INTER-SUBUNIT CONTACTS OF THE PROTEASOMAL ALPHA-SUBUNITS AS DETERMINANTS OF PARALOG GROUPS

Nikolaev S.V.*, Afonnikov D.A.
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
* Corresponding author: e-mail: nikolaev@bionet.nsc.ru

Keywords: proteasome, proteasomal subunit, proteolytic complex, protein-protein interaction, mutual information, ortholog, paralog

Summary

Motivation: Assignment of the proteasomal subunits to paralog groups has been initially based on phylogenetic analysis of amino acid sequences. The aim of the current work was to analyse the capability of the inter-subunit contact regions of the proteasomal subunits to determine the orderly arrangement of the alpha-subunits in the ring. This was because these regions are essential for the self-assembly of the proteasomes, and, hence, for the determination of the paralog groups.

Results. Based on the data on amino acid sequences and the known 3D structures of some proteasomes, we marked the positions of the alpha-subunits involved in inter-subunit contacts (the contact position, CP). There was a good match between the clustering based on the CPs and the one that was based on the complete sequences. Thus, we demonstrated that the CPs contain the information required for distinguishing the paralog groups.

Introduction

The proteasome is a multienzyme proteolytic machinery, which provides degradation of the bulk of cytoplasmic proteins up to oligopeptides. The proteasome is cylinder-shaped, composed of two identical halves, with each half arranged as two rings making up seven subunits each. The outer rings of the proteasome consist of the alpha-subunits, whereas the inner rings are composed of the beta-subunits. It is currently believed that the proteolytically active core of the proteasome (20S proteasome) results from the self-assembly of the subunits. The ring consisting of the alpha-subunits is self-assembled first, the ring composed of the beta-subunits is added after (Kopp et al., 1997). It is supposed that the orderly arrangement of the subunits in the proteasomal rings is stable. Therefore, it appeared of interest, what information on the alpha-subunit ordering in the ring is at the positions involved in the formation of the intersubunit contacts, because these positions indirectly participate in protein-protein recognition.

The accepted classification of the proteasomal alpha-subunits into paralog groups is based on phylogenetic analysis of amino acid sequences disregarding their structure and function (Bouzat et al., 2000).

The goals of this study were: 1) to clarify whether there is enough information to identify the alpha-subunit paralog groups at CPs; 2) to analyze the distribution of mutual information on the CPs.

Methods and Algorithms

Amino acid sequences of the proteasomal alpha-subunits given in Table 1 were extracted from the SWISS-Prot database, the data on the proteasomal 3D structures were extracted from the PDB. The ClustalW program (Thompson et al., 1994) was used for multiple sequence alignment and sequence clustering. A clustering tree based on the results of the multiple alignment of the entire sample was used to set up groups of the alpha-subunit orthologs (Table).

To identify the contact positions, we regarded two amino acid residues as contacting if the distance between their C-alpha atoms was not greater than 6.5 Å. This threshold distance was used to
identify the sequence positions in the alpha-subunits in contact with any residue in any other alpha- or beta-subunit. Because each group of orthologs contained two subunits with known 3D structure, to identify the CPs, the union (“Union”) of the CPs of an ortholog pair with known structure was first found, then, the intersection of the unions (“Intersection(Union)”) was determined on all paralog groups. The thus chosen positions projected upon alignment sequences were accepted as the contact positions between paralogs.

Table. Analyzed sequences and clustering results obtained by the ClustalW

<table>
<thead>
<tr>
<th>Paralog group</th>
<th>SWISS-Prot / PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1IRU:F, PSA1_HUMAN, PSA1_MOUSE, PSA1_CHICK, PSA1_DICDI, 1FNT:F, PSA1 YEAST, PSA1_SCHPO, PSA1_CAEEL, PSA1_DROME, PS11_ARATH, PS12_ARATH, PSA1 ORYSA, PSA1_TRYBR, PSA1TRYCR</td>
</tr>
<tr>
<td>2</td>
<td>1IRU:B, PSA2_HUMAN, PSA2_MOUSE, PSA2_RAT, PSA2_XENLA, PSA2_CARAU, PSA2_DROME, PSA2_ARATH, PSA2 ORYSA, 1FNT:B, PSA2_NEUCR, PSA2_SCHPO, PSA2_CAEEL, PSA2_TRYBB</td>
</tr>
<tr>
<td>3</td>
<td>1IRU:G, PSA3_HUMAN, PSA3_MOUSE, PSA3_RAT, PSA3_ARATH, PSA3_SPIOL, PSA3 ORYSA, PSA3_ACACA, PSA3 DICDI, PSA3_DROME, PSA3_CAEEL, 1FNT:G, PSA3_SCHPO</td>
</tr>
<tr>
<td>4</td>
<td>1IRU:C, PSA4_HUMAN, PSA4_MOUSE, PSA4_RAT, PSA4_DROME, PSA4_CAEEL, PSA4_DROME, 1FNT:C, PSA4_SCHPO, PSA4_ARATH, PSA4_SPIOL, PSA4_PETHY, PSA4 ORYSA, PSA4 DICDI</td>
</tr>
<tr>
<td>5</td>
<td>1IRU:E, PSA5_HUMAN, PSA5_MOUSE, PSA5_RAT, PSA5_DROME, PSA5_CAEEL, PSA51_ARATH, PSA52_ARATH, PSA5 OEBS, PSA5 ORYSA, PSA5_SCHPO, 1FNT:E, PSA5 ENTHL, PSA5 TRYBB</td>
</tr>
<tr>
<td>6</td>
<td>1IRU:A, PSA6_HUMAN, PSA6_MOUSE, PSA6_DROME, PSA61_ARATH, PSA62_ARATH, PSA6 SOYBN, PSA6 TOBAC, PSA6 ORYSA, 1FNT:A, PSA6 YEAST, PSA6_SCHPO, PSA6_CAEEL</td>
</tr>
<tr>
<td>7</td>
<td>1IRU:D, PSA7_HUMAN, PSA7_MOUSE, PSA7_RAT, PSA7 CHICK, PSA71_XENLA, PSA72_CARAU, PSA71_HUMAN, PSA71_MOUSE, PSA71_DROME, PSA71_DROVI, PSA72_DROME, PSA71_ARATH, PSA72 ARATH, PSA7 CICAR, PSA7 LYCES, PSA7 ORYSA, PSA7 DICDI, PSA71_CAEEL, PSA72_DROME, PSA73_DROVI, 1FNT:D, PSA7 YEAST, PSA7_SCHPO, PSA7 TRYBB</td>
</tr>
</tbody>
</table>

