






# A cationic motif upstream Engrailed2 homeodomain controls cell internalization through selective interaction with heparan sulfates

Received: 11 November 2021

Accepted: 29 March 2023

Published online: 10 April 2023

 Check for updates

Sébastien Cardon<sup>1,5</sup>, Yadira P. Hervis <sup>1,5</sup>, Gérard Bolbach<sup>1,2</sup>, Chrystel Lopin-Bon<sup>3</sup>, Jean-Claude Jacquinet<sup>3</sup>, Françoise Illien<sup>1</sup>, Astrid Walrant <sup>1</sup>, Delphine Ravault<sup>1</sup>, Bingwei He<sup>1</sup>, Laura Molina<sup>1</sup>, Fabienne Burlina<sup>1</sup>, Olivier Lequin <sup>1</sup>, Alain Joliot<sup>4</sup>, Ludovic Carlier <sup>1</sup> ✉ & Sandrine Sagan <sup>1</sup> ✉

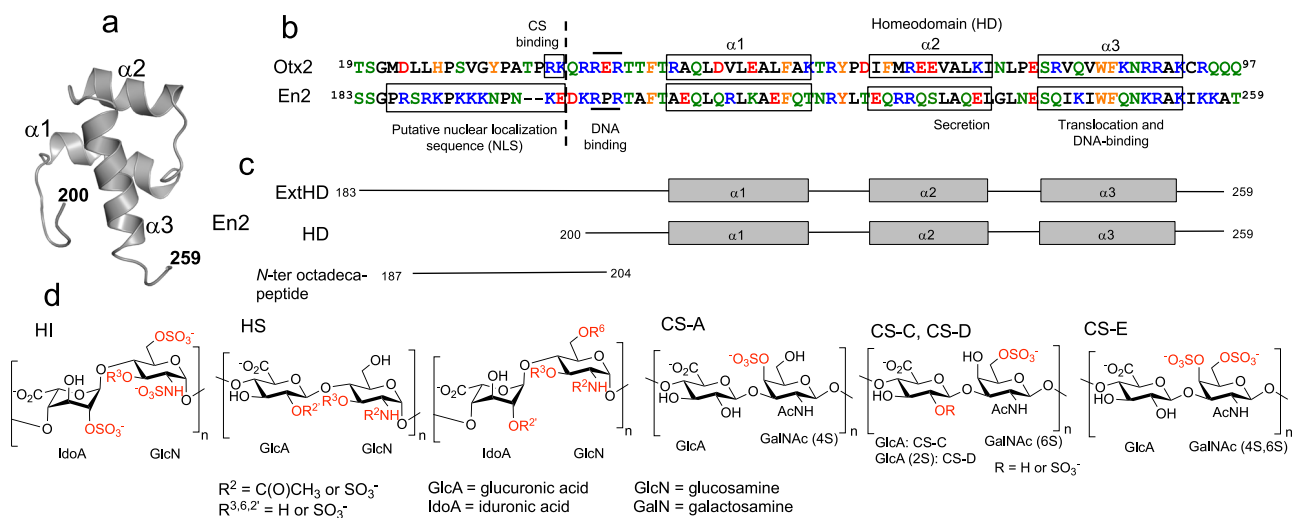
Engrailed2 (En2) is a transcription factor that transfers from cell to cell through unconventional pathways. The poorly understood internalization mechanism of this cationic protein is proposed to require an initial interaction with cell-surface glycosaminoglycans (GAGs). To decipher the role of GAGs in En2 internalization, we have quantified the entry of its homeodomain region in model cells that differ in their content in cell-surface GAGs. The binding specificity to GAGs and the influence of this interaction on the structure and dynamics of En2 was also investigated at the amino acid level. Our results show that a high-affinity GAG-binding sequence (RKPKKKNPNKEDKRPR), upstream of the homeodomain, controls En2 internalization through selective interactions with highly-sulfated heparan sulfate GAGs. Our data underline the functional importance of the intrinsically disordered basic region upstream of En2 internalization domain, and demonstrate the critical role of GAGs as an entry gate, finely tuning homeoprotein capacity to internalize into cells.

Homeoproteins (HPs) are a large family of transcription factors found in all eukaryotes, which are active during development and in adulthood. Beside regulating gene expression, HPs also act as paracrine factors thanks to their unique ability to travel from cell to cell through unconventional transfer<sup>1</sup>. Their paracrine action implies a direct access of the traveling proteins to the cytosol and nucleus of recipient cells. HP atypical paracrine activity relies on common structural features shared by all HPs<sup>2</sup>, and includes specific secretion and internalization motifs<sup>3</sup>. These motifs reside in the 60-residue DNA-binding homeodomain (HD) that defines the HP family (Fig. 1a, b). HDs are organized as three stable helices while the N- and C-

terminal ends are variable and mostly unfolded<sup>4</sup>. The third helix is responsible for the internalization of the protein while the secretion property requires a motif spanning both second and third helices. Notable is that the internalization motif identified 30 years ago within the *Drosophila Antennapedia* HD corresponds to the cationic hexadecapeptide Penetratin (RQIKIWFQNRRMKWKK), which has led to the emergence of the field of cell-penetrating peptides (CPPs)<sup>5–7</sup>. Importantly, two modes of internalization co-exist for HPs and CPPs, endocytosis and direct translocation, distinguished by their sensitivity to low temperature<sup>1,5,8,9</sup>, and by their ability to direct the protein into the cytosol<sup>10</sup>.

<sup>1</sup>Sorbonne Université, École Normale Supérieure, PSL University, CNRS, Laboratoire des Biomolécules (LBM), 75005 Paris, France. <sup>2</sup>Sorbonne Université, Mass Spectrometry Sciences Sorbonne University, MS3U platform, 75005 Paris, France. <sup>3</sup>Univ. Orléans, CNRS, ICOA, 45067 Orléans, France. <sup>4</sup>INSERM U932, Institut Curie Centre de Recherche, PSL Research University, Paris, France. <sup>5</sup>These authors contributed equally: Sébastien Cardon, Yadira P. Hervis.

✉ e-mail: [Ludovic.Carlier@sorbonne-universite.fr](mailto:Ludovic.Carlier@sorbonne-universite.fr); [Sandrine.Sagan@sorbonne-universite.fr](mailto:Sandrine.Sagan@sorbonne-universite.fr)



**Fig. 1 | Structure and recognition sequences of En2 region 183-259.** **a** Solution structure of the homeodomain of *chicken* En2<sup>1</sup> (region 200-259, PDB code 3Z0B). **b** Sequence alignment of En2 region 183-259 (*chicken* gene numbering) with Otx2 region 19-97 (numbering is identical for *chicken* and mammals). Note that these regions are strongly conserved (99% sequence identity between *chicken* and *human* En2, and 100% sequence identity between *chicken* and *human* Otx2). Positively

charged, negatively charged, polar, apolar, and aromatic amino acids are colored in blue, red, green, black, and orange, respectively. The putative nuclear localization sequence (NLS) of En2 and the RK doublet of Otx2 that was shown to be crucial for the binding to highly sulfated CS<sup>22</sup> are highlighted by rectangles. **c** Topology of En2 proteins and peptide used in this study. **d** Structure of representative disaccharide units of heparan sulfates (HS), chondroitin sulfates (CS), and heparin (HI).

CPP internalization depends on the presence of cell-surface glycosaminoglycans (GAGs) both *in vitro*<sup>8,11-14</sup> and *in vivo*<sup>15</sup>. GAGs are complex polysaccharides that are part of proteoglycans found on the cell-surface and in the extracellular matrix. GAGs bind various extracellular ligands and types of proteins at the cell-surface, which modulate their activity. The functional role of GAGs includes numerous biological processes such as embryonic development, neuronal plasticity, regulation of enzymatic activities, or regulation of cell signaling<sup>16-19</sup>. Regarding HPs, the internalization of extracellular Orthodenticle homeobox 2 protein (Otx2) is restricted *in vivo* to parvalbumin neurons of the developing mouse visual cortex. Secreted Otx2 accumulates in these cells through interactions with surrounding perineural nets (PNNs) that are enriched in disulfated chondroitin sulfate (CS) GAGs<sup>20,21</sup>. A RK doublet within Otx2 (Fig. 1b), upstream of the highly conserved homeodomain, is responsible for the specific interaction with 2,6 (CS-D) and 4,6 (CS-E) disulfated CS, leading to the restricted internalization of the protein in parvalbumin neurons<sup>22,23</sup>.

Another homeoprotein, Engrailed2 (En2), behaves differently. En2 controls the patterning of vertebrate embryos, in particular through the regulation of boundary formation in the developing brain<sup>24,25</sup>. Extracellular En2 poorly accumulates in parvalbumin neurons<sup>22</sup> and the region preceding the homeodomain strongly differs from that of Otx2 (Fig. 1b). In contrast to Otx2, this region is highly enriched in basic amino acids and has sequence similarities with nuclear localization signals (NLS).

In this context, we have questioned herein the influence of GAGs, heparan sulfate (HS) and CS, in the internalization, structure, and dynamics of En2. To decipher the role of En2 interactions with GAGs, we have compared the internalization efficacy of two En2 fragments in model cell lines differing in their cell-surface content in CS and HS. Two protein constructs were produced: the homeodomain alone (HD), containing the minimal cell-penetration Penetratin-like sequence, or extended at its *N*-terminal end with the cluster of basic aminoacids (ExtHD) (Fig. 1c). We have characterized at the molecular level the thermodynamics of interaction of both En2 fragments with selected HS and CS molecules, and studied at the amino-acid level, the impact of these interactions on the secondary and tertiary structure of the proteins. We have identified a high-affinity GAG-binding motif located upstream of the homeodomain, which controls the internalization of

En2 through selective interactions with highly sulfated HS at the cell surface.

