Fibril Formation

DOI: 10.1002/anie.201005093

Two-Dimensional Ultraviolet (2DUV) Spectroscopic Tools for Identifying Fibrillation Propensity of Protein Residue Sequences**

*Jun Jiang and Shaul Mukamel**

More than twenty neurodegenerative diseases[1–3] are associated with the formation and deposition of amyloid fibrils of misfolded proteins.[4–8] Various amyloid-forming proteins across diverse systems are known to share common structures and fibril formation kinetics.[9,10] This suggests the existence of common molecular mechanisms for different amyloid diseases. A recent article argued that "*The most common mechanism by which proteins aggregate consists of the incorporation of relatively short sequence segments into β-sheet-like assemblies*".[11] It has been conjectured that the fibrillation propensities of proteins depend strongly on their sequences.[11–13] Many theoretical tools have been proposed to describe and characterize this sequence-dependence.[11–16] Their success in predicting the fibrillation propensity of various protein sequences should help understand why proteins form fibrils.

The study of amyloid fibrils would greatly benefit from adequate tools that correlate the fibrillation propensity of proteins with physical or chemical properties accessible by both theory and experiment. However, the parameters or factors currently available in the theoretical tools for evaluating the fibrillation propensity are not accessible experimentally. For instance, the TANGO,[14] Waltz,[11] and Zyggregator[15] computational tools provide different "aggregation scores", which are not connected to physical or chemical properties accessible by experiment. The 3D profile method[13] and the PASTA algorithm[16] are based on the energy function of protein sequences. As shown in a recent work of Eisenberg et al.,[12] the ability of proteins to form fibrils can be effectively predicted by their free energies. Again, it is extremely hard to measure free energies of separated protein segments, and the required discrimination of energies within several kcal mol^{-1} is not possible by current experiments. Moreover, the above criteria normally rely on detailed structural information, which is not available for most aggregates due to the lack of suitable probes with atomic resolution.[4,6,7]

Coherent ultrafast two-dimensional (2D) optical spectroscopy is a novel tool for studying the structure and dynamical properties of biomolecules.[17–21] These techniques have been successfully employed for studying fibril structure and aggregation kinetics and mechanisms.[18,22–26] Recent progress in laser technology opens up new windows of observation[27] by extending these techniques into the ultraviolet (UV).[28–31] 2DUV may use coherent ultrafast (20 fs) pulses for the characterization of biomolecules.[32] It provides an effective way for probing the geometric and electronic structures, and the inter- and intramolecular interactions of amyloid fibrils.[24]

It was reported recently[12] that the fibrillation propensity of proteins decreases as the Rosetta free energy[33] is increased. This is computed by the software RosettaDesign[34,35] as a potential energy function, which depends on the following ingredients: 1) the Lennard-Jones potential with van der Waals radii and well depth for atoms, 2) the solvation energy computed using the Lazaridis–Karplus implicit solvation model, 3) the side-chain–main-chain hydrogen bonding term, 4) the torsion potentials estimated from protein data base (PDB) statistics, 5) a unique reference value for each amino acid type, and 6) electrostatic interactions between residues. The Rosetta energy depends strongly on through-space interactions among residues. The relationships between such interactions and the fibrillation propensity could thus be used to predict the fibril formation. Here we use advanced theoretical tools developed for 2DUV spectroscopy of protein backbone transitions, to study various protein residue sequences with known fibrillation propensity. Simulations are carried out using a QM/MM (quantum mechanics/molecular mechanics) approach (details are described in the Supporting Information).[32,36] We identify direct spectroscopic signatures of the Rosetta free energies. Combined with the established connection between the fibrillation propensity and the Rosetta energy this study demonstrates that 2DUV signals may be used to identify protein residue sequences capable of forming amyloid-like fibrils.

In a joint experimental and theoretical investigation, Eisenberg et al.[12] had shown that the fibrillation propensity of some protein segments is relatively insensitive to the residue composition but is strongly sequence-dependent. Short protein segments with four to ten residues tend to form amyloid-like fibrils provided their Rosetta energies are below −23 kcal mol^{-1}. If the sequence of a fibrillizing segment is shuffled to obtain a higher energy, it loses its tendency to form fibrils.