To evaluate the significance of a sequence position in the classification into paralog groups, we used mutual information at positions as a measure of the relation between an amino acid type at the position and a given paralog group: (Stormo et al., 1986):

\[ I_i = \sum_x \sum_y f(x,y) \cdot \log \left( \frac{f(x,y)}{f(x) \cdot f(y)} \right) \]

where:

- \( f(x_i) \) – the frequency of amino acid type \( x \) at the \( i \)-th position in multiple alignment,
- \( f(y) \) – the proportion of proteins of the \( y \)-th paralog group,
- \( f(x_i,y) \) – the frequency of the amino acid type \( x \) at the \( i \)-th position in proteins of the \( y \)-th paralog group.

The RasMol (Sayle, Milner-White, 1995) package was used for visualization, analysis and rendering of the protein structures.

Results and Discussion

Ordering of the alpha-subunits in the proteasomal ring. The proteasomal alpha-subunits extracted from the PDB (files PDB1FNT.ENT, PDB1IRU.ENT, Table) were used as markers of the alpha-subunits ordering in the ring. The clustering tree based on multiple alignment demonstrated a
subdivision of the chosen sequences with retention of non-intersecting classes of the marker subunits. A scheme for the ordering of the subunits in the alpha-ring of the proteasomes with known 3D structures and corresponding numbers of paralog groups, resulting from clustering based on multiple alignment, is given in Figure 1.

**Figure 1.** The alpha-subunit ordering in the proteasomes with known structures (the one-letter code of proteasome subunit chains in files *pdb1fat.ent, pdb1iru.ent* from the PDB) and paralog groups in which the subunits proved to be the clustering resulting from multiple alignment.

Clustering based on sequences composed of the amino acids involved in intersubunit contacts is similar to the one based on the entire sequences of the subunits. In contrast, when for the entire set of sequences, we took sequences consisting of contact positions chosen as Intersection (Union) and applied the ClustalW program, the clustering agreed well with the one obtained for the complete sequences. In fact, the paralog groups were the same as those listed in Table. This was evidence that the paralog group determinants were present at the contact positions chosen by the above method. Stating it in other words, the necessary conditions for distinguishing paralogs in protein-protein interactions at the alpha-subunit contact positions during the self-assembly of the proteasomal alpha-ring are satisfied. However, the distribution of the mutual information at contact positions of a full-size sequence is identical. For this reason, it is an open issue whether these interactions suffice to provide a prepatterned orderly arrangement of the alpha-ring subunits.

**Spatial structure of determinant distribution for the paralog groups at the contact positions.** From the subdivision into paralog groups at multiple alignment of the alpha-subunits it already followed that the paralog groups differ by sequences. Consequently, it was of interest to localize the sites where the sequences were different. The question was: How significant were the contact positions in the distinguishment of the paralog groups?

Figure 2 shows the content of mutual information sufficient for differentiating the identity of paralogs at contact positions projected upon the proteasomal alpha-ring subunit. It is apparent that, among the positions involved in the inter-subunit contacts of the alpha-ring, certain are very strong determinants. Although the number of contact positions between an alpha-subunit and a beta-subunit was small compared with the one for the alpha-alpha subunits contacts, the former proved to be also strong determinants of paralog groups. This may be taken to mean that the alpha-beta contacts are important in ordering during self-assembly of the beta-ring in conjunction with the inter-beta subunit contacts.
The reference of an alpha-subunit to a paralog group depends on not only its position in the alpha-ring, but also on its position with respect to the beta-ring subunit.

**Acknowledgements**

This work was partly supported by grants from the Russian Foundation for Basic Research (03-07-96833-p2003, 03-07-96833); Siberian Branch of the Russian Academy of Sciences (projects No. 148,119); the Russian Ministry of Industry, Sciences and Technologies (No. 43.073.1.1.1501, subcontract 38/2004), RAS Presidium Program “Molecular and Cellular Biology” (No. 10.4) and Origin and evolution of biosphere” (No. 10002-251/II-25/155-270/200404-082, the CRDF and the Ministry of Education of Russian Federation within the Basic Research and Higher Education Program (NO-008-X1, Y1-B-08-20).

**References**


