Identification of a region on hypoxia-inducible-factor prolyl 4-hydroxylases that determines their specificity for the oxygen degradation domains

Diego Villar, Alicia Vara-Vega, Manuel O. Landázuri, Luis del Peso

IDENTIFICATION OF A REGION ON HIF PROLYL 4-HYDROXYLASES THAT DETERMINES THEIR SPECIFICITY FOR THE OXYGEN DEGRADATION DOMAINS

Diego Villar*, Alicia Vara-Vega†, Manuel O. Landázuri†[2] and Luis del Peso*[1,2]

***Departamento de Bioquímica, Instituto de Investigaciones Biomédicas 'Alberto Sols', Consejo Superior de Investigaciones Científicas-Universidad Autónoma de Madrid. Arturo Duperier 4, 28029 Madrid, Spain.†Servicio de Inmunología, Hospital de la Princesa-UAM. Diego de León, 62. 28006 Madrid, Spain.**

[1] Address correspondence to: Luis del Peso. Instituto de Investigaciones Biomédicas "Alberto Sols" (CSIC-UAM), Arturo Duperier 4, 28029 Madrid, Spain. Phone. 34-91-585-4440. Fax: 34-91-585-4401. e-mail: lpeso@iib.uam.es. [2] Equal contribution.

Short Title: Identification of EGLN substrate selection determinants

Synopsis

Hypoxia inducible transcription factors (HIFs) are essential for the induction of an adaptive gene expression program under low oxygen tension. The activity of these transcription factors is mainly determined by the stability of HIFα subunit, which is regulated, in an oxygen-dependent manner, by a family of three prolyl-4-hydroxylases (EGLN1-3). HIFα contains two independent Oxygen Degradation Domains (NODD and CODD) that, upon hydroxylation by EGLNs, target HIFα for proteasomal degradation. *In vitro* studies indicate that each EGLN shows a differential preference for ODDs, however the sequence determinants for such specificity are unknown. In this work we show that while EGLN1 and 2 exerted their function upon any of these ODDs to regulate HIF1α protein levels and activity *in vivo*, EGLN3 only acted on CODD. With the aim of identifying the region within EGLNs responsible for their differential substrate preference, we investigated the activity and binding pattern of different EGLN deletions and chimerical constructs generated by domain swapping between EGLN1 and 3. These studies revealed a region of 97 residues that was sufficient to confer the characteristic substrate binding observed for each EGLN. Within this region, we identified the minimal sequence (EGLN1 residues 236-52) involved in substrate discrimination. Importantly, mapping of these sequences on EGLN1

tertiary structure indicates that substrate specificity is determined by a region relatively remote from the catalytic site.

Keywords: Hypoxia, EGLN, PHD, HIF, VHL, Hypoxia inducible factor

Abbreviations are: HIF, hypoxia inducible factor; EGLN, *EGL*-nine Homolog; ODD, Oxygen Dependent Degradation domain; NODD, N-terminal Oxygen Dependent Degradation domain; CODD, C-terminal Oxygen Dependent Degradation domain. AD, Gal4 activation domain; DBD, Gal4 DNA binding domain.

INTRODUCTION

Cells respond to a decreased oxygen tension with the induction of an adaptive gene expression program aimed to restore oxygen supply and to maintain cell viability during oxygen restriction. Most of the genes induced during low oxygen exposure are under control of the Hypoxia Inducible Factor (HIF) transcription factors [1]. HIFs are heterodimers of a constitutively expressed β subunit (HIFβ, also known as ARNT) and one of the three known oxygen-regulated α subunits (HIF1α, 2α or 3α). Both α and β subunits are basic helix-loop-helix transcription factors containing a Per-AhR-Sim (PAS) domain involved in protein-protein interaction. HIFα subunits are extremely short lived under normal oxygen tension due to efficient proteasomal degradation [2]. HIFα degradation is triggered by the hydroxylation of specific proline residues [3, 4]. This reaction is catalyzed by a family of proteins homologous to the product of the *C. elegans egl-9* gene [5-7] and are thus named EGL Nine homologues (EGLN) 1, 2 and 3 [8]. These enzymes are also widely known as Proline Hydroxylase Domain (PHD) 2, 1 and 3 respectively. Hydroxylated HIFα subunits are specifically recognized and targeted for degradation by an E3-ubiquitin ligase complex containing the von Hippel-Lindau (VHL) protein [9]. EGLNs require molecular oxygen for the hydroxylation reaction. Since the K_M values of these enzymes for oxygen are high, their activity changes within a physiologically relevant range of oxygen concentration, and they are hence thought to function as molecular oxygen sensors in the HIF regulation pathway [10, 11]. Thus, under limiting oxygen tensions, proline hydroxylation is compromised and HIFα escapes VHL binding and degradation. As a result, HIFα half life is extended resulting in its accumulation and interaction with HIFβ, forming a transcriptionally competent dimer [2].

Two independent prolines within HIF1α, Pro^{402} and Pro^{564}, have been found to be hydroxylated in an oxygen-dependent fashion and have been proposed to be substrates for the EGLNs [12]. Interestingly, the sequences containing both proline residues, as well as the equivalent proline residues on HIF2α (Pro^{405} and Pro^{531}) and HIF3α (Pro^{491}), align into a conserved LXXLAP motif [3, 4]. Moreover, this sequence motif is evolutionarily conserved in HIFα proteins from nematodes to mammals, suggesting an important role of these residues for EGLN and/or VHL binding. [3, 4]. Several studies suggested the absence of a rigid consensus sequence for EGLN binding [13, 14]. In spite of this apparent flexibility in their binding to target sequences, different EGLNs show a differential preference for the sequence containing HIF1α Pro^{402} (NODD) or HIF1α Pro^{564} (CODD) residues [10, 11, 14-16]. Both NODD and CODD are substrates *in vitro* for EGLN1 and EGLN2 although the K_M value for NODD is much lower [11]. In contrast, EGLN3 only hydroxylates CODD [10, 11, 14]. In addition to these *in vitro* studies, it has been demonstrated that both isolated ODDs are sufficient to drive protein degradation in mammalian cells [15]. However, the relative contribution of each site in the regulation of full length HIFα or the specificity of each EGLN for the different sites *in vivo* are issues that wait to be fully elucidated. Therefore, although EGLNs have different substrate preferences, both *in vivo* and

in vitro, the sequence determinants responsible for this substrate discrimination remain unclear. Moreover, in contrast to the effort dedicated to the identification of the sequence determinants in substrates, to the best of our knowledge no attention has been paid to the regions within EGLNs responsible for their different target preference.

