Polish Infrastructure for Supporting Computational Science in the European Research Space
PL-Grid

IS THE "FUZZY OIL DROP" MODEL OF GENERAL CHARACTER?

T. Jadczyk[1], M.Bubak[1,2], and I.Roterman[3]

1) Academic Computer Center CYFRONET, Nawojki 11, 30-950 Krakow, PL
2) Institiute of Computer Science, AGH, Mickiewicza 30, 30-059 Krakow, PL
3) Department of Bioinformatics and Telemedicine, Jagiellonian University – Medical College, Lazarza 16, 31-530 Krakow, PL

Introduction

The “fuzzy oil drop” model appeared to describe well the structure of the hydrophobic core in many proteins. “Fuzzy oil drop” model assumes the hydrophobicity distribution in proteins to be accordant with the 3-D Gauss function differentiating the hydrophobicity density from the highest one in the center of the molecule and zero level on the surface [1].
The question was given – Are there many proteins representing the structure of “fuzzy oil drop” character in respect to hydrophobicity distribution ?
To answer this question the large scale calculation was performed using the complete set of proteins deposited in PDB aimed to identify the proteins representing assumed structure. The structural unit was defined in two ways: the protein complexes were taken as the one unit, each chain was taken separately.

Materials and Methods

Expected (theoretical) hydrophobicity distribution (3D Gauss function):

$$\tilde{H}_i^t = \frac{1}{\sum \tilde{H}^t} \exp\left(\frac{-(x_i-\bar{x})^2}{2\sigma_x^2}\right) \exp\left(\frac{-(y_i-\bar{y})^2}{2\sigma_y^2}\right) \exp\left(\frac{-(z_i-\bar{z})^2}{2\sigma_z^2}\right)$$

$(\bar{x}, \bar{y}, \bar{z})$ Geometric center of the molecule

$(\sigma_x, \sigma_y, \sigma_z)$ Molecule size

Observed hydrophobicity (Levitt function)

$$\tilde{H}_i^e = \frac{1}{\sum \tilde{H}^e} \begin{cases} \left(H_i^r + H_j^r\right)\left(1 - \frac{1}{2}\left(7\left(\frac{r_{ij}}{c}\right)^2 - 9\left(\frac{r_{ij}}{c}\right)^4 + 5\left(\frac{r_{ij}}{c}\right)^6 - \left(\frac{r_{ij}}{c}\right)^8\right)\right), \text{for } r_{ij} \leqslant c \\ 0, \text{for } r_{ij} > c \end{cases}$$

$\left(H_i^r + H_j^r\right)$ sum of hydrophobicity (according to hydrophobic scale) of interacting residues

$c = 9\,\mathring{A}$ cutoff value, Levitt [3]

The similarity of both distribution was calculated according to Kullback-Leibler distance entropy [4]:

$$O/T = \sum_{i=1}^{N} O_i \cdot \log_2 \frac{O_i}{T_i}$$

O/R value expresses the distance between observed distribution and the random one:

$$O/R = \sum_{i=1}^{N} O_i \cdot \log_2 \frac{O_i}{R_i}, R_i = \frac{1}{N}$$

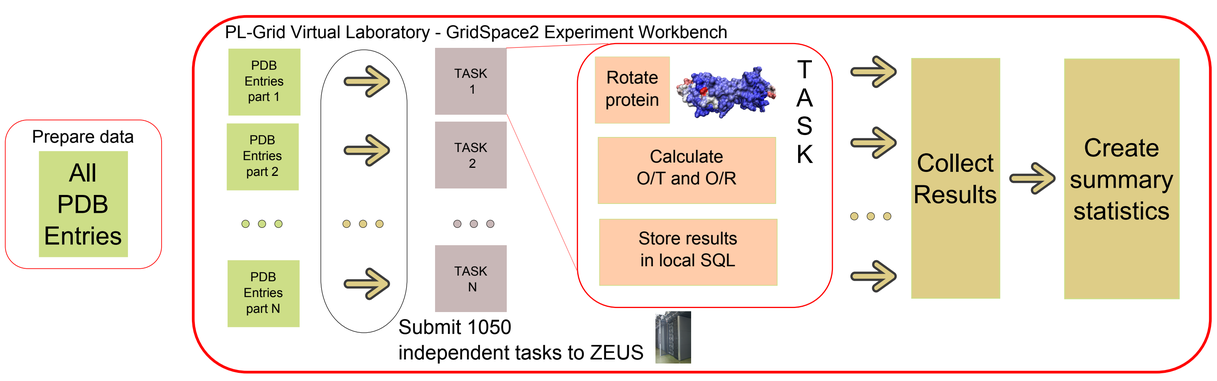
Data: Complete set of proteins present in PDB (December 2010), nucleic acid molecules excluded

Search: The protein representing the relation:

