Volume 295, number 1,2,3, 1–2 FEBS 10536 December 1991

Hypothesis

$(HX)_n$ repeats: a pH-controlled protein–protein interaction motif of eukaryotic transcription factors?

Ralf Janknecht[1], Chris Sander[2] and Olaf Pongs[3]

[1]*Institut für Molekularbiologie, Medizinische Hochschule Hannover, Konstanty-Gutschow-Strasse 8, D-3000 Hannover, Germany,* [2]*European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Germany* and [3]*Lehrstuhl für Biochemie, Ruhr-Universität Bochum, Universitätsstrasse 150, D-4630 Bochum, Germany*

Received 31 October 1991

A characteristic sequence repeat of type His-X, repeated several times in a row, is present in several eukaryotic transcription factors, e.g. HPHAHPHP in *paired* protein. Detailed molecular modelling and database searches lead to the suggestion that $(HX)_n$ repeats can mediate interaction between transcription factors in a pH-controlled fashion.

Gene regulation; Transcription factor; Dimerization; Zn-binding; Sequence analysis

$(HX)_n$ repeats were first observed in the His-Pro-rich PRD domain of the *Drosophila* paired protein [1] and aroused our interest as part of the *ets*-oncoprotein related *Drosophila 74E* gene products [2]. A subsequent systematic search for significant $(HX)_n$ repeats in the databases of known protein sequences yielded mostly further *Drosophila* transcription factors (Fig. 1): Bicoid [3], Deformed [4], Odd-Skipped [5], Daughterless [6], Cf1a [7], and E75A [8]. Other proteins found with this pattern were a chicken transcription factor (Chox-1.4 [9]), the rabbit T-cell receptor β-chain precursor (TCR [10]), the tobacco acetolactate synthase precursor (ALS-SuRA [11]), and the Herpes latency-related proteins 1 and 2 (Lrp1 and Lrp2 [12]). $(HX)_n$ repeats in proteins with polyhistidine stretches or with unusually high histidine content were not regarded as significant. Analysis of the $(HX)_n$ repeats shows that X is preferentially P (Pro), S (Ser), H (His), A (Ala) or T (Thr) and that the multiplicity of the repeat ranges from n=2 to n=11 (Fig. 1). Note that Daughterless has a nearly perfect $(HX)_{12}$ repeat, with Leu in place of His at positions 233, 235, 239 and 245.

The predominant occurrence of $(HX)_n$ repeats in *Drosophila* transcription factors hints at a common functional interpretation of this sequence motif. We suggest the hypothesis that $(HX)_n$ repeats mediate protein–protein interactions, which are known to be a necessary prerequisite for the function of transcription factors, in a pH-dependent fashion. Structurally, $(HX)_n$ repeats in an extended conformation can present linear arrays of His side chains, which can associate in pairs by coordinating metal ions like Zn^{2+} or Cu^{2+} (Fig. 2). This structure could be termed a 'histidine–metal zipper'. Consistent with this structural model is the observation in known three-dimensional structures that Pro often occurs adjacent to His in metal binding sites of enzymes [13].

Interacting $(HX)_n$ repeats could guide the formation of dimers or of networks of transcription factors. An example may be the *74E* gene products and E75A, which have been shown to bind to early and late puffs on *Drosophila* polytene chromosomes, after induction of third larval instar salivary glands with 20-hydroxyecdysone (for a review see [14]). It is tempting to speculate that this coordinated binding involves interaction via the $(HX)_n$ repeats.

An intriguing property of protein–protein interactions mediated by $(HX)_n$ would be sensitivity to pH. As the pK_a of a histidine side chain is in the range of physiological pH values, a small shift in intracellular pH could have a dramatic effect on the binding capacity of $(HX)_n$ repeats. By this mechanism extracellular inductors like growth factors, which can increase the intracellular pH by 0.1–0.3 units [15], might influence the association state of an entire network of transcription factors.

At the present stage of limited experimental evidence we cannot exclude the possibility of other functional roles of $(HX)_n$ repeats: for example, direct binding to DNA or RNA through H-bonds from His side chains to phosphate groups; or, interaction with the acidic do-

Correspondence address: C. Sander, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Germany. Fax: (49) (6221) 38 7306.

74E:	658	qqqqHPHSql	ngpHPHSHPH	SHPHSHPHAg	qHTHStiaaa	697
Paired:	544	gHAHSHHgHP	HApHPHAHPH	PqyagaHPHY	pppsssaHFm	583
Bicoid:	10	fyHHplpHTH	THPHPHSHPH	PHSHPHPHHq	HPqlqlppqf	49
Deformed:	73	mtgHPHSmHP	admvsdymaH	HHNpHSHSHS	HTHSlpHHHS	112
	113	nsaisgHHqa	saggyssnya	natppsHPHS	HPHAHPHQsl	152
Odd-Skipped:	137	naaiyqqqqq	qqqHPHHHHh	HGHPhHPHPH	PHHvrpypag	176
Daughterless:	219	HHslnHTpHA	HSHTlplpHA	lpHGHTlpHP	HHsqqnspav	258
Cf1a:	298	HHggyHPHHd	mHGspmgtHS	HSHSppmlsp	qnmqssavaa	237
E75A:	156	HHHPqqeHQp	qqqqqqHHlq	hHPHPHVmyp	HGyqqanlHH	195
Chox-1.4:	26	HSgsagssas	yHPhHPHPHA	ppppppppppp	HLHAaHPgpa	65
TCR:	48	sHTHRHSyll	HPHTHVctHT	HTctHTHIHA	stHVciHTHT	87
ALS-SuRA:	33	fpfpHHpHKt	tppplHLtHT	HIHIHSqrrr	ftisnvistn	72
Lrp1:	31	HPHSHApplp	rtptpsHPHS	rapplprapt	ptHPHSHApp	70
Lrp2:	13	HPHSHApplp	rtptpaHPHS	HApplprtpt	ptHPHSHApp	52

