A large number of diseases are caused by (or at least connected toxic when they form fibrils (Bucciantini perpendicularly to the axis of the fibril) and binds Congo red and $\beta$-fibrillar aggregation (Jimenez et al., 1999). An amyloid fibril usually has a cross-$\beta$-X-ray diffraction pattern ($\beta$-strands are oriented perpendicularly to the axis of the fibril) and binds Congo red and Thioflavin-T dyes (Rudall, 1952). Normal proteins can become toxic when they form fibrils (Bucciantini et al., 2004). Numerous studies have shown that, in addition to proteins involved in amyloid diseases, many proteins not related to any amyloid disease can aggregate into fibrils under destabilizing conditions (e.g. Chiti et al., 1999; Fandrich et al., 2001; Guirajaran et al., 1998). The experimental analysis of different proteins which form amyloid fibrils revealed that these proteins contain rather small fragments which are required for the amyloidogenesis (Ivanova et al., 2004). Mutations in these regions lead to a decrease in or to a total disappearance of the ability of the proteins to form amyloid fibrils. Moreover, these amyloidogenic fragments are able to form amyloid fibrils even if taken alone, without the rest of the original protein (Thompson et al., 2000). By now, the positions of amyloidogenic regions are experimentally localized for some proteins (for the list of proteins, see, for example, Galzitskaya et al., 2006a and references therein). Besides that, there are several hundreds of artificial peptides which are able to form fibril-like aggregates, most of the peptides can be considered as derivatives of the amyloidogenic fragments observed in natural proteins. The analysis of the available experimental data reveals that most of the amyloidogenic fragments are similar to each other: they have an elevated content (compared to the typical one for globular proteins) of hydrophobic amino acid residues (Galzitskaya et al., 2006b). However, there is an amyloidogenic region with a very distinct amino acid composition: GNNQQNY which is responsible for aggregation of yeast prion Sup35. The crystal structure of GNNQQNY was obtained earlier (Nelson et al., 2005). Hydrogen bonds (inside the protein backbone as well as involving side-chains of the amino acids) play an important role in stabilization of the formed fibril. Except for Sup35, aggregates of a similar type can be formed in numerous disease-related polyQ proteins which possess lengthy (sometimes several dozens of glutamine residues long) polyglutamine tracts the length of which varies from individual to individual and which lead to aggregation of the protein in a case when its polyQ tracts are too long. Thus, the amyloidogenic fragments can be classified (Galzitskaya et al., 2006b) into two types: type I (rich in hydrophobic amino acid residues) and type II (rich in Gln and Asn). Presently, most methods of prediction of amyloidogenic fragments are designed for searching for the regions of type I since they are much more investigated experimentally. At the moment, several methods of prediction of amyloidogenic regions are known. Two methods (Fernandez et al., 2003; Yoon and Welsh, 2004) are elaborated to find amyloidogenic regions in proteins with known spatial structures. Some other methods (Lopez de la Paz and Serrano, 2004; Tartaglia et al., 2005; Thompson et al., 2005; Thompson et al., 2001; Guijarro et al., 2001; Thompson et al., 2000) have been introduced by us to detect amyloidogenic regions in a protein sequence. We demonstrate that regions with high expected probability of the formation of backbone-backbone hydrogen bonds as well as regions with high expected packing density are mostly responsible for the formation of amyloid fibrils. Our method (FoldAmyloid) has been tested on a dataset of 407 peptides (144 amyloidogenic and 263 non-amyloidogenic peptides) and has shown good performance in predicting a peptide status: amyloidogenic or non-amyloidogenic. The prediction based on the expected packing density classified correctly 75% of amyloidogenic peptides and 74% of non-amyloidogenic ones. Two variants (averaging by donors and by acceptors) of prediction based on the probability of formation of backbone-backbone hydrogen bonds gave a comparable efficiency. With a hybrid-scale constructed by merging the above three scales, our method is correct for 80% of amyloidogenic peptides and for 72% of non-amyloidogenic ones. Prediction of amyloidogenic regions in proteins where positions of amyloidogenic regions are known from experimental data has also been done. In the proteins, our method correctly finds 10 out of 11 amyloidogenic regions. The FoldAmyloid server is available at http://antares.protres.ru/fold-amyloid/

