Article

Chaperonin Structure – The Large Multi-Subunit Protein Complex

Mateusz Banach [1,2], Katarzyna Stąpor [2,3] and Irena Roterman [1,2,*]

[1] Department of Bioinformatics and Telemedicine – Jagiellonian University, Collegium Medicum, Lazarza 16, 31-531 Krakow, Poland; E-Mail: myroterm@cyf-kr.edu.pl

[2] Faculty of Physics, Astronomy and Applied Computer Science - Jagiellonian University, Reymonta 4, 30-059 Krakow, Poland; E-Mail: mateusz.banach@uj.edu.pl

[3] Silesian Technical University, Institute of Computer Science, Akademicka 16 44-100 Gliwice, Poland; E-Mail: katarzyna.stapor@polsl.pl

* Author to whom correspondence should be addressed; E-Mail: myroterm@cyf-kr.edu.pl; Tel. +48-12-619-96-94; Fax: +48-12-619-96-93

Abstract: The multi sub-unit protein structure representing the chaperonins group is analyzed with respect to its hydrophobicity distribution. The proteins of this group assist protein folding supported by ATP. The specific axial symmetry GroEL structure (two rings of seven units stacked back to back - 524 aa each) and the GroES (single ring of seven units - 97 aa each) polypeptide chains are analyzed using the hydrophobicity distribution expressed as excess/deficiency all over the molecule to search for structure-to-function relationships. The empirically observed distribution of hydrophobic residues is confronted with the theoretical one representing the idealized hydrophobic core with hydrophilic residues exposure on the surface. The observed discrepancy between these two distributions seems to be aim-oriented, determining the structure-to-function relation. The hydrophobic force field structure generated by the chaperonin capsule is presented. Its possible influence on substrate folding is suggested.

Keywords: Protein folding, hydrophobicity, chaperonin.

1. Introduction

It has been discovered that the protein folding process is guided by additional molecules directing the structural changes toward the correct native form. Molecular chaperones are the proteins which bind and stabilize unfolded or partially folded proteins, thereby preventing them from being degraded [1-4]. Only a certain subset of cellular proteins undergo the folding process accompanied by the chaperonins, which are large protein constructs which directly facilitate the protein folding process with participation of ATP molecules. Chaperonin exists as a back-to-back linked double-ring complex. The symmetric (7-fold) rings of GroEL interact with the co-chaperonin GroES. The mechanism of ATP binding and its collaboration with internal structural changes in *cis-* (called chains A-G in this paper) and *trans*-rings (chains H-N in this paper) reveals the functioning algorithm of the folding machine. Each part is responsible for a specific element of this algorithm [1-4].

The object of the analysis was the chaperonin used as an example to search for possible mechanisms for the generation of such large constructions with a nano-machine character. The question may be asked, how do these proteins become folded? How do they influence the substrate folding?

An attempt to find answers to these questions on the basis of the "fuzzy oil drop" model [5-13] has been undertaken and is presented in this paper. The assumed model helps clarify to what extent the hydrophobicity distribution may help, support or even direct the protein folding. The specific role of chaperonins as nano-machines with two functions: first as "holder" – complexation of the folding molecule to prevent the misfolding and secondly, as a "folder" – directing the folding process to the generation of proper native structure is the object of the work presented [14].

The structure of the GroEL-GroES-(ADP)7 complex is a very good example to test the "fuzzy-oil-drop" model applicability to recognize the structural and functional specificity of the protein under consideration due to multi-subunit protein assembly comprising rings of subunits stacked back to back. The presence of ADP ligands makes possible the analysis of ligand docking to this molecule. The complete complex is presented in parts distinguishing the structural and functional fragments of this multi-subunit construction.

2. Materials and Methods

2.1. Data

The 1AON – the structure of the object under consideration has been taken from the PDB (deposit 1AON) [15].

2.2. "Fuzzy-oil-drop" Model

It is assumed that the presence of an external force field of hydrophobic character expressed by the three-dimensional Gauss function is able to direct the protein folding toward hydrophobic core

generation, with the simultaneous exposure of hydrophilic residues toward the surface of the protein molecule.

The external force field is represented by the three-dimensional Gauss function. The value of the Gauss function (traditionally interpreted as probability density value) is assumed to represent the hydrophobic density in the protein body. The hydrophobicity density can be calculated for any point the space covering the protein molecule.

The three-dimensional Gauss function is given as follows:

$$Ht_j = \frac{1}{Ht_{sum}} \exp\left(\frac{-\left(x_j - \bar{x}\right)^2}{2\sigma_x^2}\right) \exp\left(\frac{-\left(y_j - \bar{y}\right)^2}{2\sigma_y^2}\right) \exp\left(\frac{-\left(z_j - \bar{z}\right)^2}{2\sigma_z^2}\right)$$

The value of Ht_j is assumed to represent the hydrophobicity distribution at a particular point belonging to the protein body. The hydrophobicity maximum is localized in the center of the ellipsoid ($\bar{x}, \bar{y}, \bar{z}$) and decreases in a distance-dependent manner according to the Gauss function. The mean value at which the Gauss function reaches its maximum is localized at the (0.0.0) point in a coordinate system. The standard deviation values $\sigma_x, \sigma_y, \sigma_z$ calculated for each dimension (axis) separately represent the size of the drop which depends on the length of the polypeptide under consideration [7].

The *j-th* grid point, for which the hydrophobicity is calculated, represents the effective atom position (averaged position of the side chain including Cα atoms) making possible attachment of a particular hydrophobicity density to a particular amino acid.

Before the external hydrophobic force field can be defined for any protein molecule, it shall be oriented in the space according to following procedure:

1. the geometric centre of the molecule shall be localized in the center of coordinate system.

2. the longest distance between two residues (represented by the effective atom – geometric centre of side chain of the amino acid) shall overlap one of the axes (say the X-axis).

3. the molecule shall be rotated around the X-axis to orient the longest inter-projections (on the YZ plane) distance along the Y-axis.

4. the linear size (the maximum inter-atomic distance along the X, Y, and Z axes) increased by 9 Å in each direction (the cutoff distance for hydrophobic interaction) makes possible calculation of $\sigma_x, \sigma_y, \sigma_z$

This is how the geometric parameters of protein molecule can be interpreted according to the Gauss function.

Taking into account the high symmetry of the system under consideration, a user defined orientation of the coordinate system is necessary. Thus, the initial orientation determining the X-axis is defined by the position of the symmetrical units, so the averaged position of the top elements and averaged position of bottom elements (user-defined) determine the initial orientation of the molecule. The user-defined orientation of the molecule is available and necessary for any protein molecules or complexes before the "fuzzy oil drop" model can be applied. The X-axis defined this way is simultaneously the 7-fold symmetry axis.

The empirical (observed) distribution of hydrophobicity can be different than the idealized one. The empirical hydrophobicity distribution can be calculated according to Levitt function [16]:

$$\widetilde{Ho}_j = \frac{1}{\widetilde{Ho}_{sum}} \sum_{i=1}^{N} (H_i^r + H_j^r) \left\{ \begin{array}{l} \left[1 - \frac{1}{2}\left(7\left(\frac{r_{ij}}{c}\right)^2 - 9\left(\frac{r_{ij}}{c}\right)^4 + 5\left(\frac{r_{ij}}{c}\right)^6 - \left(\frac{r_{ij}}{c}\right)^8 \right) \right] \text{ for } r_{ij} \leq c \\ 0 \text{ for } r_{ij} > c \end{array} \right.$$

where Ho_j represents the empirical hydrophobicity value characteristic for the *j-th* grid point, H_i^r represents the hydrophobicity characteristic of the *i*-th amino acid, r_{ij} is the distance between the *j*-th grid point and *i*-th effective atom in the amino acid, and *c* expresses the cutoff distance, which has a fixed value of 9.0 Å following the original paper [16].

