nature biotechnology

Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins

Ana-Maria Fernandez-Escamilla[1], Frederic Rousseau[2], Joost Schymkowitz[2] & Luis Serrano[1]

We have developed a statistical mechanics algorithm, TANGO, to predict protein aggregation. TANGO is based on the physico-chemical principles of β-sheet formation, extended by the assumption that the core regions of an aggregate are fully buried. Our algorithm accurately predicts the aggregation of a data set of 179 peptides compiled from the literature as well as of a new set of 71 peptides derived from human disease-related proteins, including prion protein, lysozyme and β2-microglobulin. TANGO also correctly predicts pathogenic as well as protective mutations of the Alzheimer β-peptide, human lysozyme and transthyretin, and discriminates between β-sheet propensity and aggregation. Our results confirm the model of intermolecular β-sheet formation as a widespread underlying mechanism of protein aggregation. Furthermore, the algorithm opens the door to a fully automated, sequence-based design strategy to improve the aggregation properties of proteins of scientific or industrial interest.

A growing body of experimental data has recently identified those residues within a protein sequence that promote ordered aggregation[1–5] and amyloid formation[3,6]. On the basis of these observations, we have developed a statistical mechanics algorithm, TANGO, that identifies the β-aggregating regions of a protein sequence. Our approach considers different competing conformations: β-turn, α-helix, β-sheet, the folded state and β-aggregates, and assumes that in β-aggregates the nucleating regions will be fully buried and tend to satisfy their hydrogen-bonding potential. TANGO takes into account protein stability and physico-chemical parameters such as pH, concentration, ionic strength and trifluorethanol (TFE) concentration. Every segment of a protein can populate each of the conformational states according to a Boltzmann distribution; that is, the frequency of the population of each structural state is relative to its energy, which is derived from statistical and empirical considerations. Therefore, to predict β-aggregating segments of a peptide, TANGO simply calculates the partition function of the conformational phase-space (see Methods).

We compared the output of the TANGO algorithm with data found in the literature on the aggregation of 179 peptides corresponding to sequence fragments of 21 different proteins (**Fig. 1** shows results for 4 proteins; for more information see **Supplementary Table 1** online). We considered a peptide to be aggregating when the spectral property under investigation (mostly circular dichroism (CD) and NMR; **Supplementary Table 1** online) revealed a concentration dependence or when binding to a β-aggregation-reporting dye (thioflavine T) was observed. Sixty-seven out of 179 peptides in our set were experimentally observed to aggregate in the concentration range between 1 μM and 5 mM, whereas the remaining peptides were monomeric in this range. For each residue in a peptide, TANGO computes the percent occupancy of the β-aggregation conformation. We considered peptides as having some aggregation tendency when they possessed segments of at least five consecutive residues populating the β-aggregated conformation.

We observed two regions of confidence in our predictions. First, predictions for peptides with segments of five residues populating the aggregated state at more than 5% per residue were highly accurate, as only 9 of 62 predictions were false positives. Second, 52% of the predictions (12 out of 23) in the 0.2%–5% range were erroneous, suggesting that TANGO is unable to accurately predict low levels of aggregation propensity. However, TANGO yielded a success rate of 87%, correctly predicting 155 out of 179 peptides, with 21 false positives and 3 false negatives. This resulted in a correlation of 0.74 with a P value of 10^{-12}. When we excluded predictions within the 0.2%–5% range, the success rate rose to 92% (142 out of 156 peptides predicted correctly) with only nine false positives (representing 15% of the predicted positives) and three false negatives (3% of the predicted negatives), resulting in a correlation of 0.83.

Our guideline of five consecutive residues with a TANGO score above 5% proved to be a good predictor of aggregation, independent of the size of the protein under study. It must be emphasized, however, that TANGO cannot in principle be used quantitatively to compare peptides or proteins differing widely in sequence. Quantitative comparison of aggregation propensities is only valid when comparing mutants of a common sequence measured under identical conditions, as discussed below for Alzheimer β-amyloid peptide, transthyretin and human lysozyme.