In this work we investigated the regulation of each ODD within full length HIF1α by individual EGLNs *in vivo*. We found that, in agreement with *in vitro* data, EGLN1 and EGLN2 were able to regulate HIF by acting on either of the two ODDs while EGLN3 preferentially recognized CODD. Then we investigated which region of these enzymes was responsible for the observed substrate preference pattern. Through the analysis of the effect of EGLN deletions and chimerical constructs on HIF activity and stability, together with the utilization of yeast binding assays, we identified a region that conferred the specific EGLN substrate preference. Interestingly, this region is located relatively far from the catalytic site where proline hydroxylation occurs.

MATERIALS AND METHODS

Cells and reagents - For the yeast two hybrid assays described in this work we used the reporter yeast strain AH109 from BD Biosciences, (Palo Alto, CA). Yeast growth media reagents (YPD, SD amino acids and X-gal) were from BD Biosciences. Reporter assays were performed using the human cervical carcinoma cell line HeLa. HeLa cells were maintained in Dulbecco's Modified Eagle Medium (Cambrex, NY), supplemented with 100 units/ml penicillin, 100 µg/ml streptomycin and 5% (v/v) fetal bovine serum. Cells were grown at 37ºC and 5% CO_2 in a humidified incubator.
Anti-HIF1α monoclonal antibody was purchased from Transduction Laboratories (Lexington, KY), while anti-α-tubulin and anti-FLAG epitope (M2 clone) were from Sigma (Saint Louis, Missouri).

Generation of plasmid contructs - EGLNs cDNAs were generously provided by Steven L. McKnight [6] and Peter Ratcliffe [4]. Chimeric EGLNs were generated by PCR and the resulting coding sequences were cloned into the pCDNA-Flag plasmid for their expression in mammalian cells or into the pBRIDGE plasmid for the yeast two hybrid assays. The identity of all constructs was confirmed by sequencing.
Constructs expressing wild type and mutant forms of HIF1α (P402A, P564A and P402A/P564A) were described elsewhere [16].

Yeast transformation and interaction assays - Freshly prepared competent yeast cells were transformed with 0.1µg of each of the indicated plasmids by a modified version of the lithium acetate method [17]. The semi-quantitative interaction experiments were described elsewhere [16]. Briefly, transformed yeast were plated on minimal SD media plates in the absence of Leu and Trp. Subsequently, equal number of colonies from each transformation were transferred to saline solution (0.9% NaCl) and subjected to serial dilutions. Aliquots of each cell suspension (typically 20µl) were plated on different stringency culture media: plates lacking Leu/Trp (minimal stringency, no interaction between fusion proteins required for yeast growth), Leu/Trp/His (medium stringency, interaction required to support yeast growth) or Leu/Trp/His/Adenine (maximal stringency, interaction required to support yeast growth).

Cell transfection and reporter assays - Cells were seeded on six-well (10^5 cells/well) or twelve-well plates ($4 \cdot 10^4$ cells/well) 16 hours prior to transfection. A 9µg DNA mixture containing 1µg of p9xHRE-Luc reporter plasmid [18], 2.5µg of a construct directing the expression of HIF1α wt, HIF1α P402A, HIF1α P564A or HIF1α P402A,P564A; and 50 ng of a plasmid encoding for *Renilla* firefly luciferase under the control of a SV40 promoter was used for transfection with the calcium phosphate method [19]. The p9xHRE-Luc plasmid contains the firefly luciferase coding sequence under the control of nine consecutive copies of the VEGF Hypoxia Response Element [18].Where indicated, an

additional 200 ng of a pCDNA-Flag-EGLN construct was added to the mixture. Six hours after transfection, cells where washed and further incubated for 14 hours.

For reporter assays, cells in twelve-well plates were lysed and the firefly and *Renilla* luciferase activities of the lysate were determined using a dual luciferase system (Promega, Madison, WI). The firefly luciferase activity was normalized to that of *Renilla* luciferase. The results are presented as the average reporter activity obtained in *n* (indicated in each figure legend) independent experiments as the percentage of the activity obtained in cells expressing HIF1α protein in the absence of exogenous EGLN enzymes (none). The luciferase activity observed in cells transfected with the reporter construct alone is shown (Bkgr). The error bars in graphs showing reporter activity represent the standard error.

Western blot - Cells were washed with ice-cold tris-buffered saline (TBS) and harvested in 80 µl of NP40 lysis buffer (1% NP-40, 140 mM ClNa, 10 mM EDTA, 10% glycerol, 20mM Tris pH 8, 1mM PMSF, 1mM orthovanadate). Samples (30 µg total protein) were resolved in 10% SDS-polyacrilamide gels. Proteins were then transferred to nitrocellulose membranes (Biorad, CA), blocked with 5% non-fat dry milk in TBS-T (50mM Tris pH 7.6, 150mM ClNa, 0.1% Tween-20) and incubated overnight at 4ºC with the indicated antibodies. Immunolabeling was detected by enhanced chemiluminiscence (ECL, Amersham Pharmacia Biotech. Piscataway, NJ) and visualized with a digital luminescent image analyzer (FUJIFILM LAS-1000 CH).

Model of EGLN3 structure and generation of structure figures– A model of EGLN3 structure was generated by homology modeling using EGLN1 structure as template. The model was generated at the Technical University of Denmark CPHmodels 2.0 sever (http://www.cbs.dtu.dk/services/CPHmodels/). A similar model of EGLN3 structure was obtained using the SwissModel server (http://swissmodel.expasy.org/). PyMOL sofware (DeLano Scientific LLC, http://www.pymol.org) was used to visualize EGLN1 structure (Protein Data Bank ID 2G19) and EGLN3 model and to generate all the images on figure 6, including the calculation of electrostatic potentials.

Statistical analysis of data– Experimental data were analyzed with the Prism™ GraphPad (version 4.01) software. Data from reporter assays were analyzed by the analysis of variance test (ANOVA) followed by the Bonferroni´s multiple comparison test. For the multiple comparison the activity of each chimerical construct was compared to that of wild type EGLN3 or EGLN1 (figure 2). Significant differences ($p<0.001$) are indicated in figures by asterisks. Since the transcriptional activity of HIF in the absence of hydroxylases was normalized these values were not included in the statistical analysis.

RESULTS

Individual EGLN paralogs show differential preference for each HIF Oxygen Degradation Domain

In vitro enzymatic studies have demonstrated that, unlike EGLN1 and EGLN2, EGLN3 is unable to hydroxylate the NODD even in the context of the full length HIF1α protein [10, 11]. These results are in agreement with a report showing that EGLN1 and EGLN2, but not EGLN3, are able to target isolated NODD in mammalian cells [15]. However, others have found hydroxylation of P402 upon expression of EGLN3 in mammalian cells [20]. In order to test the EGLNs substrate specificity *in vivo* in the context of the full length HIFα protein, we studied the effect of EGLNs overexpression on the transcriptional activity and stability of different HIFα mutant proteins (figure 1).

