ISSN 0006-2979, Biochemistry (Moscow), 2009, Vol. 74, No. 2, pp. 186-193. © Pleiades Publishing, Ltd., 2009.
Original Russian Text © O. V. Galzitskaya, 2009, published in Biokhimiya, 2009, Vol. 74, No. 2, pp. 229-237.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM08-263, December 21, 2008.

Are the Same or Different Amino Acid Residues Responsible for Correct and Incorrect Protein Folding?

O. V. Galzitskaya

Institute of Protein Research, Russian Academy of Sciences, ul. Institutskaya 4, 142290 Pushchino, Moscow Region, Russia; fax: (495) 632-7871; E-mail: ogalzit@vega.protres.ru

Received August 13, 2008

Abstract—It has been shown for 20 proteins that amino acid residues included into the protein folding nucleus, determined experimentally, are often involved in the theoretically determined amyloidogenic fragments. For 18 proteins, Φ-values indicative of the extent of residue involvement into the folding nucleus are on average higher for amino acid residues within amyloidogenic regions. Amyloidogenic fragments were predicted for 20 proteins by two methods chosen from four on the basis of comparison of prediction of amyloidogenic regions known from experimental data. Since theoretical folding nuclei are detected by the protein three-dimensional structure and amyloidogenic regions by the protein chain primary structure, the detected regularity makes possible predictions of folding nucleation sites on the basis of amino acid sequence.

DOI: 10.1134/S0006297909020096

Key words: folding nucleus, amyloid fibril, globular proteins, amyloidogenic regions, FoldUnfold, Tango, Zyggregator

The fact that a protein is able to fold incorrectly or to form a partially unfolded state that later undergoes irreversible aggregation, points to the problems that emerge during protein folding. The inability of a protein chain to fold into a native structure results in change in the protein function and subsequent disease [1]. Amyloid diseases are a special group of conformational diseases. A common feature of these diseases is formation of amyloid fibrils in which separate β-sheets are oriented at right angle to the main axis of fibrils. Fibril formation is a striking example of change in the conformation of a protein molecule caused by increased fraction of β-structure [2, 3]. In general, fibril formation is independent of the protein precursor structure but largely depends on amino acid sequence of the protein.

There are various types of amyloid diseases—hereditary or familial, infectious such as prion diseases, sporadic forms like most forms of Alzheimer's disease, as well as diseases associated with medical treatment like hemodialysis, or those concomitants of different diseases like rheumatoid arthritis. Many diseases like amyloidoses caused by transthyretin can be both hereditary and sporadic. Understanding of molecular mechanisms of fibril formation would be helpful in explaining the pathogenesis of these diseases.

Results of many investigations show that amyloid fibril formation is a fundamental feature of protein molecules as a whole rather than of a narrow class of so-called amyloidogenic proteins associated with amyloid diseases, as it was thought until recent times [2]. In this case, normal proteins become toxic after fibril formation [4]. However, it is known that immature water-soluble fibrils are more cytotoxic than mature insoluble amyloid fibrils. The structure of these toxic immature fibrils is rich with β strands and should be universal, because antibodies emerging in response to precursors of amyloid Aβ peptide, associated with Alzheimer's disease, can also bind precursors of amyloid fibrils formed also of other proteins with quite different amino acid sequence [5]. It seems that the mechanism of toxicity of these precursors should also be similar for fibrils formed by different proteins.

For some proteins, sites that could be responsible for amyloid fibril formation have been determined experimentally. Experiments show that fibril formation depends on experimental conditions, and denaturing factors stimulate fibril formation; for aggregation to occur, proteins should first be completely or partially unfolded [1].

Aggregation of peptides and proteins associated with such type of amyloidoses as type II diabetes and Alzheimer's and Parkinson's diseases do not require preliminary protein unfolding because these proteins are mainly unstructured under physiological conditions [6]. However, most natively unfolded proteins do not aggregate *in vivo* [7]. This fact shows that protein unfolding is a

necessary but an insufficient condition for amyloid fibril formation. Most likely, there should be special motifs in amino acid sequence, which after exposure to the solvent are more predisposed to aggregation compared to other regions of amino acid sequence. Experimental facts support the hypothesis that small protein regions are responsible for amyloidogenic behavior [8-10].