Contact: ogalzit@vega.protres.ru

1 INTRODUCTION

A large number of diseases are caused by (or at least connected with) aggregation of incorrectly folded (misfolded) proteins. A typical variant in this case is the formation of the so-called amyloid fibrils. Amyloid fibril formation is associated with an increase in $\beta$-structure content in the structure of a protein, leading to its fibrillar aggregation (Jimenez et al., 1999). An amyloid fibril usually has a cross-$\beta$-X-ray diffraction pattern ($\beta$-strands are oriented perpendicularly to the axis of the fibril) and binds Congo red and Thioflavin-T dyes (Rudall, 1952). Normal proteins can become toxic when they form fibrils (Bucciantini et al., 2004). Numerous...
et al., 2006; Trovato et al., 2006) are designed for prediction of amyloidogenic regions from the amino acid sequence. A method previously (Galzitskaya et al., 2006b) designed by us was also able to successfully predict amyloidogenic regions from the amino acid sequence (Galzitskaya et al., 2006b). In this article, we obtained several additional scales inferred from the statistics of protein structures for predicting amyloidogenic regions from a protein sequence, and the server (named FoldAmyloid) for predicting amyloidogenic regions from a protein sequence has been created.

2 METHODS

The observed packing density and the statistics of hydrogen bonds were calculated in a database which contained spatial structures of 3769 proteins (Galzitskaya et al., 2006b). The database consists of proteins which belong to structural classification of proteins (SCOP) (Murzin et al., 1995) to four main SCOP classes (a, b, c and d which are α, all-β, α/β and α+β proteins, respectively). The proteins of the database have <25% sequence identity between each other.

The observed packing density (the observed number of contacts per residue) for each amino acid residue in the database was calculated (Galzitskaya et al., 2006b) as the number of close residues (within a given distance). A residue was considered close to the given residue if any pair of their non-hydrogen atoms was at a distance <8 Å (the neighboring residues were excluded from the consideration). Then, the observed packing densities were averaged for each of 20 types of amino acid residues.

Hydrogen bonds were searched for in the same database. We collected statistics separately for four variants of hydrogen bonds: backbone-backbone (both the donor and the acceptor are in the protein backbone), backbone-side-chain (the donor is in the protein backbone while the acceptor is in a side-chain), side-chain-backbone (the donor is in a side-chain while the acceptor is in the protein backbone) and side-chain-side-chain (both the donor and the acceptor are in side-chains). Backbone-backbone hydrogen bonds (that is, hydrogen bonds where the donor is in the NH-group of the protein backbone and the acceptor is the O-atom of the protein backbone) were analyzed by a standard program DSSP (Kabsch and Sander, 1983). For each NH-group, only one hydrogen bond (which had the best energy, according to DSSP) was taken into consideration in this case. The criterion of hydrogen bond formation was that recommended by DSSP authors (the calculated energy lower than −0.5 kcal/mol). Hydrogen bonds where the donor or the acceptor or both belonged to side-chains were searched for by our in-house program. In this case, the criterion of hydrogen bond formation was geometrical: the distance between donor and acceptor atoms below an atom-specific cutoff value. The cutoff distance for hydrogen bonds of each type (dependent on the type of the involved electronegative ‘heavy’ atoms) was taken from Stickle et al. (1992). It should be mentioned that more than one hydrogen bond per donor can be taken into consideration in this case (i.e. for hydrogen bonds involving side-chains). For acceptors, there was no restriction on the number of hydrogen bonds in all cases.

Then, the probability to form each variant of hydrogen bond by each of 20 types of amino acid residues was calculated. During the calculation, the hydrogen bonds were ‘ascribed’ to donor or to acceptor residues, resulting in two sets of probability values for each type of amino acid residues. The probability of hydrogen bond formation was calculated as the total number of hydrogen bonds of the corresponding variant (backbone-backbone or backbone-side-chain, etc.) formed by each type of amino acid residues divided by the total number of residues of this type in the database.