Overexpression of wild type HIFα results in the saturation of the EGLN/VHL degradation pathway leading to HIFα activity as a consequence of protein accumulation (figure 1A). Coexpression of any of the EGLN paralogs, together with wild type HIFα, overcomes the pathway blockage resulting in a large decrease on HIF activity and protein levels (figure 1A). In contrast, the levels and activity of a HIF1α protein containing the P402A and P564A mutations was resistant to the expression of EGLNs (figure 1D). Thus, and in accordance with the biochemistry of prolyl hydroxylases, the effect of EGLNs on HIF1α activity and stability is dependent on the presence of the target prolines. Next, we investigated the individual contribution of each ODD to the regulation of HIF by each EGLN. Transfection of HIFαP402A (figure 1B) or HIF1αP564A (figure 1C) single mutants resulted in HIF transcriptional activity and protein accumulation. Importantly, coexpression of EGLN1 or EGLN2 downregulated protein levels and suppressed the transcriptional activity induced by either of the mutant forms (figures 1B and 1C). In sharp contrast, EGLN3 overexpression was able to repress HIF1αP402A (figure 1B), containing an intact CODD, but had no detectable effect on HIF1αP564A protein accumulation or its activity (figure 1C).

In vitro studies [10, 11] showed that EGLN3 has a very high K_M value for HIF1α NODD that correlates with the low activity of EGLN3 toward HIF1αP564A observed in our assays. A high K_M value might reflect a reduced binding affinity of EGLN3 for NODD, but only when the dissociation constant for the enzyme-substrate complex is larger than Kcat. Thus, we next studied whether differential binding of the EGLN enzymes to the NODD could explain the observed activity. In order to avoid interference with the cellular endogenous regulatory machinery and determine the interaction strength semi-quantitatively, we employed the yeast two hybrid system to test these interactions [16]. To this end, EGLN1 and 3 coding sequences were cloned in frame with the activation domain of the yeast Gal4 transcription factor. The derived constructs were transformed into yeast together with a plasmid encoding for full length HIF1α as a fusion protein with the DNA binding domain from Gal4 transcription factor. Binding between these fusion proteins allows the growth of yeast in the absence of specific nutrients. As shown in figure 1E, the three EGLN-Gal4AD fusion proteins strongly bound

to full length wild type HIF1α, as demonstrated by the growth of yeast cells in the conditions of highest stringency (Interaction, figure 1E). As expected, no yeast growth was observed under stringent conditions when double mutant P402A P564A HIF1α protein was tested against any of the EGLNs (figure 1E). This indicated that binding was completely dependant on the presence of the target proline. Importantly, while EGLN1 and EGLN2 were able to bind both HIF1αP402A and P564A single mutants, EGLN3 only bound P402A mutant (figure 1E). This result strongly suggests that the lack of EGLN3 activity toward the NODD observed in mammalian cells (figure 1C) is due to reduced binding to the Pro^{402} region.

The N-terminal unique region of EGLN1 does not determine substrate specificity

Since EGLN1/2 and EGLN3 show different substrate preference, we reasoned that they might have different substrate binding sites. Alignment of the three EGLNs paralogs (figure 2) showed that they shared a highly conserved C-terminal region. In particular, 87% residue identity was observed between EGLN1 and 3 within this region. (figure 2, labelled as C1 and C3 respectively). This region contains the residues involved in iron coordination and 2-oxoglutatrate binding (figure 2, arrowheads) and is therefore thought to form the catalytic site [21]. This sequence is preceded by a region that shares a more modest conservation between family members (45% residue identity between EGLN1 and 3 within the regions labelled as N1 and N3 respectively). A similar distribution of conserved residues within these regions was observed when EGLN1 was aligned to EGLN2 (figure 2). Finally, EGLN1 contains a unique N-terminal region (labelled as U1 in figure 2) that is completely missing in EGLN3 and shows no detectable conservation with the equivalent region in EGLN2.

Thus, we focused on these regions of differential conservation in order to identify the region determining the substrate specificity of prolyl-hydroxylases. Since the unique N-terminal region (marked U in figure 2) is only present in the EGLN isoforms that recognized NODD (EGLN1 and EGLN2), we firstly addressed the role of this region on substrate selection. To this end, we compared the substrate specificity of wild type EGLN1 with that of a construct lacking the unique N-terminal region comprising residues 1 to 177 (figure 3A, ΔU1). Additionally, we investigated the activity of an EGLN1 PCR variant (figure 3A, ΔEGLN1) that lacks 100 residues in its N-terminal region, from Gly^{76} to Gly^{177}, but conserves the MYND-type Zn-finger (residues 21-58). As shown in figure 3, neither deletion of amino acids Gly^{76}-Gly^{177} nor the whole unique N-terminal region (residues 1-177) had effect on substrate selection by EGLN1. Both EGLN1 truncated forms were able to suppress both activity and protein levels of all HIF1α forms, except for the double mutant (figures 3B to 3E). Accordingly, these truncated EGLN1 forms bound wild type and single mutant HIF1α proteins similarly to full length EGLN1 protein (figure 3F). Similarly to EGLN1, deletion of the N-terminal unique region (residues 1-167) from EGLN2 had no effect on its activity on both ODDs (data not

shown). Thus, the results shown in figure 3 clearly indicate that the N-terminal unique region is not required to determine EGLN substrate preference.

Identification of an EGLN region that determines their differential substrate specificity

In order to determine the region responsible for the differential substrate specificity of each EGLN while retaining the domain architecture of the wild type enzymes, chimerical constructs were generated by swapping the regions identified by sequence comparison (figure 2) between EGLN1 and EGLN3 (figure 4A). Substitution of the N-terminal region of EGLN3 (residues 1 to 118, labelled as N3 in figure 2) by that of EGLN1 (residues 1 to 296) resulted in an enzyme (U1N1C3) with the ability to act upon NODD as well as CODD, in a pattern identical to that of EGLN1 (figures 4B to 4E). Moreover, transference of the EGLN1 unique N-terminal region (U1) to EGLN3 did not alter its substrate preference since the resulting enzyme (U1N3C3) behaved as wild type EGLN3 (figure 4). Hence, the EGLN1 residues 201-296 (N1 region) were sufficient to confer the ability to act upon both ODDs, while residues 1-200 (U1 region) were not required. Conversely, transference of the N-terminal region from EGLN3 (residues 1-118, labelled as N3) upstream the catalytic domain of EGLN1 produces an enzyme (N3C1) that suppressed wild type (figure 4B) and HIF1α P402A (figure 4C) mutant. However, N3C1 expression did not decrease the level nor activity of the P564A mutant protein (figure 4D), in a pattern identical to that of full length EGLN3. Furthermore, analysis of the binding pattern of these chimerical constructs (figure 4F) showed that enzymes harbouring residues 201-296 from EGLN1 (U1N1C3 and N1C3, figure 4E and data not shown) bound to both NODD and CODD, while constructs containing the equivalent region from EGLN3 (U1N3C3 and N3C1, figure 4F) failed to bind NODD with detectable strength, in perfect agreement with the observed activity toward ODDs (Figures 4B-E). Altogether, these results clearly identified a region responsible for the determination of EGLN substrate binding specificity, which comprises residues 201 to 296 in EGLN1, and its homologue in EGLN3.