To better estimate the extent to which false predictions are due to inaccuracies of the algorithm or to experimental error, we measured an independent set of 71 peptides derived from human disease-related proteins (prion protein, lysozyme and β2-microglobulin) and also remeasured five peptides from the literature-derived set that were wrongly predicted. We measured all peptides by CD at

[1]European Molecular Biology Laboratory, Meyerhofstrasse 1, Heidelberg D-69117, Germany. [2]SWITCH Laboratory, Flemish Institute for Biotechnology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium. Correspondence should be addressed to L.S. (serrano@embl.de).

Published online 12 September 2004; doi:10.1038/nbt1012

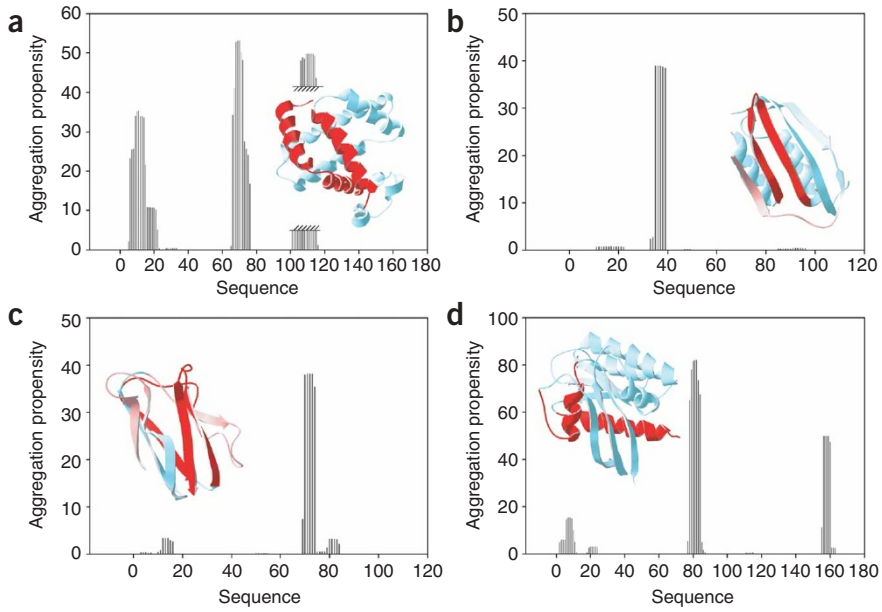

Figure 1 Schematic representation of the predicted aggregating regions in some proteins derived from the literature. For the complete set, see **Supplementary Table 1** online. The aggregation tendency according to TANGO is plotted versus the sequence of the protein. In the inset, aggregating regions are highlighted in red in the three-dimensional structure of the protein (bright red indicates strong aggregation, light red indicates mild aggregation tendency). (**a**) Sperm whale myoglobin (1duk). (**b**) Acyl phosphatase (2acy). (**c**) French bean plastocyanin (9pcy). (**d**) p21-ras (4q21).

two concentrations, 50 μM and 1 mM, in 50 mM phosphate buffer, pH 7.2, at 25 °C (**Supplementary Table 2**, **Supplementary Fig. 1** and **Supplementary Note 1** online). The performance of TANGO on the new set of peptides was approximately the same as that observed on the data extracted from the literature. TANGO correctly predicted 65 of the 71 peptides of our data set (91%) giving a correlation of 0.70, in contrast to 0.74 for the literature set (without exclusion of the 0.2%–5% range) (see **Table 1a**). Moreover TANGO turned out to correctly predict four out of the five peptides that were remeasured from the literature-derived set, suggesting that at least some of the false predictions were in fact due to experimental error.

β-sheet aggregation should not be confused with β-sheet propensity. Although the sequence requirements for both partly overlap, peptides with a high β-sheet propensity do not necessarily aggregate.