The continuity of the Gauss function allows calculation of the hydrophobicity density in any point in space (in the protein body). So any point can be treated as a grid point. It can also be the position of an effective atom in particular. This is why the index *j* may represent the position of *j*-th residue, as it is taken in this work. Ho_{sum} represents the sum of all the grid points hydrophobicity. Any hydrophobicity scale may be applied to calculate the observed density of hydrophobicity [16-21].

Since both values are standardized (the coefficient $\frac{1}{H_{sum}}$) the differences between theoretical and empirical values expressing hydrophobicity density in a particular point of space (position of the side chain of *i*-th amino acid) can be calculated according to:

$$\Delta \widetilde{H}_i = Ht_i - Ho_i$$

The profile of $\Delta \widetilde{H}_i$ (expressing the value of difference for each amino acid) reveals some maxima, which are related to hydrophobicity deficiency. The hydrophobicity deficiency ($\Delta \widetilde{H}_i > 0$) seems to represent the potentially ligand binding site. The potential ligand may adhere in this area as the complementary element compensating the hydrophobicity deficiency producing in effect the regular smoothed hydrophobicity distribution. The values of negative $\Delta \widetilde{H}_i$ represent the area of hydrophobicity higher than expected. The area of such characteristics, when localized on the surface of protein seems to represent the potential area responsible for protein-protein complex creation.

The profile of $\Delta \widetilde{H}_i$ values can show the discrepancy between the idealized and empirical hydrophobicity density distribution revealing the fragments (or individual residues) representing the hydrophobicity excess ($\Delta \widetilde{H}_i$ negative) and hydrophobicity deficiency ($\Delta \widetilde{H}_i$ positive). It is expected that the hydrophobicity irregularity versus idealized one may express localization of biological function-related area in the protein body.

2.3. Protein Partitioning

The 1AON is a quite large and complex protein molecule. This is why different approaches have been applied.

1. The complete molecule was treated as one uniform "drop" – the orientation of molecule was according to its 7-fold symmetry axis (the X-axis).
2. The chaperonin molecule represents three levels organization: two stacked rings (Gro-EL) with the third one (Gro-ES) as "cap".

3.Each ring (Gro-EL) and the "cap" (Gro-ES) is composed of seven identical polypeptide chains which are also treated as structural units (chains A-H, G-N and O-U).

4.The polypeptide chains belonging to Gro-EL evidently represent the two-domain construction. This is why each such domain is treated as independent individual part and treated as a "drop".

Figure 1. The chaperonin structure. A – Gro-ES (cap) built of the chains O-U – chain O distinguished in white, Gro-EL built of two rings: *cis* - chains A-G (chain A distinguished in red), *trans* – chains H-N (chain H in yellow); B – the axial view showing the seven-fold symmetry of the complex. The chains colored as in A.
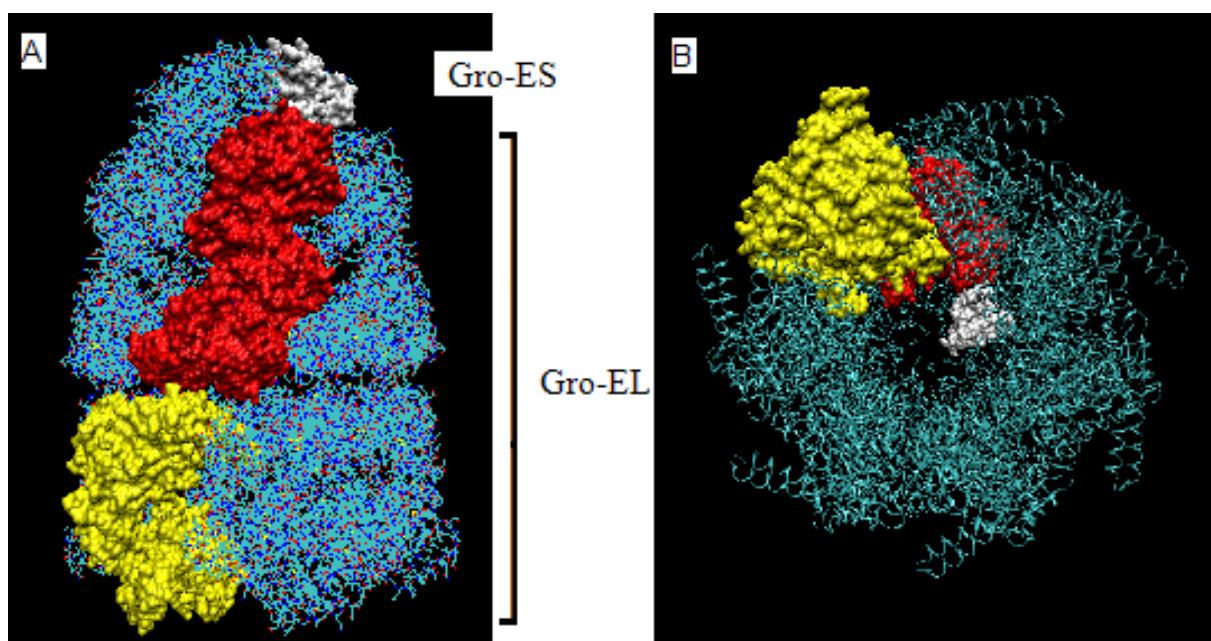
The partitioning shown above is aimed to define the folding unit and possible path leading to complex generation (Figure 1). 1AON represents the chaperonin additionally complexed with ADP and some Mg^{+2} ions. The characteristics of localization of these ligands will be analyzed with respect to the construction of the "fuzzy-oil-drop".

2.4. Identification of the Non-Bonding Interactions

The cut-off of 3.9 Å was taken to identify the residues interacting with ions, ligand (ADP) and protein (chain). The cut-off value has been taken according to the criteria applied in PDBsum [22] data base to make possible the comparison of results.

2.5. Implementation

All results have been obtained using our own program, written in Python [23] programming language. The program has been divided into multiple subroutines to ensure flexibility and diversity in dealing with PDB [24] files, which was required to work on protein partitions described above.

The first routine reads the input PDB file, removes all water residues and classifies non-empty chains into three groups: protein, nucleic and "hetero". By "hetero" we mean neither protein nor

nucleic acid chains, presumably containing only heteroatoms. All operations regarding PDB files are conducted using methods implemented in the Biopython [25] library.

The second routine is the core of the "fuzzy oil drop" evaluation: it performs an extraction of the residue subsets from the protein chains that the user is interested in and computes their effective atoms, which form the "drop". The drop is placed in the origin and rotated to achieve a desired spatial orientation, as stated in the model description. After the drop size becomes known, theoretical and observed hydrophobicity are evaluated using optimized array operators from Numpy [26] (Numerical Python) library.

Identification of interactions is conducted using a fast k-d tree algorithm implemented in the Biopython library by checking position of each atom from selected residues against all atoms in file. Each neighbor atom placed within given radius of 3.9 Å is unfolded into parental residue and then classified as either protein, nucleic, ligand or ion contact. Contacts with protein residues from same chain are ignored.

The last routine sets output values for every residue (same for each atom) in PDB file: normalizes hydrophobicity discrepancy by overwriting the beta-factor column and contact type by overwriting the occupancy column. Plotting of graphical representation of results is made using the Matplotlib library [27].

