CLUSTERING ANALYSIS OF CONFORMATIONAL STATES OF SHORT OLIGOPEPTIDES

Batsianovsky A.V.*, Vlasov P.K.
Engelhard Institute of Molecular Biology, RAS, Moscow, Russia; * Moscow State University, Moscow, Russia
* Corresponding author: e-mail: suner_s@mail.ru

Key words: peptide structure, conformation, clusters analysis

SUMMARY

Motivation: A goal of presented investigation is analysis of conformation combinations in proteins. There are no doubt about that conformations of amino acids are combining with some preferences (Vlasov et al., 2005). At now there is sufficient material to analyze thoroughly structural state of tri- and tetrapeptides as a minimal symmetrical unit of regular structures of peptide backbone. It is tentative to reveal the pattern of the alternation of conformational states characterizing by different spiral symmetry.

Results: We have performed clustering analysis, using standard clustering method “K-means”, to determine for most part of amino-acid residues the frequency of distribution of sequences of dihedral angles (φ- and ψ-angles), which are typical for protein helical conformations. It was found that the alternations amino-acid residues with different symmetry (3.6 in α-helix, 2 in β-strand) occur quite rarely in comparison to the alternations with similar symmetry. Influence of electrostatic in respect of limitations onto secondary structure alternation is discussed.

Availability: none

INTRODUCTION

Major part of a globule is formed by a set of symmetrical structures as α-helix, β-strand and others. It is obvious that formation a symmetrical network of intrinsic bonds including possible hydrogen bonds between atoms in main chain is of important for organization of protein structure and is favorable in respect of globule energetic. These symmetrical structures play essential role in protein folding. At the same time some asymmetrical elements are common in any protein structure. Fragments of such a kind wait their classification and functional annotation.

One can anticipate that majority of specific conformations must be combination of typical conformations belonging to allowed regions in Ramachandran map. These typical conformations correspond to secondary structures in general but other structural elements also possible.

Two questions arise in connection with sequence of conformations study:
- How many conformations can be constructed as a combination of typical regular protein conformations using corresponding dihedral angles? After this the symmetrical features of conformational state will be determined.
- What about of different conformation sequences occurrence?
MATERIALS AND METHODS

The method of clustering can readily applied for separation of different structural patterns. In this work a set of dihedral angles in oligopeptides of fixed length was chosen as input data for clustering. Wide-spread method of K-Means implemented in mathematical package Statistica v 6.0 was used to solving the problem. Special program has been developed for interpretation results of clustering. This approach has been applied to the data on dihedral angles in protein structures from PDBSelect. This set of protein data is useful for many aims and may be treated as representative and sufficient narrow possessing fast computations. It should be noted that these data base includes nonhomological proteins (identity is not more 30 %). Method of \( \varepsilon \)-networks was applied for improving the quality of results. This method is convenient for elimination regions with rarely density of population.

RESULTS AND DISCUSSION

Results of clustering present division of set of structures into classes. For each class we estimated average values of operation parameters, deviation of these parameters inside the clusters, the number evinces in cluster. Unfortunately, we could not achieve the absolute division. It is likely in connection with the character of data and space operation factors, and clustering method peculiarities. However, we could distinguish conservative clustering in different divisions. These clusters conserve their characteristics in all divisions. This evidently implies that these clusters are correct and reflect the real structures in proteins.

Conformations of oligopeptides from conserved clusters are combinations of conformational states of typical conformations as expected. Structures with non-typical conformational states do not form constant clusters. For example, we have found some clusters with \( \beta \)-turns, but all of them were unstable. As a result it is possible to descript these stable structures in terms of typical conformations denoting the conformation state of amino acid by region of typical conformation.

Frequencies of these structures show strong nonregularity. This effect can be explained on the basis of symmetry mechanism of compensation of electrostatic interactions. For example, pairs of adjacent amino acid residues in \( \beta \)-conformation raise frequencies of occurrence of the structure (symmetry in \( \beta \)-strand is 2), at the same time very small fragments (included one or residues) in \( \alpha \)-conformation especially in the middle of oligopeptide reduce the frequency (symmetry in \( \alpha \)-helix – 3.6). There is prohibition of some structural types. Such structures as \( \beta\alpha\beta\alpha \), \( \alpha\beta\alpha\beta \), \( \beta\alpha\alpha\beta \) are absent in all divisions. There is no compensator electrostatics in these structures. At last, structures \( \alpha\alpha\beta \) and \( \alpha\alpha\beta\beta \) occur only in some divisions. We obtained the list of stable clusters of tripeptides and tetrapeptides:

<table>
<thead>
<tr>
<th>Structure</th>
<th>No, thous</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta\beta\beta )</td>
<td>~130</td>
</tr>
<tr>
<td>( \alpha\alpha\alpha )</td>
<td>~220</td>
</tr>
<tr>
<td>( \alpha\beta\alpha )</td>
<td>~17</td>
</tr>
<tr>
<td>( \alpha\beta\beta )</td>
<td>~32</td>
</tr>
<tr>
<td>( \beta\alpha\beta )</td>
<td>~21</td>
</tr>
<tr>
<td>( \beta\alpha\alpha )</td>
<td>~35</td>
</tr>
<tr>
<td>( \beta\beta\alpha )</td>
<td>~32</td>
</tr>
</tbody>
</table>

Table 1. List of stable clusters of tripeptide and tetrapeptide structures. The symbol \( \beta \) signify conformational state corresponding to region of \( \beta \)-strand in the Ramachandran map, \( \alpha \) – \( \alpha \)-helix, \( \alpha \) – left \( \alpha \)-helix. The total number of structures in this sample is 637 thousands. Average deviation from indicated numbers in cluster is lower than 3 thousands for the large clusters, and 2 thousands for small ones.
A statistical approaches as operation of clustering is useful for finding of general rules that control structure formation in short oligopeptides. Clustering operation is partly complicated by noise effect through all Ramachandran plot, and application of \( \varepsilon \)-networks cannot overcome the noise effect. Maybe in turn the space of operation parameters is not natural for determination of structural types. However, in spite to this negative factors it was achieved clustering in reasonable and prominent groups of structures. The results demonstrate that there are no equal frequencies of various combinations of elementary conformation states. On of possible explanation consists in taking into account the electrical dipole moment of amino acid residues and possible electrostatic compensatory effects in this aspect.

REFERENCES