Thus, the amyloidogenic sequence of six amino acid residues (STVIIE), incorporated into N-terminus of a protein molecule, causes fibril formation in the water-soluble protein α-spectrin, whose SH3 domain does not form amyloid fibrils under the same conditions [11]. In this case, the fibril nucleus, inaccessible for proteases, contains short sequences of a soluble globular protein adjacent to the amyloidogenic region. Thus, short amyloidogenic regions available for intramolecular interactions enhance the self-assembly reaction by drawing the rest of the protein into the fibrillar aggregate. The reliable identification of such amyloidogenic regions in proteins makes possible their application as a target for prevention of the process of amyloid fibril formation.

A similar result was earlier obtained for mouse β_2 microglobulin that does not form amyloids. Seven amino acid residues in this protein were substituted by residues from human β_2 microglobulin (that forms amyloids) in the place where the highest difference between primary structures of these proteins is observed (residues 83-89). Such substitution in the protein results in amyloid formation *in vitro* [10]. It should be noted that the (NHVTLSQ) peptide proper forms amyloids, but the peptide composed of the same amino acids but arranged in a random order (QVLHTSN) does not [10]. The experimental data allowed the authors to propose a model of a β-zigzag-like rod decorated by the rest of the protein molecule.

Despite active investigations concerning amyloid fibril formation, detailed understanding of molecular mechanisms forming the basis for transformation of water-soluble proteins into insoluble amyloid aggregates is still an unsolved problem.

Since the protein chain is able to fold into a regular native structure as well as into an irregular one, both these processes are related with each other. It can be supposed that amino acid residues of amyloidogenic regions can play an important role both in irregular and regular protein folding. Data obtained for two proteins are contradictory. Thus, it was shown for acylphosphatase that folding and aggregation pathways are different [12]. For the other protein, β_2 microglobulin, the partially unfolded state is involved in amyloid fibril formation [13]. This suggests that the free-energy landscapes for regular and irregular folding coincide to the partially unfolded state and diverge at the level of the native-like structure. Since there are only a few experimental data with simultaneously obtained results concerning residues playing an important role in regular and irregular folding, it is impossible to draw a correct conclusion. One can clarify this problem by considering both experimental and theoretical methods to predict amino acid residues important for regular and irregular protein folding.

This work deals with 20 proteins for which residues important in regular protein folding are determined experimentally. Amyloidogenic fragments for the 20 proteins were predicted using two of four methods chosen by comparing the results of prediction of amyloidogenic regions based on known experimental data. Amino acid residues incorporated in the experimentally distinguished folding nucleus appear to be often involved in theoretically detected amyloidogenic fragments. This result suggests that amino acid residues of the folding nucleus can be important for both regular and irregular protein folding.

MATERIALS AND METHODS

Protein databases. This work deals with six proteins with experimentally detected regions responsible for formation of amyloid fibrils. PDB codes are presented in Table 1, and amino acid sequences were taken from the SWISS-PROT database [14] (http://us.expasy.org/sprot/). This database was used to compare four methods for prediction of amyloidogenic and aggregation regions in a protein chain.

Another database included 20 protein structures with known experimental Φ-values for numerous residues of the protein chain. Names of proteins with known experimental Φ-values and corresponding PDB codes are given in Table 2. The list of considered mutations is given on the site http://phys.protres.ru/resources/phi_values.htm.

It is assumed that experimental Φ-values show the extent of the residue and its "native" environmental condition in a transition state [15]. The value $\Phi = 1$ means that in the given residue in the transition state "everything is native" (the structure proper and the environment); the value $\Phi = 0$ means that the given residue in the transition state is either devoid of its native structure or of its native environment; intermediate Φ-values are usually interpreted as an indication that the environment is only partially native.

Methods used for prediction of amyloidogenic and aggregation regions in protein chain. *FoldUnfold* [16-18]. The establishment of a sufficient number of contacts between residues is necessary to compensate for a loss of conformational entropy upon protein folding. Thus, native structure is the result of balance between conformational entropy and energy of interaction between residues. Taking this fact into account, two suppositions were put forward. On one side, if the mean environmental density (the mean number of residues at a given distance) is lower than a "normal value" of this parameter for globular proteins, then the protein will not fold into a native structure and will be natively-unfolded. On the other side, if mean density of the environment exceeds a

Table 1. Amyloidogenic regions observed in experiments and predicted by different methods. Correctly predicted regions are shown in bold