2.6. Clustering Analysis

The agglomerative hierarchical clustering "ahc" algorithm has been applied to analyze the agreement between subjective interpretation of the "fuzzy oil drop" model applicability and objective discrimination of elements representing the common characteristics [28]. Clustering is the assignment of objects (feature vectors) into groups called clusters so that objects from the same cluster are more similar to each other than objects from different clusters. The short description of "ahc" algorithm is as follows. Suppose we have a data set $X = \{x_1,...,x_N\}$ of N objects to be clustered and a user-defined distance measure $d(x_i, x_j)$ to state how similar the two objects x_i and x_j are, for any i and j. The "ahc" algorithm starts off clustering this data by putting each of the data objects x_i in a singleton cluster $C_{\{i\}} = \{x_i\}$ and then keeps on joining the closest pair of clusters $C_{\{i\}} \cup C_{\{j\}} = C_{\{i,j\}}$ until there is only one large cluster $C_{\{1,2,...,N\}}$. Simultaneously, the clustering tree (dendrogram) is built from leaves towards root, where merging of clusters is depicted as a common parent for two sub-trees. The distance between clusters $C_i = C_{\{i_1,...,i_p\}}$ and $C_j = C_{\{j_1,...,j_r\}}$ can be measured in different ways. Three popular methods are single, complete and average linkage:

$$d_{\min}(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y) \qquad \text{(single)}$$

$$d_{\max}(C_i, C_j) = \max_{x \in C_i, y \in C_j} d(x, y) \qquad \text{(complete)}$$

$$d_{avg}(C_i, C_j) = \frac{1}{|C_i| \cdot |C_j|} \sum_{x \in C_i, y \in C_j} d(x, y) \qquad \text{(average)}$$

where $|C|$ denotes the number of objects in a cluster C. Thus, the distance between clusters is the minimum (or maximum or average) of the distances between one object from the first and another

from the second cluster. Once the dendrogram is generated for assumed *k* clusters, the procedure cuts the *k-1* longest links in a dendrogram. The described "ahc" algorithm was implemented as a function in the software package Matlab v.7.

First, the implemented "ahc" algorithm was used to cluster separately objects in each of the files according to the partition (section 2.3) into *k=2* groups. Each object x_i in these files is represented as a four dimensional feature vector $x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})$, where x_{i1}, x_{i2}, x_{i3} are the spatial coordinates *x,y,z* of the object x_i while x_{i4} is its estimated variable $\Delta\widetilde{H}$. For each object x_i its correct classification (being label '0' or '1' (i.e. "0" – means residue engaged in protein-protein interaction, '1' denotes the residue not engaged in the protein-protein interaction) is known (according to PDBSum criteria). The clustering result for each partitioning form (as described in Section 2.3) compared with the correct classification allowed the calculation of the clustering performance defined as the number of objects correctly assigned to a groups divided by the total number of objects in a file.

3. Results

The results may be summarized as follows:

A molecule of high complexity like 1AON is difficult to describe in a simple way.

The partitioning described in Methods part was introduced.

The chains A-H appeared identical taking the $\Delta\widetilde{H}$ profiles as the criterion (results not shown) as well as the chains G-N. The chains present in Gro-ES part also appeared to represent the identical $\Delta\widetilde{H}$ profiles. This is why the chains A, G and O were taken to represent particular rings.

### 3.1. Hydrophobicity Density Irregularity in Chaperonin Molecule

The $\Delta\widetilde{H}$ profile of chain A taken to represent the chains A-G is shown in Figure 2. The residues engaged in ligand binding are distinguished as well as the residues involved in protein-protein complexation. According to "fuzzy oil drop" the $\Delta\widetilde{H}$ minima representing the hydrophobicity excess on the surface of protein is assumed to be potential protein-protein contact area. The $\Delta\widetilde{H}$ maxima (hydrophobicity deficiency) in domains distinguished in chain A are observed in accordance with the interpretation of "fuzzy oil drop" model being engaged in ligand (ADP) binding.

The residues of local $\Delta\widetilde{H}$ minima are engaged in protein-protein interaction, particularly well seen in the domain 192-371. This accordance disappears step-wise taking the parts larger than the domain as the unit for drop construction localizing the ligand molecule even in the area of negative $\Delta\widetilde{H}$ values. Both explanations for ligand binding are possible as the complementary accordance of target molecule and ligand. This observation is important for the interpretation of the "fuzzy oil drop" model.

Analysis of the $\Delta\widetilde{H}$ profiles for chain A suggests that the construction of domains present in the chain A structure appears highly accordant to the interpretation of "fuzzy oil drop" model. The local $\Delta\widetilde{H}$ minima are engaged in protein-protein complexation (particularly in the domain 192-371, while ligand molecule occupies the position of residues of local $\Delta\widetilde{H}$ maxima. The characteristics of ligand binding residues in the Gro-EL and GroEl-ES complex get changed.

The characteristics of residues engaged in P-P or ligand binding changes due to different relative localization of these residues versus in the "oil drop" construction.

Figure 2. The $\Delta\tilde{H}$ profiles representing the irregularity of hydrophobicity density distribution in chain A calculated for units defined according to molecule partition: domains, chain, Gro-EL fragment and entire complex GroEL-GroES. The differences are the result of different orientation of chain A versus the oil drop definition. The yellow symbols represent the residues engaged in ligand binding, the pink symbols distinguish the residues engaged in protein-protein (P-P) interaction.
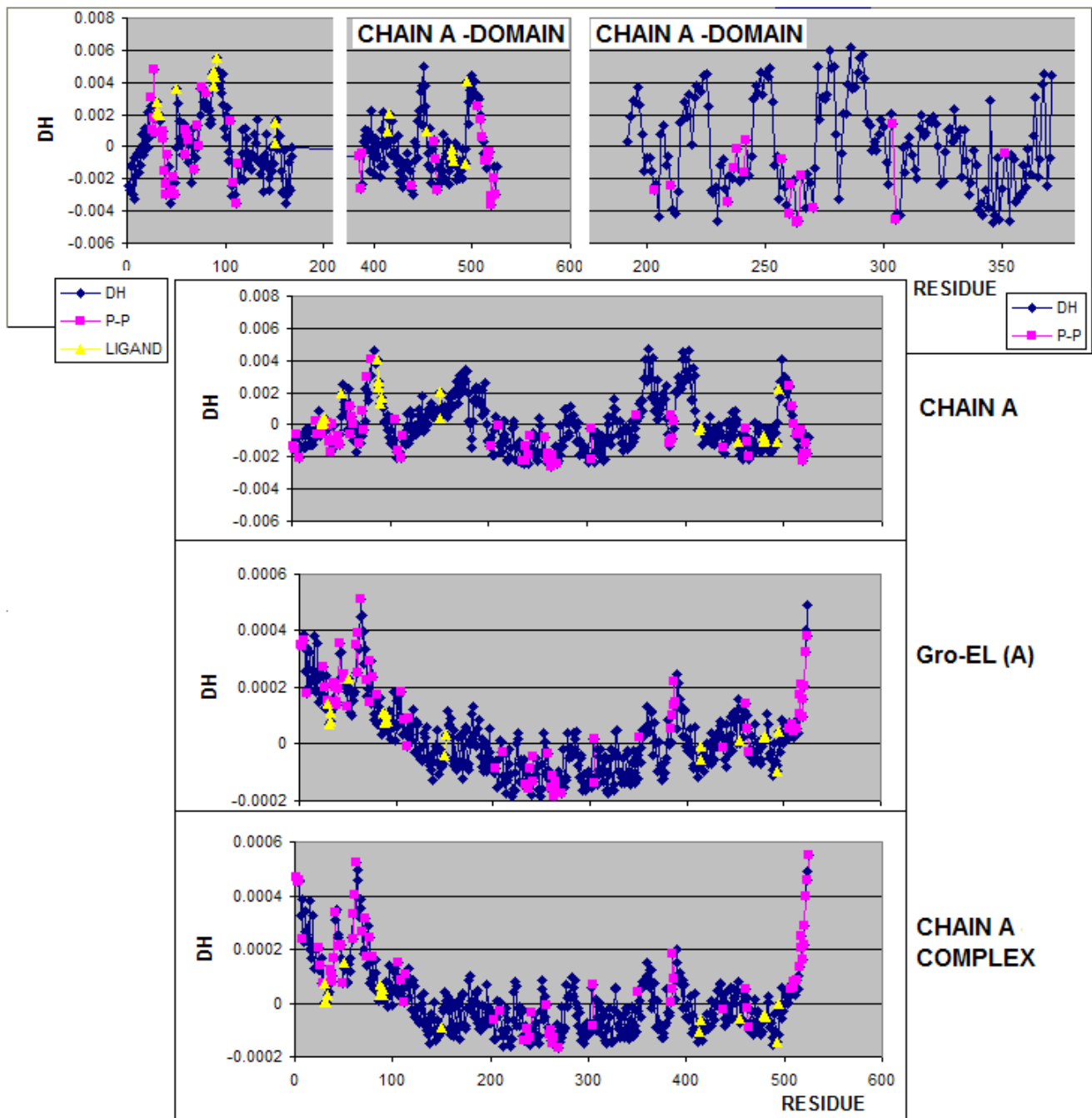
The same analysis performed for chain H representing the 7-fold system of the chains H-N is presented in Figure 3. There is no difference between the amino acids sequence in chains A-G and H-N, although some small differences of $\Delta\widetilde{H}$ profiles are observed. It is the possible result of different complexation conditions (the chains H-N have no contact with the molecule Gro-ES) and no ligand is complexed to this ring.

