Brazilian Journal of Biology
ISSN 1519-6984 (Print)
ISSN 1678-4375 (Online)

https://doi.org/10.1590/1519-6984.219608
Notes and Comments

Clustering accentuated matches and highly conserved domains between bacterial and human heat shock gene and protein

J. E. Gabriel[a]*

[a]Universidade Federal do Vale do São Francisco – UNIVASF, Campus Centro Universitário, Av. José de Sá Maniçoba, s/n, Centro, CEP 56304-917, Petrolina, PE, Brasil

*e-mail: jane.gabriel@univasf.edu.br

Received: February 6, 2019 – Accepted: August 20, 2019 – Distributed: November 30, 2020
(With 2 figures)

Heat shock proteins (Hsps) constitute a family of proteins expressed in virtually all living organisms by activating specific genes as a cellular response to exposure to several stressful conditions, including heat shock, cold, UV light, wound healing or tissue remodeling (Li and Srivastava, 2004). Among these peptides, protein HscA acts as chaperone involved specially in the maturation of the iron-sulfur cluster-containing proteins in the bacterium *Escherichia coli*, whereas its orthologous Hsp70 protein regulates the assembly of multiprotein complexes, transport/sorting of proteins into correct subcellular compartments, cell-cycle control and signaling, and protection of cells against stress/apoptosis in different species (Li and Srivastava, 2004). Within this context, the current study aimed to assess comparatively the conservation degree of residues of nucleotides of the *hscA* and *hsp70* genes and amino acids of the HscA and Hsp70 proteins of *E. coli* and *Homo sapiens*, as well as investigate variations in the conformational pattern of their secondary structures among such proteins using Bioinformatics' tools.

Sequences of genes and proteins of interest were searched and selected from biological information databases GeneBank and UniProtKB/Swiss-Prot, respectively: a) heat shock chaperone A gene (*hscA*) of *E. coli* strain K-12 (000913.3:c2658935-2657085), and heat shock protein family A gene (*hsp70*) member 1 like (HSPA1L) NG_011855.1 of *H. sapiens* on chromosome 6; and b) bacterial chaperone A protein HscA (P0A6Z1|HSCA_ECOLI) and human heat shock 70 kDa protein 1A (Hsp70) (P0DMV8|HS71A_HUMAN). Comparative structural alignments of the nucleotide and amino acid residues of the sequences of interest were performed using PipMaker program (Schwartz et al., 2000) and Clustal method (Söding, 2005), respectively. Similarity scores, match percentages, number of matches and mismatches, conservation or divergence degrees among sequences selected were also inferred. Variations in the conformational pattern of the secondary structure of the bacterial HscA and human Hsp70 proteins were predicted using SIMPA96 Secondary Structure Prediction method (Combet et al., 2000).

Under these experimental conditions, two local alignments were proposed by direct comparing bacterial and human gene sequences (453,8515-1473,9580 nucleotides and 60,4288-1473,2815 nucleotides) by PipMaker analyses. Structural alignments of the nucleotide residues of the bacterial and human heat shock genes demonstrated accentuated match percentages of 53 to 54% with similarity scores of 16479 and 13510, respectively, in both two alignments proposed. Interestingly, 586 and 803 nucleotides were perfectly matched in identical sites in both alignments (Figure 1A). Moreover, the amount of different nucleotides aligned in same position was 422 and 570 mismatches with total length of gaps corresponding to 71 and 142, respectively. On the other hand, multiple alignments of the amino acid sequences of the proteins of interest revealed a low identity of 34.077% with 229 residues situated in identical positions (Figure 1B). Moreover, 194 amino acid residues in the proteins of interest were aligned in similar positions by replacing amino acids with identical biochemical proprieties in same site, such as substitutions of valine (V) by isoleucine (I) or valine (V) by alanine (A), all amino acids having identical hydrophobic characteristics. Conformational patterns of the secondary structure of the bacterial HscA protein demonstrated particularly the presence of 274 alpha helixes (44.41%), 89 extended strands (14.42%) and 253 random coils (41%) (Figure 2A), whereas 247 alpha helixes (47%), 113 extended strands (17.6%) and 281 random coils (43.77%) were preferentially predicted from human Hsp70 protein (Figure 2B).