Table 1a Comparison of the performance of TANGO on two data sets

	Literature data set[a]	Measured data set[b]
Positives	75%	73%
Negatives	97%	95%
False positives	25%	27%
False negatives	3%	5%
Correlation	0.74	0.70

[a]Data set drawn from the literature (179 peptides). [b]Data set measured in the laboratory (71 peptides).

For example, peptide 55–64 from human lysozyme and peptide 77–86 from β-microglobulin are both part of a β-sheet in their respective crystal structures and have a high β-sheet propensity (**Supplementary Table 2** online). However, the TANGO score of both peptides is near zero and the CD spectra of these peptides show no concentration dependence between 50 μM and 1 mM. Similar results were found for a series of peptides designed elsewhere to form β-hairpins[7] and β-sheets[8]. TANGO indicates that these peptides exhibit substantial secondary structures but do not aggregate; experimentally they were found to be monomeric at concentrations up to 1 mM.

TANGO incorporates four conformational states and different energy terms, taking into account hydrophobicity and solvation energetics, electrostatic interactions and hydrogen bonding. Because aggregation has always been associated with hydrophobicity, it is conceivable that hydrophobicity and solvation terms alone are enough for accurate prediction. To assess the relative importance of the different terms included in our algorithm, we ran reduced versions of TANGO on our data set, comparing the contribution of the different elements to the accuracy of the predictions. In **Table 1b** we compare the overall prediction accuracy using partial sets in which A includes only hydrophobicity and β-sheet propensity, B adds electrostatic contributions, C adds hydrogen-bonding contributions, and the complete TANGO algorithm additionally takes into consideration the competition of α-helical and β-turn conformations with the aggregation process.

From **Table 1** it is clear that all the contributions included in the aggregation model of TANGO are important for accurate prediction

Table 1b Importance of different terms to the accuracy of the TANGO prediction[a]

	Partial set A: hydrophobicity + β-sheet propensity	Partial set B: A + electrostatics	Partial set C: B + H-bonding	TANGO: C + α-helix/β-turn competition
Positives	82%	80%	82%	75%
Negatives	75%	79%	86%	97%
False positives	18%	20%	18%	25%
False negatives	25%	21%	14%	3%
Correlation	0.57	0.59	0.68	0.74

[a]For details on the terms, see text.

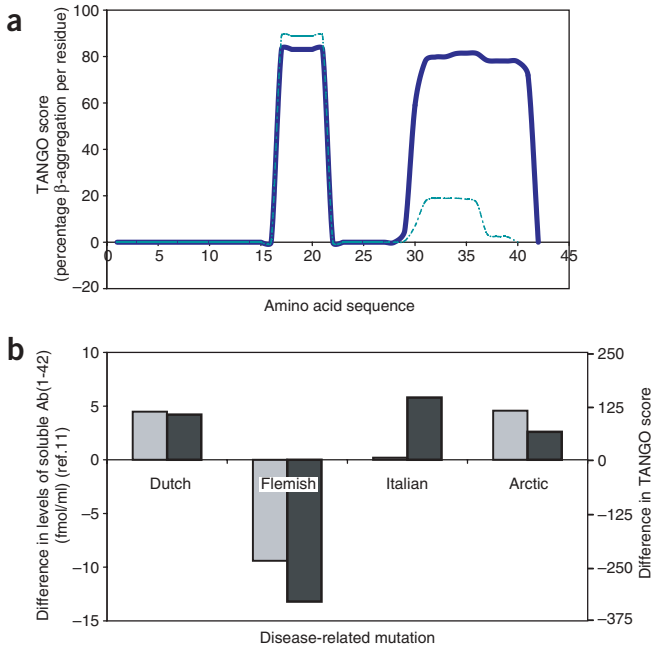

Figure 2 Prediction of aggregation in the Alzheimer β-peptide. (**a**) TANGO correctly predicts the dramatically increased aggregation propensity of the Aβ(1–42) peptide in comparison with the Aβ(1–40) version[9]. (**b**) Tango prediction versus measured *in vivo* aggregation as described[11] of the disease-related mutants of the Aβ peptide: Dutch, Flemish, Italian and Arctic.

