

Bioinformatics Analysis of Hsp20 Sequences in Proteobacteria

Michelle Heine and Sathees B.C.Chandra*

Department of Biological, Chemical and Physical Sciences, Roosevelt University, Chicago, IL-60605, USA

Abstract

Heat shock proteins are a class of molecular chaperones that can be found in nearly all organisms from Bacteria, Archaea and Eukarya domains. Heat shock proteins experience increased transcription during periods of heat induced osmotic stress and are involved in protein disaggregation and refolding as part of a cell's danger signaling cascade. Heat shock protein, Hsp20 is a small molecular chaperone that is approximately 20kDa in weight and is hypothesized to prevent aggregation and denaturation. Hsp20 can be found in several strains of Proteobacteria, which comprises the largest phyla of the Bacteria domain and also contains several medically significant bacterial strains. Genomic analyses were performed to determine a common evolutionary pattern among Hsp20 sequences in Proteobacteria. It was found that Hsp20 shared a common ancestor within and among the five subclasses of Proteobacteria. This is readily apparent from the amount of sequence similarities within and between Hsp20 protein sequences as well as phylogenetic analysis of sequences from proteobacterial and non-proteobacterial species.

Keywords: actinobacteria, common ancestry, firmicutes, heat shock proteins, Hsp20, phylogenetic analysis, proteobacteria

Introduction

Heat shock proteins are part of a cellular defensive system that is of extreme importance and can be found in nearly all living organisms. In 1974 the first heat shock protein was reported for *Drosophila melanogaster*, the universality of these proteins became a realization soon afterward when they were discovered in organisms ranging from humans to bacteria (Schleisinger, 1990). The role of heat shock proteins is to protect native proteins from misfolding and aggregating under heat in-

duced or osmotic stress (Oganesyan *et al.*, 2007). They comprise a group of several proteins in which transcription is increased when cells are exposed to heat induced or osmotic stress by heat shock factor as part of the heat shock response (Schleisinger, 1990; Suzuki *et al.*, 1998; Trotter *et al.*, 2002; Davies *et al.*, 2006; Oganesyan *et al.*, 2007). The heat shock response prompts the cell danger signaling cascade, in which heat shock proteins refold and de-aggregate proteins that have become damaged in the cell (Davies *et al.*, 2006). In addition, heat shock proteins are found as both intracellular and extracellular proteins that participate in the immune response (Davies *et al.*, 2006). Heat shock proteins are also expressed under normal conditions as molecular chaperones, to assist in protein folding and maintenance of protein integrity (Davies *et al.*, 2006; Oganesyan *et al.*, 2007). Heat shock protein Hsp20, is a family of small heat shock proteins that are approximately 20kDa in weight and identified by a conserved C terminal domain of about 100 residues (Kim *et al.*, 1998). It is hypothesized that Hsp20 proteins form large hetero-oligomeric aggregates that protect other proteins against aggregation and denaturation (Mulder *et al.*, 2007). Hsp20 was first discovered in human skeletal muscle and has been found to be expressed in cardiac muscle, stomach, intestine and bladder (Rembold *et al.*, 2000; Meeks *et al.*, 2005; Mulder *et al.*, 2007).

The phylum Proteobacteria is no exception in the sense that it too possesses heat shock proteins. It is one of the largest phyla of the Bacteria domain and is further subdivided into five categories: alpha, beta, delta, epsilon and gamma. The alpha, beta and gamma subclasses were highly supported by morphological analyses, however, the delta and epsilon subclasses were added separately and are considered to have separated earlier than the other subclasses on the phylogenetic tree (Ludwig and Klenk, 2001). All Proteobacteria are Gram negative and possess a membrane composed of lipopolysaccharides. Betaproteobacteria play a pivotal role in plant nitrogen fixation and are commonly found in environmental samples (Dedysh *et al.*, 2004). Epsilon-proteobacteria are known to inhabit the digestive tracts of animals and humans (Miroshchichenko *et al.*, 2004). Gammaproteobacteria encompass several species of medically and scientifically important groups of bacteria that include *Salmonella*, *Vibrio*, *Pseudomonas* and *Escherichia* (Lee *et al.*, 2005).