The obtained values (the average observed packing density or the probability of hydrogen bond formation) are used as a scale for constructing a profile (the packing density profile or hydrogen bond probability profile, respectively) for a protein sequence. The calculations are based on the sliding-window averaging technique. First, the expected value is determined for each residue (it is equal to the average value observed for this type of residue in spatial structure), then, the obtained values are averaged inside the window and the average is assigned to the central residue of the window. The ‘smoothed’ expected values for every position of the polypeptide chain provide the final profile which is directly used during prediction of amyloidogenic regions.

On the smoothed profile, we predict a region as an amyloidogenic one if all its residues lie above a given cutoff in the profile, and the size of the region is greater than or equal to the size of the sliding window used.

Based on the scale of the expected contacts, we predicted amyloidogenic regions as those in which the probability of hydrogen bond formation was larger than a cutoff.

Based on the scale of the expected number of hydrogen bonds (of any variant), we predicted amyloidogenic regions as those in which the probability of hydrogen bond formation was larger than a cutoff.

A peptide was predicted as amyloidogenic if it had at least one predicted amyloidogenic region.

The quality of our predictions was evaluated using the commonly used standard values: sensitivity and specificity which are the fraction of correctly predicted amyloidogenic peptides and the fraction of correctly predicted non-amyloidogenic peptides, respectively.

\[
\text{Sensitivity} = \frac{\text{TP}}{N_{\text{amyloidogenic}}}, \quad \text{ Specificity} = \frac{\text{TN}}{N_{\text{non-amyloidogenic}}},
\]

where true positives (TP) is the number of peptides correctly predicted as amyloidogenic ones, true negatives (TN) is the number of peptides correctly predicted as non-amyloidogenic ones, \( N_{\text{amyloidogenic}} \) is the total number of amyloidogenic peptides in the database and \( N_{\text{non-amyloidogenic}} \) is the total number of non-amyloidogenic peptides in the database.

3 RESULTS

The presented method (named FoldAmyloid) allows predicting amyloidogenic regions in a protein (or a peptide) starting solely from its sequence. It is based on using expected characteristics—scales: either expected packing density or the probability of formation of hydrogen bonds. The scales themselves were obtained from the statistics of spatial structures of proteins, and then the scales are used for predictions basing on amino acid sequence only. At first, we obtained these values (the expected packing density and probability of formation of hydrogen bonds) for each residue in spatial structures of proteins, then calculated average values for each of 20 types of amino acid residues, and then used these average values as the values expected for each residue of a given type in a sequence for which the prediction is made.

3.1 Analysis of packing density and hydrogen bonds observed in spatial structures of proteins

To analyze the statistics of contacts (packing density) and hydrogen bonds observed in protein structures, we have constructed (Galzitskaya et al., 2006b) a database of spatial structures of proteins with the sequence identity <25%. For each amino acid residue in this database, we have obtained the observed packing density and the observed number of hydrogen bonds.

The observed packing density was calculated (Galzitskaya et al., 2006b) for each amino acid residue as the number of residues which are close to the given one. Two residues were considered close to each other if any pair of their non-hydrogen atoms was at a distance <8 Å. Then, the mean observed packing density for each of the 20 types of amino acid residues was calculated. These mean values
are given in our previous work (Galzitskaya et al., 2006a) and in Figure 1a. The mean values were used further for predicting the packing density (thus, it is the expected packing density) from the protein sequence.

The statistics of hydrogen bonds was analyzed using the same database of 3769 three-dimensional protein structures. We searched for four variants of hydrogen bonds separately: backbone-backbone, backbone-side-chain (the donor is in the protein backbone while the acceptor is in a side-chain), side-chain-backbone (the donor is in a side-chain while the acceptor is in the protein backbone) and side-chain-side-chain (both the donor and the acceptor are in side-chains). Then, the probabilities of formation (by each type of amino acid residues) of hydrogen bonds of a given variant were calculated. During the calculation, the hydrogen bonds were assigned to donor or to acceptor residues, resulting in two sets of probability values for each type of amino acid residues. Then, this probability was used as a scale for predicting the probability of hydrogen bond formation by each residue in a protein sequence. In Figure 1b–e, one can see the obtained probabilities of formation of hydrogen bonds of different variants for the 20 types of amino acid residues. These values will be used further as scales for prediction of amyloidogenic regions.