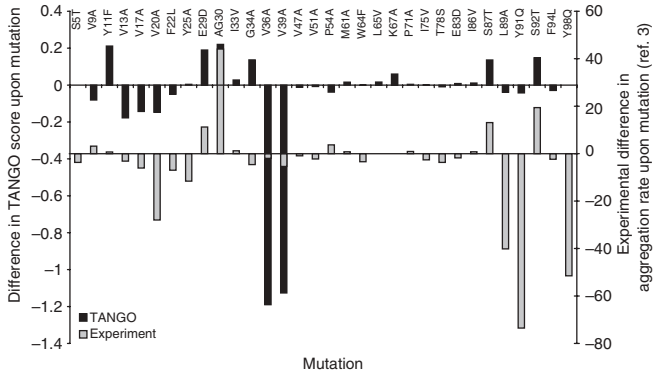

Figure 3 TANGO performance on a full-length protein. Comparison of TANGO to the aggregation prediction for full-length acyl phosphatase (AcP) with the effect of mutation on the aggregation kinetics as observed experimentally[3].

of the aggregating sequences, leaving only a small number of false negative predictions (3% of predicted negatives are false negatives). In contrast, when only hydrophobicity and β-sheet propensity are used, 25% of the predicted negatives are false negatives. However, a small drawback is that the number of false positives increases slightly (25% for TANGO in comparison with 18% when considering hydrophobicity and β-sheet propensity alone).

Using TANGO we analyzed the aggregation propensities of wild-type Aβ(1–40) and Aβ(1–42) and several familial disease-related mutants of the Aβ-peptide. The Aβ(1–42) peptide is known to have a higher aggregation propensity and to be more toxic than Aβ(1–40) (ref. 9). For both peptides, TANGO predicted two regions that participated in the aggregation process (**Fig. 2a**). The first region, situated in the middle of the peptide and encompassing residues 17–21, was also the region associated with familial disease-related mutants, whereas the second C-terminal region consisted of residues 31–36 in the Aβ(1–40) and of residues 30–42 in Aβ(1–42). These regions corresponded to regions mapped by mutagenesis experiments[10]. Furthermore, TANGO predicted a strikingly higher aggregation propensity of Aβ(1–42) over Aβ(1–40): the additional isoleucine and alanine induced a much higher aggregation propensity by recruiting a bigger part of the C terminus into aggregation.

Finally, we predicted the change in aggregation propensity between wild type and four disease-related mutants, namely the Arctic (E22G), Dutch (E22Q), Italian (E22K) and Flemish (A21G) variants, and compared these with experimental measurements of the aggregation of the Aβ-peptide *in vivo*[11] (**Fig. 2b**). All variants had a higher aggregation propensity than the wild type, except the Flemish variant, which had a lower aggregation propensity, suggesting a more intricate mechanism for pathogenesis. TANGO correctly picked up the effect of the several mutations including the Flemish variant. The Italian variant, however, was predicted to have no effect on the aggregation of the peptide, except at alkaline pH above the pKa of lysine 693.

As discussed above, TANGO accurately predicts the aggregation propensity in short peptides (95% of the peptides in the data set extracted from the literature are shorter than 25 residues). However, sequences longer than 30–40 residues can potentially introduce errors in the partition function because (i) long sequences are more likely to have more than two ordered sequences, (ii) coupling of the different ordered regions arising from tertiary interactions is also more common and (iii) intermediates in the aggregation process are likely to be present in higher proportions.

