Genomics & Informatics Vol. 3(4) 142-148, December 2005

Comparative Genomics Study of Interferon-α Receptor-1 in Humans and Chimpanzees

Il-Chul Kim[1], Seung-Wook Chi[2], Dae-Won Kim[1,3], Sang-Haeng Choi[1], Sung-Hwa Chae[1] and Hong-Seog Park[1,3]*

[1]Genome Structure Research Laboratory, [2]Protein Analysis and Design Laboratory, Korea Research Institute of Bioscience and Biotechnology, 52 Oun-dong, Yusong-gu, Daejeon 305-333, Korea, [3]University of Science and Technology (UST), Daejeon 305-333, Korea

Abstract

The immune response-related genes have been suggested to be the most favorable genes for positive selection during evolution. Comparing the entire DNA sequence of chimpanzee chromosome 22 (PTR22) with human chromosome 21 (HSA21), we have identified 15 orthologs having indel in their coding sequences. Among them, interferon-α receptor-1 gene (*IFNAR1*), an immune-response-related gene, is subjected to comparative genomic analysis. Chimpanzee *IFNAR1* showed the same genomic structure as human *IFNAR1* (11 exons and 10 introns) except the 3 bp insertion in exon 4. The sequence alignment of *IFNAR1* coding sequence indicated that "ISPP" amino acid sequence motif is highly conserved in chimpanzee and other animals including mouse and chicken. However, the human *IFNAR1* shows that one proline residue is missing in the sequence motif. The homology modeling of the *IFNAR1* structures suggests that the proline deletion in human *IFNAR1* leads to the formation of the following α-helix, whereas two sequential prolines in chimpanzee *IFNAR1* inhibit it. As a result, human *IFNAR1* may adopt a characteristic structure distinct from chimpanzee *IFNAR1*. This human specific trait could contribute to specific immune response in the most optimized manner for humans. Further molecular biological studies on the *IFNAR1* will help us to gain insights into the molecular implication of species-specific host-pathogen interaction in primate evolution.

Keywords: chimpanzee genome, comparative genomics, interferon-α receptor, evolution, insertion

*Corresponding author: Email hspark@kribb.re.kr, Tel +82-42-879-8132, Fax +82-42-879-8139
Accepted 5 November 2005

Introduction

Recently, we reported the entire DNA sequence of the euchromatic region of chimpanzee chromosome 22 (PTR22) (Watanabe *et al.*, 2004). The data allowed us to search for the differences in nucleotide sequence between human and chimpanzee. Compared with human chromosome 21 (HSA21), PTR22 differs at approximately 1.44% of their 33 million aligned nucleotides. In addition, 15 orthologs among the 231 orthologous genes on PTR22, were identified to have indels within their coding sequences. The indels on the coding sequences would potentially lead to gross structural changes on protein. It was hypothesized that indels may represent one of the major mechanisms of proteome evolution in the higher primates (Watanabe *et al.*, 2004). In terms of evolution, *IFNAR1* among the 15 orthologs is especially interesting because the immune-response-related genes have been suggested to be the most favorable genes for positive selection during evolution (Endo *et al.*, 1996; Hughes, 1997; Sawyer *et al.*, 2004).

The *IFNAR1* gene belongs to the class II cytokine receptor family (Gibbs *et al.*, 1996; Oritani *et al.*, 2001). The human interferon-α receptor is a multi-chain receptor complex which binds the multiple human type I interferons with high affinity in a species-specific manner. *IFNAR1* encodes one chain of the receptor protein (hIFN-αR), which contains an extracellular domain of 409 amino acids, a single transmembrane domain of 21 amino acids, and an IC domain of 100-amino acids. Mice in which the gene for *IFNAR1* has been inactivated by homologous recombination are unresponsive to all type I interferons and are unable to mount an effective antiviral defense (Muller *et al.*, 1994; Hwang *et al.*, 1995). These findings suggest that *IFNAR1* is a signal- transducing chain of the multi-component receptor complex that has low affinity binding for most IFN-α subspecies and is essential for type I interferon mediated responses. Type I interferons induce the transcription of interferon-stimulated genes through the formation and activation of ISRE-binding proteins. Many interferon-stimulated genes play roles in antiviral responces via innate immune mechanism (Asaoka *et al.*, 2005; Osiak *et al.*, 2005).

In this study, we analyzed and compared genomic structure of *IFNAR1* gene between human and chimpanzee and discuss its consequence on protein structure with implication of immune-response-related

genes in primate evolution.