The major aim of our present work is to study the sequence evolution of Hsp20 among all five subclasses of

*Corresponding author: E-mail schandra@roosevelt.edu
Tel +1-847-619-7968, Fax +1-847-619-8555
Accepted 12 February 2009

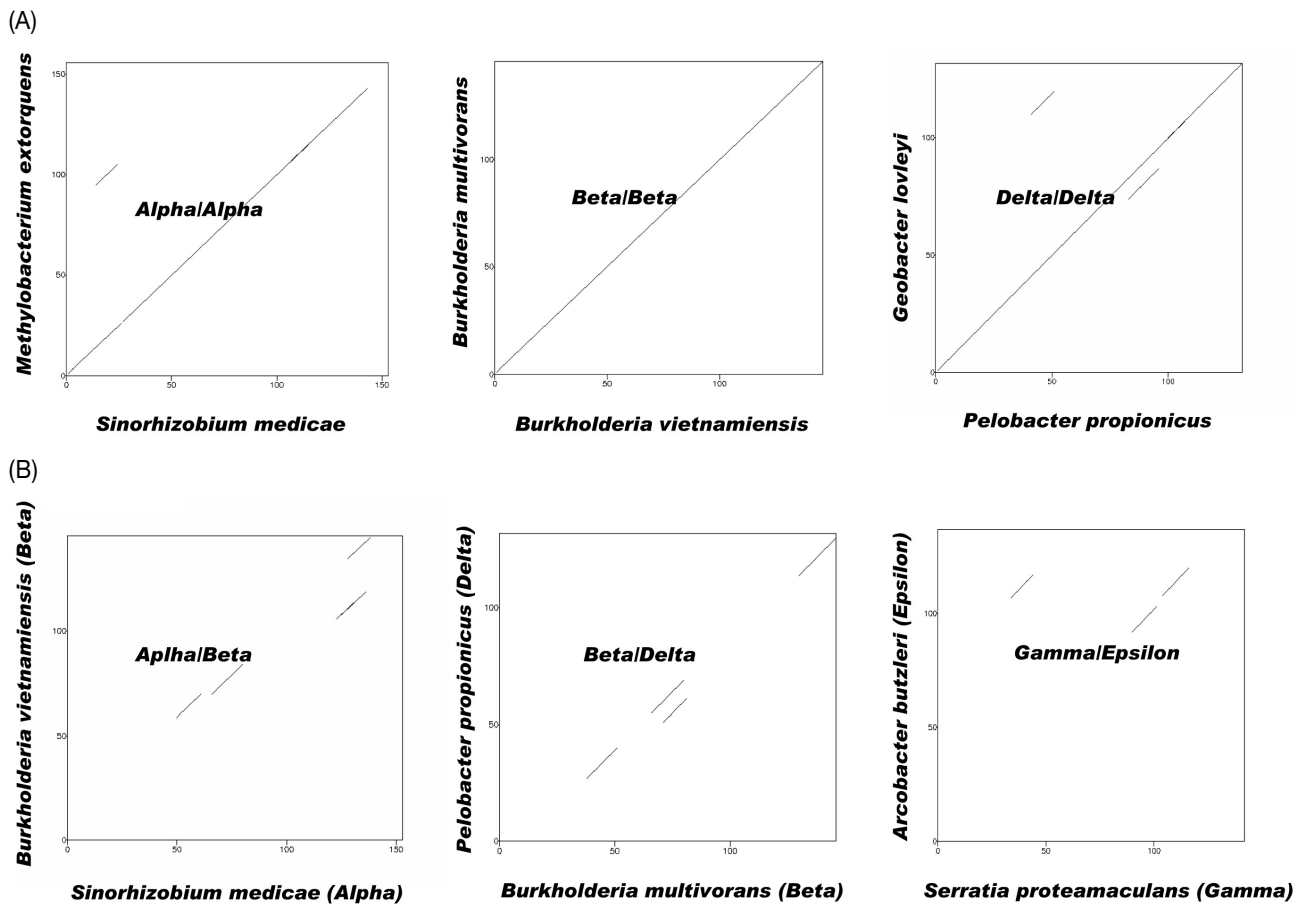


Fig. 2. Dot plots comparison of Hsp20 sequences among subclasses of phylum proteobacteria. These dot plots were computed with threshold =23 and window size =10. (A) Dot plots comparison within subclasses such as Alpha/Alpha, Beta/Beta and Delta/Delta. (B) Dot plots comparison between subclasses such as Alpha/Beta, Beta/Delta and Gamma/Epsilon.

residues, followed closely by Alphaproteobacteria with eighty-one highly conserved residues and trailed by Epsilonproteobacteria with only twenty-nine highly conserved residues. Betaproteobacteria and Deltaproteobacteria fell within the mid-range with forty-three and thirty-two highly conserved residues, respectively.

