Seminars in Cell & Developmental Biology 24 (2013) 679–686

Contents lists available at ScienceDirect

Seminars in Cell & Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Review

Revisiting the role of GCNF in embryonic development

Qin Wang, Austin J. Cooney*

Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA

ARTICLE INFO

Article history:
Available online 9 September 2013

Keywords:
GCNF
Nuclear receptor
Embryonic development
Transcriptional repressor
Embryonic stem (ES) cells
Oct4

ABSTRACT

GCNF (NR6A1) is essential for embryonic development. GCNF belongs to the nuclear receptor (NR) gene family, it is distantly related to other NRs and is the only member of subfamily 6. As the ligand for GCNF has not been identified, GCNF is designated an orphan nuclear receptor. GCNF has been found to be a transcriptional repressor, through specific binding to DR0 response elements, which is found in the Oct4 proximal promoter for example. GCNF is expressed widely in early mouse embryos, and later in the developing nervous system. GCNF knockout mouse embryos die around E10.5. GCNF is required for the restriction of Oct4 expression to primordial germ cells after gastrulation. GCNF is expressed in ES/EC cells and during their differentiation, and has been reported to be required for pluripotency gene repression during retinoic acid (RA)-induced mES cell differentiation. GCNF can interact with DNA methylation proteins, and is suggested to recruit DNA methylation complexes to repress and silence Oct4 expression. Nuclear receptor regulation in embryonic development is a complex process, as different nuclear receptors have overlapping and distinct functions. In-depth exploration of GCNF function and mechanism of action will help to comprehensively understand the nuclear receptor regulation in embryonic development.

Contents

1. Introduction

GCNF (NR6A1) is a member of nuclear receptor (NR) gene family, containing a DNA-binding domain (DBD) of NRs and a conserved ligand-binding domain (LBD) sequence characteristic of NRs. Among 48 human nuclear receptors, GCNF is the only member in the subfamily 6. GCNF is an orphan NR, because to date its ligand has not been identified. GCNF is an essential factor for normal embryonic development. GCNF knockout mice die around E10.5. The conserved DBD has been shown to be functional and binds specifically to DR0 elements (direct repeat elements with 0 spacing) to regulate genes that contain these element sequences. GCNF has been shown to be a transcriptional repressor, which recruits various corepressor complexes to repress and silence gene transcription.

2. GCNF structure and homology

GCNF was first cloned, using a low stringency hybridization strategy, from a mouse heart cDNA library in 1994, which encoded 495 amino acids [1]. Transcript expression analysis showed that this gene was most highly expressed in germ cells. Therefore it was named GCNF (germ cell nuclear receptor). Also, this gene was reported to be cloned using NR DBD degenerate primers strategy from a mouse testis cDNA library in 1995 [2]. The deduced amino acid sequence was most similar to mouse RXRs with 32–34% identity. The DBD was most similar to that of mouse RXR and RARγ, and

* Corresponding author. Tel.: +1 713 798 6250; fax: +1 713 790 1275.
E-mail address: acooney@bcm.edu (A.J. Cooney).

1084-9521/$ – see front matter © 2013 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.semcdb.2013.08.003

Table 1
Summary of mGCNF and hGCNF genes.

Chrosome	mRNA (nt)	exons	CDS (nt)	Product	Base pairs	Amino acids
Chr 2 (38.72-38.93 Mb)	6185	11	245-1732	mGCNF Isoform 1	1488	495
Chr 9 (127.28-17.53 Mb)	1903	10	179-1606	hGCNF Isoform 2	1428	475

LBD was most similar to that of mouse COUP-TFII. This gene was named RTR (retinoid receptor-related testis-associated receptor). Moreover, this gene was reported to be cloned by hybridization with a RARα genomic DNA fragment containing two DBD exons from a RA (retinoic acid) induced mouse embryonic carcinoma cell neuronal derivatives cDNA library in 1997 [3]. Transcript expression analysis showed that this gene was mainly expressed in the RA-induced neuronal development and differentiation, so it was named NCNF (neuronal cell nuclear factor). All three cloned mouse cDNAs originated from the same gene [3]. *Xenopus laevis* GCNF (xGCNF) was isolated from a frog embryonic neurula cDNA library in 1996, which encoded 435 amino acids. The N terminal domain (NTD) contained only 13 amino acids, which is much shorter than that of mGCNF and suggests it has no function, limited function or species specific functions. Its DBD was 100% identical to that of mGCNF. The whole protein except the NTD was 84% identical to that of mGCNF. Human GCNF was reported to be isolated from a human embryonic carcinoma cell line in 1996 [4] and human testis cDNA libraries in 1997 and 1998 [5–8]. *Drosophila melanogaster* DmHR4 and *Caenorhabditis elegans* CeNHR-91 are paralogs of GCNF [9,10]. Blast search shows that DmHR4 shares similarity with mGCNF: amino acids 66–155 of mGCNF with 63% identity, 76% positivity and 1% gap; amino acids 296–347 of mGCNF with 33% identity, 69% positivity and 0 gap; amino acids 383–474 of mGCNF with 24% identity, 48% positivity and 13% gap.