The findings described in current study demonstrated accentuated amounts of matches between *hscA* and *hsp70* genes, as well as highly conserved sites between bacterial HscA and human Hsp70 proteins during structural alignments of phylogenetically distant and evolutionarily quite divergent species (Figure 1). According to Ghorani-Azam et al. (2016), computational analyses of sequence homology have also demonstrated highly conserved regions from consensus sequence of human and bacterial amylase, including a SLH domain in outer surface of the protein that facilitate the binding of enzyme to cell wall. Pearson (2013) has reported that similarity searches from protein and translated-DNA sequences are much more sensitive than nucleotide searches that have between 5-10-fold shorter evolutionary look-back time than protein:protein or translated DNA:protein alignments. Moreover, DNA:DNA alignments detect rarely significant similarity after more than 200-400 million years of divergence, whereas protein:protein alignments detect routinely high similarities in sequences that last shared a common ancestor more than 2.5 billion years ago (e.g. humans to bacteria) (Pearson, 2013).

Figure 1. Comparative alignments of the sequences of the nucleotide and amino acid residues of the bacterial HscA and human Hsp70. (A) Nucleotide seq1data: 1851 bp: >NC_000913.3:c2658935-2657085 *E. coli* and seq2data: 12440 bp: >NG_011855.1 *H. sapiens* HSP70A on chromosome 6; (B) Amino acid seq1data: sites 118-233, P0A6Z1 HSCA ECOLI and seq2data: sites 106-225 PODMV8 HSP71A HUMAN.

Differentiated conformational patterns of the secondary structure of the proteins of interest revealed higher amounts of the alpha helix domain in bacterial HscA in contrast to more abundant contents of extended strand and coiled coil motifs showed in human Hsp70 (Figure 2). Several studies have focused on elucidation of the conformational pattern of the secondary structures of protein products. Flores et al. (1993) widely recognized that evolutionary divergence of protein structures occurs much less rapidly than divergence of protein sequences, indicating that selective constraints may act to preserve protein structure. Interestingly, the low identity of bacterial and human proteins revealed in current study may be resulted of differences in total length sizes of the segments selected (616 and 641 amino acids residues for bacterial HscA and human Hsp70, respectively), but also due to high total amount of 56 gaps identified under these experimental conditions. Within this context, number of insertions and deletions, conformation of loops are often very different for proteins with weak sequence identity (Flores et al., 1993), whereas helices and turns are underrepresented in conserved regions, similar numbers of loops are primarily found in conserved and random regions (Sitbon and Pietrokovski, 2007). Recently, Zea et al. (2013) found a negative correlation between conformational diversity and protein evolutionary rate, suggesting that the structural constraints underlying protein dynamism, essential for protein function, could modulate protein divergence. Postsequencing functional analyses based on alignments of complete or partial sequences of several genomes is one powerful tool for interpreting the genomic information in the biological sciences by deducing similarities and differences in gene and its protein product in distinct species (Kittler and Oliver, 2006; Srinivasan et al., 2019). Thus, the findings described herein reveal important insights into differences in thee conservation degree and organizational pattern of the secondary structural domains between bacterial and human heat shock gene and protein.