Results and Discussion

Genomic structure of *IFNAR1* gene in Chimpanzee and Human

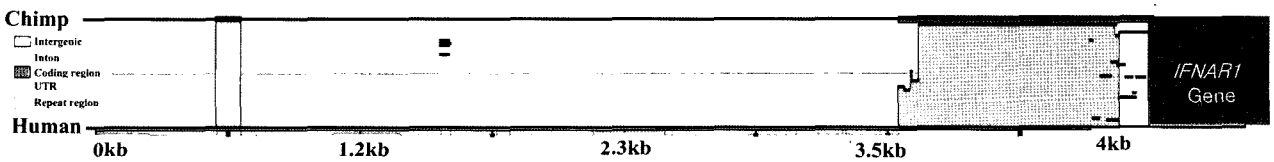
Genomic organization of chimpanzee *IFNAR1* gene was analyzed and compared with that of human. To understand the genomic characteristics of chimpanzee *IFNAR1*, we predicted genomic structure of chimpanzee *IFNAR1* (Ensenbl gene ID: ENSPTRG00000013867) using est2genome program (Mott *et al.*, 1997). The program was run by matching human Ensembl gene (ENSG00000142166) as a query sequence to chimpanzee genomic DNA sequence. The chimpanzee *IFNAR1* (ENSPTRG00000013867) showed the genomic size of 42,648 bp and the same gene structure as human *IFNAR1*. As shown in Fig.1, human and chimpanzee *IFNAR1*s are composed of 11 exons and 10 introns. The obvious difference between the two genes is that chimpanzee *IFNAR1* has about 8 kb insertion between exon 1 and 2. Since upstream sequences from the transcription start site have been known to be important for gene regulation,

Table 1. Comparative analysis of putative TFBSs (transcription factor binding sites) predicted from upstream 3 kb region of human and chimpanzee *IFNAR1*.

			Hunman		Chimp		Identity
Num.	Transcript Factor Name	D	Position	Binding Sites	Position	Binding Sites	
1	MYOGENIN_Q6	+	533-540	aaCAGCTG	530-537	aaCAGCTG	100
2	LBP1_Q6	+	535-541	CAGCTGg	532-538	CAGCTGg	100
3	CEBPA_01	+	544-557	agaTTgcctaaaaa	541-554	agaTTgcctaaaaa	100
4	POU6F1_01	+	592-602	ACATAAATTAT	589-599	ACATAAATTAT	100
5	CDXA_02	+	597-603	aaTtata	594-600	aaTtata	100
6	STAT5A_03	+	630-637	tatTTCcc	627-634	tatTTCcc	100
7	STAT6_01	+	630-637	tatTTCcc	627-634	tatTTCcc	100
8	GR_Q6_01	+	3642-3649	atTGTTCt	2154-2161	atTGTTCt	100
9	IRF1_Q6	+	3659-3665	TTCATTT	2171-2177	TTCATTT	100
10	CDXA_02	+	3666-3672	aaTtatg	2178-2184	aaTtatg	100
11	E4BP4_01	+	3666-3677	aaTTATGTAAtt	2178-2189	aaTTATGTAAtt	100
12	EN1_01	+	3672-3678	GTAATtg	2184-2190	GTAATtg	100
13	EN1_01	+	3711-3717	GTATTtg	2223-2229	GTATTtg	100
14	CDXA_02	+	3780-3786	atTactg	2292-2298	atTactg	100
15	LHX3_01	+	3791-3800	ATTTAATTAa	2303-2312	ATTTAATTAa	100
16	GATA_Q6	+	3835-3841	tGATAaa	2347-2353	tGATAaa	100
17	STAT5A_04	+	4016-4023	cacTTCct	2528-2535	cacTTCct	95
18	STAT3_02	+	4039-4046	tggTTCCc	2551-2558	tggTTCCc	100
19	AP2ALPHA_01	+	4099-4107	GCCCCGGGc	2611-2619	GCCCCGGGc	100
20	AP2GAMMA_01	+	4099-4107	GCCCCGGGc	2611-2619	GCCCCGGGc	100
21	CHCH_01	+	4190-4195	cGGGaa	2702-2707	cGGGaa	100
22	USF2_Q6	+	4218-4223	CAGGTG	2730-2735	CAGGTG	100
23	E2F1_Q3	+	4224-4231	tgTGCCGc	2736-2743	tgTGCCGc	100
24	TEF1_Q6	+	4275-4280	GGAATG	2787-2792	GGAATG	100
25	NKX25B_1	+	4287-4293	tgAAGTG	2799-2805	tgAAGTG	100
26	PU1_Q6	+	4322-4329	tGAGGAAG	2834-2841	tGAGGAAG	100
27	STAT4_01	+	4351-4358	gatTTCta	2863-2870	gatTTCta	100
28	STAT5A_04	+	4351-4358	gatTTCta	2863-2870	gatTTCta	100
29	CDXA_02	+	4352-4358	atTtcta	2864-2870	atTtcta	100
30	E2F_Q2	+	4360-4365	GGCGcg	2872-2877	GGCGcg	100
31	E2F_Q2	+	4436-4441	GGCGcg	2872-2877	GGCGcg	100
32	ETF_Q6	+	4547-4553	GGGGCGG	2981-2987	GGGGCGG	100
33	AHRHIF_Q6	+	4593-4601	cgCGTGcgc	3046-3054	cgCGTGcgc	100
34	NRF1_Q6	+	4593-4602	CGCGTGCGCg	3046-3055	CGCGTGCGCg	95
35	ETF_Q6	+	4624-4630	GCGGCGG	3135-3141	GCGGCGG	100