Protein name (source) and number of amino acid residues	PDB code	Experimentally observed amyloidogenic regions	Predicted regions			
			FoldUnfold	SecStr[a]	Tango	Zyggregator[b]
Human acylphosphatase 98	1aps	16-31 [12]; 87-98 [12]	**19-25**; **91-98**	64-65; 85-86	33-39	11; 12; **17-23**; 35-39; 44-50; **92-98**
Human β_2 microglobulin 99	1im9	20-41 [27]; 59-71 [28]; 83-89 [10]	**24-30**; **60-70**	**37-39**; 91-95	**22-27**; **60-68**	**22-28**; 54; 57; 58; **61-68**; **81-86**
Human transthyretin 127	1bm7	10-19 [29]; 105-115 [31]	**10-17**; 28-34; **105-113**	**12-17**; 26-32; 93-96; **106-110**	**12-20**; 25-30; **105-111**; 116-123	28-31; 76; 80; 81; 93-98; **108-109**; 117-121
Human lysozyme 130	1iww	26-123 [34]; 32-108 [34]	**25-33**; **54-65**; **76-82**; **107-115**	**25-31**; **82-86**; **107-112**	–	**24-32**; **37-46**; **53-67**; **75-81**; **92-93**; 124-127
Human myoglobin 153	1wla	7-18 [35]; 101-118 [36]	**8-14**; 27-33; **100-116**; 134-140	**7-14**; 26-32; 68-71; **110-115**	**10-14**; 27-34; 66-72; **101-107**	**9-12**; 31-33; 65-70; **108-114**
Human prion 253	1qm0	169-213 [40]	136-142; 159-166; **176-183**; 240-253	**175-183**; **209-214**; 242-249	239-252	126-128; **177-179**; **190**; **212-213**; 215; 232-234; 243-253

[a] Predictions for this method were done by the authors of the program.
[b] Predictions for three proteins (lysozyme, myoglobin, and prion) were taken from [21].

"normal value" of this parameter for globular proteins, then it will be more advantageous for the protein chain to form an amyloid fibril (in which all expected contacts will be able to be realized) than the native globular one. It is supposed that the packing within fibrils is tight (because it was shown in experiments that the amyloid fibril is thermostable, protease-insensitive, and rich with β structure [19]), therefore regions that could be potentially tightly packed (have high environmental density) will have the tendency to form amyloid fibrils. This approach was tested [16-18], and it was found that in fact the side chain regions characterized by increased supposed environmental density are in most cases responsible for amyloidogenic properties of proteins and peptides. We have used the slipping window of seven residues because in this case the lowest number of false predictions is obtained [18], and the region was predicted as amyloidogenic if seven or more residues are above the considered threshold level of environmental density (i.e. more than 21.4 residues drawn together in radius of 8 Å). It should be noted that neither of the considered methods uses this restriction, although it is obvious that the region of two residues is not able to exhibit amyloidogenic properties.

Zyggregator [20, 21]. This method was recently proposed for searching regions in polypeptide chains, which are important for aggregation. The method is based on plotting the profile of the aggregation properties of a region with consideration of the same factors that were proposed by the same authors for prediction of aggregation rate [22]: hydrophobicity of the region under consideration, its predisposition to formation of α-helices and β-structure, as well as absolute value of net charge. This method is available at Internet address: http://www-vendruscolo.ch.cam.ac.uk/software.html.

Tango [23]. The precursor of this program was the Agadir program [24], which predicts the probability of helical structure formation. Agadir uses the approach of statistical mechanics for calculation of the relative probability of amino acid residues existing in helical or globular conformation, taking into account several empirical