We exploited the recently described crystal structure of EGLN1 (residues 188 to 403) [21] to locate the N-region (residues 201-298) on the folded structure (figure 5). The opening to the catalytic site lies in one margin of the surface of the C region (figures 5A and 5B, coloured orange), very close to the boundary with the N-region (figures 5A, coloured in blue tones). Thus, a large fraction of the N-region surface will be accessible to the substrate approaching the catalytic site (figures 5A and 5B). Structural differences might underlie the variation in substrate selection displayed by EGLN1 compared to EGLN3. Consequently, and in order to approach the comparison of the two enzymes and further delimiting the EGLN sequence responsible for substrate specificity, we generated a model of EGLN3 tertiary structure by homology modelling [22] using EGLN1 as template (figures 5C and 5D). This model suggests that the EGLN3 overall structure is very similar to that of EGLN1 (figure 5A). However, due to their residue dissimilarities (figure 5E), the surface electrostatic potential of these molecules is markedly different at specific patches (compare figures 5C and 5D). The most prominent

difference maps on the finger-like projection within the N-region (figures 5C and 5D). While the surface of this structure in EGLN1 molecule presents alternate acidic and basic regions (figure 5D), that of EGLN3 was predicted to be uniformly charged (figure 5C). To test the role of this finger-like structure on substrate selection we generated two more chimerical constructs (figure 6A), in which the N-region was composed of sequences from EGLN1 (N1) and EGLN3 (N3) in different proportions. Importantly, N1/3A and N1/3B differ only in 16 residues (residues 236-252), corresponding to the sequence of the finger like-structure.

Analysis of the activity of these constructs (figures 6B-E), showed that both N1/3A and N1/3B were able to inhibit wild type and P402A mutant. However, only N1/3A had a significant effect over P564A mutant ($p<0.01$, N1/3A compared to U1N3C3), while the effect of N1/3B was undistinguishable from U1N3C3 ($p>0.05$, N1/3B compared to U1N3C3). These effects were even clearer at the level of binding to substrate. As shown in figure 6F, while N1/3B only bound CODD within HIF1α, the construct N1/3A recognizes CODD as well NODD, in a binding pattern identical to that of U1N1C3 and EGLN1. Thus, these results indicate that the first half of the N1 region (EGLN1 residues 201-252) is sufficient to confer the ability to bind proline 402 and that the residues 236-252 (shown in dark blue in figures 5A and 5B) are necessary. Additionally, our results strongly suggest that the finger-like structure plays an important role in substrate binding. In contrast, the residues 253-298 (shown in blue in figures 5A and 5B) did not seem to play a role in biding NODD, regardless of their localization surrounding the entrance to the catalytic site.

DISCUSSION

Several reports indicate that EGLNs have different specificity for isolated NODD and CODD sequences [10, 15, 16, 20]. In the present work we investigated the different activity of individual EGLNs toward each ODD within the context of the full length HIF1α protein. In agreement with previous data, in which isolated ODDs were used [15], we found that EGLN1 and EGLN2 work on both NODD and CODD, while EGLN3 only has significant activity upon CODD. We also demonstrated that the effect of individual EGLNs on each ODD correlates with their ability to bind them, strongly suggesting that the differential activity observed is a result of differences in binding affinity rather than differences on catalytic activity toward each target sequence. Our results and those of others [15] demonstrate that any of the sites is sufficient to regulate HIFα levels and activity. However, we found that expression of P402A mutant resulted in lower HIFα levels and activity than expression of the P564A mutant (data not shown). This probably reflects the combined effects of the lower activity of EGLN1 and EGLN2 on NODD as compared to CODD [10, 11] and the lack of effect of endogenous EGLN3 upon NODD.

Interestingly, we observed that the ability of EGLNs to downregulate HIF activity was consistently stronger than their effect on HIF1α protein levels. This was particularly apparent in the case of EGLN2 (figures 1A to 1C). Several recent reports suggest [23-25] that EGLNs might have an effect on HIF activity independent of their ability to regulate HIFα stability, suggesting an explanation for our observation.

Analysis of the substrate binding and activity of the different EGLN constructs toward HIFα mutants led us to identify a region (corresponding to residues 201 to 296 in EGLN1 sequence) responsible for the differential target preference of EGLNs. This region showed only a moderate conservation among family members, in contrast to the high conservation of the C-terminal region containing the catalytic residues of EGLN enzymes. The variation observed in this region might be required to accommodate the different substrate preferences of these enzymes. Consistently, alignment of the human EGLN paralogs revealed that, within the region responsible for substrate discrimination (EGLN1 residues 201-236), EGLN1 has higher sequence similarity with EGLN2 (56.3% identity) than with EGLN3 (44.0% identity) (figure 2). This is not a mere consequence of closer evolutionary distance between EGLN1 and EGLN2, since within the catalytic region (corresponding to residues 299-426 in EGLN1 protein), the sequence identity between these two proteins is slightly lower (81.9%) than between EGLN1 and EGLN3 (86.0%) (figure 2). The lack of detectable conservation between EGLN1 and EGLN2 within their first 200 residues together with our observation that both enzymes shared a common substrate specificity (figure 1) argues against a role for this region in substrate selection, a conclusion that is also supported by our results (figure 3). One possibility is that this region regulates EGLN1 in response to intracellular cues. In agreement to this, a recent report suggested that the unique N-terminal region of EGLN1 prevented its catalytic activity [26], opening the possibility that MYND

domain ligands could modulate EGLN1. However, we failed to detect an increased activity of EGLN1 upon deletion of this region (figure 3). Further studies will be required to establish the function of the EGLN1/2 unique amino-terminal regions. Our results (figure 3) also showed that a truncated version of EGLN1, lacking residues 76-177 from its N-terminal unique region, behaved as full length EGLN1. We believe that this conclusion is important since this construct has been extensively used to study EGLN biochemistry [6, 10, 11, 14, 16].

Another conclusion from our experiments is that the region sufficient for P402 binding (residues 201-252, shown in cyan and dark blue in figures 5A and 5B) is relatively far from the catalytic site. The involvement of residues 201-252 in HIFα binding is supported by the presence of exposed hydrophobic patches in this region that were predicted to enable interaction of EGLN with its substrate [21]. Since substrate specificity for many enzymes is generally assumed to derive from features in close proximity or inside the catalytic site, we find this observation significant. Analysis of the sequence determinants on HIF1α ODDs revealed that residues remote to the target proline are required for optimal EGLN activity [11]. Thus, while the target proline localizes at the catalytic site, some of these OOD residues are distant enough to allow them interact with EGLN regions relative remote from the catalytic site.