O/T < O/R

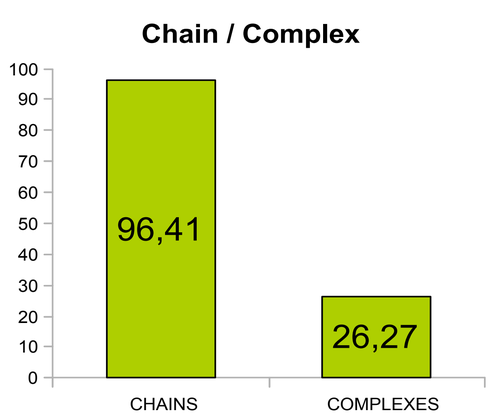
was taken as the protein representing hydrophobic core of the "fuzzy oil drop" character

Experiment

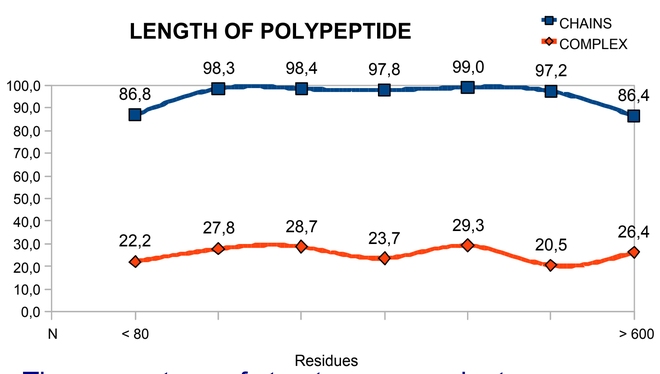


Experiment were executed in GridSpace2 Experiment Workbench. Complete PDB database (over 70000 entries) was split into 1050 subsets. Each set was an input for FOD-computing task. Tasks were run on ZEUS cluster. Each task created small database with results for analyzed set of proteins. Results from all databases were gathered and then final statistics were created.

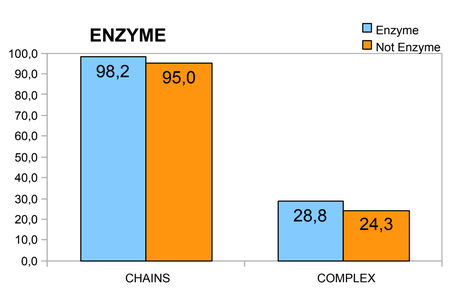
Results



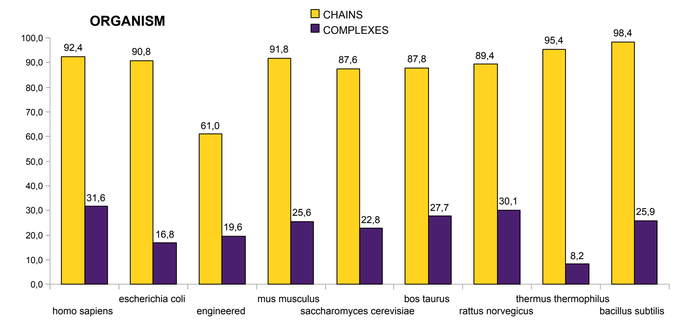
The percentage of structures accordant with the assumed model



The percentage of structures accordant with the assumed model for proteins classified according to the polypeptide chain length



The enzymes characteristics - the percentage of structures accordant with the assumed model



The percentage of proteins of the structure accordant with the assumed model classified for selected organism

Conclusions

The unexpectedly large percentage of proteins representing the structures accordant with assumed model is probably due to redundancy of PDB database.
The non-redundant set shall be extracted form PDB to make the frequency of “fuzzy oil drop” model structures reliable

References

[1] Konieczny L, Brylinski M, Roterman I. (2006) Gauss-function-Based model of hydrophobicity density in proteins. In Silico Biol. 6: 15-22
[2] Kauzman W. (1959) Some factors in the interpretation of protein denaturation. Adv Protein Chem 14: 1–63
[3] Levitt M, (1976) A simplified representation of protein conformations for rapid simulation of protein folding. J. Mol. Biol. 104 (1): 59–107
[4] Nalewajski RF. (2006) Information theory of molecular systems. Amsterdam [etc.]: Elsevier, ISBN 978-0-444-51966-5
[5] E. Ciepiela, D. Harezlak, J. Kocot, T. Bartynski, M. Kasztelnik, P. Nowakowski, T. Gubała, M. Malawski, M. Bubak (2010) Exploratory Programming in the Virtual Laboratory. Proceedings of the International Multiconference on Computer Science and Information Technology pp. 621–628

INNOVATIVE ECONOMY
NATIONAL COHESION STRATEGY

The Project is co-funded by the European Regional Development Fund as part of the Innovative Economy program.