Table 2. Experimental Φ-values and predicted amyloidogenic regions

Protein name (PDB code). Number of residues in the protein (number of experimentally studied residues)	Regions predicted by the SecStr program[a]	Regions predicted by FoldUnfold program	<Φ> inside amyloidogenic region/<Φ> outside amyloidogenic region	Residues with Φ-values above 0.5 (Φ-value)[c]
B1 domain of protein G (1pgb) 56 (20)	28-32	1-7[b]	0.35/0.32	*32 (0.55)*; 46(0.96); 47 (0.67); 49 (0.84)
B domain of protein A (1bdc) 60 (15)	**13-18**; 31-34	**14-21**	0.69/0.31	***14 (0.65); 17 (1.0); 18 (1.0);*** *32 (0.57)*
SH3 domain of α-spectrin (1shg) 57(13)	**8-13**; **30-34**; **42-46**	**8-14**; **31-35**; **40-44**[b]	0.21/0.42	47 (0.63); 52 (0.58); 53 (0.61); 55 (0.53)
src SH3 domain (1srm) 56 (26)	**9-13**; 31-34; **42-46**	**9-16**; **41-47**	0.45/0.24	30 (0.62); *32 (0.55)*; 35 (0.77); ***44 (0.54)***; **47 (0.95)**; 48 (0.72); 50 (0.86); 53 (0.68); 55 (0.56); 56 (0.71)
Sso7d (1bf4) 63 (15)	**18-23**	**22-26**; 30-34[b]	0.51/0.29	**25 (0.65)**; 45 (0.59); 57 (0.54); 59 (0.60)
CI2 (2ci2) 64 (34)	19-23; **39-53**	29-35; **39-52**	0.23/0.20	18 (0.70); ***49 (0.53)***
B1 IgG-binding domain of protein L (2ptl) 64 (38)	18-26	7-13	0.39/0.28	4 (0.51); **7 (0.62)**; **8 (0.53)**; 14 (0.85); *21 (0.75)*
Im9 (1imp) 86 (19)	**15-19**; **33-37**	**15-22**; **34-40**	0.36/0.31	13 (0.98); ***15 (0.57)***; ***16 (0.52)***
Im7 (1cei, PCA) 85 (19)	**33-36**; 41-44; 67-70	**34-44**	0.51/0.43	3 (0.52); 7 (0.56); 16 (0.72); 18 (0.52); 27 (0.56); ***37 (0.51)***; ***44 (1.0)***; *68 (0.86)*; 77 (0.86); 78 (0.78)
TI I27-domain of titin (1tiu) 89 (22)	9-13; **57-59**; **64**-70	**56-64**	0.55/0.37	23 (0.82); 49 (0.66); ***56 (0.52)***; ***58 (0.79)***; **60 (0.67)**; 71 (0.63); 75 (0.63)
Barstar (1btb) 89 (22)	2-5; **37-41**; **46-52**; 70-74	**35-56**	0.54/0.52	*5 (0.63)*; 9 (0.72); 18 (0.69); 25 (0.68); **36 (0.70)**; ***37 (0.59)***; ***50(0.77)***; 68 (0.52); *72 (0.81)*; 77 (0.90); 79 (0.63); 85 (0.51)
TNfn3-domain of tenascin (1ten) 90 (25)	**17-23**; 47-49; 68-71	**19-26**	0.39/0.28	36 (0.53); *48 (0.67)*; *70 (0.54)*
FNfn10-domain of fibronectin (1ttf) 94 (16)	8-11	28-35; 69-75	0.47/0.43	*8 (0.55)*; **34 (0.61)**; 36 (0.52); **72 (0.51)**; **74 (0.65)**; 92 (0.79)
U1A (1urn) 96 (10)	**38-43**; **53-58**	9-15; **30-45**; **53-59**	0.37/0.19	***44 (0.60)***
S6 (1ris)	24-27; **48-51**; 59-63; 85-90	1-9; **45-51**	0.45/0.27	**6 (0.52)**
Acylphosphatase (1aps, NMR) 97 (10)	64-65; 85-86	19-25; 91-98	0.47/0.30	11 (0.93); 47 (0.54); 54 (0.98); **94 (0.76)**
FKBP12 (1fkb) 107 (22)	53-56; **96-104**	**96-107**	0.36/0.28	23 (0.52); 27 (0.85); 63 (0.51); ***101 (0.57)***
Barnase (1rnb) 110 (19)	32-35	87-98; 103-109	0.66/0.26	13 (0.60); 15 (0.80); 57 (1.0); **87 (0.80)**; **90 (0.9)**; **91 (1.0)**; **96 (0.40)**
Villin 14T (2vil) 126 (18)	**18-23**; 80-84; 90-93; 105-111	**18-24**; 42-49; 58-66	0.53/0.51	***23 (0.65)***; 28 (0.58); **44 (0.85)**; **46 (0.69)**; **48 (0.62)**; 73 (0.69); 77 (0.52); 78 (0.56); *81 (0.75)*; *84 (0.68)*
CheY (3chy) 128 (9)	5-10; **18-24**; 67-71; **82-86**	**16-22**; 51-57; **80-86**	0.30/0.31	35 (0.75); 37 (0.60); 41 (0.68)

Note: Residues with high Φ-values, which fall in predicted amyloidogenic fragments, are shown in bold for the FoldUnfold method and in italics for the SecStr. In this case, similar amyloidogenic regions predicted by both methods are shown in bold.