Further analysis of the EGLN1 residues 201-252 led us to the identification of a minimal sequence (residues 236-252) necessary for NODD binding. This region contains the highest density of dissimilar residues between EGLN1 and 3 (Figure 5E) and was previously suggested to have a role on substrate discrimination [21]. Since several of the differential residues are polar, the surface electrostatic potential of the finger-like region is very different for EGLN3 (mostly positive) and EGLN1 (containing positive and negative surface potential patches). Interestingly, sequences alignment reveals that N-terminal to the Pro402 there is an evolutionarily conserved cluster of basic residues (K389, K391 and K392) that is missing from the equivalent sequence upstream Pro564. It is tempting to suggest that the EGLN1 ability to bind Pro402 is mediated through the interaction of the negatively charged surface of its finger-like structure with this basic region. Conversely, since this negative surface potential seems to be absent from the equivalent region of EGLN3, as suggested by homology modelling, the interaction with Pro402 would be hindered in EGLN3. Hence, the combination of structural data and our results would support a model in which the presence of a mixed surface potential would allow interaction with both NODD and CODD in EGLN1, while the positive surface potential in the corresponding region of EGLN3 would favour CODD binding.

Given their role on HIF regulation the EGLNs are considered as potential drug targets. In fact, several inhibitors based on analogues of the reaction cosubstrate 2-oxoglutarate have been described [27, 28] and are widely used as research tools. However, the development of inhibitors based on analogues of 2-oxoglutarate that are specific for the individual EGLNs may be difficult to obtain given their high degree of similarity within the catalytic site [21]. In this regard, our results open the possibility of

devising specific inhibitors for the prolyl hydroxylases paralogs based on their diferential subtrate binding site identified herein.

A remaining open question is why EGLNs evolved different subtrate binding sites. It is tempting to speculate that different EGLNs, particularly EGLN1/2 versus EGLN3, might in have a distinct substrate repertoire. In this regard, it is noteworthy that HIF1α, which is thought to be the evolutionary ancestor of HIFs gene family, has evolved two hydroxylation sites, while human HIF3α retains only CODD. On the other hand, EGLN1, which displays dual substrate specificity, seems to be the evolutionary origin of prolyl hydroxylases, while EGLN3 has evolved to recognize only CODD. It is therefore conceivable that this apparent selective pressure on certain EGLN-HIF isoforms could reflect specialized functions of EGLN paralogs in HIF regulation. Although this topic falls out of the scope of the preset study, an increased knowledge on the specifics of HIFs regulation by the three EGLN paralogs, especially *in vivo*, should help in elucidating these appealing issues.

ACKNOWLEDGEMENTS

We thank María Calzada for critical reading of this manuscript and her valuable suggestions. This work was supported by grants from Ministerio de Ciencia y Tecnología (SAF2002-02344 and SAF2005-00180 to L. P. and SAF2004-0824 to M. O. L.), from Red Cardiovascular (RECAVA) and Comunidad Autónoma de Madrid (S-SAL-0311_2006).

REFERENCES

1 Wenger, R. H., Stiehl, D. P. and Camenisch, G. (2005) Integration of oxygen signaling at the consensus HRE. Sci. STKE **306**, re12

2 Schofield, C. J. and Ratcliffe, P. J. (2004) Oxygen Sensing by HIF hydroxylases. Nat. Rev. Mol. Cell. Biol. **5**, 343-354

3 Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J. M., Lane, W. S. and Kaelin, W. G., Jr. (2001) HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. Science **292**, 464-468

4 Jaakkola, P., Mole, D. R., Tian, Y. M., Wilson, M. I., Gielbert, J., Gaskell, S. J., Kriegsheim, A., Hebestreit, H. F., Mukherji, M., Schofield, C. J., Maxwell, P. H., Pugh, C. W. and Ratcliffe, P. J. (2001) Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. Science **292**, 468-472

5 Epstein, A. C., Gleadle, J. M., McNeill, L. A., Hewitson, K. S., O'Rourke, J., Mole, D. R., Mukherji, M., Metzen, E., Wilson, M. I., Dhanda, A., Tian, Y. M., Masson, N., Hamilton, D. L., Jaakkola, P., Barstead, R., Hodgkin, J., Maxwell, P. H., Pugh, C. W., Schofield, C. J. and Ratcliffe, P. J. (2001) C. elegans EGL-9 and Mammalian Homologs Define a Family of Dioxygenases that Regulate HIF by Prolyl Hydroxylation. Cell **107**, 43-54

6 Bruick, R. K. and McKnight, S. L. (2001) A conserved family of prolyl-4-hydroxylases that modify HIF. Science **294**, 1337-1340

7 Ivan, M., Haberberger, T., Gervasi, D. C., Michelson, K. S., Gunzler, V., Kondo, K., Yang, H., Sorokina, I., Conaway, R. C., Conaway, J. W. and Kaelin, W. G., Jr. (2002) Biochemical purification and pharmacological inhibition of a mammalian prolyl hydroxylase acting on hypoxia-inducible factor. Proc. Natl. Acad. Sci. USA **99**, 13459-13464

8 Taylor, M. S. (2001) Characterization and comparative analysis of the EGLN gene family. Gene **275**, 125-132

9 Maxwell, P. H., Wiesener, M. S., Chang, G. W., Clifford, S. C., Vaux, E. C., Cockman, M. E., Wykoff, C. C., Pugh, C. W., Maher, E. R. and Ratcliffe, P. J. (1999)

The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. Nature **399**, 271-275

10 Hirsilä, M., Koivunen, P., Günzler, V., Kivirikko, K. I. and Myllyharju, J. (2003) Characterization of the Human Prolyl 4-Hydroxylases That Modify the Hypoxia-Inducible Factor HIF. J. Biol. Chem. **278**, 30772-30780

11 Koivunen, P., Hirsila, M., Kivirikko, K. I. and Myllyharju, J. (2006) The length of peptide substrates has a marked effect on hydroxylation by the HIF prolyl 4-hydroxylases. J. Biol. Chem. **281**, 28712-28720

12 Masson, N., Willam, C., Maxwell, P. H., Pugh, C. W. and Ratcliffe, P. J. (2001) Independent function of two destruction domains in hypoxia-inducible factor-alpha chains activated by prolyl hydroxylation. EMBO J. **20**, 5197-5206

13 Huang, J., Zhao, Q., Mooney, S. M. and Lee, F. S. (2002) Sequence Determinants in Hypoxia-inducible Factor-1a for Hydroxylation by the Prolyl Hydroxylases PHD1, PHD2, and PHD3. J. Biol. Chem. **277**, 39792-39800