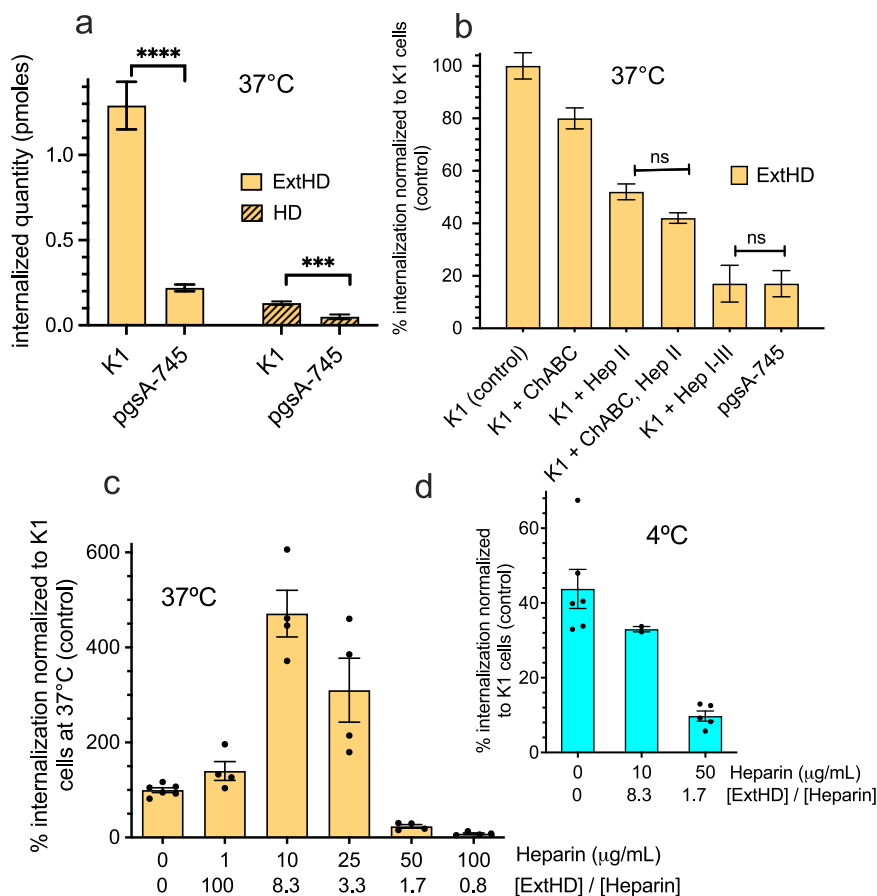
## Results

### ExtHD internalization depends on cell surface HS

First, the uptake efficacies of both En2 fragments were analyzed, using an adapted MALDI-TOF-based internalization assay<sup>26,27</sup>, in two ovarian cell lines that differ in their content of cell-surface glycosaminoglycans. These CHO cell lines present the huge advantage over cultured brain cells that are a mixture of neurons, inter-neurons and glial cells with potential inter-individual variations, of being homogenous in terms of cell phenotype and GAGs content. CHO-K1 cells contain HS and mono-sulfated CS (in a ratio about 50:50)<sup>28,29</sup>. The modified cell line CHO-pgsA-745 (GAG<sup>deficient</sup>) has a genetic deficiency in xylosyl transferase and expresses 10- to 30-fold less HS and CS compared to the parent K1 cell line<sup>28,29</sup>. In K1 cells, HD internalized at very low levels compared to ExtHD (Fig. 2a). After one-hour incubation with each protein added at 7  $\mu$ M, intracellular HD concentration reached 0.13  $\mu$ M (assuming a total volume of 1  $\mu$ L for 10<sup>6</sup> cells). In contrast, ExtHD was internalized 10-fold better. In GAG-deficient cells, internalization of ExtHD was decreased by sixfold, while HD internalization was only decreased by twofold. These results demonstrate that the presence of the *N*-terminal basic extension in ExtHD significantly increases the efficacy of internalization through a process that depends on the presence of cell-surface GAGs. To further analyze the respective contributions of HS and CS on ExtHD internalization, we used different mixtures of enzymes to selectively deplete cell-surface glycosaminoglycans in K1 cells.

Internalization of ExtHD was slightly sensitive to chondroitinase ABC (ChABC) treatment (Fig. 2b). By contrast, heparinase II (HepII) treatment decreased protein internalization by 50%. Combining HepII and ChABC treatments further decreased ExtHD internalization only by 10%. Finally, when cells were treated with mixed heparinases I-III, ExtHD internalization efficacy dramatically dropped to 15% compared to untreated cells, down to levels similar to the ones measured in GAG-deficient cells.

Altogether, the presence of the 17-amino acid extension rich in basic residues that precedes the homeodomain strongly stimulates ExtHD internalization provided that HS are present at the cell-surface,



**Fig. 2 | Quantitation of ExtHD and HD internalization in cells and impact of GAGs content.** **a** Quantity (pmoles) of internalized ExtHD and HD incubated at 7  $\mu\text{M}$  with  $10^6$  CHO-K1 or GAG<sup>deficient</sup> pgsA-745 cells for 1 h at 37°C ( $n = 27$  and 18 for CHO-K1 and pgsA-745 cells, respectively). CHO-K1 express similar levels of HS and mono-sulfated chondroitins<sup>68</sup>; pgsA-745 express 10–30-fold lower GAGs levels than CHO-K1<sup>28</sup>. **b** Quantity of internalized ExtHD incubated at 7  $\mu\text{M}$  with cells for 1 h at 37°C ( $n = 12$ –18). Data were normalized to the quantity of internalized ExtHD in control CHO-K1 cells. CHO-K1 cells were treated with enzymes degrading HS (heparinases I, II, III) or CS (chondroitinase ABC). **c**, **d** Quantity of internalized

ExtHD in CHO-K1 cells in the absence and presence of increasing amounts of extracellular heparin ( $n = 2$ –6). One million cells were incubated with 7  $\mu\text{M}$  protein in DMEM-F12 medium for 1 hr either at 37°C (**c**) or 4°C (**d**). Data were normalized to the intracellular quantity of ExtHD in control CHO-K1 cells at 37°C in the absence of heparin. The ExtHD/heparin molar ratio is indicated for each heparin concentration. Data are presented as mean values  $\pm$  SEM. One-way ANOVA and unpaired  $t$  tests were used in **a**, **b** to determine if differences between values were significant: \*\*\*\* $P < 0.0001$ , \*\*\* $P = 0.0005$ , not significant (ns). The list of  $P$  values is provided in Supplementary Table 3.

while the GAG-dependence of HD internalization appears less pronounced. The importance of the 17-amino acid extension for ExtHD internalization was confirmed with the SH-SY5Y neuronal cell line that contains similar content of cell-surface HS and CS<sup>30</sup>. As in CHO-K1, ExtHD indeed internalized ~4-fold better than HD in SH-SY5Y cells, and its internalization decreased ~3-fold with heparinase treatment of the cells (Supplementary Fig. 16). It should be noted that although our MALDI-TOF-based internalization assay could not be applied to the full-length En2 protein (see details in the Methods section), the ability of this protein to internalize into CHO-K1 cells was however evidenced qualitatively by fluorescence imaging obtained by confocal microscopy (carboxyfluorescein-labeled full-length En2 is visualized inside cells in images of Supplementary Fig. 19a, b).

### ExtHD recognizes only highly sulfated HS motifs

At the cell surface, HS is a heterogeneous linear polymer that contains highly sulfated regions (so-called S-domains) interspersed with less sulfated and non-sulfated ones<sup>31,32</sup>. To identify if specific structures within HS were responsible for protein binding, we performed a microarray analysis to probe the interaction of ExtHD with 52 structurally defined HS oligosaccharides of different lengths (from 5-mer to 18-mer) and sulfation patterns (Supplementary Fig. 1). As shown in Supplementary Figs. 2 and 3, the oligosaccharides that bind the more

efficiently ExtHD are the 12-mer compounds **18** and **19**, which have the greatest number of sulfates and the highest charge densities (i.e., number of sulfates per disaccharide unit).

Notable is the presence of two fully sulfated disaccharide units (*N*-, 2-*O*, 3-*O* and 6-*O* sulfation) in compound **18** and one in compound **19**. A significant fluorescent signal was also observed for the 12-mer compound **22** that contains 4 trisulfated disaccharide units lacking 3-*O* sulfation. By contrast, the 12-mer compounds containing only one or two sulfate groups per disaccharide unit (compounds **23** to **25**) did not yield significant signals, regardless of sulfate distribution. Investigating oligosaccharide length, only low-intensity signals were observed for the sulfated compounds of the 6-, 7- and 9-mer series, including the highly sulfated compound **16** (6-mer, 8 sulfates), indicating that ExtHD better recognizes longer HS fragments ( $\geq 12$  saccharide residues), which is supported by the significant fluorescence signal recorded for compound **27** (Supplementary Fig. 3). Overall, this screening profile shows that ExtHD binds selectively to long and highly charged HS, with a strong preference for trisulfated and tetrasulfated disaccharide units.

### Soluble HI discriminates the two internalization modes of ExtHD

To further characterize the role of HS in ExtHD internalization, the effect of exogenous heparin addition was analyzed on ExtHD internalization. The pharmaceutical heparin used in this study is a highly

**Table 1 | Interaction thermodynamics of the *N*-ter octadecapeptide and proteins with heparin (12 kDa, net charge about -55), studied by ITC**

protein (z+, z-)	NaCl (mM)	$K_d$ (nM)	$\Delta H$ (kJ/mol)	$-\Delta S$ (kJ/mol)	$n$ (prot/poly-sacch.)	Complex net charge
Full length En2 (44+, 31-)	100	570 ± 360	-51 ± 8	10 ± 9	2.9 ± 0.4	39 + /55-
<i>N</i> -ter octadecapeptide (10+, 2-)	100	300 ± 67	-81 ± 7	44 ± 5	7 ± 1	56 + /55-
HD (14+, 6-)	0	870 ± 120	-86 ± 10	51 ± 10	4.3 ± 0.5	38 + /55-
	100	2090 ± 10	-46 ± 8	14 ± 7	5.3 ± 0.5	
	200	nd	nd	nd	nd	
ExtHD (21+, 7-)	0	50 ± 10	-91 ± 2	50 ± 8	3.5 ± 0.5	45 + /55-
	100	120 ± 18	-41 ± 2	1 ± 1	2.9 ± 0.6	
	200	620 ± 130	-19 ± 2	-16 ± 2	3.4 ± 0.1	

The proteins were titrated with the polysaccharide at 25 °C in 50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4. Data are the mean ± SD ( $n=3$ ). nd not determined.

sulfated polymer of ~12 kDa, mostly composed of trisulfated GlcNS,6S-IdoA2S which is found in *N*-sulfated domains of HS<sup>33,34</sup> and is reported as a key structure for protein recognition<sup>35</sup>. Heparin exhibited bimodal concentration-dependent effects at 37 °C, either promoting or inhibiting ExtHD internalization (Fig. 2c). At concentrations lower than 25 µg/mL (protein/heparin molar ratio >3), heparin increased intracellular ExtHD up to 5-fold while concentrations above 50 µg/mL (protein/heparin molar ratio <2) resulted in a substantial decrease of ExtHD internalization, almost to complete inhibition at 100 µg/mL. Interestingly, analysis by dynamic light scattering (DLS) of the mixed heparin/ExtHD solutions (Supplementary Fig. 4) revealed the formation of aggregated species (up to 40%) of µm size at low heparin concentration (10 µg/mL), but neither with the protein alone nor when heparin and ExtHD were in stoichiometric quantities. In contrast, we did not observe aggregated species with the heparin fragment dp20 (Supplementary Fig. 4) and concomitantly did not measure any increase in ExtHD internalization in the presence of dp20, in contrast to heparin (Supplementary Fig. 18). Since the formation of GAG/protein aggregates was previously shown to promote endocytosis of Penetratin-like CPPs<sup>13,14,36</sup>, we analyzed the effect of decreasing temperature on ExtHD internalization efficacy in CHO-K1 cells.

At low temperature (4 °C), endocytosis pathways are totally inhibited, while direct crossing of the lipid bilayer, which is not strictly energy-dependent, still occurs<sup>8,37,38</sup>. In the absence of exogenous heparin, ExtHD internalized in CHO-K1 at 4 °C with 44 ± 6% efficacy compared to 37 °C control condition (Fig. 2d), indicating that the protein efficiently internalizes through direct translocation in cells. Contrasting with 37 °C conditions, at 4 °C the presence of 10 µg/mL heparin in the culture medium resulted in the inhibition of ExtHD internalization in K1 cells (Fig. 2d). Thus, low concentration of soluble heparin can fulfill, even outperform, the requirement of membrane-associated forms of HS for ExtHD endocytosis likely through formation of large complexes between heparin and ExtHD. Conversely, translocation would specifically require interaction with membrane-associated HS, which is antagonized by soluble heparin. Corroborating these results, ExtHD translocation (at 4 °C) was impaired in GAG-deficient CHO psgA-745 cells that have reduced levels of membrane-bound HS (Supplementary Fig. 17).