Protein sequences within and between the five subclasses of Proteobacteria were analyzed by Dot-plot comparison to align protein by protein between two microorganisms (Fig. 2). These comparisons allow for visualization of sequence breaks and co-linearity among the sequences analyzed (Chandra *et al.*, 2006). The dot-plots performed within the five subclasses revealed a high degree of co-linearity (Fig. 2A). This is readily apparent among Alphaproteobacteria, Betaproteobacteria, and Gammaproteobacteria. The dot-plots performed between the five subclasses revealed very little to no regions of co-linearity (Fig. 2B).

Comparative analysis of the 24, Hsp20 protein sequences by phylogenetic analysis resulted in a tree with

three major clades (Fig. 3). The first clade consisted of subclasses Delta and Epsilon group. The second clade consisted of micro organisms representing Beta proteobacterial group only. The third major clade consisted of subclasses Alpha and Gamma group. The Alpha and Gamma clade had a bootstrap value of 100, but the Delta and Epsilon clade had a bootstrap value of 76. A rectangular cladogram constructed using both proteobacterial and non-proteobacterial species resulted in possible common ancestry of proteobacterial Hsp20 proteins (Fig. 4).

Discussion

In view of the multiple sequence alignment performed, it is reasonable to conclude that the Hsp20 found in all 24 Proteobacteria sequences shares a common ancestor. This becomes even more apparent when considering the multiple sequence alignments within five subclasses of Proteobacteria. The largest amount of con-