14 Li, D., Hirsila, M., Koivunen, P., Brenner, M. C., Xu, L., Yang, C., Kivirikko, K. I. and Myllyharju, J. (2004) Many amino acid substitutions in a HIF-1alpha -like peptide cause only minor changes in its hydroxylation by the HIF prolyl 4-hydroxylases: Substitution of 3,4-dehydroproline or azetidine-2-carboxylic acid for the proline leads to a high rate of uncoupled 2-oxoglutarate decarboxylation. J. Biol. Chem. **279**, 55051-55059

15 Appelhoff, R. J., Tian, Y.-M., Raval, R. R., Turley, H., Harris, A. L., Pugh, C. W., Ratcliffe, P. J. and Gleadle, J. M. (2004) Differential function of the prolyl hydroxylases PHD1, PHD2 and PHD3 in the regulation of the hypoxia inducible factor. J. Biol. Chem. **279**, 38458-38465

16 Landazuri, M. O., Vara-Vega, A., Viton, M., Cuevas, Y. and del Peso, L. (2006) Analysis of HIF-prolyl hydroxylases binding to substrates. Biochem. Biophys. Research Comm. **351**, 313-320

17 Gietz, D., St Jean, A., Woods, R. A. and Schiestl, R. H. (1992) Improved method for high efficiency transformation of intact yeast cells. Nucl. Acid Res. **20**, 1425

18 Aragones, J., Jones, D. R., Martin, S., San Juan, M. A., Alfranca, A., Vidal, F., Vara, A., Merida, I. and Landazuri, M. O. (2001) Evidence for the involvement of

diacylglycerol kinase in the activation of hypoxia-inducible transcription factor 1 by low oxygen tension. J. Biol. Chem. **276**, 10548-10555

19 Chen, C. and Okayama, H. (1987) High-efficiency transformation of mammalian cell by plasmid DNA. Mol. Cell. Biol. **7**, 2745-2752

20 Chan, D. A., Sutphin, P. D., Yen, S.-E. and Giaccia, A. J. (2005) Coordinate regulation of the oxygen-dependent degradation domains of hypoxia inducible factor 1a. Mol. Cell. Biol. **25**, 6415-6426

21 McDonough, M. A., Li, V., Flashman, E., Chowdhury, R., Mohr, C., Lienard, B. M. R., Zondlo, J., Oldham, N. J., Clifton, I. J., Lewis, J., McNeill, L. A., Kurzeja, R. J. M., Hewitson, K. S., Yang, E., Jordan, S., Syed, R. S. and Schofield, C. J. (2006) Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). Proc. Natl. Acad. Sci. USA **103**, 9814-9819

22 Baker, D. and Sali, A. (2001) Protein Structure Prediction and Structural Genomics. Science **294**, 93-96

23 Ozer, A., Wu, L. C. and Bruick, R. K. (2005) The candidate tumor suppressor ING4 represses activation of the hypoxia inducible factor (HIF). PNAS **102**, 7481-7486

24 To, K. K. W. and Huang, L. E. (2005) Suppression of Hypoxia-inducible Factor 1{alpha} (HIF-1{alpha}) Transcriptional Activity by the HIF Prolyl Hydroxylase EGLN1. J. Biol. Chem. **280**, 38102-38107

25 Shen, C., Shao, Z. and Powell-Coffman, J. A. (2006) The Caenorhabditis elegans rhy-1 gene inhibits HIF-1 hypoxia-inducible factor activity in a negative feedback loop that does not include vhl-1. Genetics, genetics.106.063594

26 Choi, K.-O., Lee, T., Lee, N., Kim, J.-H., Yang, E. G., Yoon, J. M., Kim, J. H., Lee, T. G. and Park, H. (2005) Inhibition of the Catalytic Activity of Hypoxia-Inducible Factor-1{alpha}-Prolyl-Hydroxylase 2 by a MYND-Type Zinc Finger. Mol. Pharmacol. **68**, 1803-1809

27 Mole, D. R., Schlemminger, I., McNeill, L. A., Hewitson, K. S., Pugh, C. W., Ratcliffe, P. J. and Schofield, C. J. (2003) 2-Oxoglutarate analogue inhibitors of hif prolyl hydroxylase. Bioorganic & Medicinal Chemistry Letters **13**, 2677-2680

28 Warnecke, C., Griethe, W., Weidemann, A., Jürgensen, J. S., Willam, C., Bachmann, S., Ivashchenko, Y., Wagner, I., Frei, U., Wiesener, M. and Eckardt, K.-U. (2003)

Activation of the hypoxia-inducible factor pathway and stimulation of angiogenesis by application of prolyl hydroxylase inhibitors. FASEB J. **17**, 1186-1188

29 Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. and Lipman, D. J. (1990) Basic Local Alignment Search Tool. J. Mol. Biol. **215**, 403-410

FIGURE LEGENDS

Figure 1. Effect of EGLN expression on HIF1α protein levels and activity. A, HeLa cells were transfected with a HRE-driven reporter construct together with a plasmid encoding for HIF1α (none) or a combination of plasmids encoding for HIF1α and the indicated EGLN enzymes. The graphs represent the normalized average reporter activity obtained in three independent experiments (n=3). The HIF1α and α-tubulin (loading control) protein levels were determined in HeLa cell cultures transfected with the indicated combination of plasmids as detailed above. B-D, cells were treated as in A, except that P402A (B), P564A (C) or P402A/P564A (D) mutant forms of HIF1α were used for the assays. The asterisks (***) indicate that mean values for EGLN1 and EGLN2 are significantly different to those obtained for EGLN3. E, Yeast cells were transformed with AD-HIF1α constructs together with DBD-EGLN1, DBD-EGLN2 or DBD-EGLN3 proteins. Serial dilutions of transformed clones were grown on plates lacking Leu and Trp (Control, no interaction required between fusion proteins for yeast growth) or lacking Leu, Trp, His and adenine (Interaction, the interaction between fusion proteins is required for yeast growth). WT, full length wild type HIF1α; P402A, HIF1α P402A mutant (intact CODD); P564A, HIF1α P564A mutant (intact NODD); PP, HIF1α P402A/P564A mutant. The results shown are representative of at least three independent experiments.

Figure 2. Sequence conservation among EGLNs paralogs. Human EGLN1 (NP_071344) was pair wise aligned with EGLN3 (NP_071356) or EGLN2 (NP_444274) protein sequences using the BLAST algorithm [29]. The graph represents the percentage of sequence identity within a 20 residue sliding window. Positions correspond to EGLN1 sequence. Diagrams above the graph represent the aligned EGLN sequences. Boxes correspond to regions of none (U1/U2), moderate (N1-N3) or high (C1-C3) sequence conservation that were swapped for the generation of chimeras. Black arrowheads indicate the location of the conserved triad His, Asp, His of residues involved in iron binding at the catalytic centre. Numbers indicate amino acid residue positions.