### The high binding affinity of ExtHD for HI relies on the *N*-ter peptide

To further understand the molecular basis of En2 recognition by highly sulfated HS, we used ITC to determine the thermodynamic parameters of heparin interaction with HD and ExtHD proteins (Table 1 and Supplementary Fig. 5). The two protein fragments differed in their affinities for heparin, ExtHD exhibiting a 20-fold greater affinity than HD (Table 1 and Supplementary Fig. 5). These interactions were enthalpy-driven, showing that electrostatic interactions are involved in the

formation of the complex between either protein and heparin, as supported by the strong dependence of binding affinity on salt concentration. Interestingly, the synthetic octadecapeptide corresponding to the cationic region that precedes the helical core of En2 homeodomain (aa 187–204; Fig. 1c) had similar affinity for heparin as the entire ExtHD protein, indicating that this motif is the main determinant of HS interaction within ExtHD. The ExtHD fragment can accommodate 6–7 disaccharides (20 disaccharides in heparin / 3 bound proteins) while HD and the *N*-ter octadecapeptide can bind 4 (20/5) and 3 (20/7) disaccharides, respectively. As the stoichiometry (polysaccharide/protein or 1/ $n$  determined in ITC experiments) of interaction between ExtHD and heparin corresponds to the sum of the stoichiometry of interaction measured with the *N*-ter octadecapeptide and HD alone respectively, cooperative events in this interaction appear unlikely. Altogether, these results indicate that the *N*-ter octadecapeptide confers a unique property to the ExtHD protein in terms of binding to highly sulfated HS. It should be noted that the full-length En2 protein interacts with heparin (Table 1 and Supplementary Fig. 6) with an affinity about ~5-fold lower than that of ExtHD, the stoichiometry being identical for the two proteins (3 proteins bound per HI polymer). This indicates the absence of additional heparin binding site in the full-length En2 protein. The lower affinity of full-length En2 for heparin arises from an unfavorable entropy contribution, likely because of the presence of the large *N*-terminal disordered region that partly compensates the strong favorable enthalpy contribution (see discussion in caption of Supplementary Fig. 7).

### Similar binding affinity of ExtHD and HD for phospholipids

Since protein-lipid interactions play a key role in membrane-translocation properties of homeodomains<sup>10</sup>, we next investigated whether enhanced affinity with phospholipids also contribute to the higher internalization efficacy of ExtHD. In particular, we previously showed that the interaction of En2 homeodomain with anionic lipid vesicles induces a major rearrangement of its tertiary structure<sup>4</sup>. Herein we used ITC to probe the interaction of HD and ExtHD with 100 nm large unilamellar vesicles (LUVs) that we prepared either from 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (POPG), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), or a combination of both. The presence of one saturated and one unsaturated fatty acyl chains makes these phospholipids good mimics of mammalian cell membrane composition.

With pure zwitterionic lipids (Table 2), no interaction occurred with either of the two proteins used at 20 µM as previously reported for the HD construct<sup>4</sup>. Addition of anionic POPG to POPC to obtain LUVs with 50/50 POPG:POPC and 70/30 POPG:POPC composition, had no effect. Significant ExtHD and HD interactions were observed only with pure POPG LUVs, showing that negatively charged partners are mandatory for interaction of the proteins with lipid membranes. The



dissociation constant value determined for POPG was the same for the two proteins ( $\approx 7 \mu\text{M}$ ; Table 2). Thus, the presence of the cationic *N*-terminal extension does not strengthen the interaction of ExtHD with anionic phospholipids.

### ExtHD contains two independent GAG-binding sites

We next investigated the molecular basis of the interaction between En2 and highly sulfated GAGs at the single residue level using heteronuclear single-quantum correlation (HSQC) NMR experiments. To identify GAG-binding regions of ExtHD, we performed chemical shift perturbation (CSP) experiments by collecting  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra on uniformly  $^{15}\text{N}$ -labeled samples in the presence of increasing amounts of unlabeled oligosaccharides. Our initial attempts to titrate  $^{15}\text{N}$ -labeled ExtHD with 12 kDa heparin resulted in dramatic line broadening of NMR signals (Supplementary Fig. 9), in agreement with the formation of a high molecular weight complex as determined by ITC (3:1 protein:GAG complex of 40.5 kDa). Since both the glycan microarray analysis and ITC indicated that ExtHD can accommodate oligosaccharide fragments containing  $\sim 6$ – $7$  trisulfated disaccharides, we selected a heparin dodecasaccharide (dp12) fragment for NMR titrations. As shown in Fig. 3a, the addition of stoichiometric amounts of heparin dp12 induced large chemical shift changes in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC of ExtHD, consistent with a fast exchange regime on the chemical shift time scale and a dissociation constant in the micromolar–millimolar range. An increase in the line width of the NMR cross-peaks

was only observed for a few residues, indicating that the binding to the heparin fragment does not induce oligomerization of the globular homeodomain.

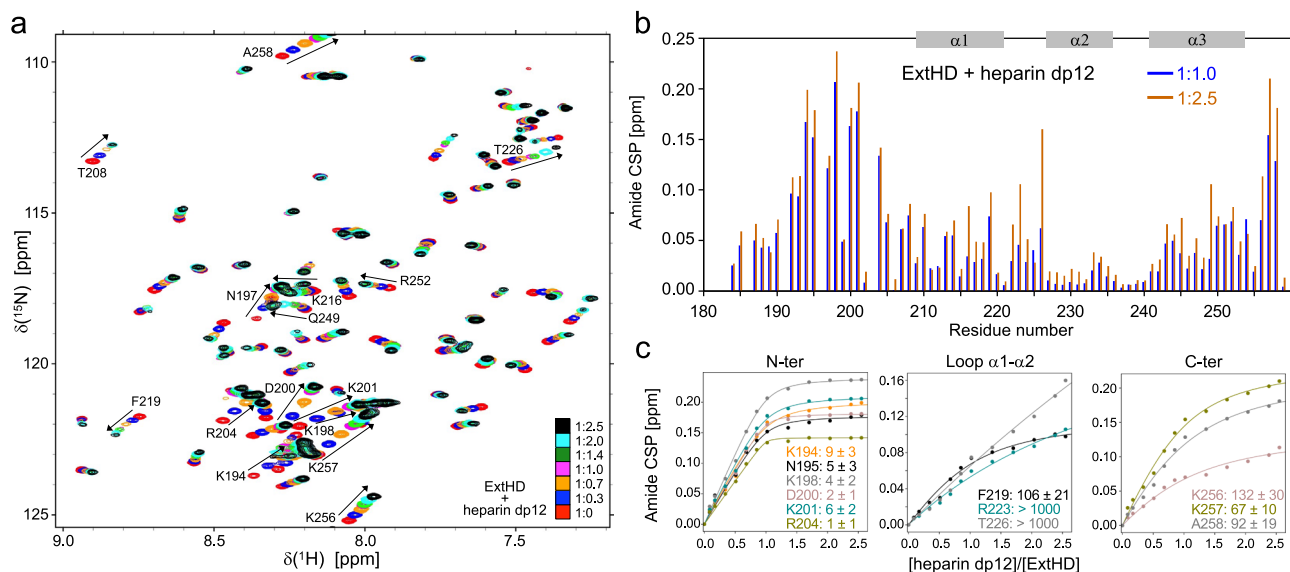
The highest perturbations were mostly localized in the cationic region  $^{189}\text{RKP}^{\text{KKKNPNKEDKR}}^{\text{RPR}204}$ , which includes the basic residues preceding helix  $\alpha 1$  known to interact with DNA (Fig. 3b). Large CSP values were also observed at the *C*-terminal tail of the homeodomain that encompasses three positively charged residues ( $^{254}\text{KIKKAT}^{259}$ ). By contrast, residues of the helical domain displayed rather modest chemical shift changes in the presence of 1.0 molar equivalent of heparin dp12. Additional large perturbations are nevertheless observed beyond a 1:2 protein:GAG ratio in helix  $\alpha 1$  (K216 and F219), the loop  $\alpha 1$ - $\alpha 2$  (R223 and T226), and helix  $\alpha 3$  (Q249, N250, K251, and R252). Apparent thermodynamic affinities ( $K_d^{\text{app}}$ ) were determined at the residue level from saturation curves (Fig. 3c). Variable  $K_d^{\text{app}}$  values in the low micromolar range ( $\sim 4$ – $10 \mu\text{M}$ ) were obtained in the *N*-terminal region, the highest binding affinities being observed in the basic epitope 189–204 (Supplementary Table 2). The *C*-terminal residues 256–258 exhibit  $K_d^{\text{app}}$  values in the range of  $70$ – $130 \mu\text{M}$  and bound heparin dp12 with a significantly lower affinity. In the case of residues R223 and T226, the binding curves were characteristic of very weak interactions with  $K_d^{\text{app}}$  values in the millimolar range. As a result, the loop  $\alpha 1$ - $\alpha 2$  did not contribute significantly to the binding to heparin despite the large CSP values measured in this region above a protein:GAG ratio of 1:2.

Taken together, our data reveal that the helical core of En2 homeodomain is flanked by two heparin-binding motifs with residues 192–204 forming the major binding site, and the *C*-terminal residues 254–259 forming an additional, low-affinity binding region. To confirm that these two motifs do not bind heparin in a synergistic manner, we titrated heparin dp12 at increasing concentrations into a solution of  $^{15}\text{N}$ -labeled HD. The absence of residues 183–199 in the HD construct does not modify significantly the perturbation profile of residues 200–259, which CSP values were remarkably similar to the ones measured with the longer ExtHD construct (Supplementary Fig. 10).

**Table 2 | Thermodynamics of interaction of ExtHD and HD with POPG LUVs (100 nm), determined by ITC**

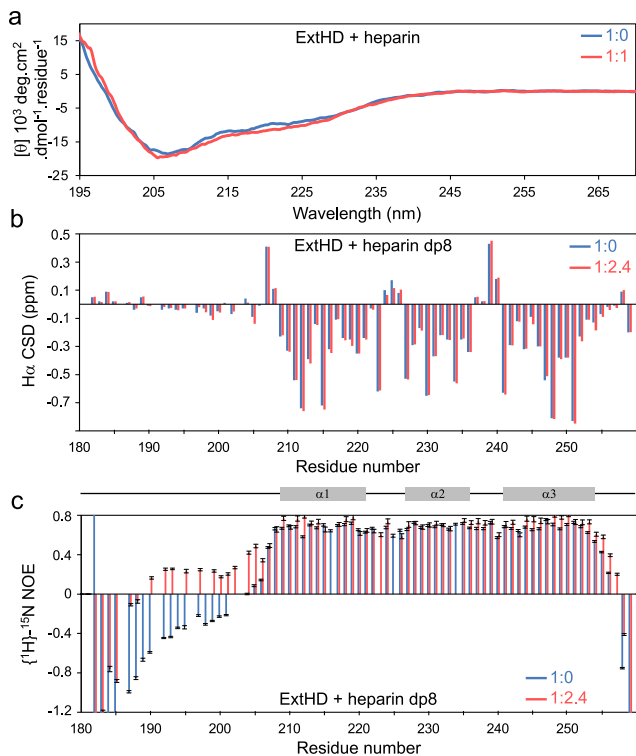
Protein	$K_d^{\text{apparent}}$ ( $\mu\text{M}$ )	$\Delta H$ (kJ/mol)	$-\Delta S$ (kJ/mol)	$n$ (lipids/protein)
HD	$7.5 \pm 2.5$	$1.6 \pm 0.1$	$-30 \pm 5$	$7.6 \pm 0.8$
ExtHD	$6.8 \pm 2.3$	$3.1 \pm 1.3$	$-29 \pm 3$	$8.7 \pm 2.3$

The proteins were titrated by LUVs at  $25^\circ\text{C}$  in 50 mM  $\text{NaH}_2\text{PO}_4$  buffer, pH 7.5. Data are the mean  $\pm$  SD ( $n = 3$ ).