we identified 35 conserved and 35 aligned transcription factor binding sites (TFBS) by using rVista(limited veterbate TF, Matrix simility 0.95)

* D : direction of transcription



The rectangles of left side represent intergenic sequence, intron, coding region, UTR (untranslated region), repeat region, respectively. The sequence comparison and prediction of TFBS (transcription factor binding sites) were done using rVista2.0 (http://rvista.dcode.org/; Loots *et al.* 2004).

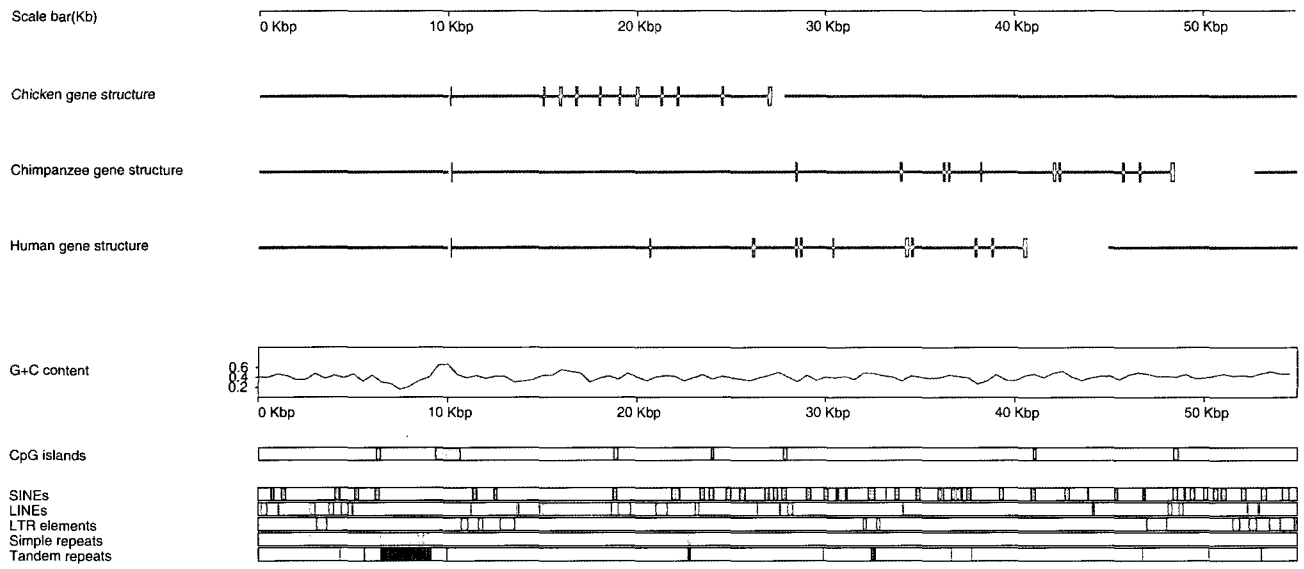


Fig. 1. Comparative analysis of *IFNAR1* gene in Human and Chimpanzee.
Human and chimpanzee mapped gene structure, G+C content, position of CpG islands, interspersed repeats (SINEs, LINEs, LTR elements, Simple repeats), tandem repeats are represented. The UTR (untranslated region) is colored in yellow. The direction of transcription is indicated from left to right side. The window size of G+C content is 500 bp. CpG islands at the middle of gene annotation section are predicted using a modification of the CpG program (Takai *et al.*, 2002). The interspersed repeats are identified using the RepeatMasker program (Smit & Green, unpublished data). Tandem repeats are predicted by TRF program (Benson. 1999).