Figure 3. The $\Delta\widetilde{H}$ profile representing the irregularity of hydrophobicity density distribution in chain H calculated for units defined according to molecule partition: domains, chain, Gro-EL fragment and entire complex GroEL-GroES. The differences are the result of different orientation of chain H versus the oil drop definition.
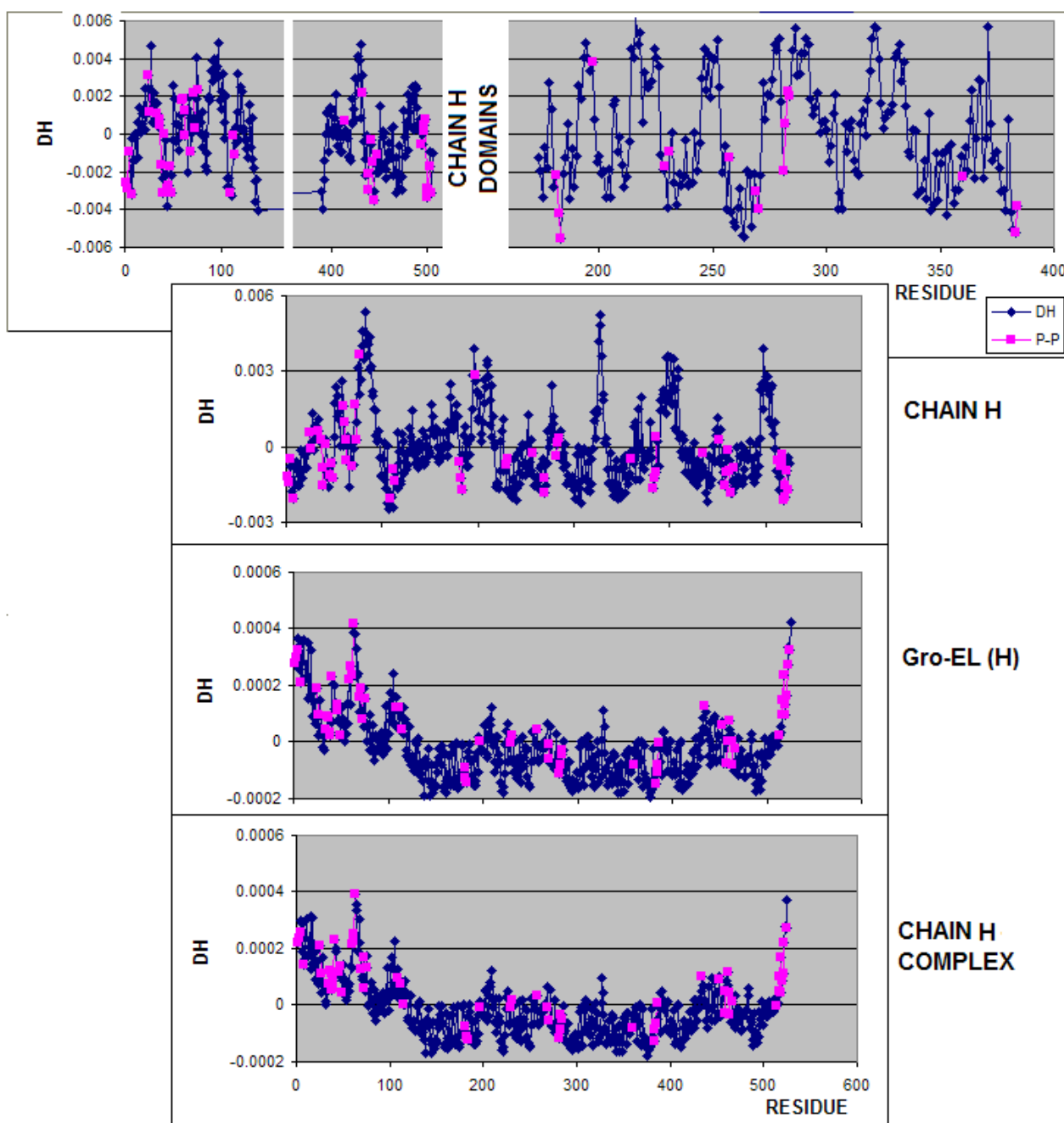
The accordance of protein-protein interaction area with expectations based on "fuzzy oil drop" model is lower in chain H than in chain A.