**Fig. 3 | NMR characterization of ExtHD interaction with heparin dp12 at the amino acid level.** Interaction of ExtHD with heparin dp12 is probed in 50 mM  $\text{NaH}_2\text{PO}_4$  (pH 7.4) and 100 mM NaCl. **a** Overlay of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra (500 MHz) obtained for  $^{15}\text{N}$ -ExtHD in the absence and presence of increasing amounts of heparin dp12. Stoichiometry between ExtHD and heparin dp12 is: 1:0 (red), 1:0.3 (dark blue), 1:0.7 (orange), 1:1 (pink), 1:1.4 (green), 1:2 (light blue) and 1:2.5 (black).

Amino acid residues exhibiting the highest perturbations (arrows) are indicated in the spectra. **b** Amide CSP of ExtHD induced by the addition of 1.0 and 2.5 molar equivalents of heparin dp12. **c** Saturation curves corresponding to the binding to heparin dp12 are shown for residues displaying the highest perturbations with experimental data represented by points and non-linear curve fitting by lines.  $K_d^{\text{app}}$  values are given in micromolar.



**Fig. 4 | The GAG-binding region 189–204 of Engrailed2 remains disordered upon binding to heparin or heparin fragments.** a CD spectra of ExtHD in the absence (blue) and presence (red) of 1 molar equivalent 12 kDa heparin. Chemical shift deviations (CSD) and  $\{^1\text{H}\}^{15}\text{N}$  NOEs of ExtHD recorded at 500 MHz in the absence (blue) and presence (red) of 2.4 molar equivalents heparin dp8 are shown in b, c, respectively. Error bars in c indicate the standard error for each value based on peak intensity relative to background noise (see the “Methods” section).

This observation is consistent with the existence of independent interacting regions that bind heparin with differential affinities.

To identify basic residues within the *N*-terminal region that predominantly contribute to heparin binding, we produced three ExtHD mutants in which three basic pairs in the region  $^{189}\text{RKPKKKNPKNKEDKRPR}^{204}$  were individually replaced by serine pairs. ITC data revealed that all three mutants lose affinity for heparin compared to the parent protein, from 5-fold for the R189S/K190S mutant up to 8.5-fold for the K193S/K194S one (Supplementary Table 1 and Supplementary Fig. 5). Since the two mutants K193S/K194S and K201S/R202S exhibited the lowest binding affinities, in the  $\mu\text{M}$  range, the basic residues within the region 192–204 are mostly responsible for the binding to heparin.

### The *N*-ter GAG-binding motif bound to HI fragments remains disordered

We next investigated the influence of heparin binding on the conformational properties of En2 region 183–259. For that purpose, we examined the site-specific secondary structure and dynamics of ExtHD in the absence and presence of heparin fragments using  $^1\text{H}\alpha$  chemical shift deviations (CSDs) and heteronuclear  $^1\text{H}^{15}\text{N}$  NOEs. The stability of the NMR samples containing stoichiometric amounts of heparin dp12 was not long enough to record these experiments. Therefore, we used the shorter heparin dp8 fragment, which provided more stable samples and binds ExtHD with a similar CSP profile (Supplementary Fig. 10) albeit with a slightly weaker affinity (Supplementary Table 2). CSDs are defined as the differences between measured chemical shifts and corresponding random coil values for each residue<sup>39</sup>. Previous studies have shown that residues 183–206 upstream of the homeodomain are highly disordered in an

En2 fragment encompassing residues 146–259<sup>40,41</sup>. Accordingly, small near-zero values characteristic of disorder were observed in the region 183–206 of ExtHD in the absence of heparin dp8 (Fig. 4b). By contrast, the three  $\alpha$ -helices of the homeodomain were well identified by three continuous stretches of negative CSDs. The addition of 2.4 molar equivalents of heparin dp8 did not induce significant changes in the  $^1\text{H}\alpha$  chemical shifts of ExtHD, and the CSD diagram of the bound form was virtually identical to that of the free form. Together with the fact that the presence of heparin fragments only induced selective perturbations of the  $^1\text{H}^{15}\text{N}$  chemical shifts, our data indicate that the binding to heparin dp8 does not modify the secondary and tertiary structure of ExtHD. To verify that the conformational properties of ExtHD do not depend on oligosaccharide chain length, we recorded circular dichroism (CD) spectra in the absence and presence of 1 molar equivalent of 12 kDa heparin. The presence of heparin did not modify significantly the CD spectrum (Fig. 4a), confirming that ExtHD, and in particular the *N*-terminal residues 192–204 do not undergo major structural rearrangement upon GAG binding. We measured heteronuclear  $\{^1\text{H}\}^{15}\text{N}$  nuclear Overhauser effects (NOEs) to probe backbone motions of ExtHD that occur on the picosecond-to-nanosecond timescale (Fig. 4c). In the free form, the disordered nature of the *N*- and *C*-terminal residues flanking the helical core of the homeodomain was confirmed by negative NOEs that are indicative of high flexibility. Upon binding to heparin dp8, NOE values were found to increase in the *N*-terminal region 190–205 and at the *C*-terminal extremity, evidencing that the interaction with GAGs significantly restricts internal motions in these regions. However, the amplitude of NOEs in the bound form remained weak in the *N*-terminal region ( $<0.25$ ) when compared to the rigid helical core that exhibits NOEs  $>0.7$ . As a result, the backbone of the *N*-terminal GAG-binding site retained substantial mobility when bound to heparin, which supports a binding mode mostly driven by electrostatic contacts.

### ExtHD selectively interacts with HI compared to disulfated CS

As evidenced above, ExtHD preferentially interacts with highly sulfated regions of heparan sulfates. Since other GAG types can harbor high sulfate content, such as sulfated chondroitin (CS) enriched in the PNNs of developing cortex, we investigated whether ExtHD also recognizes structurally defined CS oligosaccharides. The NMR titrations of ExtHD with CS-A and CS-C dp6 fragments did not result in significant modifications of protein chemical shifts. By contrast, the CSP pattern induced by CS-E dp6 showed strong similarities with that obtained with heparin dp6 (Supplementary Fig. 11). Large CSP were observed in both the *N*- and *C*-terminal regions flanking the helical domain and the loop  $\alpha 1$ – $\alpha 2$ . Similar to what was observed with heparin fragments, the greatest apparent binding affinities were obtained in the *N*-terminal region 192–204 (Supplementary Fig. 11 and Supplementary Table 2). These data suggest that heparin and CS-E oligosaccharides share similar ExtHD-binding sites. However, in comparison with heparin dp6, the CSP magnitude and apparent binding affinities are weaker in both the *N*-terminal region and the *C*-terminal extremity.

To confirm the lower affinity of ExtHD for CS-E, we used ITC and titrated HD and ExtHD proteins with a commercially available CS-E fragment of 72-kDa (Supplementary Fig. 8). Because the size of this polysaccharide differs from that of heparin, i.e. 12 kDa ( $\sim 20$  disaccharides) for heparin and 72 kDa ( $\sim 135$  disaccharides) for CS-E, the binding affinity for each oligosaccharide type was determined by dividing the measured free energy by the number of bound protein molecules per polysaccharide determined by ITC (see Supplementary Information for further explanation). Our results indicate that ExtHD displays a 126-fold greater affinity for heparin than for CS-E (Table 3). In addition, ExtHD has 17-fold better affinity than HD for heparin whereas the two proteins have similar affinity for CS-E. These data confirm that the presence of the cationic *N*-terminal extension confers to the

**Table 3 | Interaction thermodynamics of the N-ter octadecapeptide and proteins with heparin and CS-E, studied by ITC**

Protein	Polysacch.	$\Delta G_n$ (kJ/mol)	<i>n</i> protein/ polysacch.	$\Delta G_n/n$ (kJ/mol)	$K_d$ mM
N-ter	heparin	-37	7	-5.3	117
	CS-E	-43	55	-0.78	730
HD	heparin	-32	5	-6.4	75
	CS-E	-44	33	-1.3	591
ExtHD	heparin	-40	3	-13.3	4.5
	CS-E	-41	29	-1.4	568

The proteins were titrated with the polysaccharides at 25 °C in 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 100 mM NaCl, pH 7.4. The affinity of peptide or protein for the polysaccharide is determined by dividing the free Gibbs energy by *n* (stoichiometry of the complexes) and calculated as  $K_d = e^{\Delta G_n/nRT}$ . Data are presented as the mean (*n* = 3).

homeodomain a higher selectivity of recognition for trisulfated HS (heparin over CS-E: selectivity of 126-fold for ExtHD versus 8-fold for HD construct).

## Discussion

The internalization pathways of homeoproteins and of CPPs derived from the third helix of homeodomains still need to be fully characterized<sup>7,42</sup>. Similarly to the requirement of disulfated GAGs for Otx2 internalization demonstrated *in vivo*, internalization of Penetratin (the highly conserved third  $\alpha$ -helical hexadecapeptide of homeodomains) depends on the presence of cell-surface GAGs<sup>8,12-15,43,44</sup>. Therefore, one could assume that Penetratin-like CPPs, HDs, and HPs share at least some steps of their mechanisms of internalization<sup>45</sup>. Using a combination of quantitative methods, we explored the role of specific cell-surface GAGs in the recognition and internalization process of the homeoprotein Engrailed2 (En2), and intended to identify protein motifs involved in the interaction.

Our data show that the ExtHD fragment encompassing En2 residues 183–259 enters cells principally through interaction with glycosaminoglycans of heparan sulfate (HS) type, thanks to the presence of a unique GAG-binding motif <sup>189</sup>RKP<sup>44</sup>K<sup>43</sup>KNPNKEDK<sup>41</sup>RP<sup>204</sup> that selectively binds highly sulfated HS. Truncation of this high affinity GAG-binding motif that precedes the helical core of the homeodomain drastically reduces its internalization in the CHO cell model. NMR studies reveal that the C-terminal extremity of the homeodomain, also interacts with heparin, however with about two orders of magnitude less affinity. As demonstrated by ITC and NMR data, the two main GAG-binding regions flanking the helical domain interact with heparin independently from each other and not synergistically.