Figure 3. Effect of different N-terminal deletions on EGLN1 activity. A, Schematic diagram representing the constructs generated by deletion of the indicated regions of EGLN1. B, HeLa cells were transfected with a HRE-driven reporter construct together with a plasmid encoding for HIF1α (none) or a combination of plasmids encoding for HIF1α and the indicated EGLN1 forms. ΔU1, truncated EGLN1 lacking residues 1-177. ΔEGLN1, truncated EGLN1 lacking residues 76-177. The graphs represent the normalized average reporter activity obtained in three independent experiments (n=3). The HIF1α and α-tubulin (loading control) protein levels were determined in HeLa cell

cultures transfected with the indicated combination of plasmid as detailed above. C-E, cells were treated as in A, except that P402A (C), P564A (D) or P402A/P564A (E) mutant forms of HIF1α were used for the assays. F, Yeast cells were transformed with AD-HIF1α constructs together with wild type DBD-EGLN1 or the indicated deletion mutants fused to DBD. Serial dilutions of transformed clones were grown on plates lacking Leu and Trp (Control) or plates lacking Leu, Trp and His (Interaction). The results shown are representative of at least three independent experiments. Abbreviations and symbols were described in figure 1.

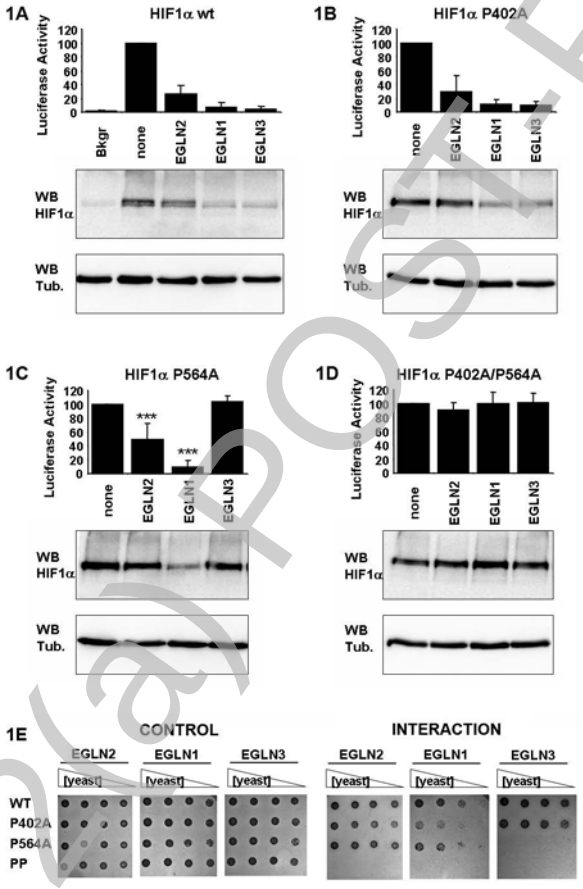
Figure 4. Effect of chimerical EGLN constructs on HIF1α protein levels and activity. A, Schematic diagram representing the chimerical constructs generated by exchange of the indicated regions between EGLN1 and EGLN3. B, HeLa cells were transfected with a HRE-driven reporter construct together with a plasmid encoding for HIF1α (none) or a combination of plasmids encoding for HIF1α and the indicated wild type or chimerical enzymes. A description of the domain structure of the chimerical constructs structure is given in figure 2. The graphs represent the normalized average reporter activity obtained in three independent experiments (n=3). The asterisks (***) indicate that mean values for those constructs are significantly different to those obtained for EGLN3. The HIF1α and α-tubulin (loading control) protein levels were determined in HeLa cell cultures transfected with the indicated combination of plasmids as detailed above. C-E, cells were treated as in A, except that P402A (C), P564A (D) or P402A/P564A (E) mutant forms of HIF1α were used for the assays. F, Yeast cells were transformed with AD-HIF1α constructs together with DBD-EGLN1 or DBD-EGLN3 proteins. Serial dilutions of transformed clones were grown on plates lacking Leu and Trp (Control) or plates lacking Leu, Trp, His and Adenine (Interaction). The results shown are representative of at least three independent experiments. Abbreviations and symbols were described in figure 1.

Figure 5. Localization and electrostatic potential of the N-region on EGLN1 structure and EGLN3 model. A,B and D, images of the molecular surface of EGLN1 (Protein Data Bank, 2G19). Colour code in A and B: Cyan, residues 188-236. Dark blue, residues 237-252. Light blue, residues 253-298. Orange: residues 297-403. The wire representation of the 2-oxoglutarate competitive inhibitor compound A is show in red and the iron atom is represented as a yellow sphere. A front views of the molecule showing the deep catalytic site. B, 180º rotation of the molecule as depicted on A. C and D, representation of the qualitative electrostatic surface potential (PyMol) for the EGLN1 structure (D) and a EGLN3 model (C). Shades of red and blue are used to represent negative and positive potentials respectively. E, Alignment of human EGLN1 and EGLN3 sequences. The numbers

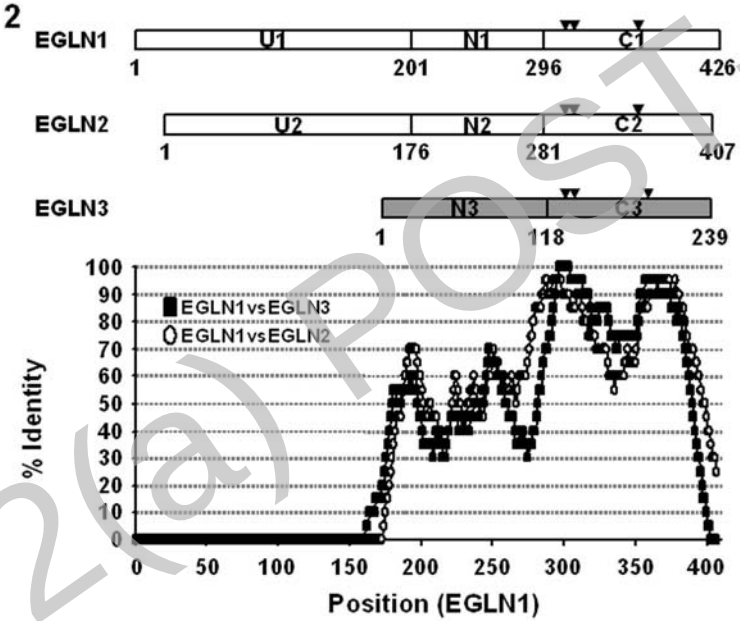
correspond to residue positions according to EGLN1 sequence. Identical residues are shown in black background and conservative substitutions in grey.