we analyzed transcription factor binding sites (TFBS) from the 3 kb upstream transcription start site of *IFNAR1* in both species using rVista (limited vertebrate TF, Matrix similarity 0.95). We had identified 35 TFBS from the upstream sequence of human *IFNAR1* and confirmed all of the TFBSs are conserved in chimpanzee genome (Table 1) suggesting that the regulation of *IFNAR1* expression might is similar in both the species. Next, we tried to speculate the changes in coding sequence (CDS) of human *IFNAR1* caused by the genomic change. As shown in Fig. 2, human *IFNAR1* has 9 base substitution and 3 bp insertion in exon 4. The 3 bp insertion in exon 4 resulted in the addition of amino acid proline at amino acid position 149. The 4 out of 9 base substitutions led to amino acid substitutions of Val4Ala, Thr68Pro, Ala425Val and Ser478Asn. *IFNAR1* provides a unique example of cross-species orthologs in that they are from evolutionarily closed primates. As most of gene structures are almost identical for primates, sequences with such striking difference are available only in a few cases. In this context, the 3 bp indel in the orthologs of *IFNAR1* is very exceptional. Thus, species-specific, rapidly diverging insertions or deletions may contribute to gene diversification during evolution.

Proline indel results in structural difference between human and chimpanzee *IFNAR1*s

To determine whether the proline indel affects the three-dimensional structure of *IFNAR1*, we carried out the homology modeling of human and chimpanzee *IFNAR1*s, using the program SwissModel. The template was derived from the crystal structure of the extracellular domain (ECD) in human Interferon γ receptor α chain (PDB code 1FYH:E) (Randal *et al.*, 2001). The Interferon γ receptor α chain shows 28% identity and 52% similarity of amino acid sequence with human *IFNAR1* for the residues 60-128. It also has 22% identity and 44% similarity of sequence with chimpanzee *IFNAR1* for the residues 60-205. The comparison of the modeled structures reveals the obvious structural difference in the vicinity of Pro149 in the ligand binding domain of chimpanzee *IFNAR1* (Fig. 4). Fig. 4A shows the crystal structure of the ECD in human Interferon γ receptor α chain. In the structure, Pro131 of human Interferon γ receptor α chain corresponds to the site of the proline insertion in chimpanzee *IFNAR1*. In the human Interferon γ receptor α chain, Pro131 serve as connecting the preceding β-strand and the following α-helix, kinking the running direction of the backbone. In human *IFNAR1*, Pro148 corresponds to Pro131 of human Interferon γ receptor α chain, playing the same role (Fig. 4B). On the other hand, in chimpanzee *IFNAR1*, Pro149 is inserted at the same position and thus two sequential prolines (Pro148 and Pro149) are present (Fig. 4C). It is interesting to find that the following α-helix is not formed due to the presence of two sequential prolines. We