In Table 1, correlations between some obtained scales are listed. One can see that there is no correlation between the probability of formation of backbone-backbone hydrogen bonds and the probabilities of formation of hydrogen bonds of other types. On the other hand, there is a significant correlation (the correlation coefficient is 0.85) between the probability of formation of backbone–backbone hydrogen bonds (assigned to acceptors) and the average number of contacts (the ‘contact scale’).

### 3.2 Prediction of amyloidogenic regions in peptides by the FoldAmyloid program

The expected values (the expected packing density or probability of hydrogen bond formation) are used as a scale for constructing a profile (the packing density profile or hydrogen bond probability profile, respectively) for a protein sequence. First, the corresponding expected value is ascribed to each residue of the sequence (it equals to the average value observed for this type of residue in spatial structure); then, these numbers are averaged inside the window and the average is assigned to the central residue of the window.

The ‘smoothed’ expected value for every position of the polypeptide chain provides the final profile (the profile of the expected packing density or the profile of probability of hydrogen bonds formation, depending on the used scale) which is directly used during prediction of amyloidogenic regions. On the smoothed profile, we predict a region as amyloidogenic if all its residues lie above the given cutoff in the profile, and the size of the region is greater than or equal to the size of the sliding window used.

To evaluate the quality of our predictions as well as to obtain the optimal values for two adjustable parameters (the cutoff value and the size of the sliding window) for each method (based on each scale), we constructed a database of amyloidogenic and non-amyloidogenic peptides by merging the data published by Fernandez-Escamilla et al. (2004) and Thompson et al. (2006). The dataset of Thompson et al. (AmylHex) is a database of fibril-forming and non-fibril-forming hexapeptides. The dataset of Fernandez-Escamilla et al. contains peptides of different sizes (typically larger than six residues) and in fact consists of two parts: the first part was derived by the authors from literature while the second part was experimentally investigated by the authors in standard conditions. The whole merged database consisted of 407 peptides; 144 of these were amyloidogenic while 263 were non-amyloidogenic. The database is available at http://antares.protres.ru/fold-amyloid/amyloid_base.html.

We attempted to discriminate between amyloidogenic peptides and non-amyloidogenic ones in this merged database. A peptide was predicted to be amyloidogenic if at least one amyloidogenic region was predicted in it.
The results of the predictions and their comparison with experiment are presented in the form of receiver operator characteristic (ROC) curves: sensitivity and specificity are plotted for different sliding window sizes (different curves) at varying cutoff values (along the curve). Figure 2a represents the results of prediction using the scale of the expected packing density. In Figure 2b, the results are shown for the scale of the probability of formation of backbone–backbone hydrogen bonds assigned to donors (further termed a ‘donor scale’). Figure 2c displays the results obtained with the scale of the probability of formation of backbone–backbone hydrogen bonds assigned to acceptors (further termed an ‘acceptor scale’).

For all three scales, the quality of predictions was approximately the same for the sliding window size of three and five amino acid residues. When the size of the sliding window was seven amino acid residues and especially when the size of the sliding window was one amino acid residue (that is, without averaging), the quality of predictions was worse (see Fig. 2). Thus, we can use the sliding window size of either three or five residues. Taking into account our previously reported data (Galzitskaya et al., 2006b) and the sliding window size of five amino acid residues gave better results on the prediction of amyloidogenic fragments in a database of amyloidogenic proteins, we decided in favor of the sliding window size of five amino acid residues for all three scales.