[a] Predictions by this method were done by the authors of the program.

[b] For these proteins amyloidogenic regions can be predicted at the slipping window of five residues. For all other proteins, the window was of seven residues.

[c] If several Φ-values are known for a given residue (several different mutations were introduced for this residue), then Φ-values were averaged.

parameters. Tango uses the same approach as the Agadir program, but considers already four states: unfolded, helix, β turn, and aggregation. This approach is available at the address: http://tango.crg.es/protected/academic/calculation.jsp.

SecStr [25]. This method is based on the supposition that regions with high predisposition to formation both of α-helices and β-structure will tend to form amyloid fibrils. In other words, these regions will work as conformation switchers. The method uses six programs for prediction of secondary structure and issues the total result as the probability of formation for the assigned amino acid sequence of three states: α-helices, β-structure, and coil. The name of the method corresponds to that of the program that calculates secondary structure. The program is available at the address: http://biophysics.biol.uoa.gr.

RESULTS AND DISCUSSION

Comparison of four methods for prediction of amyloidogenic and aggregation regions in a protein chain. Since we should choose a method making the best predictions for proteins in which amyloidogenic regions are known from experimental works, four methods were compared. The database of the amyloid fibril-forming proteins with experimentally detected amyloidogenic regions was used (Table 1). Usually amyloids are called fibrils; they bind dyes Congo red and Thioflavin T and have X-ray diffraction patterns similar to that of β structure [26].

Calculations by four methods were done to compare the ability of each for six proteins. The regions detected experimentally and predicted using these methods are shown in Table 1.

The following acylphosphatase regions important for amyloid fibril formation were detected in experiments: regions 16-31 [12] and regions 87-98 [12]. Only FoldUnfold and Zyggregator predict these two regions. It should be noted that this protein is not involved in diseases associated with formation of amyloid fibrils.

The following β_2-microglobulin regions important for amyloid fibril formation were detected in experiments: residues 20-41 [27], 59-71 [28], and 83-89 [10]. Tango and FoldUnfold predict two regions (the third fragment is predicted by our method for the window of five residues). Zyggregator predicts all three regions.

For transthyretin the most amyloidogenic fragments are two regions: residues 10-19 of A strand of internal β sheet, which quickly form fibrils in water at low pH [29, 30], and residues 105-115 acquiring in the amyloid fibril the extended conformation similar to that in the native protein [31]. Methods FoldUnfold, SecStr, and Tango correctly predict these two regions (10-19 and 105-115), while Zyggregator predicts only the second region (Table 1).

An experimentally detected amyloidogenic fragment of lysozyme corresponding to residues 49-64 [32, 33] is included in a β-structural part of the native protein. It has been recently shown that amyloid fibrils are formed at low pH. These fibrils were analyzed by limited proteolysis and infrared spectroscopy. After fibril cleavage by pepsin, fragments 26-123 and 32-108 were found [34]. Three methods except Tango predict sites, which fall into the region of residues 26-123.

It should be noted that human myoglobin is not involved in diseases associated with formation of amyloid fibrils. Myoglobin regions including residues 7-18 [35] and 101-118 [36] were found experimentally to be important for amyloid fibril formation. These fragments are found by all four methods.

Despite the huge amount of experimental data on prion aggregation regions, it is difficult to determine just which regions are amyloidogenic in this protein. It was shown that the first α-helix of human prion (residues 144-153) plays an important role in amyloid fibril formation [37, 38]. On the other side, peptides corresponding to the prion α-helical fragments have been synthesized (the first helix being residues 144-153, the second residues 178-193, and the third residues 198-218) [39] and it was shown that only the second α-helix (residues 180-193 and 178-193) forms amyloids. Only the recent experimental work on the hydrogen–deuterium exchange has shown that the region 169-213 corresponds to a structural nucleus of amyloid fibril [40]. Three methods (excluding Tango) predict regions within the sequence 169-213.