Figure 4. Frequency of the values of $\Delta\widetilde{H}$ representing the residues in chain A engaged in protein-protein (P-P) interaction (the dark blue and pink curves – interaction of two domains) and in ligand complexation. The negative $\Delta\widetilde{H}$ characterize the residues engaged in this type of interaction. Ligand complexation is represented by residues of positive $\Delta\widetilde{H}$ values when the domain is taken as the unit.
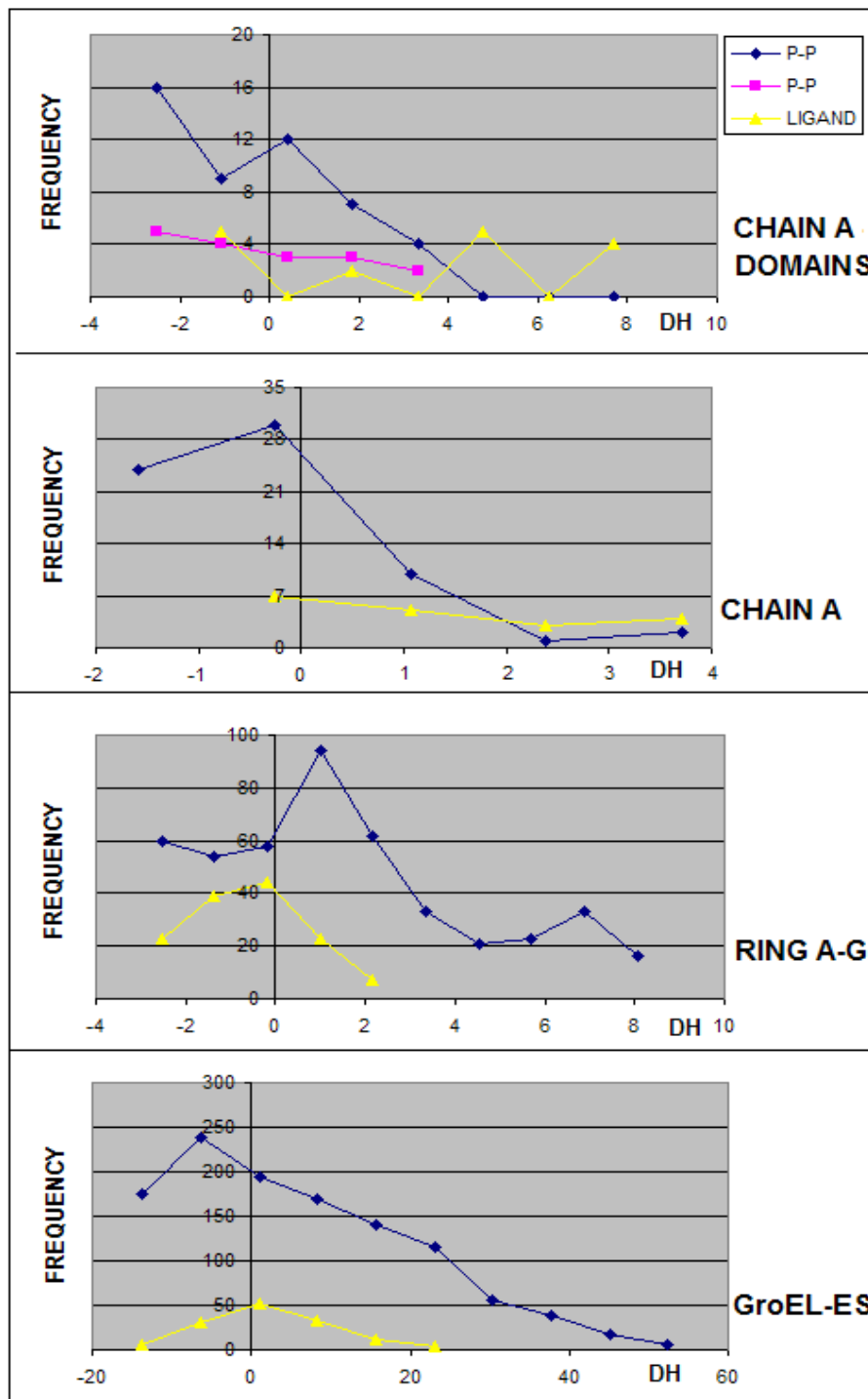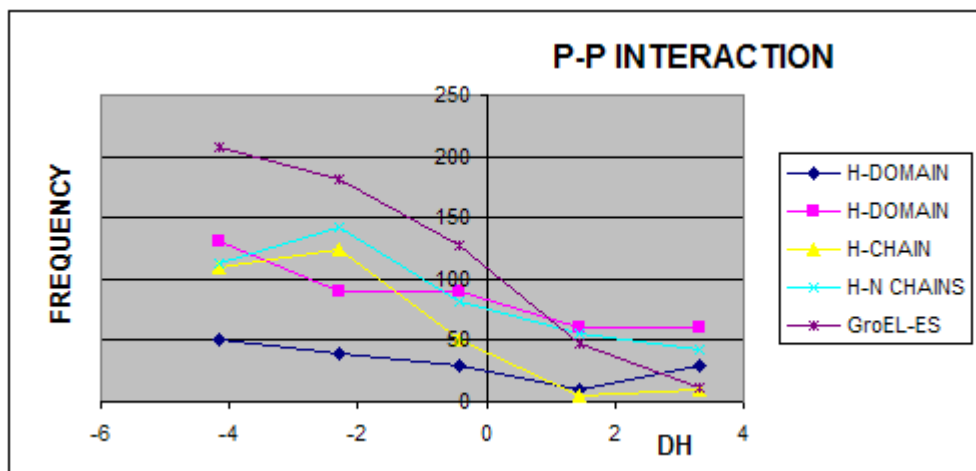
Figure 5. Frequency of the values of $\Delta\widetilde{H}$ of residues in the chain H engaged in protein-protein (P-P) interaction. The negative $\Delta\widetilde{H}$ are characteristic for this type of interaction.

According to expectations the residues involved in the protein-protein interaction represent the hydrophobicity excess ($\Delta\widetilde{H}$ <0). The accordance is also found for the protein-ligand interaction which engages the residues of $\Delta\widetilde{H}$ >0 hydrophobicity deficiency, which may be seen in Figures 4 and 5. The frequency of residues engaged in protein-protein complexation decreases together with the $\Delta\widetilde{H}$ value (except the ring A-H which reaches the maximum for positive values of $\Delta\widetilde{H}$).

Figure 6. The $\Delta\widetilde{H}$ profile for chain O as calculated for the entire Gro-ES fragment. The pink positions distinguish the residues engaged in protein-protein interaction, the yellow positions show the residues engaged in the interaction with the chain A of *cis* ring in the Gro-EL complex.
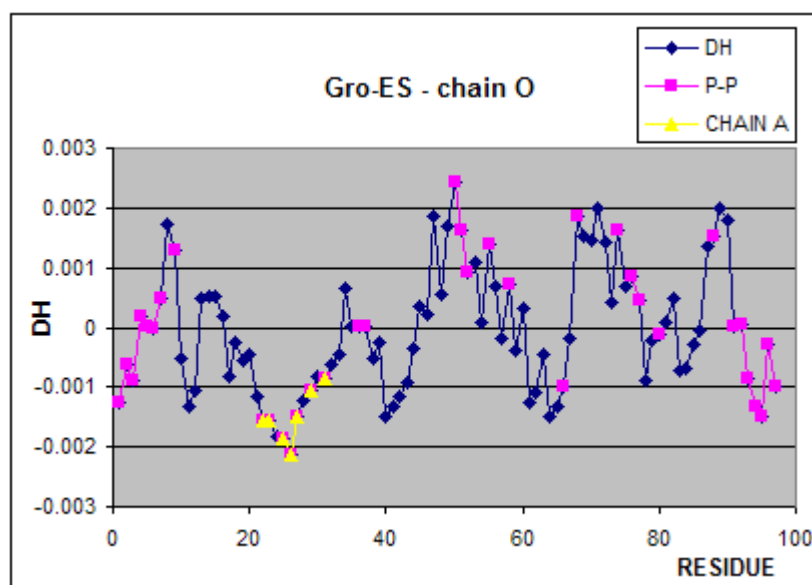
The $\Delta\widetilde{H}$ profile of the chain O as calculated for the Gro-ES fragment reveals rather large fragments of hydrophobicity deficiency in this part of the chaperonin (Figure 6). The residues engaged in the interaction with the chain A are of special importance. Representing the very low values of $\Delta\widetilde{H}$

(hydrophobicity excess) the residues of chain O fit well with the residues in chain A engaged in the interaction with this chain (residues 233-269), also representing the low values of $\Delta\widetilde{H}$. A high accordance can be seen in this case.

The residues of low $\Delta\widetilde{H}$ values at the N- and C-terminal fragments are engaged in the interaction with chains of the *cis*-ring. This interaction seems to be of the form of hydrophobic interaction although the fragments of hydrophobic deficiency character are also engaged in protein-protein interaction.