Table 1. Correlation coefficients between the scales obtained from the statistics observed in protein structures

<table>
<thead>
<tr>
<th></th>
<th>Contact scale</th>
<th>Backbone-backbone scale</th>
<th>Backbone-side-chain scale</th>
<th>Side-chain-backbone scale</th>
<th>Side-chain-side-chain scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Donors</td>
<td>Acceptors</td>
<td>Donors</td>
<td>Acceptors</td>
<td>Donors</td>
</tr>
<tr>
<td>Contact scale</td>
<td>0.62</td>
<td>0.85</td>
<td>−0.25</td>
<td>−0.22</td>
<td>−0.16</td>
</tr>
<tr>
<td>Backbone-backbone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scale</td>
<td>Donors</td>
<td>0.62</td>
<td>0.81</td>
<td>0.35</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Acceptors</td>
<td>0.85</td>
<td>−0.12</td>
<td>−0.21</td>
<td>−0.12</td>
</tr>
<tr>
<td>Backbone-side-chain</td>
<td>−0.25</td>
<td>0.35</td>
<td>−0.12</td>
<td>0.44</td>
<td>0.25</td>
</tr>
<tr>
<td>scale (donors)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Side-chain-backbone</td>
<td>−0.22</td>
<td>−0.06</td>
<td>−0.21</td>
<td>0.44</td>
<td>0.85</td>
</tr>
<tr>
<td>scale (donors)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Side-chain-side-chain</td>
<td>−0.16</td>
<td>0.00</td>
<td>−0.12</td>
<td>0.25</td>
<td>0.85</td>
</tr>
<tr>
<td>scale (donors)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Correlation coefficients higher than 0.60 are marked bold.

The correlation between the previously reported data (Galzitskaya et al., 2006b) and the results of this study are also shown in Table 1.

The results of predictions and their comparison with experiment are presented in the form of receiver operator characteristic (ROC) curves: sensitivity and specificity are plotted for different sliding window sizes (different curves) at varying cutoff values (along the curve). Figure 2a represents the results of prediction using the scale of the expected packing density. In Figure 2b, the results are shown for the scale of the probability of formation of backbone–backbone hydrogen bonds assigned to donors (further termed a ‘donor scale’). Figure 2c displays the results obtained with the scale of the probability of formation of backbone–backbone hydrogen bonds assigned to acceptors (further termed an ‘acceptor scale’).

For all three scales, the quality of predictions was approximately the same for the sliding window size of three and five amino acid residues. When the size of the sliding window was seven amino acid residues and especially when the size of the sliding window was one amino acid residue (that is, without averaging), the quality of predictions was worse (see Fig. 2). Thus, we can use the sliding window size of either three or five residues. Taking into account our previously reported data (Galzitskaya et al., 2006b) and the sliding window size of five amino acid residues gave better results on the prediction of amyloidogenic fragments in a database of amyloidogenic proteins, we decided in favor of the sliding window size of five amino acid residues for all three scales.

Basing on the ROC curves, we selected the cutoff values optimal for amyloidogenic predictions. The optimal vales were the following (with the optimal size of the sliding window): 0.697 for the method based on the donor scale, 0.671 for the method based on the acceptor scale and 21.4 contacts per residue for the method based on the scale of the expected packing density. With the optimal parameters, the scale of the expected packing density correctly predicts 108 of 144 (75%) amyloidogenic peptides and 194 of 263 (74%) non-amyloidogenic ones. The donor scale correctly predicts 100 of 144 (69%) amyloidogenic peptides and 206 of 263 (78%) non-amyloidogenic ones. The acceptor scale correctly predicts 111 of 144 (77%) amyloidogenic peptides and 194 of 263 (74%) non-amyloidogenic ones. Thus, all three scales allow predicting the status of peptides (amyloidogenic or non-amyloidogenic) with a comparable efficiency.

The other three considered scales (the probabilities of formation of backbone-side-chain, side-chain-backbone and side-chain-side-chain hydrogen bonds, see Table 1) were unsuccessful in predictions (data not shown). All the obtained ROC-curves lay below the diagonal. That is, non-amyloidogenic peptides in fact have a larger probability to form hydrogen bonds of these types compared to amyloidogenic peptides. However, it should be noted that our database contains no amyloidogenic peptides of the second type (with a ‘polar’ template) which could be predicted by the last of the mentioned scales. The problem is that a ‘polar’ template is rather rare. Thus, we cannot presently make a large-scale test of the prediction methods for the amyloidogenic fragments of the second type. However, since our method based on the scale of side-chain–side-chain hydrogen bonds finds both amyloidogenic fragments of Sup35 (data not shown), we hope that it will work also when more amyloidogenic fragments of the second type are known.