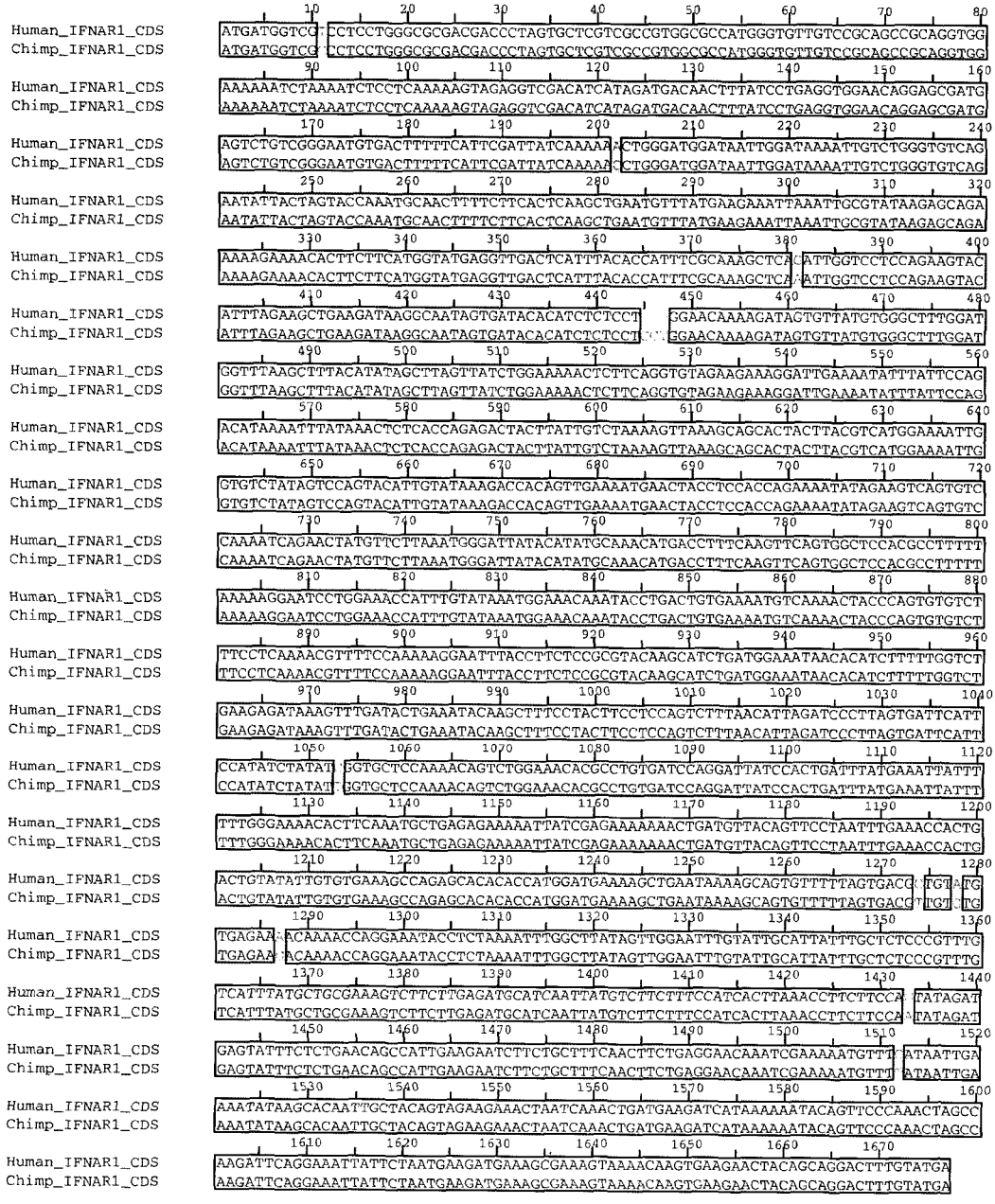


Fig. 2. Comparison of *IFNAR1* CDS between human and chimpanzee.

suggest that the proline deletion in human *IFNAR1* results in the formation of the following α-helix, changing the conservation of secondary structure pattern.

In addition to the proline indel, there are 4 amino acid substitutions between human and chimpanzee *IFNAR1*s. The most noticeable one is the change from Thr68 in human *IFNAR1* to Pro68 in chimpanzee *IFNAR1* (Thr68Pro). The residue 68 in *IFNAR1* corresponds to Tyr49 of human Interferon γ receptor α chain. The side-chain oxygen atom of Tyr49 was observed to form the intermolecular hydrogen bond with the side-chain oxygen of Glu312 in interferon γ (Randal *et al.*, 2001). Therefore, it is likely that the substitution at the position effects on interferon γ binding. We suggest that Thr68 in human *IFNAR1* potentially forms the same hydrogen bond as in human Interferon γ receptor but Pro68 in chimpanzee *IFNAR1* cannot since it has no side-chain oxygen atom capable of making hydrogen bonds. In addition, Thr68 may form the part of β-strand like in human Interferon γ receptor but Pro68 cannot since it

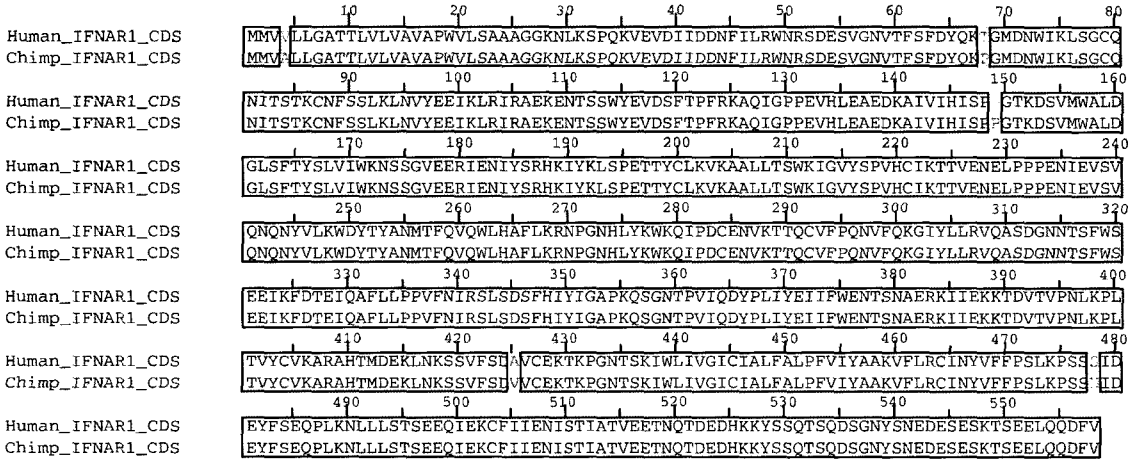


Fig. 3. Comparison of primary structure of *IFNAR1* between human and chimpanzee.