Despite the simplicity of the TANGO model for aggregation, we tested it on full-length proteins. TANGO predictions were compared with experimental measurements on the rate of aggregation of 33 mutants of acyl phosphatase[3] (AcP). **Figure 3** plots the ratio of mutant and wild-type aggregation rates at different positions along the sequence and compares them with the ratio of the TANGO scoring indexes. We found that TANGO predicts that the same regions are involved in aggregation as those observed experimentally. Moreover, TANGO can also predict whether the mutations increase or decrease aggregation. However, we would expect a better correlation between the amplitudes of the experimental and predicted changes in aggregation. We can think of several causes for this weak correlation. First, although AcP is in principle unfolded under the conditions of the experiment, there might still be residual structure and tertiary interactions complicating the picture. Second, although AcP is observed to aggregate via single exponential kinetics, the mechanism for aggregation of AcP might be more elaborate. CheY and suc1, which are both small globular proteins for which a simple aggregation behavior might be expected on the basis of the AcP data, aggregate by complex kinetics under similar conditions (data not shown). Finally, we compared experimental rates of aggregation with predictions for the amplitude of aggregation. Despite all these approximations, TANGO was able to pick up the aggregation pattern of unfolded proteins.

Predicting the aggregation tendencies of denatured proteins and peptides is relevant to the optimization of recombinant protein production. With regard to human disease, however, it is often more important to predict the aggregation effect of point mutations in folded proteins. To do this, we included the folded state in the TANGO partition function, thus allowing for competition between the folded state and the other structural states implemented in TANGO. This procedure allows detection of mutations that will increase the

Table 2 Predicted effect of pathogenic and protective mutations on protein stability and aggregation in mutants of transthyretin[12,16] and human lysozyme[13]

Protein	Mutant	Stability free-energy change upon mutation calculated using the Fold-X force field[22,a]	TANGO score for aggregation
Human lysozyme	Wild type		0
(involved in nonneuropathic	F57I	3.72	0.6
systemic amyloidosis)	I56T	3.10	0.3
	D67H	3.88	0.8
	W64R	3.40	0.4
Transthyretin	Wild type		0.70
(involved in amyloidotic	D18G	14.10	42.80
polyneuropathy)	P24S	8.10	24.50
	V20I	0.60	1.70
	A45T	0.09	0.90
	D18G	9.11	42.80
	G53R	15.20	40.90
	L12P	10.80	40.90
	L55P	10.70	40.80
	L111M	1.40	5.70
	V30G	12.80	36.60
	V30M	4.35	18.40
Transthyretin, protective			
mutations	R104H	−5.15	0.11
	T119M	−2.70	0.00
	V122I	−1.70	0.10

In contrast to peptides and unfolded proteins, both stability and aggregation effects need to be considered to detect pathogenic mutations in folded proteins. The T119M and V122I mutations in transthyrethin were reported to protect patients carrying common pathogenic mutations[14–17], such as V30G, from disease and are predicted to stabilize the protein as well as reduce aggregation slightly.
[a]ΔΔG values calculated using FoldX agree well with reported midpoints of temperature denaturations[12,17,19].

aggregation tendency of the denatured state, as discussed above, and also mutations that will increase the aggregation tendency of the protein by destabilizing the folded state. In human lysozyme and transthyretin, pathogenic mutations identified in patients[12,13] were shown to be related to increased aggregation and changes in the stability of the proteins[14–21]. Mutations in the human lysozyme gene cause autosomal dominant hereditary amyloidosis[13], whereas amyloid deposits of transthyretin are frequently found in patients suffering from familial amyloidotic polyneuropathy[12]. Apart from the pathogenic mutations that were identified for both proteins, mutations were also identified in transthyretin that protect patients from the devastating effect of certain disease-related mutations[14,15]. The free-energy change of unfolding upon mutation was calculated using the Fold-X force field[22] and correlated very well with the changes in midpoint of thermal denaturation reported in the literature[17,21]. **Table 2** shows the predicted effect of pathogenic and protective mutation on stability (Fold-X) and aggregation (TANGO). TANGO correctly predicted the behavior of the mutants in the two proteins when both stability and aggregation effects were considered, reflecting the two scenarios outlined above by which mutations can cause increased aggregation in folded proteins. TANGO, in combination with the Fold-X force field, is capable of predicting the effect of point mutations on the stability of proteins and can be used to identify mutations that cause increased aggregation and may hence lead to disease.