In conclusion one may say that the H and A chain domains (especially the domain containing the residues 192 – 371) are highly accordant with the model. The domains generated by N-terminal and C-terminal polypeptide fragments seem also to be accordant with the "fuzzy oil drop" model.

3.2. Clustering

To make the interpretation of $\Delta\widetilde{H}$ profiles more objective the clustering analysis was applied (as described in 2.6.) was performed. The results are shown in Table 1.

Table 1. Accordance between expected clustering and observed one for classification between amino acids engaged in protein-protein complexation versus all others (not engaged in protein-protein interaction) – third column and amino acids engaged in protein-protein complexation versus those which are not in protein-protein contact (the residues engaged in ligand or ion binding excluded). P-P denotes the residues engaged in protein-protein interaction.

PROTEIN	METHOD	ACCORDANCE	
		P-P versus all others	P-P versus not complexed
A-domain 1-191 371-524	HIERARCHY DMIN	**_0.8404_**	**_0.8293_**
	HIERARCHY DMAX	**0.7590**	**0.7422**
	HIERARCHY AVG	0.6254	0.5993
A-domain 192-371	HIERARCHY DMIN	**_0.9000_**	**_0.9000_**
	HIERARCHY DMAX	**0.7167**	**0.7167**
	HIERARCHY AVG	**0.7667**	**0.7667**
Chain A	HIERARCHY DMIN	**_0.8702_**	**_0.8651_**
	HIERARCHY DMAX	0.5553	0.5377
	HIERARCHY AVG	0.5668	0.5496
Chains ABCDEFG	HIERARCHY DMIN	**_0.8760_**	**_0.8710_**
	HIERARCHY DMAX	0.5461	0.5445
	HIERARCHY AVG	0.5774	0.5808
Chain O	HIERARCHY DMIN	0.6598	0.6598
	HIERARCHY DMAX	0.5773	0.5773
	HIERARCHY AVG	0.5979	0.5979
COMPLEX	HIERARCHY DMIN	**0.8404**	**0.8554**
	HIERARCHY DMAX	0.5145	0.5069
	HIERARCHY AVG	0.5062	0.5563

The highest accordance between expected (according to the "fuzzy oil drop" model) and observed (according to the PDBSum criteria) was obtained for the domain 192-371 in chain A. This observation supports the assumption that this domain could be the first one spontaneously folded according to "fuzzy oil drop" model exposing on the surface the hydrophobic residues in contact with other polypeptide chains. The chain H, although representing the identical sequence, displays some differences *versus* the A chain, suggesting it to be folded under other conditions than the chain A. Lower accordance between expected and observed classification of all other fragments of the complex (partitioning) suggest that the proper unit to be applied for protein folding in the environment simulated by the "fuzzy oil drop" is the domain 192-371 of the chain A. The relatively high accordance observed for the entire complex (chaperonin molecule) may be interpreted as the reliability of the "fuzzy oil drop" model. It suggests that the protein-protein interaction in the complete molecule may be recognized by the minima of $\Delta\widetilde{H}$ in the profile. The results given in Table 1 seem to support the observations presented in Figures 4 and 5.

3.3. Structure-to-function Characteristics

The "fuzzy-oil-drop" model may be also applied for biological function recognition of the protein under consideration [10-12]. The role of chaperonin molecule (complex) is to create the environment for folding proteins. Thus, the internal surface characteristics of the capsule seems to be of special importance in respect to the biological function of this molecule.

The $\Delta\widetilde{H}$ values of residues localized on the internal surface of the complex are presented in Figure 7 and the 3-D representation in Figure 8. These two figures show the residues ordered according to the X-axis localized on the internal surface of the capsule. The high excess of hydrophobicity in Gro-ES suggests the high participation of this type of interaction in substrate binding. The *cis*-ring (chains A-H), in contrast to the GroES part, presents highly differentiated characteristics expressing excess/deficiency hydrophobicity, although biased significantly toward the hydrophobicity deficiency. It may be interpreted as a specific distribution for stronger/weaker interactions with a substrate molecule. The *trans*-ring (H-N) presents rather low differentiation of hydrophobicity excess/deficiency with $\Delta\widetilde{H}$ values close to zero (particularly in the end area). This suggests high accordance of the hydrophobicity distribution with the expected one (accordant with "fuzzy oil drop" model).

Figures 7 and 8 present the hydrophobicity irregularity of the internal channel where the folding reaction takes place. These two pictures show the characteristics of the external force field of hydrophobic character. Its specific deformation (in the sense of irregularity versus the idealized Gauss function distribution) seems to represent the localization of possible anchorage for folding protein in the chaperonin capsule.

The discussion of hydrophobicity based structure–to-function relationships also concerns other parts of the complex presenting highly irregular hydrophobicity distributions. Particularly the excess hydrophobicity areas not engaged in protein-protein interactions (responsible for complex generation) seem to represent the regions potentially ready to interact under changed circumstances during the action of the chaperonin. The structural changes are reported to be a large deviation from the 7-fold

symmetry of Gro-EL rings [14]. The possible protein-protein interaction can be simulated linking two hydrophobicity-excess areas of interacting chains. This can be seen in Figure 9.

Figure 7. The $\Delta\widetilde{H}$ profile of residues exposed on the internal surface ordered along the X-axis (7-fold symmetry). The Gro-ES (chains O – T) followed by the Gro-EL part: *cis*-ring generated by the chains A-G and the *trans*-ring created by the chains H-N.
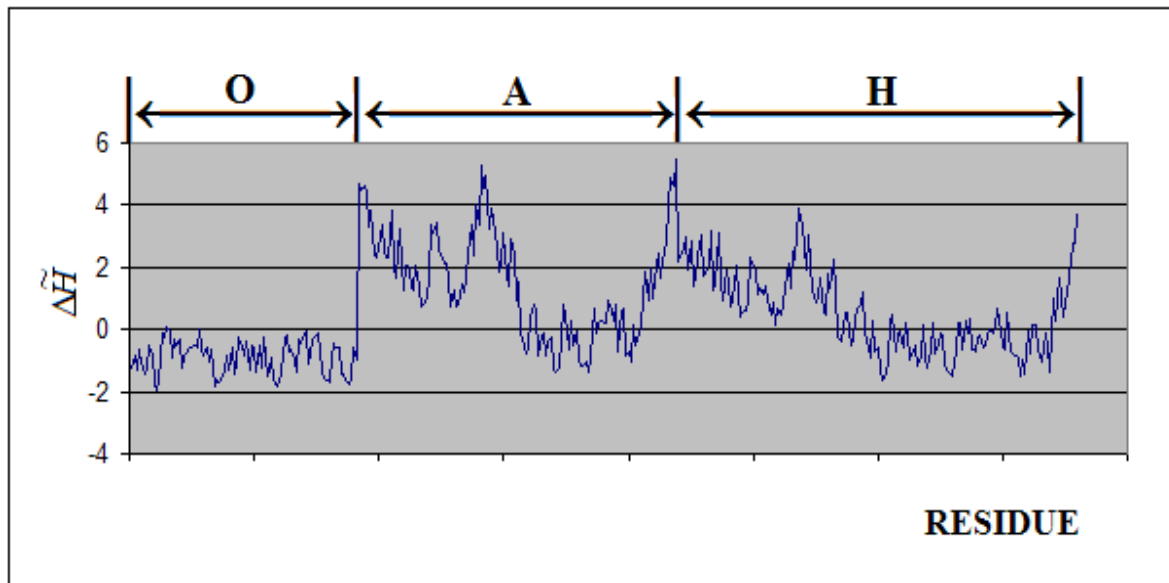

