Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 244 (2007) 127–132
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Adaptation of proteins to different environments: A comparison of proteome structural properties in *Bacillus subtilis* and *Escherichia coli*

Sayed-Amir Marashi[a,*], Reza Behrouzi[a,1], Hamid Pezeshk[b]

[a]*Department of Biotechnology, University College of Science, University of Tehran, Enghelab Ave., Tehran, Iran*
[b]*Center of Excellence in Biomathematics, School of Mathematics, Statistics and Computer Sciences, University College of Science, University of Tehran, Enghelab Ave., Tehran, Iran*

Received 8 May 2006; received in revised form 10 July 2006; accepted 21 July 2006
Available online 24 July 2006

Abstract

We compared amino acid solvent accessibilities and helix propensities in data sets of *Escherichia coli* and *Bacillus subtilis* proteins. These species reside in very different environments and hold very different physiological properties. From the observations, it was proposed that the cytoplasm of *B. subtilis* is more ion-rich compared to the cytoplasm of *E. coli*, which might be more hydrophobic; therefore, during evolution these differences have resulted in different protein folding tracks. Such inherent differences imply that the results of bioinformatic analyses of protein structures might depend on the species from which the proteins are picked. It is also suggested that different cytoplasmic environments cause *E. coli* and *B. subtilis* to be appropriate for expression of distinct types of proteins.

Keywords: Protein evolution; Accessible surface area; Helix propensity; Protein expression hosts

1. Introduction

It is well-known that environmental conditions (e.g. temperature, pH, ionic strength, etc.) can affect peptide and protein structures in vitro. The effect of environmental conditions has been compensated for by evolution of sequences and structures that are best fitted to the living condition. For example, the frequency of certain amino acids can be significantly different in thermophilic and mesophilic organisms (Singer and Hickey, 2003), although a general pattern is not yet suggested (Vieille and Zeikus, 2001).

Each cell (or cellular compartment) can be considered as an ''island'' of biological macromolecules, enclosed by a membrane, which separates it from the surrounding environment. The environmental conditions inside the cells are dependent on cell types and specifically membrane proteins; hence, it is not odd to observe different evolutionary conditions applied on proteomes in different cell types. While the effect of temperature on the evolution of proteins has been studied vigorously (Vieille and Zeikus, 2001; Facchiano et al., 1998), to the best of our knowledge the effect of other environmental conditions has little been considered so far.

Here, we compared protein structure properties in two different microorganisms: *Bacillus subtilis* and *Escherichia coli*. They are both prokaryotes and lack internal compartments (e.g. nucleus, lysosome, etc.); this property makes them suitable for our study, since the internal organelles might hold different internal environments. Moreover, optimal growth temperatures of *B. subtilis* and *E. coli* are close (38.5 °C vs. 37 °C, respectively). Furthermore, *B. subtilis* is Gram-positive, while *E. coli* is Gram-negative, and their phylogenetic distance is substantial; hence, it is reasonable to expect differences between their cytoplasmic environments. For both organisms, there are enough resolved protein structures available to enable us to perform statistical analyses.

*Corresponding author. Tel./fax: +98 21 6649 1622.
E-mail addresses: marashie@khayam.ut.ac.ir (S.-A. Marashi), behroozi@khayam.ut.ac.ir (R. Behrouzi), pezeshk@khayam.ut.ac.ir (H. Pezeshk).
[1]Present address: New Ideas Research Institute, 11 Proshat Ave., Motahari St., Tehran, Iran.

0022-5193/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2006.07.021

2. Materials and methods

2.1. Non-redundant protein data sets

Eighty-four protein structures determined with resolutions of <2.5 Å and with sequence identity $<30\%$, were chosen by searching for *B. subtilis* in the SOURCE section of PDB entries, followed by a culling procedure using PISCES server (Wang and Dunbrack, 2003; available from: http://www.fccc.edu/research/labs/dunbrack/pisces). The method was repeated for *E. coli*, resulting in 456 proteins.

2.2. Extraction of amino acid properties

Secondary structures and accessible surface areas (*ASA*) of individual amino acids in selected proteins were extracted from the DSSP databank (Kabsch and Sander, 1983; available from: http://www.sander.ebi.ac.uk/dssp/). Amino acid frequencies in the proteins and in α-helices were counted. If there was no α-helix in a protein, its amino acid count was excluded from calculations.