In conclusion, it is important to emphasize that TANGO predicts β-sheet aggregation and not amyloid formation. Although we found a good correlation between the aggregation tendency predicted by TANGO and the increase or decrease of amyloid formation in mutants of amyloidogenic peptides (Alzheimer peptides) and proteins (lysozyme and transthyretin), this does not mean that TANGO captures the specific sequence contribution for amyloid formation. Rather, it is probable that those peptides and proteins have amyloidogenic sequences and that β-sheet aggregation favors amyloidosis by promoting multimeric aggregated precursors.

METHODS

For more details on the methods and the data sets used for the calibration and testing of the algorithm, please refer to **Supplementary Notes 1** and **2** online.

The TANGO model for protein aggregation. The model used by the TANGO algorithm is designed to predict β-aggregation in peptides and proteins and consists of a phase-space encompassing the random coil and the native conformations as well as other major conformational states, namely β-turn, α-helix and β-aggregate. Every segment of a peptide can populate each of these states according to a Boltzmann distribution. Therefore, to predict β-aggregating segments of a peptide, TANGO simply calculates the partition function of the phase-space.

α-helical propensities. The parameters used in the latest version of AGADIR (AGADIR-1s[23]) have been used to determine the helical propensity of the amino acid sequences. The only modification has been the implementation of a two-window approximation (see **Supplementary Note 2** online).

β-turn propensities. β-turn propensity is calculated by considering three energy contributions (see **Supplementary Note 2** online): (i) an amino-acid specific cost in conformational entropy for fixing that residue in a β-turn compatible conformation, (ii) interactions of each amino acid with the turn structure in a position-dependent manner and (iii) a single H-bond between the main chains of residues i and i + 3 of the turn. We have considered only four types of turns for which we could obtain significant statistical data, types I, I′, II and II′.

β-sheet aggregation. To estimate the aggregation tendency of a particular amino acid sequence, we have made the following assumptions: (i) In an ordered β-sheet aggregate, the main secondary structure is the β-strand. (ii) The regions involved in the aggregation process are fully buried, thus paying full solvation costs and gains, full entropy and optimizing their H-bond potential (that is, the number of H-bonds made in the aggregate is related to the number of donor groups that are compensated by acceptors. An excess of donors or acceptors remains unsatisfied). (iii) Complementary charges in the selected window establish favorable electrostatic interactions, and overall net charge of the peptide inside but also outside the window disfavors aggregation.

The effect of physico-chemical conditions on aggregation. The effect of pH, temperature and ionic strength on electrostatic interactions was taken into account as described in AGADIR2-1s[23]. Similarly the dependence of entropy, H-bonds and hydrophobic interactions on temperature and ionic strength are taken into consideration as described in AGADIR2-1s[23].

TFE dependence. The effect of TFE on the stability of the different structural conformations considered here has been taken into consideration in the following way. First, we assume a general increase of the H-bond contribution to the energies of the helical, turn and aggregated conformations[24]. Second, we consider a change in the helical propensities of the amino acids based on published experimental results[25]. We assume that the effect of TFE is linear with concentration up to 40% where no further changes are considered. This is based on the empirical observation that for many peptides analyzed experimentally above 40% there are few, if any, changes (49).

Assumptions. We have opted for a two-window sampling approximation that assumes that the probability of finding more than two ordered segments in the

same polypeptide chain is too low to be considered (the simple one-window sampling will deviate too much from reality for peptides with >50 residues). Further, we assume there is no energetic coupling between the two windows. Finally, we do not consider aggregation intermediates. This means that we consider aggregates as a single molecular species or structural state in competition with the folded protein, the β-turn and α-helical conformations (see Methods). We do not consider β-hairpins as aggregating motifs, as no satisfactory method exists to predict β-hairpin stability.