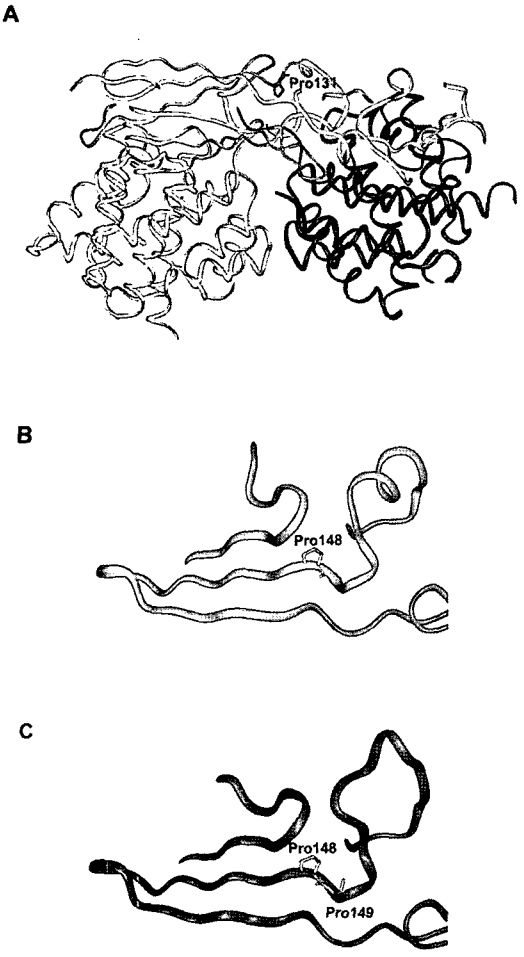


Fig. 4. Modeled structures of human and chimpanzee *IFNAR1s*. (A) The crystal structure of the extracellular domain in human Interferon γ receptor α chain (PDB code 1FYH:E) (B) Modeled structure of the ligand binding domain in human *IFNAR1*. (C) Modeled structure of the ligand binding domain in chimpanzee *IFNAR1*.

has no main-chain amide proton. The other substitutions contain Val4Ala, Ala425Val, and Ser478Asn. These substitutions are not likely to induce significant alteration in the secondary structure as the properties of changed amino acids are similar.

Implication for molecular function and evolution of immune-response-related genes

Multiple sequence alignment of *IFNAR1* from human, chimpanzee, and other animals including mouse, rat and chicken, indicated that "ISPP" motif in the ligand binding domain of *IFNAR1* is highly conserved in all of the other animals from chicken to chimpanzee (Fig. 5). The human *IFNAR1* showed that one proline residue is missing in the segment, implying that the proline might be deleted after divergence from common ancestor of human and chimpanzee during evolution. It is striking to notice that this conserved sequence motif is likely to be the only region that bears significant difference between the human and nonhuman hominoids. Our homology modeling study reveals the structural difference between human and chimpanzee *IFNAR1*s. The proline deletion in human *IFNAR1* leads to the formation of the following α-helix, whereas two sequential prolines in chimpanzee *IFNAR1* inhibit it (Fig. 4). At present, the functional implication of these differences in the *IFNAR1* structure is not clear. However, it has been reported that *IFNAR1* expression is related to the effectiveness of interferon therapy in hepatitis C virus (HCV)-associated chronic liver disease (CLD) (Fukuda *et al.*, 1996). Furthermore, it was also reported that great apes could be infected with human hepatitis viruses, but usually do not progress to cirrhosis or hepatocellular carcinomas as often seen in human (Muchmore *et al.*, 1988). Therefore, it would be