Propensity of amino acid X for structure S is defined as the frequency of its occurrence in the structure $S(f^S)$ divided by its frequency in all protein structures (f^{all}):

$$P_S(X) = f^S(X)/f^{all}(X). \quad (1)$$

Eq. (1) is conventionally employed for a data set of proteins. Here, we used it at the level of single proteins; then we averaged $P_S(X)$ values over the whole data set. Fig. 1 illustrates how good the average of such a distribution correlates with conventional propensity in our data sets. Applying a distribution for $P_S(X)$ enables us to employ statistical inference methods to study the significance of (probable) differences in propensity values of an amino acid for a certain structure in different organisms. Furthermore, it is reasonable to assume that the mean value is a better estimator of in vivo and in vitro propensity values, since overall data set propensity calculation clearly results in a considerable information loss.

2.3. Statistical inference

All of the statistical inferences were performed using MINITAB® 14. For the comparison of distributions, in most cases a two-sample t-test is appropriate to examine whether the means of the two distributions are significantly different (the differences were assumed to be significant if $p \leqslant 0.05$). However, in case of highly skewed distributions, Mann–Whitney test, which is a non-parametric analysis, was used to check the significance of difference between median values. We took $p \leqslant 0.1$ in the Mann–Whitney test to conclude whether the differences were significant.

3. Results and discussion

3.1. Amino acid ASAs are generally greater in B. subtilis, suggesting the existence of a more stabilizing cytoplasm in this organism

During evolution, proteins might become more flexible in a stabilizing environment compared to the proteins in destabilizing condition, which are selected to resist such condition; the latter proteins are expected to be more compact with a higher packing in their core (Li et al., 1998). As a result of their flexibility/rigidity states, their average magnitudes of amino acid solvent accessibilities are expected to be different (see Knapp et al., 1999): in an in vitro structure determination system, the more the flexibility of proteins, the more the observed *ASA* of residues. In other words, for two protein data sets, if calculated *ASA* of amino acids is significantly higher in one data set, one may conclude that the proteins of this data set originate from a more stabilizing environment.

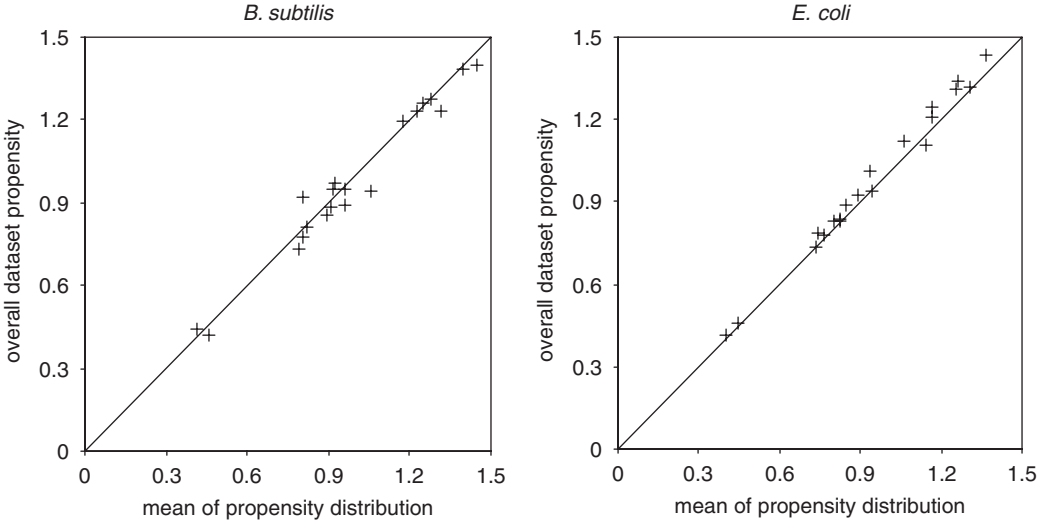

Fig. 1. Overall data set propensities (conventional propensity) vs. the mean of propensity distributions for *B. subtilis* and *E. coli* protein data sets.