The high-affinity HS-binding sequence lies in a highly flexible region of the protein<sup>40,41</sup>, which remains disordered upon binding to GAGs. It contains the BBXB motif (where B is a basic residue) previously described as a consensus HS-binding sequence<sup>46</sup>. It also holds the KKK sequence identified as one of the most significant tripeptides (99th percentile) found among 437 heparin-binding proteins<sup>47</sup>. Accordingly, mutation of basic pairs within the <sup>192</sup>KKK<sup>194</sup> and <sup>201</sup>KRPRT<sup>205</sup> sequences lead to a significant reduction in the binding affinity for heparin. Mutation of the basic pair R189/K190 also decreases the affinity for heparin, showing that the full hexadecapeptide region 189–204 is required to maintain high-affinity binding to heparin. We can thus assume that the full HS binding motif of En2 corresponds to the sequence (B)<sub>2</sub>X(B)<sub>3</sub>(X)<sub>3</sub>BAABBXXB where A is an acidic residue (B and X being respectively a basic and any AA). Interestingly, this motif corresponds to an evolutionary conserved region in engrailed gene sequences.

An important feature of cell-surface HS is the high variability in size, composition, and sulfate distribution of the sulfated regions (S-domains) in which disaccharide units can be sulfated at 4 distinct

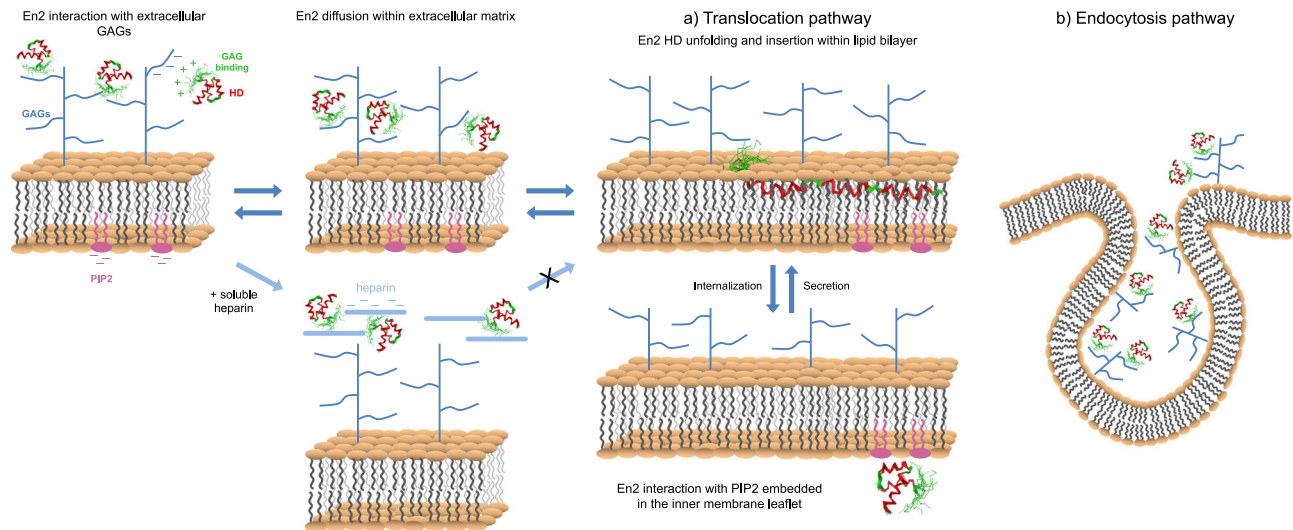
positions. In addition, the disaccharide composition of HS is known to vary from one cell type to the other<sup>35,48</sup>. Using a glycan microarray analysis, we showed that En2 region 183–259 binds preferentially to long (>dp12) and highly sulfated HS oligosaccharides. The correlation between binding affinity and HS sulfate density, together with the strong dependence of the interaction on salt concentration supports a charge-based binding mode. Accordingly, the high-affinity HS-binding motif of En2 can also accommodate the highly sulfated polysaccharide CS-E, albeit with a lower affinity in comparison with the binding to heparin. In addition, the presence of soluble CS-E in the extracellular medium was shown to either promote or inhibit En2 internalization in K1 cells, depending on the GAG concentration, as observed with soluble heparin (Supplementary Fig. 18). These data suggest that membrane surface CS-E could also play a role in En2 internalization.

A specific feature of the internalization of HP and HP-derived domains (HD, Penetratin) is the co-existence of two mechanisms, endocytosis and translocation. For all, internalization indeed occurs at 4 °C while endocytosis is prevented, and mutations that decrease Penetratin uptake efficiency, such as the replacement of the critical Trp residue by Phe, also prevent HD and HP internalization<sup>5,49</sup>. Translocation implies a transient perturbation/disorganization of the lipid bilayer<sup>3,37,45,50</sup> and it has been recently shown in cell culture that direct interactions between En2 and the anionic phospholipid PIP2 promote the translocation of the protein towards the cytosol<sup>10</sup>. Moreover, Carlier et al. demonstrated that interaction of En2 homeodomain with anionic phospholipids is required to induce a conformational change in its 3D structure, leading to the insertion of the third helix within the acyl chains of the lipid bilayer<sup>4</sup>. The plasma membrane is a highly asymmetric structure. In particular, negatively charged lipids are under-represented in the outer membrane leaflet facing the extracellular space, raising the issue of the initiation of the translocation event leading to internalization.

Our data herein support a role of HS as a critical actor of En2 internalization not only through endocytosis<sup>51</sup>, but also through direct translocation across the plasma membrane. En2 translocation (measured at 4 °C) is reduced in GAG-deficient cells compared to wild-type cells. Based on previous reports<sup>4</sup> and the data in the present work, we propose that sulfated GAGs anchored by proteins at the cell surface would be good candidates as translocation initiators (Fig. 5). In this model, the highly sulfated S-domains of HS attract En2 at the cell-surface through electrostatic interactions with the HS-binding motif that precedes the homeodomain. The dynamic nature of these interactions allows a fast diffusion of En2 within the extracellular matrix leading to its stepwise accumulation at the close vicinity of the plasma membrane. Indeed, the selective inhibitory action of low concentrations of soluble heparin on translocation but not on endocytosis, supports the need for a proximal distance between GAGs and the lipid bilayer to evoke translocation. This initial interaction with HS would bring the homeodomain in close contact to the acyl part of the lipid bilayer<sup>52</sup>, leading to the modification of its conformation and the insertion of the third helix within the lipid bilayer. The interaction of En2 with PIP2 in the inner leaflet, shown to regulate the bidirectional exchanges of the protein between the cytosol and the lipid bilayer<sup>10</sup>, would eventually result in the transfer of the protein within the cytosol. Following this model, we propose that the bidirectional transfer of En2 is regulated by a dynamic equilibrium between protein-lipid or protein-GAG interactions depending on the side of the plasma membrane. It is interesting to note that our DLS and NMR data do not support the oligomerization of ExtHD when bound to heparin fragments from dp4 to dp20. Therefore, En2 oligomerization does not seem to be a prerequisite for membrane translocation, as already proposed for its secretion<sup>53</sup>.

The interaction of GAGs with extracellular HPs has important consequences at the physiological level, as the paracrine action of HPs relies on their ability to be internalized. For instance, the plasticity of





**Fig. 5 | Hypothetical model of GAG-dependent bidirectional transfer of HPs across the plasma membrane.** The role of highly sulfated HS in cell internalization of En2 through translocation and endocytosis pathways is illustrated in panels **a** and **b**, respectively.

the visual cortex relies on the ability of secreted Otx2 to internalize into parvalbumin neurons through interactions with the surrounding extracellular matrix enriched in disulfated GAGs<sup>22,23</sup>. It is rather interesting however to mention that the molecular composition and glycan structure of CSPGs vary in PNNs<sup>54</sup>. This heterogeneity of PNNs is crucial since it impacts the localization of the homeoprotein Otx2, thus controls the functional maturation of PV cells, but it makes it more challenging to define a CS-E structure representative of that found in PNNs. The axonal guidance of retinal neurons also depends on the paracrine action of HPs, Vax1 at the optic chiasma and later, En2 in the tectum<sup>55,56</sup>. Vax1 internalization requires the presence of HS on retinal neurons, and for En2, its delivery into the cytosol is required to exert its paracrine action that involves the translational activation of Ephrin-mediated axon guidance<sup>55,57</sup>. Since a graded distribution of En2 in the extracellular matrix of the tectum is necessary for the proper positioning of retinal axons in the developing brain<sup>58</sup>, GAGs might also act to restrict HP diffusion in the extracellular space<sup>33,59</sup>, protecting them from enzymatic degradation (Supplementary Fig. 12).

A critical component of HP paracrine action is the restriction of their internalization to specific cell types. Due to their widely heterogeneous structures, GAGs are attractive candidates to modulate the cellular tropism of HP internalization. Compared to En2, the binding specificity for GAGs is markedly different in the case of Otx2 homeoprotein, which contains only 5 positively charged residues, including a RK doublet, in the *N*-terminal extension preceding the homeodomain helical core (Fig. 1). In vivo, an Otx2 fragment encompassing the homeodomain and the RK doublet is selectively internalized into cortical neurons surrounded by perineuronal nets similarly to the full protein, while the corresponding AA double-mutant is not. Mutation of the RK doublet by alanines in an Otx2 pentadecapeptide (<sup>36</sup>RKQRRERTTFTRAQL<sup>50</sup>) fully abolishes peptide binding to CS-D and CS-E oligosaccharides<sup>22</sup>. These results show that the RK doublet upstream of the homeodomain is responsible for the specific interaction of Otx2 with cell-surface CS and its subsequent internalization in PNN-positive cells. Interestingly, the GAG-interacting domains within En2 and Otx2 strongly differ, the RK doublet that recognized disulfated CS in Otx2 being replaced by a KE doublet in En2, which is frequently found in HS-binding proteins<sup>47</sup>. In addition, En2 region 183-259 (ExtHD) interacts more efficiently with heparin than disulfated CS-E, while Otx2 displays higher affinity for CS-E and CS-D as compared to heparin<sup>22</sup>. Accordingly, En2 is significantly less internalized in PNN-positive cells than Otx2<sup>22</sup>. Overall, these findings support the idea that the cationic motifs preceding the homeodomain region control the spatial

distribution of extracellular homeoproteins through a fine tuning of their binding specificity for GAGs. This underlines the importance of these GAG-binding motifs that define a functional region involved in HP trafficking, acting in concert with internalization/secretion motifs.