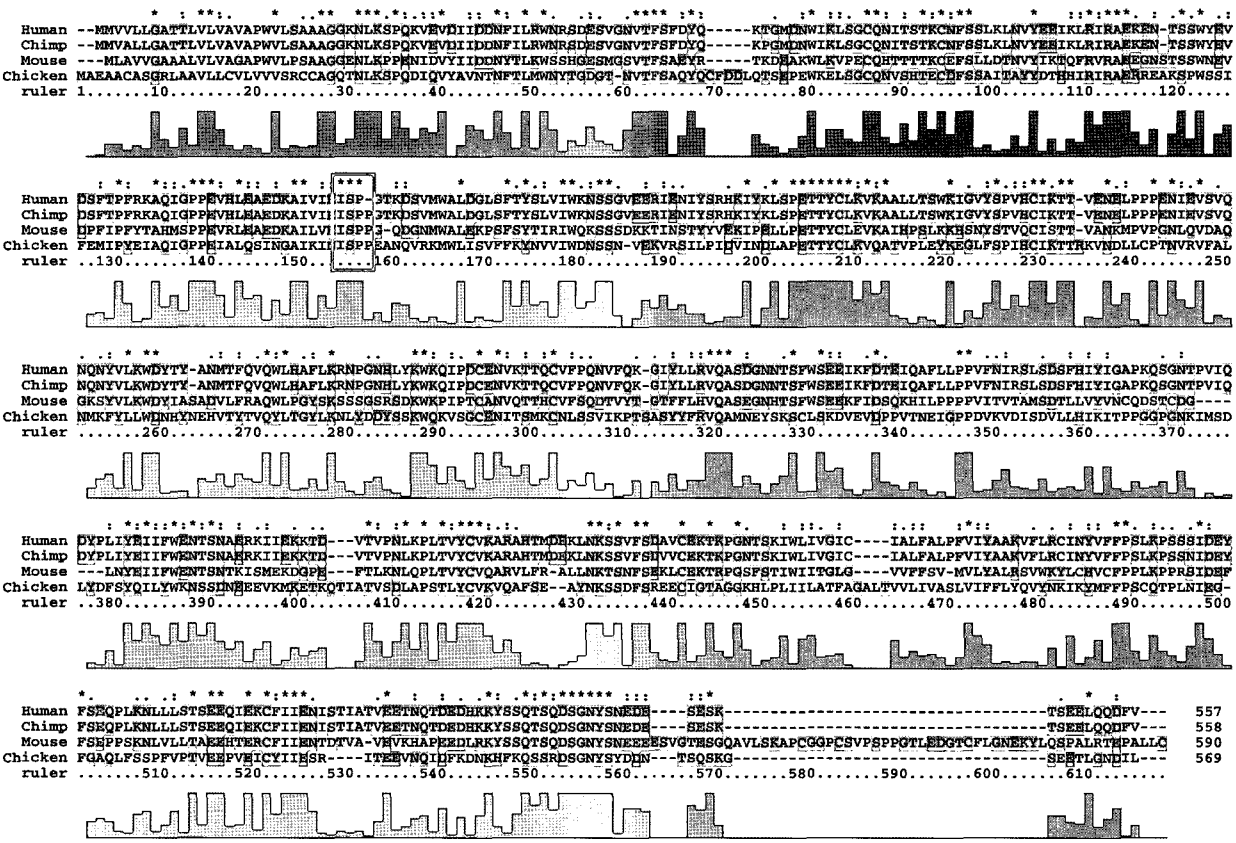


Fig. 5. Multiple alignment with IFNAR1 coding sequences of human, chimpanzee, mouse and chicken. Multiple alignment was done using clustalX program (Thompson, *et al.*, 1994). Red rectangle indicate 'ISPP' motif of domain in ligand binding domain. Ruler is the position of base pair.

interesting to examine a possibility that functional consequences of *IFNAR1* differences may related with the differences in efficacy of interferon therapy or cirrhotic progression in both the species during chronic hepatitis C virus infection. Although the establishment of biological significance of human specific variations has a long way to go, this kind of genomic change might be the most important in the development of human traits. Human specific traits, such as those changes in *IFNAR1* may contribute to distinctive biological properties of immune-response to pathogens and specific adaptations in the human lineage.

In this study, we showed that human *IFNAR1* has 3 base pair deletion in genomic DNA, compared with chimpanzee ortholog. As a consequence, human *IFNAR1* may adopt a characteristic structure distinct from chimpanzee *IFNAR1* and this human specific trait could contribute to specific immune response in the most optimized manner for humans. Further molecular biological studies on the *IFNAR1* will help us to gain insights into the molecular implication of species specific host-pathogen interaction in primate evolution.

Acknowledgements

This work was supported by the ministry of Science and Technology, and the Korea Research Institute of Bioscience and Biotechnology, Korea.