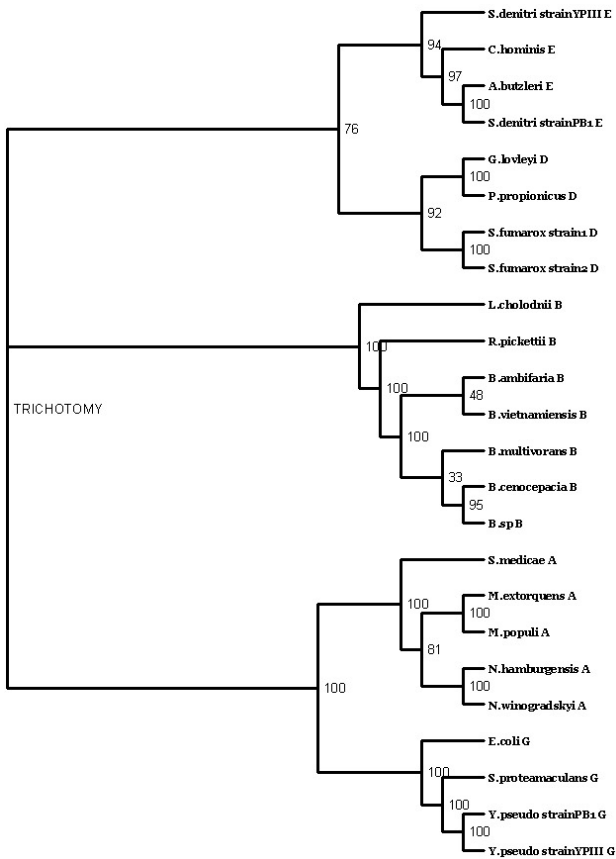


Fig. 3. Amino acid sequence conservation in all 24 species belonging to the five subclasses of Proteobacteria (* -highly conserved, : -conserved, . -semi-conserved). The capital letter A (following the name of the species) represents Alpha-proteobacteria, B represents Betaproteobacteria, D represents Deltaproteobacteria, E represents Epsilonproteobacteria and G represents Gammaproteobacteria. The accession numbers are indicated in the parenthesis. *Arcobacter butzleri* (YP_001491167.1), *Sulfurimonas denitrificans* strain PB1 (YP_393028.1), *Campylobacter homini* (YP_001406197.1), *Sulfurimonas denitrificans* strain YPIII (YP_393029.1), *Geobacter lovleyi* (YP_001950437.1), *Pelobacter propionicus* (YP_900006.1), *Syntrophobacter fumaroxidans* strain 1 (YP_845865.1), *Syntrophobacter fumaroxidans* strain 2 (YP_846481.1), *Burkholderia ambifaria* (YP_773034.1), *Burkholderia vietnamiensis* (YP_001119064.1), *Burkholderia cenocepacia* (YP_001764532.1), *Burkholderia sp.* (YP_368646.1), *Burkholderia multivorans* (YP_001580237.1), *Ralstonia pickettii* (YP_001900983.1), *Leptothrix cholodnii* (YP_001792398.1), *Methylobacterium extorquens* (YP_001641995.1), *Methylobacterium populi* (YP_001927702.1), *Nitrobacter hamburgensis* (YP_576445.1), *Nitrobacter winogradskyi* (YP_319554.1), *Sinorhizobium medicae* (YP_001328380.1), *Yersinia pseudotuberculosis* strain PB1 (YP_001874525.1), *Yersinia pseudotuberculosis* strain YPII (YP_001718779.1), *Serratia proteamaculans* (YP_001476287.1), *Escherichia coli* (YP_001723032.1).

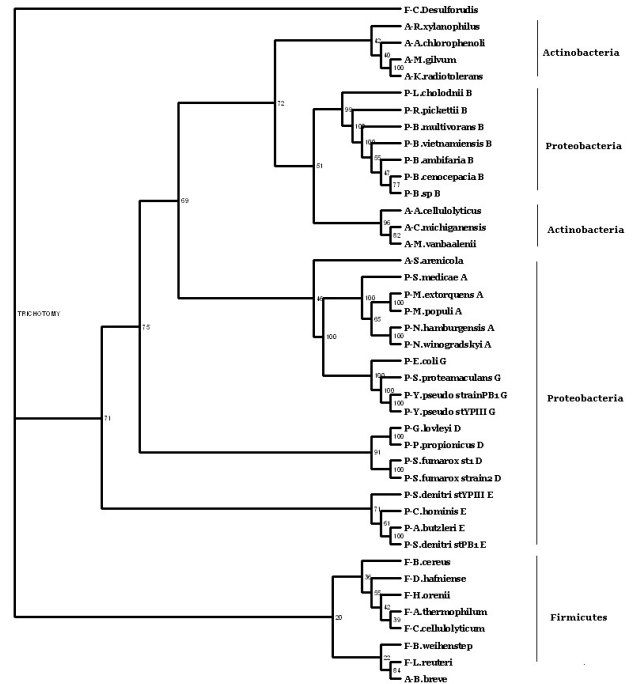


Fig. 4. Phylogenetic analysis of 24 proteobacterial and 17 non-proteobacterial (Firmicutes and Actinobacteria) Hsp20 sequences. The phylogenetic analysis was performed using neighbor-joining algorithm in CLUSTALX. The first letter in each species indicates whether it is a proteobacterial species, P refers to Proteobacteria, F refers to Firmicutes and A refers to Actinobacterial species. The capital letter A (following the name of the species) represents Alphaproteobacteria, B represents Betaproteobacteria, D represents Deltaproteobacteria, E represents Epsilonproteobacteria and G represents Gammaproteobacteria. Please refer to legends of Figure1 for complete genus and species name. For complete genus and species name and accession numbers of proteobacterial group, see figure legends of Fig. 3. Accession numbers and full name of non-proteobacterial species are given below: **Firmicutes:** *Anaerocellum thermophilum* (YP_002573380.1), *Clostridium cellulolyticum* (YP_002507212.1), *Desulfitobacterium hafniense* (YP_002458177.1), *Candidatus Desulforudis* (ACA59000.1), *Lactobacillus reuteri* (YP_001271901.1), *Bacillus cereus* (YP_001374964.1), *Bacillus weihenstephannensis* (ABY43298.1), *Halothermothrix orenii* (ACL70671.1) **Actinobacteria:** *Bifidobacterium breve* (ABL75149.1), *Clavibacter michiganensis* (CAN02159.1), *Arthrobacter chlorophenolicus* (YP_002487772.1), *Mycobacterium gilvum* (YP_001131670.1), *Salinispora arenicola* (ABW00145.1), *Mycobacterium vanbaalenii* (YP_955721.1), *Rubrobacter xylanophilus* (YP_643457.1), *Kineococcus radiotolerans* (ABS01796.1), *Acidothermus cellulolyticus* (YP_871921.1).

served protein residues was found among the Alpha, Delta, Epsilon and Gamma subclasses. From our sequence analysis, it is reasonable to conclude that the Alpha and Gamma subclasses evolved together. In addition the Delta and Epsilon subclasses must have evolved together leaving the Beta subclass to evolve on its own.