## Methods

### Expression and purification of recombinant proteins

Regions 200-259 (HD) and 183-259 (ExtHD) of *chicken* Engrailed2 (Uniprot accession number Q05917) as well as the full-length protein (residues 1-289) were expressed as (His)<sub>6</sub>-tagged Cherry fusion proteins using pScherry1 vectors and the *E. coli* SE1 strain as expression host (Eurogentec, Seraing, Belgium). ExtHD mutants K189S/K190S, K193S/K194S, and K201S/R202S were expressed as (His)<sub>6</sub>-tagged GST fusion proteins using pGEX-6P plasmids (Genscript, Piscataway, NJ, USA) and the *E. coli* BL21-CodonPlus-RIL expression strain (Agilent Technologies, Santa Clara, CA, USA). A PreScission protease cleavage site (LEVLFG/GP) was introduced prior to the En2 sequence in all constructs. As a result of PreScission cleavage, final proteins have four extra residues at the *N*-terminus (GPAS). For internalization assays, an additional Cys residue was introduced at the *N*-terminus by site-directed mutagenesis (leading to GPCAS). Unlabeled fusion proteins were produced from cells cultured in LB rich medium. Uniformly labeled <sup>15</sup>N and <sup>15</sup>N,<sup>13</sup>C proteins were overexpressed in M9 minimum media using the procedure developed by Marley et al. for preparing high yields of uniformly labeled proteins<sup>60</sup>. Protein expression was induced by adding 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) to the cell cultures followed by a 4-h incubation period at 37 °C under agitation. Cells were pelleted by centrifugation (3000 × g for 20 min), resuspended in a 20 mM NaH<sub>2</sub>PO<sub>4</sub> buffer (pH 7.4) containing 500 mM NaCl, 30 mM imidazole and 1 mM DTT, and lysed by sonication. Soluble fractions containing En2 proteins were separated from cell debris by centrifugation (30,000 × g for 40 min at 4 °C), and loaded on a Ni-NTA column (GE Healthcare, Chicago, IL, USA). Proteins were purified using IMAC standard protocols with a 30–250 mM imidazole gradient. Eluted fractions were pooled and dialyzed against PBS buffer, and (His)<sub>6</sub>-tagged PreScission protease was added to cleave fusion proteins. A second step of purification was further performed using Ni-NTA chromatography to remove the cleaved (His)<sub>6</sub>-tagged fusion partner and the (His)<sub>6</sub>-tagged PreScission protease. Purity of protein fractions was controlled by SDS-PAGE analysis, and protein identity was checked by MALDI-TOF analysis. Fractions containing pure target proteins were finally concentrated by ultrafiltration using a 3-kDa cutoff membrane (Millipore). Protein folding was controlled by



NMR. In particular, the  $^1\text{H}$ ,  $^{15}\text{N}$  chemical shifts of the homeodomain region in the *E. coli*-expressed ExtHD (Fig. 3a) and HD (Supplementary Fig. 10) proteins were very similar to those previously reported for the native helix-turn-helix fold of En2 homeodomain<sup>4</sup>, indicating that the region of the homeodomain adopts its native fold in our *E. coli*-expressed proteins. The synthetic *N*-ter octadecapeptide used in ITC experiments (biotin sulfone-G4-RSRKPKKKNPNKEDKRPR-CONH<sub>2</sub>) was synthesized by the Boc strategy, purified by RP-HPLC (purity >95%) and checked by MALDI-TOF mass spectrometry (expected mass: 2744.5 Da; measured mass: 2744.9 Da).

### Preparation of $^{14}\text{N}$ , $^{15}\text{N}$ biotin-labeled recombinant proteins

Biotin-labeled proteins were used in internalization assays to fish the internalized species out the cell lysates through specific binding to streptavidin-coated magnetic beads. The two recombinant proteins expressed in *E. coli* contain an additional Cys on the *N*-terminal end. Purified proteins were transferred to a buffer containing 50 mM sodium phosphate (pH 7.0), 150 mM NaCl, 10 mM EDTA, and 400  $\mu\text{M}$  TCEP. The protein solution was degassed for 30 min under vacuum before addition of 5 equivalents of maleimide-PEG2-biotin (MW = 525) (Sigma Aldrich, Saint-Louis, MO, USA). The reaction was carried out under inert atmosphere (N<sub>2</sub>) to prevent oxidation of the maleimide-PEG2-biotin and decrease of the reaction yield. The reaction solution was protected from the light and stirred for 90 min at room temperature. The reaction was then repeated with addition of fresh maleimide-PEG2-biotin solution. Proteins were transferred into a buffer containing 20 mM NaH<sub>2</sub>PO<sub>4</sub> (pH 7.2), 20 mM NaCl, and stored at  $-80^\circ\text{C}$ . The yield of protein biotinylation was evaluated to be >90% for HD and ExtHD, obtained from MALDI-TOF MS analyses (Supplementary Fig. 14). In the case of the full-length En2 protein, we encountered many difficulties in obtaining the biotin-labeled protein in a sufficient yield (>50%) to perform the quantification assays, even when trying different strategies. In addition, we could not recover the intact protein after incubation with cells, likely because the large *N*-terminal disordered region is unstable and partially subject to in-cell proteolysis during the assays. Other attempts to introduce one biotin to the full-length protein, either to the native Cys178 or by addition of an extra Cys at the *C*-terminal end of the protein were also unsuccessful: the yield of biotinylation was too low (~20%) in the first case and biotin was found inaccessible to streptavidin-coated beads, in the second one.

### Cell culture

Wild type Chinese Hamster Ovary CHO-K1 (<https://www.atcc.org/products/ccl-61>), GAG-deficient CHO-pgsA745 cells (GAG-deficient, <https://www.atcc.org/products/crl-2242>) and SH-SY5Y (<https://www.atcc.org/products/crl-2266>) (ATCC, LGC Standards S.a.r.l. - France) were cultured in DMEM-F12 medium (Life Technologies, Carlsbad, CA, USA) supplemented with 10% fetal calf serum (FCS), penicillin (100,000 IU/L), streptomycin (100,000 IU/L), and amphotericin B (1 mg/L) in a humidified atmosphere containing 5% CO<sub>2</sub> at 37 °C. SH-SY5Y cells were cultured in DMEM-F12 medium supplemented with 10% FCS and differentiated by a sequential treatment with retinoic acid (RA) and brain derived neurotrophic factor (BDNF). The protocol of differentiation consisted in the addition of 10  $\mu\text{M}$  RA (Sigma Aldrich, Saint-Louis, MO, USA) in DMEM-F12 containing 15% FCS, 2 days after seeding. Five days later, cells were washed three times with DMEM-F12 and incubated with 50 ng/mL BDNF (Sigma Aldrich, Saint-Louis, MO, USA) in serum-free medium for 24 h.

### Enzymatic removal of glycosaminoglycans from cell surface

Cells were washed with PBS and incubated with enzymes, alone or in mixture: heparinase I, heparinase II, heparinase III or/and chondroitinase ABC (Sigma Aldrich, Saint-Louis, MO, USA) at a final concentration at 0.1 IU/mL in PBS pH 7.4 (containing calcium and

magnesium) for 1 h at 37 °C. Cells were then washed with HBSS, then with DMEM F12 before further incubation with the proteins.

### Quantification of protein internalization by mass spectrometry

A MALDI-TOF-based internalization assay was previously shown to accurately quantify the cellular uptake of biotinylated cell-penetrating peptides<sup>26</sup>. We have adapted this protocol to proteins (Supplementary Fig. 13). The method relies on the calibration of the internalized unlabeled protein with the addition of a known amount of its isotopically  $^{15}\text{N}$ -labeled counterpart just prior cell lysis and before affinity capture through a biotin bait introduced on a cysteinyl residue added at the *N*-terminus (Supplementary Fig. 14). The quantification is achieved through the ratio of the respective integrated peak area of the non-labeled and  $^{15}\text{N}$  labeled protein, here used as internal standard. The rate of labeling was calculated from the ion signals (m/z) to be >90%. Despite this high rate of labeling, the  $^{14}\text{N}$  and  $^{15}\text{N}$  protein ions have overlapped m/z signals in MALDI-TOF MS spectra. To overcome this overlapping, we developed a software allowing the deconvolution of the signals and an accurate fit of the ratio between the  $^{14}\text{N}$  non-labeled and  $^{15}\text{N}$  labeled species (see below for details). By comparing the experimental mass spectra to those obtained from the mixing of the individual spectra respectively for the labeled and the non-labeled protein, we were able to determine the absolute quantity of protein internalized inside cells (Supplementary Fig. 15).

Briefly, adherent and confluent cells (10<sup>6</sup> cells/well in 12-well plates) were incubated with 500  $\mu\text{L}$  of a  $^{14}\text{N}$ -protein in culture medium at 7  $\mu\text{M}$  for 60 min at 37 °C. After washing, trypsin treatment of cells permits to digest membrane-bound protein species and to detach cells. Cells are lysed with a hypertonic detergent solution containing a known amount of the  $^{15}\text{N}$ -labeled proteins and boiled. After centrifugation to remove cell debris, the lysate is incubated with streptavidin-coated magnetic beads for 1 h to extract  $^{14}\text{N}$ - and  $^{15}\text{N}$ -proteins. After several washing steps, proteins are eluted from beads in 3  $\mu\text{L}$  of CHCA matrix at room temperature for 15 min. We found CHCA the most suitable matrix for elution, incorporation in the matrix crystals, and MALDI-TOF analysis. We used purified CHCA and classical calibrant mixtures, such as Peptide Mix3 and Protein Mix2, from LaserBio Labs (Sofia Antipolis, France). In all MALDI-TOF experiments, we worked with saturated CHCA solution in 1:1 ACN:H<sub>2</sub>O (0.1% TFA). The samples are then analyzed by MALDI-TOF MS (positive ion linear mode) on a Voyager-DE Pro mass spectrometer (Applied Biosystems, Foster City, CA, USA). Experiments were first done to determine the accurate amount of  $^{15}\text{N}$ -protein to be added to obtain as close as possible a 1:1 ratio between the internalized  $^{14}\text{N}$  and  $^{15}\text{N}$  internal standard proteins. This first step allows one to get accurate, robust and reproducible MS quantification results<sup>61</sup>. Each experiment was done in triplicate and independently repeated at least 3 times.

### Software developed to quantify the cellular uptake of proteins by mass spectrometry

In a previous study<sup>26</sup>, MALDI-TOF was found to be well adapted to quantify the cellular uptake of biotinylated cell-penetrating peptides. This method was based on the addition of a known and appropriate amount of the labeled peptide ( $^2\text{H}$  deuterium labeling) to the cell lysis buffer before extraction of the two peptide species ( $^2\text{H}$ -labeled and non-labeled). As in conventional internal standard experiment the quantification is achieved through integrating the peaks area of both labeled and non-labeled peptides. For proteins with higher mass than peptides, as those studied herein, we have developed a software allowing an adjustment of the ratio *r* between non-labeled ( $^{14}\text{N}$ )/labeled ( $^{15}\text{N}$ ) proteins, by comparing the experimental mass spectra profile to that obtained from the mixture of separate labeled and non-labeled protein (Supplementary software file). The validity of such an approach originates from the properties of both proteins: same sequence implying identical ionization yield, similar incorporation in

the MALDI matrix, same peak broadening due to attachment of alkaline and matrix adducts, identical detection efficiency due to the small mass difference. In addition, if experimental mass spectra of  $^{14}\text{N}$ ,  $^{15}\text{N}$  and the mixture ( $^{14}\text{N}$  and  $^{15}\text{N}$ ) from internalization are obtained under the same experimental conditions (delayed ion extraction, ion detection and laser fluence), it is possible after mass calibration to calculate a theoretical mass spectrum for a mixture with a ratio  $r$  and to compare it to the experimental mass spectrum. The software calculates, normalizes and displays the experimental and simulated mass spectra and a fast adjustment of the ratio  $r$  can be obtained. Baseline correction on the calculated mass spectrum can be also implemented to get a better adjustment simulated vs experimental. Studying the different charge states allows to estimate the accuracy of the ratio.

The non-complete labeling of the protein can be also studied by comparison of  $^{14}\text{N}$  and  $^{15}\text{N}$  mass spectra through the calculated convolution of the  $^{14}\text{N}$  ion peak (+1, +2, +3...) with a Gaussian distribution of  $^{15}\text{N}$  atoms by adjusting the mean value  $\langle 15\text{N} \rangle$  and the standard deviation ( $\sigma$ ). The convolution operates over the  $m/z$  range in which the  $^{14}\text{N}$  peak is defined. The display of the convoluted and  $^{15}\text{N}$  experimental peak allows one to deduce the value of  $\langle 15\text{N} \rangle$  and  $\sigma$ . If needed, experimental data can be integrated before processing the convolution.