References

Asaoka, K., Ikeda, K., Hishinuma, T., Horie-Inoue, K., Takeda, S., and Inoue, S. (2005). A retrovirus restriction factor TRIM5alpha is transcriptionally regulated by interferons. *Biochem. Biophys. Res. Commun.* 338, 1950-1956.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573-580.

Endo, T., Ikeo, K., and Gojobori, T. (1996). Large-scale search for genes on which positive selection may operate. *Mol. Biol. Evol.* 13, 685-690.

Florea, L., Hartzell, G., Zhang, Z., Rubin, G.M., and Miller, W. (1998). A computer program for aligning a cDNA

sequence with a genomic DNA sequence. *Genome Res.* 8, 967-974.

Fukuda, R., Ishimura, N., Ishihara, S., Tokuda, A., Satoh, S., Sakai, S., Akagi, S., Watanabe, M., adn Fukumoto, S. (1996). Expression of interferon-alpha receptor mRNA in the liver in chronic liver diseases associated with hepatitis C virus: relation to effectiveness of interferon therapy. *J. Gastroenterol.* 31, 806-811

Gibbs, V.C., Takahashi, M., Aguet, M., and Chuntharapai, A. (1996). A negative regulatory region in the intracellular domain of the human interferon-alpha receptor. *J. Biol. Chem.* 271, 28710-28716.

Hughes, A.L. (1997). Rapid evolution of immunoglobulin superfamily C2 domains expressed in immune system cells. *Mol. Biol. Evol.* 14, 1–5.

Hwang, S.Y., Hertzog, P.J., Holland, K.A., Sumarsono, S.H., Tymms, M.J., Hamilton, J.A., Whitty, G., Bertoncello, I., and Kola, I. (1995). A Null Mutation in the Gene Encoding a Type I Interferon Receptor Component Eliminates Antiproliferative and Antiviral Responses to Interferons α and ß and Alters Macrophage responses. *Proc. Natl. Acad. Sci. USA* 92, 11284-11288.

Loots, G. and Ovcharenko, I. (2004) rVista 2.0: evolutionary analysis of transcription factor binding sites. *Nucleic Acids Res.* 32(Web Server Issue), W217-W221.

Mott, R. (1997). EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. *Comput. Appl. Biosci.* 13, 477-478.

Muchmore, E., Popper, H., Peterson, D.A., Miller, M.F., and Lieberman, H.M. (1988). Non-A, non-B hepatitis-related hepatocellular carcinoma in a chimpanzee *J. Med. Primatol.* 17, 235-246.

Muller, U., Steinhoff, U., Reis, L.F.L., Hemmi, S., Pavlovic, J., Zinkernagel, R.M., and Aguet, M. (1994). Functional role of type I and type II interferons in antiviral defense. *Science* 264, 1918-1921.

Oritani, K., Kincade, P.W., Zhang, C., Tomiyama, Y., and Matsuzawa, Y. (2001). Type I interferons and limitin: a comparison of structures, receptors, and functions. *Cytokine Growth Factor Rev.* 12, 337-348.

Osiak, A., Utermohlen, O., Niendorf, S., Horak, I., and Knobeloch, K.P. (2005). ISG15, an interferon-stimulated ubiquitin-like protein, is not essential for STAT1 signaling and responses against vesicular stomatitis and lymphocytic choriomeningitis virus. *Mol. Cell. Biol.* 15, 6338-6345.

Randal, M. and Kosslakoff, A.A. (2001). The Structure and Activity of a Monomeric Interferon-γ:α-Chain Receptor Signaling Complex. *Structure* 9, 155.

Sawyer, S.L., Emerman, M., and Malik, H.S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biology.* 2, e275.

Takai, D. and Jones, P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. USA* 99, 3740-3745.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680.

Watanabe, H., Fujiyama, A., Hattori, M., Taylor, T.D., Toyoda, A., Kuroki, Y., Noguchi, H., BenKahla, A., Lehrach, H., Sudbrak, R., Taenzer, S., Galgoczy, P., Platzer, M., Scharfe, M., Nordsiek, G., Blocker, H., Hellmann, I., Khaitovich, P., Paabo, S., Reinhardt, R., Zheng, H.J., Zhang, X.L., Zhu, G.F., Wang, B.F., Fu, G., Ren, S.X., Zhao, G.P., Chen, Z., Lee, Y.S., Cheong, J.E., Choi, S.H., Wu, K.M., Liu, T.T., Hsiao, K.-J., Kim, C.G., Oota, S., Kitano, T., Kohara, Y.,Saitou, N., Tsai, S.-F., Park, H.S., Wang, S.-Y., Yaspo, M.-L., and Sakaki, Y. (2004). DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature* 429, 382-388.