35° versus p120 α1c/α2c +35°, E/F −15° vs α5c/α6c −25°, A/E +15° vs α1c/α5c −20° (Δ=−35°), B/F −20° vs α2c/α6c −55° (Δ=−35°)).

The crystal structures of p120 and p50 GAP interacting domains thus appear to be two stable states of a same fold devoted to GAP function. This plasticity of a fully α-helical domain is in accordance with a recent and important insight relative to α-helices, discussing preferential packings based on the distribution of interaxial angles between packed α-helices [19].

This structural diversity seems to be closely related to the functions and specificity of these activating domains. A rigid manual docking of the complexes of p120 and p50 with small G proteins from separate available coordinates (PDB identifiers 1WER, 1RGP, 121P) as well as from published close contacts between these partners [10,13] has been performed (data not shown). It appears that relative to the common referential p120 α2c-α3c-α4c/p50 B-C-D helices, the small GTPases follow the position of the p120 α6c helix with respect to that of the p50 F one. The angle between these two helices is close to 50° and is also approximately equal to the rotation between the two complexed GTPases around the GAP cradle fold domain.

Interestingly, Fig. 4C shows that the p120 and p50 loops presenting the essential arginines p120 R789 (also named finger loop [20,21]) and p50 R282 (R85 in the structural numbering of [11]) have a similar geometry distribution. The same is true for the overall shape of the functionally important α6c and F helices, which also possess the two important conserved amino acids R903 (p120) and N391 (p50) that may stabilize the essential arginine by a direct interaction.

In conclusion, Ras- and Rho-specific domains have probably diverged from a common template, as supported by detectable sequence similarities among the considered segments and by obvious 3D relationships. In contrast, the also fully α-helical activator domain of a regulator of the α-subunit of G


A
S2
P120:L797
P50:I317
P120:Y801
P50:F321
α2,B
P50:S3
P120:S3
P50:L354
P50:V350
P120:L845
P120:V841
P50:L357
P120:L848
α4,D
P120:L814
P50:L334
P120:I818
P50:V338
α3,C
S2
A'
B
α5
α1
A
E
α6
F
α0
A1
B
α2
α7
D
G
α4
B
α2
C
α3
C
P50:R282
N
P120:R789
P50
P120
C
P50
P120
N
C
P50:N391
P120:R903
P50:P398

protein signalling (RGS) such as RGS4, exhibits only a remote 3D overall similarity with the above described GAP interacting domains; there is no sequence similarity and a different sequential topology is observed, probably due to convergent evolutionary functional constraints. Nevertheless, these two families share conserved key residues since like p50, RGS4 supplies an asparagine (N128) into the Gα catalytic site as a stabilizing residue [22].

References

[1] Macara, I.G., Lounsbury, K.M., Richards, S.A., Mckiernan, C. and Bar-Sagi, D. (1996) FASEB J. 10, 625–630.
[2] Lowy, D.R. and Willumsen, B.M. (1993) Annu. Rev. Biochem. 62, 851–891.
[3] Ridley, A.J. (1996) Curr. Biol. 6, 1256–1264.
[4] Machesky, L.M. and Hall, A. (1997) J. Cell Biol. 138, 913–926.
[5] Bourne, H.R., Sanders, D.A. and McCormick, F. (1991) Nature 349, 117–127.
[6] Boguski, M.S. and McCormick, F. (1993) Nature 366, 643–654.
[7] Cerione, R.A. and Zheng, Y. (1996) Curr. Opin. Cell Biol. 8, 216–222.
[8] Lamarche, N. and Hall, A. (1994) Trends Genet. 10, 436–440.
[9] Scheffzek, K., Lautwein, A., Kabsch, W., Ahmadian, M.R. and Wittinghofer, A. (1996) Nature 384, 591–596.
[10] Scheffzek, K., Ahmadian, M.R., Kabsch, W., Wiesmüller, L., Lautwein, A., Schmitz, F. and Wittinghofer, A. (1997) Science 277, 333–338.
[11] Barrett, T. et al. (1997) Nature 385, 458–461.
[12] Rittinger, K., Walker, P.A., Eccleston, J.F., Smerdon, S.J. and Gamblin, S.J. (1997) Nature 389, 758–762.
[13] Rittinger, K., Walker, P.A., Eccleston, J.F., Nurmahomed, K., Owen, D., Laue, E., Gamblin, S.J. and Smerdon, S.J. (1997) Nature 388, 693–697.
[14] Geyer, M. and Wittinghofer, A. (1997) Curr. Opin. Struct. Biol. 7, 786–792.
[15] Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J.P. (1987) FEBS Lett. 224, 149–155.
[16] Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B. and Mornon, J.-P. (1997) Cell. Mol. Life Sci. 53, 621–645.
[17] Woodcock, S., Mornon, J.P. and Henrissat, B. (1992) Protein Eng. 5, 629–635.
[18] Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. and Mornon, J.P. (1990) Biochimie 72, 555–574.
[19] Bowie, J.U. (1997) Nat. Struct. Biol. 4, 915–917.
[20] Noel, J.P. (1997) Nat. Struct. Biol. 4, 677–680.
[21] Ahmadian, M.R., Stege, P., Scheffzek, K. and Wittinghofer, A. (1997) Nat. Struct. Biol. 4, 686–689.
[22] Tesmer, J.J.G., Berman, D.M., Gilman, A.G. and Sprang, S.R. (1997) Cell 89, 251–261.

←

Fig. 4. A. Atomic superimposition of the α2c-α3c-α4c/B-C-D cradle scaffold (S2+S3 segments of Figs. 1 and 2). The two corresponding sequences are: P50: 313-LPAV**I**LKT**F**LRELPEPLLTFD**L**YPH**V**VGFLNIDESQR**V**PAT**L**QV**L**QTL-360; P120: 793-LAST**L**MEQ**Y**MKATATQFVH-NA**L**KDS**I**LKIMESKQSCELSPSKLEKNED**V**NTN**L**TH**L**LNIL**S**ELVEK**I**F-860, with amino acids represented in atomic details shown in bold. These highlighted residues correspond to key markers in hydrophobic clusters, as shown in Fig. 1. The underlined 26 amino acids are best superimposed with a resulting root mean square distance (rms) on Cα of 2.6 Å. A′ view is orthogonal to A (90° clockwise rotation along a vertical axis). B: Separate superimposition of the p120 and p50 modules 1 (left) and modules 2 (right). Same conventions and colors as in Fig. 3. Two orthogonal views (up and down) are displayed. Essential arginines R789 and R282 are depicted in a stick rendering. C: Local superimposition of two functionally important structures of the Ras and Rho GAP interacting domains. The two 'finger' loops (rms on Cα 1.2 Å) presenting the essential arginines R789 (p120) and R282 (p50) (P120: 786-TLF**R**ATT-792, P50: 279-GIF**R**RSA-285) (left). The p120 α6c helix (899-FVFL**R**LIC**P**AILN-911) and the p50 helix F (390-T**N**LAVVFG**P**NLLW-402) (rms on Cα 2.1 Å) present two important residues (p120 R903 and p50 N391). The conserved position of prolines p120 P907 and p50 P398 is mentioned (right).