Figure 8. The O-A-H chain system (space filling presentation) in the company of two neighbors chains (T+P, B+G, I+N) in lines presentation. The color scale expresses the $\Delta\widetilde{H}$ magnitude (the higher $\Delta\widetilde{H}$ the more red color, the lower $\Delta\widetilde{H}$ the more blue color). The green color represents the area of hydrophobicity accordant with the idealized 'fuzzy oil drop" model. The molecule shown in red – the ADP molecule complexed to the protein.
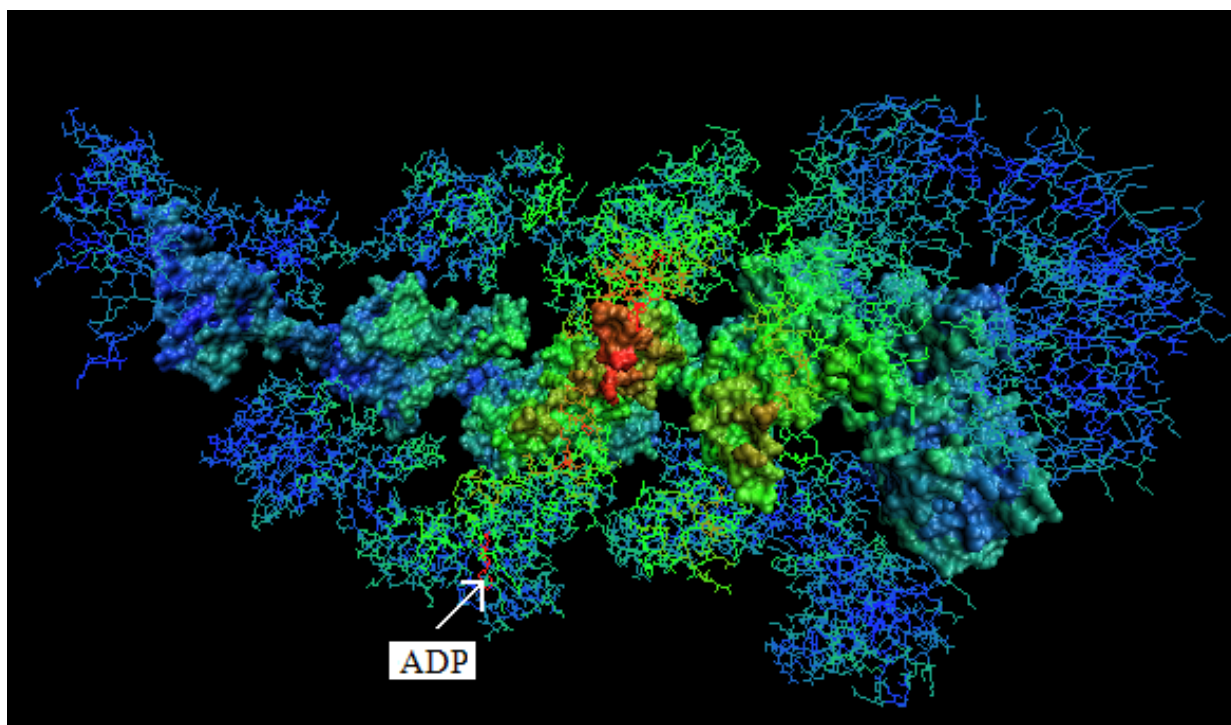
Figure 9. Relation between residue localization (distance) versus the 7-fold symmetry axis of residues and its $\Delta\tilde{H}$ values expressing the degree of hydrophobicity density irregularity. The dark blue symbols represent all residues present in Gro-EL Gro-ES complex. The pink symbols distinguish the residues of Gro-ES fragment. The yellow symbols show the characteristics of residues engaged in ligand binding. The light green symbols show resides engaged in ion binding. The light blue symbols in lower picture represent the residues engaged in protein-protein interaction. The residues in the area distinguished by red circle represent residues on the chaperonin surface potentially ready for hydrophobic protein-protein interaction. The residues (belonging to the set distinguished by green circle) localized on the internal cylinder surface of high $\Delta\tilde{H}$ are potentially ready (not being engaged in any other interaction in the complex) to interact according to non-bonding interaction category. P-P denotes the protein-protein contact, ES-EL – the entire complex, Gro-ES the fragment ES, LIG – interaction with ADP and ION – residues engaged in ion binding.
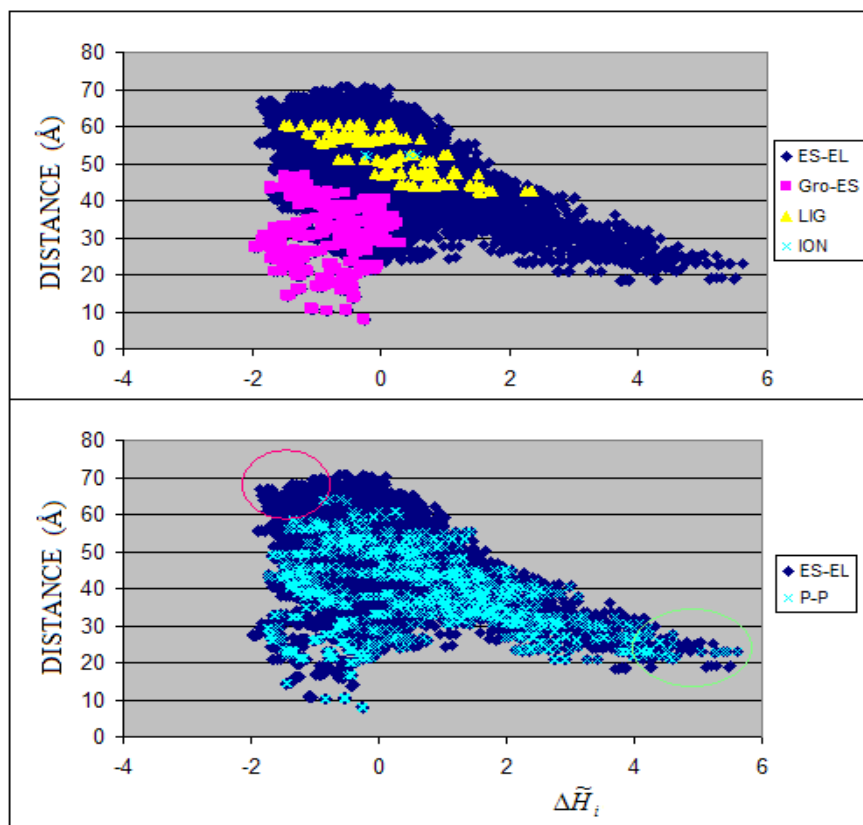
The large–scale structural changes observed as accompanying the folding process engage residues potentially ready to interact. The residues localized on the surface (large distance versus the 7-fold symmetry axis) carrying highly negative $\Delta\tilde{H}$ values seem to be ready for hydrophobic interaction (distinguished by red circle in Figure 9). The residues localized closely versus the 7-fold symmetry axis representing large positive $\Delta\tilde{H}$ values are ready for non-bonding interaction (distinguished by green circle in Figure 9). The first possibility seems to be related to structural changes in the chaperonin molecule while the second one seems to be related to the interaction with the folding protein molecule (internal surface of the capsule) [29].