It appears that Tango finds six fragments of 11 (Table 1, correctly predicted fragments are shown in bold). Zyggregator and FoldUnfold find 10 of 11 experimentally detected regions and SecStr finds seven regions. An important characteristic of the programs is also the number of additional, false predictions. FoldUnfold and Tango have the lowest number of additional predictions (six regions) compared to the other considered methods: eight for SecStr and 16 for Zyggregator. It should be noted that FoldUnfold works better than the other methods.

It is seen that different methods give correct prediction in most cases, but several methods should be used for correct prediction of amyloidogenic regions to distinguish the reliable predictions. We shall consider two methods, FoldUnfold and SecStr, for prediction of amyloidogenic regions in 20 proteins in which residues important for protein folding have been experimentally determined.

Search for common amino acid residues incorporated in experimentally distinguished folding nucleus, and residues involved in theoretically determined amyloidogenic fragments. Table 2 shows data for 20 proteins along with residues and Φ-values that exceed 0.5, which indicates that the residue plays an important role in folding of a given protein. Residues with high Φ-values, which fall in predicted amyloidogenic fragments, are shown in bold

in Table 2 for the FoldUnfold method and in italics for SecStr. In this case, similar amyloidogenic regions predicted by both methods are shown in bold. For 15 proteins there is broad agreement for predicted regions, while for two (SH3 domain of α-spectrin and Im9) they completely coincide. It is seen that the two methods used together give more matches with the folding-important residues than either alone (Table 2). The mean of Φ-values for residues incorporated and not incorporated into predicted amyloidogenic regions were calculated using the FoldUnfold program. For 18 proteins, the first value exceeded the second. It should be noted that for two proteins no agreement of residues important for regular and irregular protein folding (CheY and SH3 domain of α-spectrin) was observed, and the mean levels of Φ-values for residues incorporated into amyloidogenic regions predicted using the FoldUnfold program do not exceed mean levels of Φ-values for residues not incorporated into predicted amyloidogenic regions. For the FoldUnfold method the mean level of Φ-values for residues incorporated into amyloidogenic regions for all 20 proteins was 0.41 ± 0.02, whereas the mean level of Φ-values for residues not incorporated into amyloidogenic regions was 0.33 ± 0.01. This result shows that amino acid residues of the folding nucleus may be important for both regular and irregular protein folding.

It should be noted that theoretical folding nuclei are identified by the three-dimensional structure of the pro-

Table 3. Changes caused by mutations in amyloidogenic proteins and peptides

Protein name	Mutation [22]	Changes caused by mutation			
		in observed aggregation rate [22] ($\ln(v_{mut}/v_{wt})$)	in environmental density	in hydrophobicity (kcal/mol) [42]	in predisposition to insertion into β-structure [43]
Amylin	N22A	0.70	−1.40	2.30	0.08
	F23A	−2.65	7.29	−1.88	−0.86
	G24A	−0.05	−2.78	0.39	1.20
	I26A	−2.40	5.82	−1.43	−0.82
	L27A	−0.95	5.47	−1.43	−0.51
	S20G	1.00	1.08	1.24	−1.90
	correlation with $\ln(v_{mut}/v_{wt})$		−0.83	0.91	0.10
Peptides obtained from prion	H111A	0.60	1.83	3.26	−0.02
	H111K	−0.25	4.05	0.10	0.25
	A117V	1.50	−4.04	0.91	1.00
	V210I	0.85	−1.78	0.52	−0.18
	correlation with $\ln(v_{mut}/v_{wt})$		−0.96	0.16	0.50
α-Synuclein	A53T	1.20	0.08	−1.39	1.10
	A76E	−2.70	2.43	−3.30	0.01
	A76R	−0.90	−1.14	−4.34	0.27
	correlation with $\ln(v_{mut}/v_{wt})$		−0.61	0.67	0.97
Amyloid β peptide	A21G	−0.07	2.78	−0.39	−1.20
	E22K	0.90	−0.21	0.14	0.26
	E22Q	2.90	−1.77	1.61	0.22
	E22G	2.05	0.35	2.91	−1.21
	D23N	3.95	−1.08	1.90	0.86
	F19T	−2.50	7.37	−3.27	0.24
	correlation with $\ln(v_{mut}/v_{wt})$		−0.93	0.91	0.23

Note: v_{mut}, aggregation rate for mutated protein or peptide; v_{wt}, aggregation rate for wild type of protein or peptide.

tein [41] and amyloidogenic regions by its primary structure [16-25]. The detected regularity makes possible prediction of nucleation folding sites on the basis of amino acid sequence.