Figure 6. Effect of N-region chimerical constructs on HIF1α activity. A, Schematic diagram representing the chimerical constructs generated by exchange of the indicated regions between EGLN1 and EGLN3. Numbers indicate amino acid residue positions relative to EGLN1 sequence. B, HeLa cells were transfected with a HRE-driven reporter construct together with a plasmid encoding for HIF1α (none) or a combination of plasmids encoding for HIF1α and the indicated chimerical enzymes. The graphs represent the normalized average reporter activity obtained in four independent experiments (n=4). C-E, cells were treated as in A, except that P402A (C), P564A (D) or P402A/P564A (E) mutant forms of HIF1α were used for the assays. The asterisks (***) indicate that mean values for N1/3A and U1N1C3 are significantly different to those obtained for U1N3C3. Differences between N1/3B and U1N3C3 were not statistically significant ($p>0.05$). F, Yeast cells were transformed with AD-HIF1α constructs together with DBD-EGLN chimerical proteins. Serial dilutions of transformed clones were grown on plates lacking Leu and Trp (Control) or plates lacking Leu, Trp, His and Adenine (Interaction). The results shown are representative of two independent experiments. Abbreviations and symbols were described in figure 1.

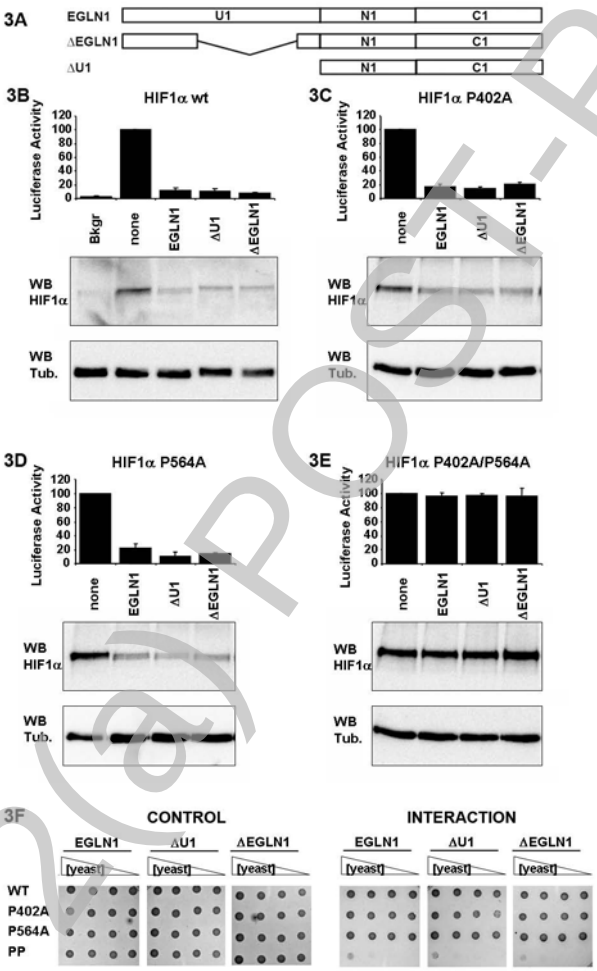
Biochemical Journal Immediate Publication. Published on 28 Aug 2007 as manuscript BJ20071052



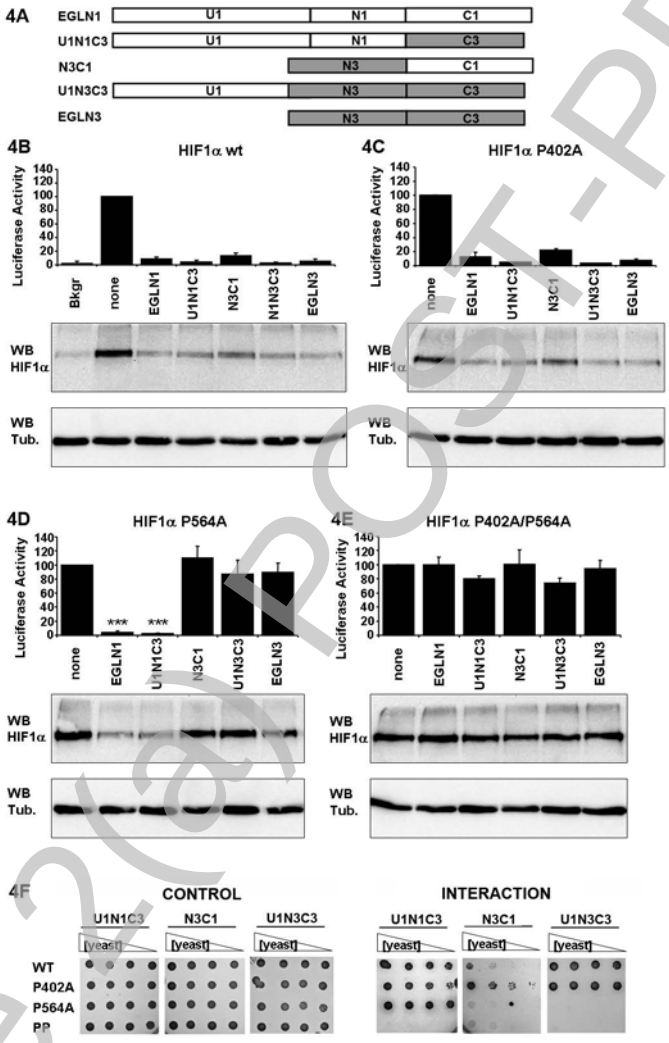
2



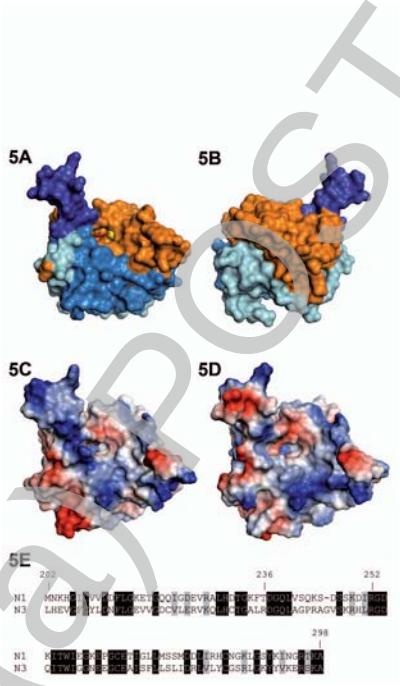
Biochemical Journal Immediate Publication. Published on 28 Aug 2007 as manuscript BJ20071052



Biochemical Journal Immediate Publication. Published on 28 Aug 2007 as manuscript BJ20071052



Biochemical Journal Immediate Publication. Published on 28 Aug 2007 as manuscript BJ20071052



Biochemical Journal Immediate Publication. Published on 28 Aug 2007 as manuscript BJ20071052