4. Conclusions

The "fuzzy oil drop" model was assumed to identify the area of excess hydrophobicity on the protein surface as a potential area for protein-protein interaction(s). The analysis of 1AON was treated as an example allowing the estimation of the limits of the applicability of this model for protein-protein interaction areas and ligand binding predictability. The CAPRI [30] initiative is oriented on protein-protein complexation blind prediction. The presented model was assumed to apply for protein-protein interaction recognition. According to the results shown in this paper, some fragments of polypeptide representing local $\Delta\tilde{H}$ minima in the profile can be treated as potential regions engaged in protein-protein complexation, particularly when calculated for domains present in the protein structure. The specificity of individual domains is able to determine the protein-protein complexation. They may be treated as the original source for this process. The ligand localization appeared also to be accordant with expectations – in the fragments of high $\Delta\tilde{H}$ values. Additionally the analysis of histograms particularly calculated of domains (which are assumed to be folded independently) perfectly well supports this interpretation (see Figure 4 and Figure 5.).

The fragments of low $\Delta\tilde{H}$ values not engaged in protein-protein interaction in the complex under consideration seem to represent fragments potentially ready for this type of interaction. The large structural deformations experimentally observed in chaperonin molecule during the protein (substrate) folding seem to be possible in molecule with hydrophobic areas on the protein surface, which is what is observed in 1AON [14].

This observation seems to be additionally supported by the well defined correlation between the hydrophobicity of a side-chain and the logarithm of the folding rate that has been reported in [31], where almost perfect linear correlation has been found for $\Delta\Delta G$ versus the change in hydrophobicity plots observed for few proteins. This is why the analysis of hydrophobicity distribution in protein bodies seems to be of high importance.

The "fuzzy oil drop" model was generated to represent the external force field to generate the environment for folding process assumed to direct the hydrophobic residues toward the center of the molecule and exposure of hydrophilic residues on the surface. The specific irregularity ($\Delta\tilde{H}$ profile) appeared to be biological function related [12]. This observation is assumed to support the postulated hypothesis of the necessary specific ligand participation in folding process to ensure the generation of highly specific cavity (ligand binding) [6-12].

The influence of external force field seems to be obvious in the case of the folding process assisted by a chaperonin molecule. This molecule is assumed to create the proper environment for folding polypeptide chains [31]. The hydrophobicity based characteristics of the interior of the capsule of chaperonin molecule seems to be able to direct the folding process in the form of controlled hydrophobicity excess/deficiency distribution in the folding molecule. The fragment of high positive $\Delta\tilde{H}$ (hydrophobicity deficiency) values fixes the non-bonding interactions and the fragments of low $\Delta\tilde{H}$ (hydrophobicity excess) constraints the hydrophobicity based interactions keeping the hydrophobic residues on the surface of the folding polypeptide if necessary. Assuming that the interior of the Gro-EL chamber really introduces the restraints of this character, the folded molecule shall represent the structure of $\Delta\tilde{H}$ distribution on the protein surface complementary to the internal

surface of the chaperonin chamber (shown in Figures 8 and 9). This hypothesis is currently in the focus of the analysis.

References

1. Motojima, F.; Chaudhry, C.; Fenton, W.A.; Farr, G.W.; Horwich, A.L. Substrate polypeptide presents a load on the apical domains of the chaperonin GroEL. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15005-15012.
2. Rye, H.S.; Roseman, AM.; Chen, S.; Furtak, K.; Fenton, W.A.; Saibil, H.R.; Horwich, A.L. GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings *Cell* **1999**, *97*, 325-338.
3. Xu, Z.; Sigler, P.B. GroEL/GroES: Structure and function of a two-stroke folding machine. *J. Struct. Biol.* **1998**, *124*, 129-141.
4. Fenton, W.A.; Horwich, A.L. GroEL-mediated protein folding. *Protein Sci.* **1997**, *6*, 743-760.
5. Konieczny, L.; Brylinski, M.; Roterman, I. Gauss-function-Based model of hydrophobicity density in proteins. *Silico Biol.* **2006**, *6*, 15-22.
6. Brylinski, M.; Konieczny, L.; Roterman, I. Hydrophobic collapse in late-stage folding (in silico) of bovine pancreatic trypsin inhibitor. *Biochimie* **2006**, *88*, 1229-1239.
7. Brylinski, M.; Konieczny, L.; Roterman, I. Fuzzy oil drop hydrophobic force field - a model to represent late-stage folding (*in silico*) of lysozyme. *J. Biomol. Struct. Dyn.* **2006**, *23,* 519-528.
8. Brylinski, M.; Konieczny, L.; Roterman, I. Hydrophobic collapse in (*in silico*) protein folding. *Comput. Biol. Chem.* **2006**, *30*, 255-267.
9. Brylinski, M.; Konieczny, L.; Roterman, I. Is the protein folding an aim-oriented process? Human haemoglobin as example. *Int. J. Bioinform. Res. Appl.* **2007**, *3*, 234-260.
10. Brylinski, M.; Prymula, K.; Jurkowski, W.; Kochanczyk M.; Stawowczyk E.; Konieczny, L.; Roterman, I. Prediction of functional sites based on the fuzzy oil drop model. *PLoS Comput. Biol.* **2007**, *3*, e94.
11. Brylinski, M.; Kochanczyk, M.; Broniatowska, E.; Roterman, I. Localization of ligand binding site in proteins identified *in silico*. *J. Mol. Model.* **2007**, *13*, 665-675.
12. Brylinski, M.; Konieczny, L.; Roterman, I. Ligation site in proteins recognized *in silico*. *Bioinformation* **2006**, *1*, 127-129.
13. Brylinski, M.; Kochanczyk, M.; Konieczny, L.; Roterman, I. Sequence-structure-function relation characterized *in silico*. *In Silico Biol.* **2006**, *6*, 589-600.
14. Taguchi, H. Chaperonin GroEL meets the substrate protein as a "load" of the rings. *J. Biochem.* **2005**, *137*, 543-549.
15. Xu, Z.; Horwich, A.L.; Sigler, P.B. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* **1997**, *388*, 741-750.
16. Levitt, M. A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **1976**, *104*, 59-107.
17. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105-132.

18. Engelman, D.M.; Steitz, T.A.; Goldman, A. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* **1986**, *15*, 321-353.

19. Hopp, T.P.; Woods, K.R. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 3824-3828.

20. Rose, G.D.; Geselowitz, A.R.; Lesser, G.J.; Lee, R.H.; Zehfus, M.H. Hydrophobicity of amino acid residues in globular proteins. *Science* **1985**, *229*, 834-838.

21. Wolfenden, R.; Andersson, L.; Cullis, P.M.; Southgate, C.C. Affinities of amino acid side chains for solvent water. *Biochemistry* **1981**, *20*, 849-855.

22. http://www.ebi.ac.uk/pdbsum/, accessed 20 January 2009.

23. http://python.org, accessed 20 January 2009.

24. http://www.rcsb.org, accessed 20 January 2009.

25. http://biopython.org, accessed 20 January 2009.

26. http://numpy.scipy.org, accessed 20 January 2009.

27. http://matplotlib.sourceforge.net, accessed 20 January 2009.

28. Jain, A.K.; Dubes, R.C. *Algorithms for clustering data*; Prentice Hall: Englewood Cliffs NJ, USA, 1988.

29. Clare, D.K.; Bakkes, P.J.; van Heerikhuizen, H.; van der Vies, S.M.; Saibil, H.R. Chaperonin complex with a newly folded protein encapsulated in the folding chamber. *Nature* **2009**, *457*, 107-110.

30. http://www.ebi.ac.uk/msd-srv/capri, accessed 20 January 2009.

31. Anil, B.; Sato, S.; Cho, J-H.; Raleigh, D.P. Fine structure analysis of a protein folding transition state; Distinguishing between hydrophobic stabilization and specific packing. *J. Mol. Biol.* **2005**, *354*, 693-705.