Protein Aggregation

DOI: 10.1002/anie.200802345

Identification of Aggregation-Prone Elements by Using Interaction-Energy Matrices**

Jennifer M. Bui, Andrea Cavalli, and Jörg Gsponer**

The ability of proteins and peptides to convert from their soluble forms into amyloid aggregates has been linked to a variety of severe human pathologies. Amyloid aggregates, which are found as intra- or extracellular deposits in diseased tissues, contain numerous fibrils that have a characteristic X-ray diffraction pattern of a cross-β structure, in which β strands run perpendicular to the fibril axis.[1] The finding that cross-β fibrils can also be formed by nonpathological proteins promoted the idea that ordered aggregation and fibril formation is an inherent property of polypeptide chains.[2]

Experimental as well as theoretical approaches have been used to identify the specific chemical properties that promote the formation of ordered aggregates and the segments within proteins that have those properties.[3,4] It was found that very often short segments in a polypeptide chain—so called aggregation-prone elements—drive the aggregation process. An experimental method developed recently by Tessier and Lindquist enables polypeptide chains to be screened systematically for such aggregation-prone elements.[5] Tessier and Lindquist arrayed a complete library of over 100 overlapping denatured 20-mer peptides derived from the prion domains of Sup35 from *S. cerevisiae* (ScNM) and *C. albicans* (CaNM) on glass slides. Aggregation-prone peptides (elements) were then identified by quantifying the fluorescence intensities of labeled full-length ScNM and CaNM bound to the arrays after an incubation time of several hours or days.

Herein, we introduce a new computational method based on first principles, peptide interaction matrix analyzer (PIMA), for the identification of β-strand aggregation-prone elements within proteins. In analogy to the peptide-array experiments of Tessier and Lindquist, interaction energies between all possible *n*-mer pairs of a protein are first calculated, and the most aggregation-prone *n*-mers are then identified on the basis of their interaction energies. This new method enables the detection of aggregation-prone elements with high accuracy and can be used to determine the likelihood of coaggregation within or between proteins.

The primary assumptions of our approach are that aggregation-prone peptides, which show high fluorescence intensities in the peptide-array experiment, 1) form in-register parallel or antiparallel β sheets when interacting with themselves or other peptides and 2) display favorable interaction free energies. The first assumption is based on the experimental finding that most peptides adopt an in-register parallel or antiparallel β-sheet conformation in regular aggregates,[1] and the second is the thermodynamic requisite for the high reproducibility of the results of the peptide-array experiments. Moreover, if a similar loss in conformational entropy upon β-sheet formation is assumed for short peptides of the same length, interaction energy differences should enable accurate comparison of the likelihoods of β-sheet aggregation of peptide *n*-mers.

In this method, polypeptide chains are first divided into overlapping *n*-mers (adjacent *n*-mers overlap by one residue) that cover their entire length, and then all possible *n*-mer pairs *ij* ($j \geq i$) are threaded onto canonical in-register parallel or antiparallel β sheets (Figure 1a). Next, the β-sheet-interaction energies are calculated as $\Delta E_{ij}^{\mathrm{eff}} = E_{ij} - E_{i,j}$, in which E_{ij} is the minimized energy of the *n*-mer pair *ij* and $E_{i,j}$ the minimized energy of *n*-mers *i* and *j* at infinite separation (Figure 1b). In both cases, the side chains are modeled using the Dunbrack rotamer library,[6] and energies are calculated in the CHARMM potential with PARAM19[7] and the implicit solvation model SASA.[8] In the next step, all *n*-mer pairs with a $\Delta E_{ij}^{\mathrm{eff}}$ value higher than a global interaction energy threshold $\Delta E_{ij}^{\mathrm{agg}}$ are filtered out, because only *n*-mers that form low-energy pairs can be considered as aggregation prone. Analogously, of all interaction pairs initially formed during the incubation period of the array experiment, only those with favorable low energies, irrespective of whether their orientation is parallel or antiparallel, are likely to resist the washing step and contribute eventually to the fluorescence signal. The filtering process leaves an interaction matrix that contains only *n*-mer pairs with $\Delta E_{ij}^{\mathrm{eff}} \leq \Delta E_{ij}^{\mathrm{agg}}$ (Figure 1c). In the final step, this filtered interaction matrix is used to determine the aggregation intensity of each *n*-mer according to Equation (1):

$$I_i^{\mathrm{agg}} = \frac{\sum_j \Delta E_{ij}^{\mathrm{eff}}}{\max \sum_j \Delta E_{ij}^{\mathrm{eff}}} \quad (1)$$