Changes in aggregation rate induced by mutations in proteins and peptides. The fact that the folding-important residues are also important for irregular folding is also confirmed by data showing that the change of an average number of residues drawn together in globular state, used in the FoldUnfold technique, correlates with changes in the rate of aggregation or amyloid fibril formation upon mutation in the studied protein or peptide. Since it is assumed that high predisposition of residues to β-structure formation and high hydrophobicity are important factors for amyloid fibril formation, two additional scales were considered (hydrophobicity [42] and the residue predisposition to β-structure formation [43]) and changes in aggregation rates of mutant proteins were compared with those for the wild-type proteins as well as with changes in the studied parameters (environmental density or the average number of residues drawn together in globular state in the given distance, predisposition of residues to β-structure formation, and hydrophobicity). These changes for four proteins and peptides are shown in Table 3. It is seen that among three considered parameters environmental density exhibits the best correlation with the experimental data.

In addition to "normal", native spatial structure, some proteins are able to form alternative incorrectly folded structures. For 20 proteins with known residues important for "normal" folding (so-called folding nuclei), regions important for amyloid fibril formation were predicted. It appeared that amino acid residues incorporated in the experimentally isolated folding nucleus are often involved in the theoretically identified amyloidogenic fragments. This result suggests that amino acid residues of the folding nucleus may be important for both regular and irregular protein folding.

This work was supported by the Russian Foundation for Basic Research (grant No. 08-04-00561) and the Russian Science Support Foundation.