Fig. 1. Partial sequences of proteins, primarily transcription factors, containing significant $(HX)_n$ repeats. The sequences are not strictly aligned HX sequences are in capital letters. Protein codes are explained in the text.

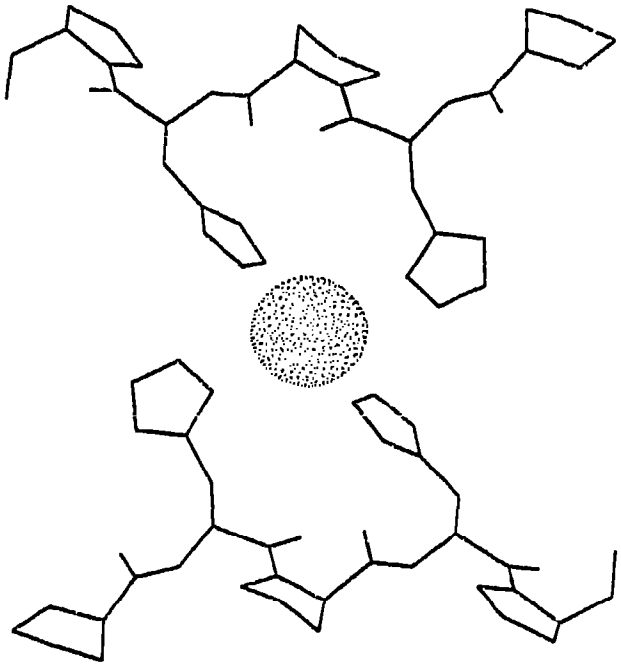

Fig. 2. Approximate molecular model of two PHPHP units interacting via a bound metal ion. Only two HX units arranged in antiparallel are shown here. Longer, zipper-like repeating structures can be generated without violation of stereochemistry by simple lateral displacement in the direction of the polypeptide backbone. The association of two $(HX)_n$ repeats is likely to be pH-dependent in the relatively narrow range of intracellular pH values available to physiological control.

mains of other transcription factors, which are thought to pass stimulatory signals to the transcription machinery by interacting with TFIID or TFIIB [16].

We hope that our hypothesis of pH-dependent protein–protein interaction between transcription factors mediated by $(HX)_n$ repeats will stimulate specific experiments.

REFERENCES

[1] Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. and Noll, M. (1986) Cell 47, 735–746.
[2] Janknecht, R., Taube, W., Lüdecke, H.-J. and Pongs, O. (198 Nucleic Acids Res. 17, 4455–4464.
[3] Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M. and Nüsslein-Volhard, C. (1988) EMBO J. 1749–1756.
[4] Regulski, M., McGinnis, N., Chadwick, R. and McGinnis, W. (1987) EMBO J. 6, 767–777.
[5] Coulter, D.E., Swaykus, E.A., Beran-Koehn, M.A., Goldberg D., Wieschaus, E. and Schedl, P. (1990) EMBO J. 8, 3795–380
[6] Cronmiller, C., Schedl, P. and Cline, T.W. (1988) Genes Dev. 1666–1676.
[7] Johnson, W.A. and Hirsh, J. (1990) Nature 343, 467–470.
[8] Segraves, W.A. and Hogness, D.S. (1990) Genes Dev. 4, 204–21
[9] Sasaki, H., Yokoyama, E. and Kuroiwa, A. (1990) Nucleic Acids Res. 18, 1739–1747.
[10] Angiolillo, A.L., Lamoyi, E., Bernstein, K.E. and Mage, R.G (1985) Proc. Natl. Acad. Sci. USA 82, 4498–4502.
[11] Lee, K.Y., Townsend, J., Tepperman, J., Black, M., Chui, C.F Mazur, B., Dunsmuir, P. and Bedbrook, J. (1988) EMBO J. 1241–1248.
[12] Wechsler, S.L., Nesburn, A.B., Zwaagstra, J. and Ghiasi, H (1989) Virology 168, 168–172.
[13] Chakrabarti, P. (1990) Prot. Engineering 4, 57–63.
[14] Thummel, C.S. (1990) BioEssays 12, 561–568.
[15] Moolenaar, W.H. (1986) Ann. Rev. Physiol. 48, 363–376.
[16] Sharp, P.A. (1991) Nature 351, 16–18.