The Dot-plot comparisons performed within the five subclasses revealed that the Hsp20 sequences showed a high degree of similarity. This result is particularly apparent among the Alpha, Beta and Gamma subclasses compare to Delta and Epsilon subclasses. Again, this data highly supports the work performed by Ludwig & Klenk (2001) in that the Alpha, Beta and Gamma subclasses are separated from the Delta and Epsilon subclasses. In addition, it is becoming highly apparent that the Hsp20 heat shock proteins are divergent between the subclass of the Proteobacteria yet they all share a common ancestor. Dot-plot analyses performed between the five subclasses either did not result in any co-linearity or there were several sequence breaks present in the comparison. The Dot-plot comparison performed between Gammaproteobacteria and Alphaproteobacteria revealed some sequence similarity with several sequence breaks. Despite the sequence breaks, it appears that Alphaproteobacteria and Gammaproteobacteria share a more recent last common ancestor than any of the other five subclasses.

The phylogenetic tree clearly supports the distinctions of Hsp20, among the five subclasses; this is evident from the multiple sequence alignments and the Dot-plot comparisons as well. There is a clear separation of Betaproteobacteria from the rest of the subclasses and but not among the subclass itself. This suggests that Hsp20 in Betaproteobacteria evolved separately from the other four subclasses. It is then possible to hypothesize that Betaproteobacteria was the first of the five subclasses of Proteobacteria to have evolved. In addition, the Gammaproteobacteria and the Alphaproteobacteria clades are joined together with a bootstrap value between them of 100. The grouping of Alphaproteobacteria with Gammaproteobacteria is consistent with the near equal multiple sequence alignment conservation that was observed and the Dot-plot comparison between the two subclasses, suggesting that the Hsp20 heat shock proteins among Alphaproteobacteria and Gammaproteobacteria share a more recent last common ancestor than do any of the other subclasses. In addition, the grouping of Deltaproteobacteria and Epsilonproteobacteria suggests that Hsp20 heat shock proteins among these two subclasses share a more recent common ancestor than they do with any of other Proteobacteria subclasses. In spite of the separation of

the subclass clades among each other, the separation of Hsp20 between the subclasses is consistent with the Dot-plot comparisons. This strongly suggests that Hsp20 evolved separately within the five subclasses to adapt to the needs of the particular subclass. Furthermore, analysis of proteobacterial and non-proteobacterial sequences corroborated the possible common ancestry of proteobacterial Hsp20 proteins by producing unique clades.

Future Perspectives

It is clear that Hsp20 plays an important role among Proteobacteria from the fact that it is not only present in several strains of Proteobacteria but also that the Hsp20 protein sequences are highly conserved among the five subclasses of Proteobacteria. In addition, it is evident that Hsp20 shares a common ancestor within and between the five subclasses of Proteobacteria. However, it is unclear at this point what the last common ancestor among Proteobacteria Hsp20 heat shock proteins was and therefore further research is needed to answer vital questions in this regard. Our future work may focus on finding a link between Hsp20 in Proteobacteria and Hsp20 in the rest of the bacteria phylum, or going even farther, and finding a relation among Archaeobacteria and Eukaryotes. This could be achieved through further bioinformatic analyses that encompass large number of sequences from Bacteria, Archaea and Eukarya domains.