REFERENCES

1. Dobson, C. M. (1999) *Trends Biochem. Sci.*, **24**, 329-332.
2. Fandrich, M., Fletcher, M. A., and Dobson, C. M. (2001) *Nature*, **410**, 165-166.
3. Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., and Dobson, C. M. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 3590-3594.
4. Bucciantini, M., Calloni, G., Chiti, F., Formigli, L., Nosi, D., Dobson, C. M., and Stefani, M. (2004) *J. Biol. Chem.*, **279**, 31374-31382.
5. Kayed, R., Head, E., Thompson, J. L., McIntire, T. M., Milton, S. C., Cotman, C. W., and Glabe, C. G. (2003) *Science*, **300**, 486-489.
6. Rochet, J. C., and Lansbury, P. T., Jr. (2000) *Curr. Opin. Struct. Biol.*, **10**, 60-68.
7. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) *Proteins*, **41**, 415-427.
8. Tenidis, K., Waldner, M., Bernhagen, J., Fischle, W., Bergmann, M., Weber, M., Merkle, M. L., Voelter, W., Brunner, H., and Kapurniotu, A. (2000) *J. Mol. Biol.*, **295**, 1055-1071.
9. Von Bergen, M., Friedhoff, P., Biernat, J., Heberle, J., Mandelkow, E. M., and Mandelkow, E. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 5129-5134.
10. Ivanova, M. I., Sawaya, M. R., Gingery, M., Attinger, A., and Eisenberg, D. (2004) *Proc. Natl. Acad. Sci. USA*, **101**, 10584-10589.
11. Esteras-Chopo, A., Serrano, L., and Lopez de la Paz, M. (2005) *Proc. Natl. Acad. Sci. USA*, **102**, 16672-16677.
12. Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., et al. (2002) *Nature Struct. Biol.*, **9**, 137-143.
13. Jahn, T. R., Parker, M. J., Homans, S. W., and Radford, S. E. (2006) *Nature Struct. Mol. Biol.*, **13**, 195-201.
14. Bairoch, A., and Apweiler, R. (2000) *Nucleic Acids Res.*, **28**, 45-48.
15. Matouscheck, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M., and Fersht, A. R. (1990) *Nature*, **346**, 440-445.
16. Galzitskaya, O. V., Garbuzynskiy, S. O., and Lobanov, M. Yu. (2006) *Mol. Biol. (Moscow)*, **40**, 821-828.
17. Galzitskaya, O. V., Garbuzynskiy, S. O., and Lobanov, M. Yu. (2006) *J. Bioinform. Comput. Biol.*, **4**, 373-388.
18. Galzitskaya, O. V., Garbuzynskiy, S. O., and Lobanov, M. Yu. (2006) *PLoS Comput. Biol.*, **2**, e177.
19. Kajava, A. V., Baxa, U., Wickner, R. B., and Steven, A. C. (2004) *Proc. Natl. Acad. Sci. USA*, **101**, 7885-7890.
20. Pawar, A. P., Dubay, K. F., Zurdo, J., Chiti, F., Vendruscolo, M., and Dobson, C. M. (2005) *J. Mol. Biol.*, **350**, 379-392.
21. Tartaglia, G. G., Pawar, A. P., Campioni, S., Dobson, C. M., Chiti, F., and Vendruscolo, M. (2008) *J. Mol. Biol.*, **380**, 425-436.
22. Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C. M. (2003) *Nature*, **424**, 805-808.
23. Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004) *Nature Biotech.*, **22**, 1302-1306.
24. Munoz, V., and Serrano, L. (1994) *Nature Struct. Biol.*, **1**, 399-409.
25. Hamodrakas, S. J., Liappa, Ch., and Iconomidou, V. A. (2007) *Int. J. Biol. Macromol.*, **41**, 295-300.
26. Rudall, K. M. (1952) *Adv. Protein Chem.*, **7**, 253-290.
27. Kozhukh, G. V., Hagihara, Y., Kawakami, T., Hasegawa, K., Naiki, H., and Goto, Y. (2002) *J. Biol. Chem.*, **277**, 1310-1315.
28. Jones, S., Manning, J., Kad, N. M., and Radford, S. E. (2003) *J. Mol. Biol.*, **325**, 249-257.
29. Chamberlain, A. K., MacPhee, C. E., Zurdo, J., Morozova-Roche, L. A., Hill, H. A., Dobson, C. M., and Davis, J. J. (2000) *Biophys. J.*, **79**, 3282-3293.
30. MacPhee, C. E., and Dobson, C. M. (2000) *J. Mol. Biol.*, **297**, 1203-1215.
31. Jaroniec, C. P., MacPhee, C. E., Astrof, N. S., Dobson, C. M., and Griffin, R. G. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 16748-16753.
32. Krebs, M. R., Wilkins, D. K., Chung, E. W., Pitkeathly, M. C., Chamberlain, A. K., Zurdo, J., Robinson, C. V., and Dobson, C. M. (2000) *J. Mol. Biol.*, **300**, 541-549.

33. Frare, E., Polverino de Laureto, P., Zurdo, J., Dobson, C. M., and Fontana, A. (2004) *J. Mol. Biol.*, **340**, 1153-1165.
34. Frare, E., Mossuto, M. F., Polverino de Laureto, P., Dumoulin, M., Dobson, C. M., and Fontana, A. (2006) *J. Mol. Biol.*, **361**, 551-561.
35. Picotti, P., de Franceschi, G., Frare, E., Spolaore, B., Zambonin, M., Chiti, F., de Laureto, P. P., and Fontana, A. (2007) *J. Mol. Biol.*, **367**, 1237-1245.
36. Fandrich, M., Forge, V., Buder, K., Kittler, M., Dobson, C. M., and Diekmann, S. (2003) *Proc. Natl. Acad. Sci. USA*, **100**, 15463-15468.
37. Morrissey, M. P., and Shakhnovich, E. I. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 11293-11298.
38. Speare, J. O., Rush, III, T. S., Bloom, M. E., and Caughey, B. (2003) *J. Biol. Chem.*, **278**, 12522-12529.
39. Thompson, A., White, A. R., McLean, C., Masters, C. L., Cappai, R., and Barrow, C. J. (2000) *J. Neurosci. Res.*, **62**, 293-301.
40. Lu, X., Wintrode, P. L., and Surewicz, W. K. (2007) *Proc. Natl. Acad. Sci. USA*, **104**, 1510-1515.
41. Galzitskaya, O. V., and Finkelstein, A. V. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 11299-11304.
42. Fauchere, I. I., and Pliska, V. (1983) *Eur. J. Med. Chem.-Chim. Ther.*, **18**, 369-375.
43. Minor, D. L., Jr., and Kim, P. S. (1994) *Nature*, **371**, 264-267.