Definition of and criteria for aggregation. The aggregation behavior of the peptides shown in **Supplementary Table 1** online was taken from a range of publications, mostly reporting CD and NMR studies on these proteins. Aggregation was mainly judged from the protein concentration dependence of CD or NMR signals in a range of 50 μM to 5 mM. When concentration dependence was reported in the range studied, we assigned the peptide in question to the aggregation group. Alternatively, for some peptides, aggregation was measured by fluorescence using thioflavine S or T that specifically bind to β-aggregates[26].

CD spectroscopy of peptides. All measurements were performed in a Jasco spectropolarimeter using peptides purchased from Jerini GmbH (peptides were synthesized using the Pepspot technology).

Statistical evaluation of the performance of the prediction. To evaluate the correlation between our predictions and the experimental results, we used the Mathews correlation coefficient[27], which takes into account the number of true positives (*tp*), true negatives (*tn*), false positives (*fp*) and false negatives (*fn*) of the prediction, and can be calculated as follows:

$$corr^{Mathews} = \frac{(tp * tn) - (fp * fn)}{\sqrt{(tn + fn)(tn + fp)(tp + fn)(tp + fp)}}$$

To estimate the probability that the correlation between our predictions and the experimental results would occur by change, we calculated a so-called *P* value as follows: Let *P*(S) be the probability of a successful (S) prediction of the aggregation propensity of a randomly chosen peptide. Because we only distinguish between two possible outcomes, aggregation A or no aggregation $\overline{A}$, we can express *P*(S) as

$$P(S) = P(S|A)\ P(A) + P(S|\overline{A})\ P(\overline{A}),$$

where *P*(S|A) and *P*(S|$\overline{A}$) are the probabilities of a successful prediction conditioned on whether the peptide actually aggregates or not. Let $P(A) = \theta$ be the fraction of aggregating peptides in the sample. If θ_p is the probability for predicting aggregation by guessing at random, the probability of success would be:

$$P(S) = \theta_p\ \theta + (1 - \theta_p)(1 - \theta)$$

The best random strategy is obtained by maximizing *P*(S) with respect to θ_p. When $\theta < 1/2$, *P*(S) is maximized by choosing $\theta_p = 0$, implying $P(S) = 1 - \theta$. Consequently, in a sample consisting of *N* peptides, the number of successful random predictions, *X*, will be a stochastic variable with a binomial distribution, $X \sim b(N\ ;\ 1 - \theta)$. The relevant *P* value, *p*, for our results is obtained by computing the probability of obtaining the same or larger number of successful predictions, as obtained by TANGO, when a random prediction scheme is used instead. If N_S denotes the actual number of successful predictions, we get

$$p = P(X \geq N_s) = \sum_{j=N_s}^{N} \frac{N!}{j!(N-j)!}(1-\theta)^j \theta^{N-j}$$

Website. TANGO can be accessed on the world wide web at http://tango.embl.de/. Access is free for academic use, but users are requested to register.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank Jesper Borg for helpful discussions. We are also grateful to Fabrizio Chiti for pointing us to data on amyloid forming sequences. J.S. and F.R. held a Prize Traveling Fellowship of the Wellcome Trust while at EMBL. A.M.F.E. was funded through a Marie Curie Fellowship from the European Union. This work was partly supported by an EU Training and Mobility of Researchers grant (EU Network on amyloid fibril formation, CT2-00241).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 20 February; accepted 21 July 2004
Published online at http://www.nature.com/naturebiotechnology/