Experimental MS spectra are extracted as ASCII files ( $m/z$ , intensity) from the processing software (Data Explorer) of the MALDI-TOF apparatus (DE-Pro and 4700 Proteomics Analyzer both from Applied Biosystems, Foster City, CA, USA). The software to determine the ratio  $r$  and also the labeling distribution, if any, has been written in Visual Basic (V 6.0). The simplest case is that of a complete labeling for which non-labeled and labeled peaks are superimposable and thus the  $m/z$  shift of the non-labeled peak overlay that of labeled peak using an intensity ratio (Supplementary Fig. 15). To take into account a labeling distribution, each experimental point of the unlabeled peak is shifted in  $m/z$  with a relative intensity modulated by the probability of the shift. A new peak is formed after adding all the contribution for a value  $m/z$  within a chosen window (1-5 Da depending on the protein mass).

In Supplementary Fig. 15, an example is shown with the ExtHD protein as unlabeled ( $^{14}\text{N}$ ) and labeled ( $^{15}\text{N}$ ) versions. No  $^{15}\text{N}$  distribution was necessary to fit the +1 and +2 protein peaks showing that the labeling was complete. The experimental data of  $^{15}\text{N}$  were well fitted from the  $^{14}\text{N}$  data shifted by 128 Da for both +1 (lower mass spectrum) and +2 (upper mass spectrum) charge states.

### Microarray of structurally defined HS compounds

A custom HS-microarray was designed by the company Glycan Therapeutics (<https://www.glycantherapeutics.com/services/microarray-analysis>), as described<sup>62</sup>. 52 structurally defined HS oligosaccharides were synthesized and immobilized on a microarray chip. The biotinylated ExtHD protein was incubated at 100 nM with the microarray slide for 1 hr in a Tris/phosphate buffer pH 7.50 containing 137 mM NaCl and 10% bovine serum albumin (BSA)<sup>62</sup>. After washing with the Tris/phosphate buffer, the microarray slide was treated with 10  $\mu\text{g}/\text{mL}$  Alexa-Fluor 488-labeled Avidin. The wash process was repeated before scanning fluorescence of the array slide at 488 nm.

### Isothermal titration analysis (ITC)

Experiments were done with a nano-ITC calorimeter (TA Instruments, New Castle, DE, USA). Measurements were done at 25 °C in 50 mM  $\text{NaH}_2\text{PO}_4$  pH 7.4 and various NaCl concentrations (0, 100, 200 mM). Titration of the proteins in the ITC cell was done by 25 successive injections of 10  $\mu\text{L}$  of polysaccharide (heparin, heparin-derived, CS-E, obtained from Iduron, Cheshire, UK) with 5 min injection intervals. Proteins, peptide and polysaccharides were used in different concentrations to get accurate fitted curves (between 20-80  $\mu\text{M}$  for

proteins and peptide and 40-120  $\mu\text{M}$  for GAG). Control experiments were obtained by injections of buffer into proteins and injections of GAGs into buffer. Data were analyzed with NanoAnalyze software provided by TA Instruments.

LUVs were prepared as described<sup>63</sup>. Briefly, lipid films were prepared by dissolving phospholipids into chloroform or a mixture of chloroform and methanol (2/1, vol/vol). Formation of the lipid film was achieved by evaporating solvents under nitrogen flux, then drying in a vacuum chamber for at least 1 hr. Films were then hydrated with 10 mM sodium phosphate buffer pH 7.5 and extensively vortexed at a temperature superior to the lipid phase transition temperature to obtain MLVs. To form LUVs, the MLVs were subjected to five freeze/thawing cycles. The homogeneous lipid suspension was then passed 19 times through a mini-extruder (Avanti Polar Lipids, Avanti Alabaster, AL, USA) equipped with two stacked 100 nm polycarbonate membranes at a temperature above the phase transition temperature of the phospholipid. Titrations were performed by injecting aliquots of LUVs (1 mg/mL) into the calorimeter cell containing the protein solution (0.1 mM), with 5 min intervals between injections.

### Circular dichroism (CD)

CD spectra were measured on a Jasco J-815 spectropolarimeter (JASCO Corporation, Tokyo, Japan) over the wavelength range 190–270 nm, by using a 0.1-cm path-length quartz cell (internal volume 200  $\mu\text{L}$ ) from Hellma (Muellheim, Germany). Measurements were carried out at 25 °C with a 1 nm/min scan speed and a band width of 1 nm. Proteins and oligosaccharides were dissolved in a buffer containing 20 mM sodium phosphate (pH 7.4) and 100 mM NaF. Four scans were accumulated and averaged for each sample. All spectra were corrected by subtraction of the background obtained for each protein-free mixture. Protein concentration was 10  $\mu\text{M}$  in the absence and presence of 1 molar equivalent heparin.

### NMR data collection and analysis

NMR experiments were acquired at 298 K on a Bruker Avance III 500 MHz spectrometer equipped with a 5-mm TCI cryoprobe. NMR spectra were processed and analyzed with the software programs NMRPIPE<sup>64</sup> and NMRFAM-SPARKY<sup>65</sup>. All experiments were performed with NMR samples containing  $^{15}\text{N}$  or  $^{15}\text{N},^{13}\text{C}$  uniformly labeled HD or ExtHD, previously dissolved in a buffer containing 40 mM sodium phosphate (pH 6.3), 100 mM NaCl, and 7.5%  $\text{D}_2\text{O}$ . 0.1 mM Sodium 2,2-dimethyl-2-silapentane-5-sulfonate- $d_6$  (DSS) was added in the NMR samples as an internal  $^1\text{H}$  chemical shift reference.  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts were referenced indirectly to DSS, using the absolute frequency ratios. Sequence-specific assignments of backbone  $^{15}\text{NH}$ ,  $^1\text{HN}$ ,  $^{13}\text{C}\alpha$ ,  $^1\text{H}\alpha$ ,  $^{13}\text{C}\text{O}$ , and side-chain  $^{13}\text{C}\beta$ ,  $^1\text{H}\beta$  resonances were obtained from the analysis of a series of three-dimensional triple resonance experiments (HNCA, HNCO, HN(CA)CO, HNCACB, CBCA(CO)NH, HNHA, and HBHA(CO)NH), recorded on samples containing 600  $\mu\text{M}$   $^{15}\text{N},^{13}\text{C}$  ExtHD. Chemical shift perturbation experiments (CSP) were performed by collecting  $^1\text{H}-^{15}\text{N}$  HSQC spectra on  $^{15}\text{N}$  uniformly labeled HD or ExtHD (100-150  $\mu\text{M}$ ) in the absence and presence of increasing amounts of heparin fragments (obtained from Iduron, Cheshire, UK) or structurally defined CS-E dp6 that was chemically synthesized<sup>66</sup>, previously dissolved in the NMR buffer. CSP values were calculated for each residue from the differences in  $^1\text{HN}$  ( $\Delta\delta_{\text{H}^1}$ ) and  $^{15}\text{N}$  ( $\Delta\delta_{\text{N}^{15}}$ ) chemical shifts between the free and the bound states using the relation:

$$\text{CSP} = \sqrt{(\Delta\delta_{\text{H}^1})^2 + (0.1\Delta\delta_{\text{N}^{15}})^2}$$
 Apparent thermodynamic affinities ( $K_d^{\text{app}}$ ) at the residue level were calculated for residues in the fast exchange regime from a nonlinear least squares curve fitting of CSPs upon oligosaccharide addition using in-house Python scripts<sup>67</sup>. Steady state  $^{15}\text{N}-\{^1\text{H}\}$  NOE values in the absence and presence of heparin dp8 (Iduron, Cheshire, UK) were determined as the ratio of peak heights in

$^{15}\text{N}$ - $^1\text{H}$  spectra collected with and without proton saturation during the recycle delay.  $^1\text{H}$  saturation was achieved by applying a train of 120° pulses spaced at 5 ms interval for a period of 4.5 s. Uncertainty in the  $^{15}\text{N}$ - $^1\text{H}$  NOE values were calculated from the root mean square baseline noise of both spectra estimated with NMRPIPE.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Additional data that support this study are available from the corresponding authors upon request. Source data are provided with this paper.

### Code availability

The software developed to quantify the cellular uptake of proteins by mass spectrometry is available in the Supplementary Software file and requires Visual Basic to run it.

