Available online at www.sciencedirect.com

SCIENCE DIRECT

Genomics 85 (2005) 551–556

GENOMICS

www.elsevier.com/locate/ygeno

Mammalian SP/KLF transcription factors: Bring in the family

Guntram Suske[a], Elspeth Bruford[b], Sjaak Philipsen[c,*]

[a]*Institut fuer Molekularbiologie und Tumorforschung, Philipps-Universitaet Marburg, Emil-Mannkopff-Strasse 2, D-35037 Marburg, Germany*

[b]*HUGO Gene Nomenclature Committee and The Galton Laboratory, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK*

[c]*Erasmus MC Department of Cell Biology, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands*

Received 5 January 2005; accepted 18 January 2005

Abstract

The advent of the genome projects has provided new avenues to explore the question of how DNA sequence information is used appropriately by mammalian cells. Regulation of transcription is not the only, but is certainly a very important, mechanism involved in this process. We can now identify all the genes encoding transcription factors belonging to a certain class and study their biological functions in unprecedented detail through the use of an array of biomolecular tools. It is important to use rigorous and uniform definitions for the classification of transcription factors, because this helps us to comprehend the functions of transcription factor families in biological networks. Here, we propose an unambiguous nomenclature for the members of the Specificity Protein/Krüppel-like Factor (SP/KLF) transcription factor family.

Keywords: Transcription factors; Zinc fingers; Krüppel-like; Specificity proteins; Buttonhead

GC and GT boxes (5′-GGGGCGGGG-3′ and 5-GGTGTGGGG-3′) are recurring motifs in promoters and more distal regulatory elements of mammalian genes. A protein interacting with these motifs was first identified in the 21-bp repeats of the SV40 early promoter [1] and termed SP1, for Specificity Protein 1. Molecular cloning revealed that the DNA binding domain of SP1 is composed of three abutting zinc fingers of the classical Cys_2–His_2 type [2]. Closely related, but distinct, factors were later identified and called SP2, SP3, and SP4 [3,4]. The linkage of each factor to a *HOX* gene cluster further emphasized their evolutionary relationships (Supplemental Fig. 1; [5]), as did the discovery of SP-related factors in *Drosophila* [6,7]. SP-related factors are also found in relatively simple multicellular organisms such as the nematode *Caenorhabditis elegans* but not in unicellular organisms such as baker's yeast, *Saccharomyces cerevisiae*.

The *SP1–4* genes are closely linked to the genes encoding the other SP factors (*SP5–9*), usually in pairwise combinations (Supplemental Fig. 1; [8,9]). A characteristic hallmark of SP factors is the presence of the Buttonhead (BTD) box CXCPXC, just N-terminal to the zinc fingers (Fig. 1; [10]). The function of the BTD box is unknown, but the fact that it is also present in *Drosophila* and *C. elegans* SP factors suggests an important physiological role. Another feature of most SP factors is the presence of a conserved amino acid stretch, the so-called SP box, located close to the N-terminus (Supplemental Fig. 2; [10]).

The first mammalian Krüppel-like factor was cloned from erythroid cells and therefore called erythroid Krüppel-like factor or EKLF (KLF1; [11]). This was soon followed by the discovery of a number of related factors, and the KLF nomenclature was first introduced by Turner and Crossley following a proposal from the HUGO Gene Nomenclature Committee [12]. The absence of the BTD box is the most distinguishing feature between the SP and the KLF sub-

* Corresponding author. Fax: +31 10 4089468.
E-mail address: j.philipsen@erasmusmc.nl (S. Philipsen).