[*] Dr. J. M. Bui, Dr. A. Cavalli
Chemistry Department, University of Cambridge
Lensfield Road, Cambridge, CB1 1EW (UK)
Fax: (+44) 1223-336362
E-mail: amc82@cam.ac.uk
Dr. J. Gsponer
Structural Studies Division, MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 0QH (UK)
Fax: (+44) 1223-213556
E-mail: jgsponer@mrc-lmb.cam.ac.uk

[**] This research was supported by the NSF (Postdoctoral Fellowship in Biological Sciences, J.M.B.), the Hughes Hall Fellowship, the University of Cambridge (J.M.B.), and the Medical Research Council (J.G.). We are grateful to Dr. P. Tessier and Dr. S. Lindquist for providing fluorescence intensities, and to Dr. C. M. Dobson and Dr. S. Radford for critical reading of the manuscript.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200802345.

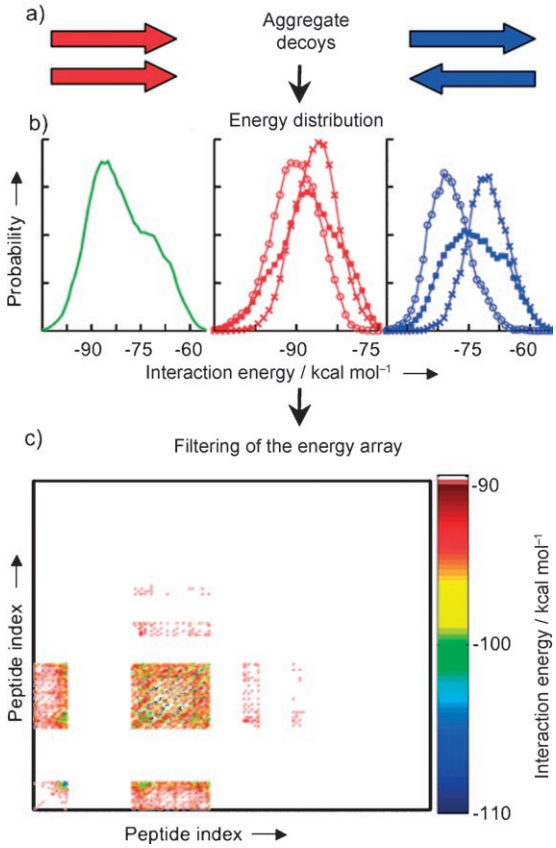

Figure 1. The framework of PIMA. a) Canonical in-register parallel (red) or antiparallel β sheets (blue). b) Interaction-energy distributions for parallel (red), antiparallel (blue), and pooled (green) 20-mer pairs. Energy distributions for ScNM, CaNM, and cross-species 20-mer pairs are indicated by spheres, squares, and crosses, respectively. c) Filtered interaction-energy matrix of CaNM 20-mers.

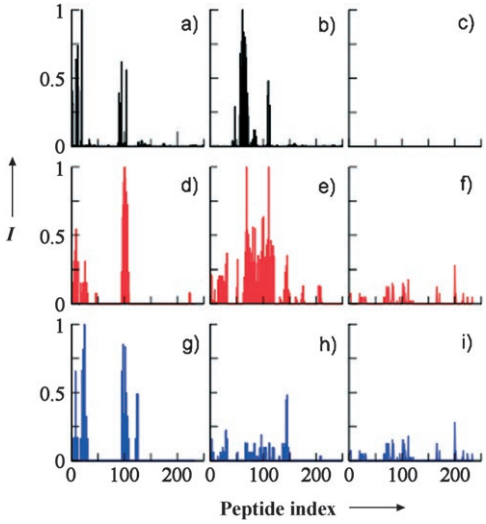

Figure 2. Comparison of the relative aggregation intensities of 10-mers for ScNM (d,g), CaNM (e,h), and the cross-species (f,i) with the corresponding experimental peptide-array data (a–c). The data for the parallel conformers are shown in red (d–f), the data for the antiparallel conformers in blue (g–i).