1. Chiti, F. *et al.* Studies of the aggregation of mutant proteins *in vitro* provide insights into the genetics of amyloid diseases. *Proc. Natl. Acad. Sci. USA* **99**, 16419–16426 (2002).
2. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C.M. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**, 805–808 (2003).
3. Chiti, F. *et al.* Kinetic partitioning of protein folding and aggregation. *Nat. Struct. Biol.* **9**, 137–143 (2002).
4. Dobson, C.M. The structural basis of protein folding and its links with human disease. *Philos. Trans. R. Soc. Lond. B* **356**, 133–145 (2001).
5. Dobson, C.M. Protein-misfolding diseases: getting out of shape. *Nature* **418**, 729–730 (2002).
6. de la Paz, M.L. & Serrano, L. Sequence determinants of amyloid fibril formation. *Proc. Natl. Acad. Sci. USA* **101**, 87–92 (2004).
7. Ramirez-Alvarado, M., Blanco, F.J., Niemann, H. & Serrano, L. Role of beta-turn residues in beta-hairpin formation and stability in designed peptides. *J. Mol. Biol.* **273**, 898–912 (1997).
8. Lopez de la Paz, M., Lacroix, E., Ramirez-Alvarado, M. & Serrano, L. Computer-aided design of beta-sheet peptides. *J. Mol. Biol.* **312**, 229–246 (2001).
9. Vassar, R. *et al.* Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science* **286**, 735–741 (1999).
10. Caughey, B. & Lansbury, P.T. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26**, 267–298 (2003).
11. Nilsberth, C. *et al.* The 'Arctic' APP mutation (E693G) causes Alzheimer's disease by enhanced Abeta protofibril formation. *Nat. Neurosci.* **4**, 887–893 (2001).
12. McCutchen, S.L., Lai, Z., Miroy, G.J., Kelly, J.W. & Colon, W. Comparison of lethal and nonlethal transthyretin variants and their relationship to amyloid disease. *Biochemistry* **34**, 13527–13536 (1995).
13. Pepys, M.B. *et al.* Human lysozyme gene mutations cause hereditary systemic amyloidosis. *Nature* **362**, 553–557 (1993).
14. Almeida, M.R., Alves, I.L., Terazaki, H., Ando, Y. & Saraiva, M.J. Comparative studies of two transthyretin variants with protective effects on familial amyloidotic polyneuropathy: TTR R104H and TTR T119M. *Biochem. Biophys. Res. Commun.* **270**, 1024–1028 (2000).
15. Hammarstrom, P., Schneider, F. & Kelly, J.W. Trans-suppression of misfolding in an amyloid disease. *Science* **293**, 2459–2462 (2001).
16. Quintas, A., Saraiva, M.J. & Brito, R.M. The amyloidogenic potential of transthyretin variants correlates with their tendency to aggregate in solution. *FEBS Lett.* **418**, 297–300 (1997).
17. Shnyrov, V.L. *et al.* Comparative calorimetric study of non-amyloidogenic and amyloidogenic variants of the homotetrameric protein transthyretin. *Biophys. Chem.* **88**, 61–67 (2000).
18. Canet, D. *et al.* Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nat. Struct. Biol.* **9**, 308–315 (2002).
19. Morozova-Roche, L.A. *et al.* Amyloid fibril formation and seeding by wild-type human lysozyme and its disease-related mutational variants. *J. Struct. Biol.* **130**, 339–351 (2000).
20. Chamberlain, A.K., Receveur, V., Spencer, A., Redfield, C. & Dobson, C.M. Characterization of the structure and dynamics of amyloidogenic variants of human lysozyme by NMR spectroscopy. *Protein Sci.* **10**, 2525–2530 (2001).
21. Booth, D.R. *et al.* Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385**, 787–793 (1997).
22. Guerois, R., Nielsen, J.E. & Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369–387 (2002).
23. Lacroix, E., Viguera, A.R. & Serrano, L. Elucidating the folding problem of alpha-helices: Local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* **284**, 173–191 (1998).
24. Blanco, F.J. & Serrano, L. Folding of protein G B1 domain studied by the conformational characterization of fragments comprising its secondary structure elements. *Eur. J. Biochem.* **230**, 634–649 (1995).
25. Luo, P. & Baldwin, R.L. Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry* **36**, 8413–8421 (1997).
26. Naiki, H. & Nakakuki, K. First-order kinetic model of Alzheimer's beta-amyloid fibril extension *in vitro*. *Lab. Invest.* **74**, 374–383 (1996).
27. Mathews, B.W. Comparison of predicted and observed secondary structure of t4 phage lysozyme. *Biochim. Biophys. Acta* **405**, 442–451 (1975).