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.01.005

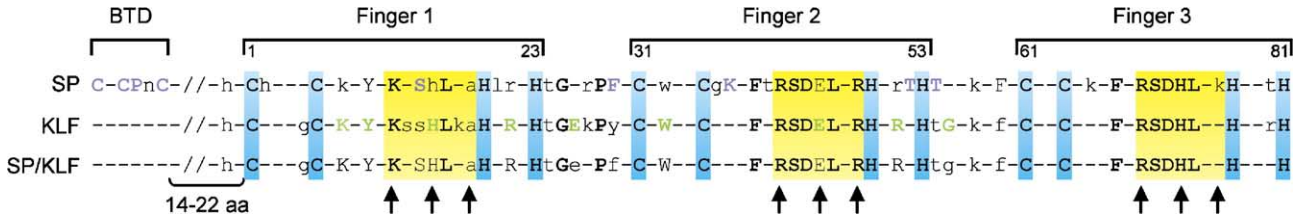

Fig. 1. Characteristic hallmarks of SP/KLF family members. Consensus sequences for the zinc finger domains of all the SP and KLF factors in human (25 factors), *Drosophila* (9 factors), and *C. elegans* (6 factors) are shown for the SP factors, the KLF factors, and the entire family. All the DNA binding domains are 81 aa in length, with the exception of those of Ce-Y40B1A.4 (Finger 1: CXXC instead of CXXXXC), Ce-T22C8.5 (Finger 3: HXXXXH instead of HXXXH), and D-BTD (Finger 2: CXXC instead of CXXXXC). The highly conserved BTD box N-terminal to the zinc fingers is a unique feature of the SP factors. Bold capital letters indicate residues that are 100% conserved between all family members (black), between all KLF factors (green), or between all SP factors (purple). Capital letters indicate >90% conservation, lowercase letters >75% conservation. Blue bars indicate the cysteine and histidine residues involved in zinc coordination, yellow boxes residues thought to contact DNA. The arrows point at the residues that probably determine the recognition specificity of the fingers through specific contacts with DNA bases.

families (Fig. 1). Based on this criterion, KLFs are also found in *Drosophila* and *C. elegans*. Outside the zinc finger domain, there is usually very little homology between SP/KLFs. This makes the analysis of functional domains particularly challenging.

Three fingers keep the family together

The array of three zinc fingers is the most outstanding feature of the SP/KLF family members. Without exception, the finger domain of mammalian SP/KLFs consists of 81 amino acids (Fig. 1). This strongly suggests that the fingers act as a single unit, with heavy constraints not only on the amino acids of the Cys_2–His_2 units, but also on those of the interfinger domains. Indeed, the interfinger domains are highly conserved (Fig. 1). By comparison to the crystal structure of the Zif268–DNA complex [13], it can be inferred that the SP/KLF finger domain interacts with the G-rich strand of the 9-bp recognition sequence in a 3′-to-5′ fashion, e.g., finger 3 with the first three nucleotides, finger 2 with the three nucleotides in the middle, and finger 1 with the most 3′ nucleotides. The residues thought to make contact with the DNA and conferring specific base recognition are among the most conserved parts of the finger domain (Fig. 1).

Don't call me names

The discovery of novel genes and their corresponding proteins is one of the most exciting aspects of molecular biology. Naturally, investigators have been clever in creating names for their newly found factors, loosely guided by field-specific gentlemen's agreements on nomenclature. Owing to the complexity of the genomes of higher vertebrates, many genes have been discovered a number of times and given different names. These names are often based on the experimental systems used, resulting in a sometimes confusing array of different names for a single factor. The mammalian SP/KLF family has also suffered from inadequate nomenclature (Table 1). With such highly conserved factors, one might draw inspiration from invertebrate model organisms such as *Drosophila* to resolve this issue. However, the orthologous relationships between mammalian and *Drosophila* factors are not easily established, even when functional assays are used. For instance, transgenic rescue of the *Drosophila* btd mutant by mouse Sp1 and Sp8 is incomplete. Yet, Sp8 was dubbed the mouse Btd orthologue because it rescued the phenotype slightly better [14]. However, phylogenetic analysis does not support the notion that Sp8 is the mouse Btd homologue. By this analysis SP8 is rather more closely related to D-SP1 than to BTD (Fig. 2). *Drosophila* appears to have only four SP factors, as opposed to nine in mammals. The orthologous relationships between mammalian and *Drosophila* SP factors are therefore not easily described on a one-to-one basis. Thus, for reasons of clarity, a straightforward nomenclature for the mouse and human SP/KLF factors, independent of the nomenclature used for invertebrate genes, is highly desirable. Transgenic rescue experiments, such as those described by [14], can then be judged on their own merits in the context of the orthologous relationships between the factors.

An unambiguous nomenclature for mouse and human SP/KLF transcription factors

The availability of near-complete genome sequences of mouse and human greatly facilitates the unambiguous assignment of names to all the members of the SP/KLF family. We propose to base the subgroup of mammalian SP factors on the presence of the BTD domain just N-terminal to the zinc fingers. The remaining factors are placed in the KLF subgroup, characterized by the presence of the highly conserved 81-amino-acid DNA binding domain. This subdivision is supported by phylogenetic analysis based on the BTD/zinc finger domains (Fig. 2). The proposed