Like other NR members, GCNF has a variable NTD, a DBD with two C4-type zinc fingers, a hinge domain and a putative LBD predicted to comprise 12 helices (H1 to H12). Unlike most NR members, GCNF does not have a conserved AF2 (activation function 2) domain at the C terminal end of the LBD [7]. The UCSC Genome Browser shows that mouse GCNF is located on chromosome 2 (38.72–38.93 Mb). The NCBI RefSeq database shows that mouse GCNF has three transcript variants differing in 5′ UTR and coding sequence, resulting in three protein isoforms, 495 amino acids, 438 amino acids and 437 amino acids, having distinct N terminal domains. We list the 495-amino acid isoform, which is most used in the scientific community, in Table 1 and Fig. 1. For human GCNF, the UCSC Genome Browser shows that it is located on chromosome 9 (127.28–127.53 Mb). The NCBI RefSeq database shows that it has two transcript variants from alternative splice sites, resulting in two protein isoforms, 480 amino acids and 475 amino acids, having the same N terminus and C terminus. The 475-amino acid hGCNF isoform is listed in Table 1 and Fig. 1 to compare with mGCNF. Human GCNF is very similar to mGCNF, except for 19 amino acids fewer after amino acid 48 in the NTD and one amino acid, Ser, fewer after its amino acid 194 in the hinge domain (Fig. 1). GCNF and SF-1 (NR5A1) are neighboring genes on the chromosomes of the mouse, human and rat [11]. GCNF and SF-1 might have evolved from a tandem duplication of an ancestral gene [12]. The first exon of mouse SF-1 is located only 13 kb downstream of the last exon of GCNF [13]. However, the expression patterns of GCNF and SF-1 are very distinct. An insulator apparatus has been reported to be located between the two loci [13].

3. DNA binding, response element recognition and dimerization of GCNF

Recombinant GCNF can bind DNA as homodimers with specificity for DR0 elements AGGTCAAGGTCA [1,3]. The DBD of GCNF is most similar to that of RXR, so it was postulated that GCNF may bind to a DR element. The DR0 to DR8 elements were tested in EMSA assays with *in vitro* translated GCNF, and GCNF only bound as homodimers to a DR0 sequence. The binding was very specific, as it can be abolished by specific point mutations of the DNA elements, competed by excess cold DNA elements and supershifted by GCNF antibodies or tagged antibodies. In a similar study using PCR-based binding site selection, GCNF was shown to bind as homodimers with highest affinity to conRTRE (consensus RTR response element), $TCA(AG(G/T)TCA)_2$, which is a DR0 element preceded by TCA [14], although to date an *in vivo* target has not been identified containing this extended sequence. GCNF also bound as homodimers or monomers with lower affinity to the extended half-sites, a DR0 half site preceded by TCA [14]. GCNF has not been found to bind DNA elements with other NRs as heterodimers [3,15]. Deletions and mutation analysis showed that the H3 and H12 regions of the LBD were important for GCNF to bind DNA [16]. The double point mutation of V484D/L485P in H12 greatly reduced the binding of GCNF to DNA elements (Fig. 1) [16].

Recombinant GCNF over-expressed in COS1 cells specifically binds to DR0 elements as homodimers, while cell extracts of RA-treated EC and ES cells specifically bind to DR0 elements as a significantly larger complex, called TRIF (transiently RA-induced factor) complex that migrates much slower in EMSA [17–20]. The mixture of *in vitro* translated GCNF, P19 mouse EC cell extracts and the Oct4 DR0 formed a homodimer complex and a TRIF complex, in which when the amounts of *in vitro* translated GCNF were increased, the amounts of the TRIF complexes were increased, suggesting that the cellular proteins may bind *in vitro* translated GCNF to form TRIF complexes in P19 cells [20].

Initially, it was thought that the TRIF complex was a GCNF homodimer associated with a co-repressor. However, the TRIF complex does not contain other cellular proteins, such as the corepressors NCoR or SMRT or any other interacting partners identified since then [20]. Biochemical analysis found that the TRIF complex was a GCNF oligomer, best estimated a hexamer [21]. As far as we know this is a unique feature with the NR gene family. Evolutionary trace analysis of NR LBD sequences found evolutionarily conserved amino acids in the H3 and H11 regions. Point mutations of E308A, A/K318, and K319W in H3, and L459K in H11 disrupted both recombinant GCNF homodimer complex formation and P19 TRIF complex formation (Fig. 1). H3 and H11 may be the interaction surfaces for GCNF oligomerization. A previous study implicated H3 in interacting with the co-repressor NCoR [22]. However, the loss of interaction with NcoR and SMRT may have been secondary to loss of dimerization. Western blot analysis of protein native gel shows that the presence of DR0 elements is required for the GCNF complex formation, suggesting that the GCNF DBD first binds DR0 elements and then oligomerizes [21]. Fast protein liquid chromatography (FPLC) analysis with a gel filtration column of RA-treated P19 cell nuclear extracts indicated that among all the GCNF proteins, only those in the high molecular weight fractions could bind the DR0 element to form TRIF complexes, and the GCNF proteins in the low molecular weight fractions, which are the majority, failed to bind DNA [23]. Future crystallographic studies will definitively define the mechanism of GCNF dimerization, although crystals generated with recombinant materials may not reflect what is observed *in vivo*. Also the fact that so much endogenous GCNF is non-functional with respect to DNA binding suggests that there is a rate-limiting step in