References

- Chandra, S., Vlk, J., and Kapatral, V. (2006). Comparative insect mitochondrial genomes: Differences despite conserved genome synteny. *African Journal of Biotechnology* 5, 1308-1318.
- Davies, E.L., Baclear, M.M.F.V.G., Marshall, M.J., Andrew, S.M., and Williams, J.H.H. (2006). Heat shock proteins form part of a danger signal cascade in response to lipopolysaccharide and GroEL. *Clinical and Experimental Immunology* 145, 183-189.
- Dedysh, S., Ricke, P., and Liesack, W. (2004). NifH and NifD phylogenies: an evolutionary basis for understanding nitrogen fixation capabilities of methanotrophic bacteria. *Microbiology* 150, 1301-1313.
- Kim, K.K., Kim, R., and Kim, S.H. (1998). Crystal structure of a small heat-shock protein. *Nature* 394, 595-599.
- Landes, C., Henaut, A., and Risler, J. (1992). Dot-plot comparison by multivariate analysis (DOCMA): a tool for classifying protein sequences. *Bioinformatics* 9, 191-196.
- Lee, K., Liu, C., Anzai, Y., Kim, H., Aono, T., and Oyaizu, H. (2005). The hierarchical system of the 'Alphaproteobacteria'; Xanthobacteraceae fam. nov. and Erythrobacteraceae fam. nov. *International Journal of Systematic*

- and *Evolutionary Microbiology* 55, 1907-1919.
- Ludwig, W., and Klenk, H.P. (2001). Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics. In Bergey's Manual of Systematic Bacteriology. Garrity, G.M. (ed.). New York, USA: Springer, pp.49-66.
- Meeks, M., Ripley, M., Jin, Z., and Rembold, C. (2005). Heat shock protein 20-mediated force suppression in forskolin-relaxed swine carotid artery. *American Journal of Physiological - Cell Physiology* 288, C633-C639.
- Miroschnichenko, M., L' Haridon, S., Schumann, P., Spring, S., Bonch-Osmolovskaya, E., Jeanthon, C., and Stackebrandt, E. (2004). *Caminibacter profundus* sp. Nov., a novel thermophile of Nautiliales ord. nov. within the class 'Epsilonproteobacteria', isolated from a deep-sea hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology* 54, 41-45.
- Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Buillard, V., Cerutti, L., Copley, R., Courcelle, E., Das, U., Daugherty, L., Dibley, M., Finn, R., Fleischmann, W., Gough, J., Haft, D., Hulo, N., Hunter, S., Kahn, D., Kanapin, A., Kejariwal, A., Labarga, A., Langendijk-Genevaux, P.S., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Nikolskaya, A.N., Orchard, S., Orengo, C., Petryszak, R., Selengut, J.D., Sigrist, C.J., Thomas, P.D., Valentin, F., Wilson, D., Wu, C.H., and Yeats, C. (2007). New developments in the InterPro database. *Nucleic Acids Res.* 35, D224-D228.
- Oganesyan, N., Ankoudinova, I., Kim, S., and Kim, R. (2007). Effect of osmotic stress and recombinant protein overexpression and crystallization. *Protein Expression Purification* 52, 280-285.
- Page, R. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12, 357-358.
- Rembold, C., Foster, D., Strauss, J., Wingard, C., and Van Eyk, J. (2000). cGMP-mediated phosphorylation of heat shock protein 20 may cause smooth muscle relaxation without myosin light chain dephosphorylation in swine carotid artery. *The Journal of Physiology* 524, 865-878.
- Rice, P., and Longden, I. (2000). Emboss: the European Molecular Open Software Suite. *Trends in Genetics* 16, 276-277.
- Schlesinger, M. (1990). Minireview: Heat Shock Proteins. *The Journal of Biological Chemistry* 265, 12111-12114.
- Suzuki, A., Sugiyama, Y., Hayashi, Y., Nyu-I, N., Yoshida, M., Nonaka, I., Ishiura, S., Arahata, K., and Ohno, S. (1998). MKBP, a novel member of the heat shock protein family, binds and activates the myotonic dystrophy protein kinase. *The Journal of Cell Biology* 140, 1113-1124.
- Thompson, J., et al. (1994). CLUSTAL W: improving sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680.
- Thompson, J., Gibson, T., Plewniak, F., Jeanmougin, F., and Higgins, D. (1997). The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876-4882.
- Trotter, E., Kao, C., Berenfeld, L., Botstein, D., Petsko, G., and Gray, J. (2002). Misfolded proteins are competent to mediate a subset of the responses to heat shock in *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry* 277, 44817-44825.