Thus, there are three scales which allow predicting amyloidogenic fragments (or rather, the capability of a peptide to be amyloidogenic): the scale of the packing density (contact scale), and two scales of the probability to form backbone–backbone hydrogen bonds (assigned to donor and to acceptor residues, termed donor and acceptor scales, respectively). To take into consideration the above-mentioned scales simultaneously, we have constructed several ‘hybrid’ scales by merging the individual scales with equal weights. For this purpose, we have normalized each scale to obtain the average (over 20 values) value equal to zero and standard deviation equal to unity, and then we averaged the normalized scales to obtain a hybrid scale. In Figure 3, ROC curves for predictions by FoldAmyloid with some hybrid scales are shown (the size of the sliding window was five). As shown in the figure, the hybrid scales also give the quality of predictions comparable to that of the two original scales. The hybrid scale which included all three scales with equal weights (contacts+donors+acceptors which is shown as a solid line in Fig. 3) correctly predicts 80% of amyloidogenic peptides.
Fig. 2. ROC curves for prediction of amyloidogenic regions in peptides by the FoldAmyloid method with different scales obtained from the statistics of protein structures: (a) with the scale of expected packing density, (b) with the scale of probability of formation of backbone-backbone hydrogen bonds (donors), (c) with the scales of probability of formation of backbone-backbone hydrogen bonds (acceptors).

(115 of 144 peptides) and 72% of non-amyloidogenic ones (189 of 263 peptides) with cutoff value 0.062.

We have compared the quality of predictions by our method (with different scales) with the predictions of the other methods. We took TANGO (Fernandez-Escamilla et al., 2004) and PASTA (Trovato et al., 2007) methods which also allow predicting amyloidogenic regions from protein sequence. To correctly compare different methods, their performance should be assessed on the same database.

Trovato et al. (2007) compared the work of their method (PASTA) with that of TANGO on a base which makes in fact a third of our database [these were 179 peptides which were originally derived by Fernandez-Escamilla et al. (2004) from literature]. Thus, we predicted amyloidogenic regions separately for this set. The quality of prediction on the same database by our program FoldAmyloid with different scales [the expected packing density scale, the scale of probability of backbone-backbone hydrogen bonds, and the hybrid scale (1:1)] as well as by TANGO and PASTA are given in Table 2.

Table 2. Performance of different methods of prediction of amyloidogenic regions on a database of 179 peptides

<table>
<thead>
<tr>
<th>Method</th>
<th>Scale</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>FoldAmyloid</td>
<td>Contact scale (Gažitkaya et al., 2006b)</td>
<td>0.74</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Hydrogen bond scale (donors)</td>
<td>0.73</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>Hydrogen bond scale (acceptors)</td>
<td>0.74</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>Hybrid scale (contacts+donors+acceptors)</td>
<td>0.79</td>
<td>0.74</td>
</tr>
<tr>
<td>TANGO</td>
<td>(Fernandez-Escamilla et al., 2004)</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>PASTA</td>
<td>(Trovato et al., 2007)</td>
<td>0.78</td>
<td>0.78</td>
</tr>
</tbody>
</table>

One can see that all analyzed methods are of a comparable efficiency. It should be noted that the quality of TANGO and PASTA predictions remains unknown if to consider the merged database. Besides that, our method with the side-chain-side-chain scale is able to predict not only the ‘hydrophobic’ template but also the ‘polar’ one.
3.3 Prediction of amyloidogenic regions in proteins by the FoldAmyloid method