### References

- Dupont, E., Prochiantz, A. & Joliot, A. Identification of a signal peptide for unconventional secretion. *J. Biol. Chem.* **282**, 8994–9000 (2007).
- Lee, E. J. et al. Global analysis of intercellular homeodomain protein transfer. *Cell Rep.* **28**, 712–722.e3 (2019).
- Spatazza, J. et al. Homeoprotein signaling in development, health, and disease: a shaking of dogmas offers challenges and promises from bench to bed. *Pharmacol. Rev.* **65**, 90–104 (2013).
- Carlier, L. et al. Investigation of homeodomain membrane translocation properties: insights from the structure determination of engrailed-2 homeodomain in aqueous and membrane-mimetic environments. *Biophys. J.* **105**, 667–678 (2013).
- Derossi, D., Joliot, A. H., Chassaing, G. & Prochiantz, A. The third helix of the Antennapedia homeodomain translocates through biological membranes. *J. Biol. Chem.* **269**, 10444–10450 (1994).
- Dupont, E., Prochiantz, A. & Joliot, A. Penetratin story: an overview. *Methods Mol. Biol.* **1324**, 29–37 (2015).
- Walrant, A., Cardon, S., Burlina, F. & Sagan, S. Membrane crossing and membranotropic activity of cell-penetrating peptides: dangerous liaisons? *Acc. Chem. Res.* **50**, 2968–2975 (2017).
- Jiao, C.-Y. et al. Translocation and endocytosis for cell-penetrating peptide internalization. *J. Biol. Chem.* **284**, 33957–33965 (2009).
- Perez, F. et al. Antennapedia homeobox as a signal for the cellular internalization and nuclear addressing of a small exogenous peptide. *J. Cell. Sci.* **102**(Pt. 4), 717–722 (1992).
- Amblard, I. et al. Bidirectional transfer of homeoprotein EN2 across the plasma membrane requires PIP2. *J. Cell Sci.* **133**, jcs244327 (2020).
- Fuchs, S. M. & Raines, R. T. Pathway for polyarginine entry into mammalian cells. *Biochemistry* **43**, 2438–2444 (2004).
- Ziegler, A. Thermodynamic studies and binding mechanisms of cell-penetrating peptides with lipids and glycosaminoglycans. *Adv. Drug Deliv. Rev.* **60**, 580–597 (2008).
- Bechara, C. et al. Tryptophan within basic peptide sequences triggers glycosaminoglycan-dependent endocytosis. *FASEB J.* **27**, 738–749 (2013).
- Bechara, C. et al. Massive glycosaminoglycan-dependent entry of Trp-containing cell-penetrating peptides induced by exogenous sphingomyelinase or cholesterol depletion. *Cell. Mol. Life Sci.* **72**, 809–820 (2015).
- Nakase, I., Konishi, Y., Ueda, M., Saji, H. & Futaki, S. Accumulation of arginine-rich cell-penetrating peptides in tumors and the potential for anticancer drug delivery in vivo. *J. Control. Release* **159**, 181–188 (2012).
- Xu, D. & Esko, J. D. Demystifying heparan sulfate-protein interactions. *Annu. Rev. Biochem.* **83**, 129–157 (2014).
- Smith, P. D., Coulson-Thomas, V. J., Foscari, S., Kwok, J. C. F. & Fawcett, J. W. 'GAG-ing with the neuron': the role of glycosaminoglycan patterning in the central nervous system. *Exp. Neurol.* **274**, 100–114 (2015).
- Mikami, T. & Kitagawa, H. Chondroitin sulfate glycosaminoglycans regulate distinct cell surface receptor-mediated neuronal functions. *Trends Glycosci. Glycotechnol.* **33**, E11–E16 (2021).
- Vallet, S. D., Berthollier, C. & Ricard-Blum, S. The glycosaminoglycan interactome 2.0. *Am. J. Physiol. Cell Physiol.* **322**, C1271–C1278 (2022).
- Deepa, S. S. et al. Composition of perineuronal net extracellular matrix in rat brain: a different disaccharide composition for the net-associated proteoglycans. *J. Biol. Chem.* **281**, 17789–17800 (2006).
- Carulli, D. et al. Composition of perineuronal nets in the adult rat cerebellum and the cellular origin of their components. *J. Comp. Neurol.* **494**, 559–577 (2006).
- Beurdeley, M. et al. Otx2 binding to perineuronal nets persistently regulates plasticity in the mature visual cortex. *J. Neurosci.* **32**, 9429–9437 (2012).
- Spatazza, J. et al. Choroid-plexus-derived Otx2 homeoprotein constrains adult cortical plasticity. *Cell Rep.* **3**, 1815–1823 (2013).
- Joyner, A. L. Engrailed, Wnt and Pax genes regulate midbrain–hindbrain development. *Trends Genet.* **12**, 15–20 (1996).
- Rampon, C. et al. Control of brain patterning by Engrailed paracrine transfer: a new function of the Pbx interaction domain. *Development* **142**, 1840–1849 (2015).
- Burlina, F., Sagan, S., Bolbach, G. & Chassaing, G. Quantification of the cellular uptake of cell-penetrating peptides by MALDI-TOF mass spectrometry. *Angew. Chem. Int. Ed. Engl.* **44**, 4244–4247 (2005).
- Illien, F. et al. Quantitative fluorescence spectroscopy and flow cytometry analyses of cell-penetrating peptides internalization pathways: optimization, pitfalls, comparison with mass spectrometry quantification. *Sci. Rep.* **6**, 36938 (2016).
- Esko, J. D., Stewart, T. E. & Taylor, W. H. Animal cell mutants defective in glycosaminoglycan biosynthesis. *Proc. Natl Acad. Sci. USA* **82**, 3197–3201 (1985).
- Piva, M. B. R. et al. Glycosaminoglycans affect heparanase location in CHO cell lines. *Glycobiology* **25**, 976–983 (2015).
- Huynh, M. B. et al. Glycosaminoglycans from Alzheimer's disease hippocampus have altered capacities to bind and regulate growth factors activities and to bind tau. *PLoS ONE* **14**, e0209573 (2019).
- Casu, B. in *Chemistry and Biology of Heparin and Heparan Sulfate* (eds Garg, H. G., Linhardt, R. J. & Hales, C. A.) 1–28 (Elsevier Science, 2005).
- Karlsson, R. et al. Dissecting structure-function of 3-O-sulfated heparin and engineered heparan sulfates. *Sci. Adv.* **7**, eabl6026 (2021).
- Maccarana, M., Sakura, Y., Tawada, A., Yoshida, K. & Lindahl, U. Domain structure of heparan sulfates from bovine organs. *J. Biol. Chem.* **271**, 17804–17810 (1996).
- Turnbull, J. E. & Gallagher, J. T. Distribution of iduronate 2-sulphate residues in heparan sulphate. Evidence for an ordered polymeric structure. *Biochem. J.* **273**(Pt. 3), 553–559 (1991).
- Lindahl, U., Kusche-Gullberg, M. & Kjellén, L. Regulated diversity of heparan sulfate. *J. Biol. Chem.* **273**, 24979–24982 (1998).
- Rullo, A., Qian, J. & Nitz, M. Peptide-glycosaminoglycan cluster formation involving cell penetrating peptides. *Biopolymers* **95**, 722–731 (2011).
- Joliot, A., Pernelle, C., Deagostini-Bazin, H. & Prochiantz, A. Antennapedia homeobox peptide regulates neural morphogenesis. *Proc. Natl Acad. Sci. USA* **88**, 1864–1868 (1991).



38. Derossi, D. et al. Cell internalization of the third helix of the Antennapedia homeodomain is receptor-independent. *J. Biol. Chem.* **271**, 18188–18193 (1996).
39. Wishart, D. S. et al. <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **6**, 135–140 (1995).
40. Augustyński, R., Balayssac, S., Ferrage, F., Bodenhausen, G. & Lequin, O. <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N resonance assignment of a 114-residue fragment of Engrailed 2 homeoprotein, a partially disordered protein. *Biomol. NMR Assign.* **5**, 229–231 (2011).
41. Khan, S. N. et al. Distribution of pico- and nanosecond motions in disordered proteins from nuclear spin relaxation. *Biophys. J.* **109**, 988–999 (2015).
42. Bechara, C. & Sagan, S. Cell-penetrating peptides: 20 years later, where do we stand? *FEBS Lett.* **587**, 1693–1702 (2013).
43. Yang, H. et al. Chondroitin sulfate as a molecular portal that preferentially mediates the apoptotic killing of tumor cells by penetratin-directed mitochondria-disrupting peptides. *J. Biol. Chem.* **285**, 25666–25676 (2010).
44. Amand, H. L. et al. Cell surface binding and uptake of arginine- and lysine-rich penetratin peptides in absence and presence of proteoglycans. *Biochim. Biophys. Acta* **1818**, 2669–2678 (2012).
45. Sagan, S. et al. Homeoproteins and homeoprotein-derived peptides: going in and out. *Curr. Pharm. Des.* **19**, 2851–2862 (2013).
46. Cardin, A. D. & Weintraub, H. J. Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* **9**, 21–32 (1989).
47. Rudd, T. R., Preston, M. D. & Yates, E. A. The nature of the conserved basic amino acid sequences found among 437 heparin binding proteins determined by network analysis. *Mol. Biosyst.* **13**, 852–865 (2017).
48. Kato, M., Wang, H., Bernfield, M., Gallagher, J. T. & Turnbull, J. E. Cell surface syndecan-1 on distinct cell types differs in fine structure and ligand binding of its heparan sulfate chains. *J. Biol. Chem.* **269**, 18881–18890 (1994).
49. Joliot, A. et al. Identification of a signal sequence necessary for the unconventional secretion of Engrailed homeoprotein. *Curr. Biol.* **8**, 856–863 (1998).
50. Joliot, A. H., Triller, A., Volovitch, M., Pernelle, C. & Prochiantz, A. alpha-2,8-Polysialic acid is the neuronal surface receptor of antennapedia homeobox peptide. *N. Biol.* **3**, 1121–1134 (1991).
51. Belting, M. Heparan sulfate proteoglycan as a plasma membrane carrier. *Trends Biochem. Sci.* **28**, 145–151 (2003).
52. Jiao, C.-Y. et al. Exploiting benzophenone photoreactivity to probe the phospholipid environment and insertion depth of the cell-penetrating peptide penetratin in model membranes. *Angew. Chem. Int. Ed. Engl.* **56**, 8226–8230 (2017).
53. Di Nardo, A. A., Joliot, A. & Prochiantz, A. Homeoprotein transduction in neurodevelopment and physiopathology. *Sci. Adv.* **6**, eabc6374 (2020).
54. Miyata, S., Nadanaka, S., Igarashi, M. & Kitagawa, H. Structural variation of chondroitin sulfate chains contributes to the molecular heterogeneity of perineuronal nets. *Front. Integr. Neurosci.* **12**, 3 (2018).
55. Brunet, I. et al. The transcription factor Engrailed-2 guides retinal axons. *Nature* **438**, 94–98 (2005).
56. Kim, N. et al. Regulation of retinal axon growth by secreted Vax1 homeodomain protein. *Elife* **3**, e02671 (2014).
57. Stettler, O. et al. Engrailed homeoprotein recruits the adenosine A1 receptor to potentiate ephrin A5 function in retinal growth cones. *Development* **139**, 215–224 (2012).
58. Wurst, W. & Bally-Cuif, L. Neural plate patterning: upstream and downstream of the isthmus organizer. *Nat. Rev. Neurosci.* **2**, 99–108 (2001).
59. Wizenmann, A. et al. Extracellular Engrailed participates in the topographic guidance of retinal axons in vivo. *Neuron* **64**, 355–366 (2009).
60. Marley, J., Lu, M. & Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *J. Biomol. NMR* **20**, 71–75 (2001).
61. Burlina, F., Sagan, S., Bolbach, G. & Chassaing, G. A direct approach to quantification of the cellular uptake of cell-penetrating peptides using MALDI-TOF mass spectrometry. *Nat. Protoc.* **1**, 200–205 (2006).
62. Yang, J. et al. Construction and characterisation of a heparan sulphate heptasaccharide microarray. *Chem. Commun.* **53**, 1743–1746 (2017).
63. Walrant, A. et al. Different membrane behaviour and cellular uptake of three basic arginine-rich peptides. *Biochim. Biophys. Acta* **1808**, 382–393 (2011).
64. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
65. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).
66. Jacquinet, J.-C., Lopin-Bon, C. & Vibert, A. From polymer to size-defined oligomers: a highly divergent and stereo-controlled construction of chondroitin sulfate A, C, D, E, K, L, and M oligomers from a single precursor: part 2. *Chemistry* **15**, 9579–9595 (2009).
67. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).
68. van der Steen, S. C. H. A. et al. Prognostic significance of highly sulfated chondroitin sulfates in ovarian cancer defined by the single chain antibody GD3A11. *Gynecol. Oncol.* **140**, 527–536 (2016).

## Acknowledgements

This research has received fundings from the Agence Nationale de la Recherche (grants No. ANR-BLAN2016-CROSS, A.J., S.S., L.C., A.W., F.B., B.H., F.I., D.R., Y.P.H., O.L., S.C., C.L.-B.; ANR-20-CE44-0018-GLYCO-TARGET, S.S., A.W., F.I., D.R., L.C., O.L., Y.P.H., C.L.-B., A.J.).

## Author contributions

L.C. and S.S. designed research; S.C., G.B., Y.P.H., C.L.-B., J.-C.J., F.I., A.W., D.R., B.H., L.M., A.J., L.C., and S.S. performed research; S.C., G.B., Y.P.H., F.B., O.L., L.C., and S.S. analyzed data. G.B. C.L.-B. and J.-C.J. contributed reagents/analytical tools; L.C. and S.S. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-37757-6>.

**Correspondence** and requests for materials should be addressed to Ludovic Carlier or Sandrine Sagan.

**Peer review information** *Nature Communications* thanks Romain Vives, Marco Giampà and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023