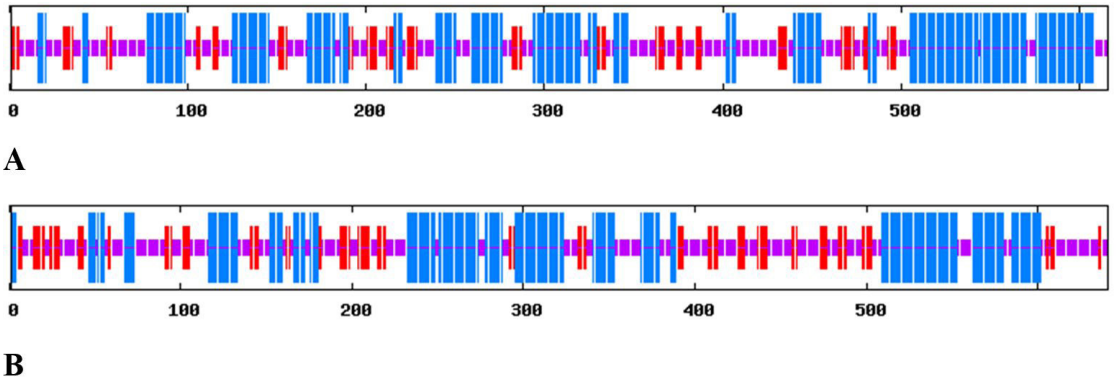

Figure 2. Differentiated conformational pattern of the secondary structures of heat shock proteins in distinct species. (A) Protein HscA of the bacterium *Escherichia coli*; (B) Human Hsp70 protein. Alpha helices (Hh) and extended strands (Ee) are represented in blue and red colors, respectively.

References

COMBET, C., BLANCHET, C., GEOURJON, C. and DELÉAGE, G., 2000. NPS@: network protein sequence analysis. *Trends in Biochemical Sciences*, vol. 25, no. 3, pp. 147-150. http://dx.doi.org/10.1016/S0968-0004(99)01540-6. PMid:10694887.

FLORES, T.P., ORENGO, C.A., MOSS, D.S. and THORNTON, J.M., 1993. Comparison of conformational characteristics in structurally similar protein pairs. *Protein Science*, vol. 2, no. 11, pp. 1811-1826. http://dx.doi.org/10.1002/pro.5560021104. PMid:8268794.

GHORANI-AZAM, A., MOHAJERI, S.A., RIAHI-ZANJANI, B., SEPAHI, S. and SEPAHI, S., 2016. Human and bacterial amylases: computational analysis of sequence homology. *Journal of Genes and Cells*, vol. 2, no. 3, pp. 21-26. http://dx.doi.org/10.15562/gnc.40.

KITTLER, J.T. and OLIVER, P.L., 2006. Genomic and post-genomic tools for studying synapse biology. In: J.T. KITTLER and S.J. MOSS, eds. *The dynamic synapse: molecular methods in ionotropic receptor biology*. Boca Raton: CRC Press/Taylor & Francis, chapt. 15. http://dx.doi.org/10.1201/9780203486283.ch15.

LI, Z. and SRIVASTAVA, P., 2004. Heat-shock proteins. *Current Protocols in Immunology*, vol. 58, no. 1, pp. A.1T.1. PMid:18432918.

PEARSON, W.R., 2013. An introduction to sequence similarity ("homology") searching. *Current Protocols in Bioinformatics*, vol. 42, no. 1, pp. 3.1.1-3.1, 8. http://dx.doi.org/10.1002/0471250953.bi0301s42. PMid:23749753.

SCHWARTZ, S., ZHANG, Z., FRAZER, K.A., SMIT, A., RIEMER, C., BOUCK, J., GIBBS, R., HARDISON, R. and MILLER, W., 2000. PipMaker- a web server for aligning two genomic DNA sequences. *Genome Research*, vol. 10, no. 4, pp. 577-586. http://dx.doi.org/10.1101/gr.10.4.577. PMid:10779500.

SITBON, E. and PIETROKOVSKI, S., 2007. Occurrence of protein structure elements in conserved sequence regions. *BMC Structural Biology*, vol. 7, no. 1, pp. 3. http://dx.doi.org/10.1186/1472-6807-7-3. PMid:17210087.

SÖDING, J., 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics*, vol. 21, no. 7, pp. 951-960. http://dx.doi.org/10.1093/bioinformatics/bti125. PMid:15531603.

SRINIVASAN, K., DEY, S. and SENGUPTA, J., 2019. Structural modules of the stress-induced protein HflX: an outlook on its evolution and biological role. *Current Genetics*, vol. 65, no. 2, pp. 363-370. http://dx.doi.org/10.1007/s00294-018-0905-x. PMid:30448945.

ZEA, D.J., MONZON, A.M., FORNASARI, M.S., MARINO-BUSLJE, C. and PARISI, G., 2013. Protein conformational diversity correlates with evolutionary rate. *Molecular Biology and Evolution*, vol. 30, no. 7, pp. 1500-1503. http://dx.doi.org/10.1093/molbev/mst065. PMid:23564939.