It should be underlined here the different situation in aggregation process between peptides and globular proteins: in the first case peptide is accessible to the solvent, but in the second case the native state of the globular protein needs to be destabilized in order to the amyloidogenic regions of an amino acid sequence should be exposed to the solvent in order to promote aggregation (Tartaglia et al., 2008). Therefore, we tried to predict amyloidogenic regions in proteins. From our previous database of amyloidogenic proteins (Galzitskaya et al., 2006b) where positions of amyloidogenic regions is now determined by experimental data, we took four proteins with globular structure (β2-microglobulin, transthyretin, lysozyme and myoglobin) as well as completely unstructured τ-protein and human prion protein which has both globular and unstructured domains. Of the globular proteins, two (transthyretin and β2-microglobulin) are partly β-structural while myoglobin is α-helical protein and lysozyme contains both α-helices and β-strands. In Figure 4, the results of the predictions are shown for three individual scales which have demonstrated good performance during predictions of amyloidogenic status of peptides. These are the scale of expected contact density and two scales of probability of backbone–backbone hydrogen bonds: donor and acceptor scale.

Totally, four considered globular proteins have eight amyloidogenic fragments found experimentally. There are two amyloidogenic fragments for transthyretin [residues 10–19 (Chamberlain et al., 2000) and 105–115 (Jaroniec et al., 2002)], three amyloidogenic fragments for β2-microglobulin [residues 20–41 (Kozhukh et al., 2002), 59–71 (Jones et al., 2003) and 83–89 (Ivanova et al., 2004)], one amyloidogenic fragment for lysozyme [residues 26–123 (Franz et al., 2006)] or, by another data, residues 49-64 (Krebs et al., 2000) and two amyloidogenic fragments for myoglobin [residues 7–18 (Picotti et al., 2007) and 101–118 (Fandrich et al., 2003)]. Of these eight amyloidogenic fragments, the expected packing density scale correctly predicts seven fragments (except for the first fragment in myoglobin). The scales of probability of backbone–backbone hydrogen bonds formation show a comparable efficiency: the donor scale allows finding six fragments (the fragment in lysozyme and one of the fragments in β2-microglobulin are missed), and the acceptor scale allows finding all eight fragments.

In the case of unstructured τ-protein, only contact scale finds the experimentally detected (von Bergen et al., 2000) region (306–311). In the partially unstructured human prion protein, all three scales predict both regions detected by experiment [residues 132–160 (Torrent et al., 2005) and 178–193 (Thompson et al., 2000)].

4 THE FOLDAMYLOID SERVER

The FoldAmyloid web server is available at http://antares.protes.ru/fold-amyloid. The server takes an amino acid sequence (in the FASTA format) as an input and calculates a profile of the requested type [depending on the used scale: of the expected number of contacts, of the expected number of backbone-backbone hydrogen bonds (donors), of the hybrid scale (contacts and donor), of the expected number of backbone-backbone hydrogen bonds (acceptors), of the hybrid scale (contacts+donors+acceptors)], along the sequence. If five or more residues in the profile lie above the given cutoff, we predict this region as amyloidogenic. The user can change the following values: the sliding window size (the default value is five amino acid residues), the reliable frame size (the default value is also five amino acid residues) and the cutoff value (the default values are 21.4 for the packing density scale, 0.697 for the donor scale, 0.09 for the twice hybrid scale, 0.671 for the acceptor scale and 0.062 for the triple hybrid scale). One can also add one’s
own scale (the user’s scale) by printing 20 values (corresponding to 20 types of amino acid residues) in the offered field. In the default variant, the server gives a list of amyloidogenic regions in the query sequence, but the user can also obtain the profile from which this prediction is done. In the output, the values of all parameters used during the prediction are listed. In addition, the amino acid composition of the protein is also reported. Minimal, maximal and average values for the profile are also given.

**Funding:** This work was supported by the Russian Academy of Sciences [‘Molecular and Cell Biology’ (01200959110) and ‘Fundamental Science to Medicine’ programs], by the Russian Foundation for Basic Research (08-04-00561), by the ‘Russian Foundation for Science Support Foundation’ program (NSH-2791.2008.4), by the Federal Agency for Science and Innovations (02.740.11.295) and by the Howard Hughes Medical Institute (grant 55005607).

**Conflict of Interest:** none declared.

REFERENCES