This protocol was first tested for 7-, 10-, and 20-mers of the prion domains ScNM and CaNM. The interaction energies $\Delta E_{ij}^{\mathrm{eff}}$ of all *n*-mer pairs from ScNM, CaNM, and the cross-species (one strand from ScNM and the other from CaNM) show a Gaussian distribution (shown for 20-mers in Figure 1b). Interestingly, a deconvolution of this distribution reveals that cross-species aggregates are significantly less favorable (with a *p* value of less than 10^{-16}) than intraspecies aggregates. Only *n*-mers that contribute to the end region of the left-hand tail of the energy distribution can be considered as aggregation prone. Therefore, the interaction energy threshold was defined as $\Delta E_{ij}^{\mathrm{agg}} = \langle \Delta E_{ij}^{\mathrm{eff}} \rangle - \sigma$, in which $\langle \Delta E_{ij}^{\mathrm{eff}} \rangle$ and σ are the mean and standard deviation of the pooled interaction energies (Figure 1b, green distribution). Such a global threshold should allow the concomitant identification of the most aggregation-prone parts of ScNM and CaNM. Indeed, the computed relative interaction intensities I_i^{agg} correlated well with the fluorescence intensities that were measured for ScNM and CaNM after incubation for 2.5 days (Figure 2; see Figure 1 in the Supporting Information).

Interestingly, the intensity profile calculated for the 10-mers shows the best agreement with the experimental data (Figure 2). It identifies the aggregation-prone regions encompassing residues 5–40 and 95–120 of ScNM and residues 70–90 and 110–125 of CaNM. The 20-mer profile also identifies the most aggregation-prone parts of both proteins, whereas the results for the 7-mer are less accurate (see Figure 1 in the Supporting Information), as the energy differences between different β-sheet 7-mer pairs are too small to be discriminative. The finding that the intensities observed for the 10-mers are in better agreement with the experimental data than those observed for the 20-mers, although the latter were used in the experiment, may indicate that not all 20 residues were involved in binding in the array experiment. On the other hand, the in vivo formation of a β sheet that consists of fully extended strands of 20 residues in length is rather unlikely. Overall, the PIMA protocol enables the identification of the aggregation-prone elements in ScNM and CaNM with good accuracy.

To test the general applicability of PIMA, aggregation-prone elements were identified for a selection of proteins that have been linked to amyloid diseases. The same global interaction threshold, $\Delta E_{ij}^{\mathrm{agg}}$, and a 10-mer screening protocol were used. The highest-intensity segments were close or identical to those shown experimentally to form fibrils (see Figure 2 in the Supporting Information). For example, PIMA captured with high accuracy the aggregation-prone segment 60–70 of human β2-microglobulin that is known to promote fibril formation.[9]

One of the exciting features of PIMA is its ability to demonstrate the relative interaction strength between different parts of the same or even different proteins. Figure 1c reveals that the high intensities observed for peptide segments located at residues 70–120 in CaNM are not only the result of energetically favorable pairings among these segments, but also of interactions with peptides located at the N terminus of CaNM. In addition, all cross-species 10-mer pairs formed with one strand from ScNM and the other from

CaNM had unfavorably high interaction energies (Figure 2), an observation that is consistent with the results of the array experiment.

In short, we have introduced a new method that uses first principles to detect not only β-strand aggregation-prone elements within proteins but also the likelihood of coaggregation between these elements. PIMA is based on the assumption that only peptides that are energetically complementary are likely to form aggregates. This assumed behavior is reminiscent of the self-complementarity of side-chain interactions that was revealed in recent amyloid structures.[10] Most excitingly, in contrast to previously introduced aggregation-prediction methods,[4] this method can be used to identify coaggregating segments in the same or different proteins. Therefore, it should be an important tool for the systematic screening of the proteome for proteins that are likely to be cosequestered in specific fibrillar aggregates.

Received: May 20, 2008
Published online: August 8, 2008

Keywords: aggregation · computational methods · interaction energy · peptides · proteins

[1] R. Nelson, D. Eisenberg, *Curr. Opin. Struct. Biol.* **2006**, *16*, 260.
[2] F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **2006**, *75*, 333.
[3] M. T. Pastor, A. Esteras-Chopo, M. Lopez de La Paz, *Curr. Opin. Struct. Biol.* **2005**, *15*, 57.
[4] A. Caflisch, *Curr. Opin. Chem. Biol.* **2006**, *10*, 437.
[5] P. M. Tessier, S. Lindquist, *Nature* **2007**, *447*, 556.
[6] R. L. Dunbrack, Jr., M. Karplus, *J. Mol. Biol.* **1993**, *230*, 543.
[7] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187.
[8] P. Ferrara, J. Apostolakis, A. Caflisch, *Proteins Struct. Funct. Genet.* **2002**, *46*, 24.
[9] G. W. Platt, K. E. Routledge, S. W. Homans, S. E. Radford, *J. Mol. Biol.* **2008**, *378*, 251.
[10] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. Wiltzius, H. T. McFarlane, A. O. Madsen, C. Riekel, D. Eisenberg, *Nature* **2007**, *447*, 453.