The Rosetta free energy consists primarily of three terms: the intrinsic free energy of the protein backbones and amino acids, solvation energy, and interaction energy between

[*] Dr. J. Jiang, Prof. Dr. S. Mukamel
Chemistry Department, University of California
Irvine, CA (USA)
Fax: (+1) 949-824-8571
E-mail: smukamel@uci.edu

[**] This research was supported by the National Institutes of Health (grants GM059230 and GM091364) and the National Science Foundation (grant CHE-0745892). We thank Prof. Yi Luo for helpful discussions.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201005093.

© 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

residues. The former two only depend on the residue composition and are not sensitive to the sequence. Interactions among residues vary with their couplings and should depend on sequence shuffling. We argue that the variation of the Rosetta energy associated with the shuffle of protein sequences should be induced by the change of interactions among residues. These interactions in turn affect the excitonic energy splitting of protein electronic transitions. In an earlier simulation of protein backbone UV spectra,[36] we found that residue interactions split the distribution of protein transition energies into several peaks. The two strong transitions at 48000 cm^{-1} (210 nm) and 53000 cm^{-1} (190 nm) result from the splitting of the $\pi \rightarrow \pi^*$ backbone transitions. Thus, the transition frequency patterns of proteins reflect the strength of the residue–residue interactions (RRIs), which affect the Rosetta energy.

Eisenberg et al. have performed a systematic study of the fibrillation propensity of the bovine pancreatic ribonuclease A (RNase A) protein and other 13 short protein segments,[12] for which the theoretical prediction of the Rosetta energy/fibrillation propensity relation has been validated by experiment. We have therefore used these proteins, to explore 2DUV characteristics associated with the Rosetta energy and fibrillation propensity. We have labeled the eight RNase A segments used in Ref. [12] as R1-F, R2-F, R3-F, R4-F, R5-F, R6-F, R7-NF, and R8-NF. F (NF) denotes fibril forming (non-forming) sequences with Rosetta free energy lower (higher) than -23 kcal mol^{-1}. The distributions of the backbone transition frequencies and UV spectra were calculated by using 500 molecular dynamic (MD) snapshots. The distributions of the $\pi \rightarrow \pi^*$ backbone transition frequencies (which dominate the UV signals of proteins[36]) are displayed in the far-UV region (46000–56000 cm^{-1}) for six RNase A segments in Figure 1. The four left (F) panels of Figure 1 contain four or five peaks. The two right (NF) panels contain fewer (two or three) peaks. This trend survives the differences in residue compositions. We have also calculated the same distributions for three segments of the wild-type RNase A and their shuffled sequences (Figure S1 in the Supporting Information). The shuffled sequences of R5-NF, R6-NF, and R8-NF are found to contain fewer and narrower peaks than the corresponding F sequences.

Ordinary one-dimensional ultraviolet (1DUV) spectroscopy of proteins provides information about the distribution of electronic transition frequencies of singly excited states which depend on RRIs.[36] UV resonant Raman spectra show characteristic vibrational features associated with the fibrillation.[37,38] We have first simulated the 1DUV spectra (linear absorption and circular dichroism) of several proteins, and looked for the fibrillation propensity/1DUV relations. We found that the differences of spectral broadening induced by the change of RRI are not easily observable in these highly congested spectra (Figures S2 and S3).

2DUV photon echo signals depend on both singly and doubly excited electronic states, and thus contain richer spectral features associated with the geometric and electronic structures of proteins than 1DUV. 2DUV signals probe doubly excited states, and additional 2DUV peaks imply that excitations of the protein segments have more diverse transition energies. 2DUV spectra are very sensitive to variations in protein backbone transitions, which in turn depend on RRI.

The fibrillation propensity/transition frequency relation was examined for both non-chiral (*xxxx*) and chirality-induced (*xxxy*) pulse polarization configurations. We found the 2DUV *xxxx* spectra of RNase A segments to be insensitive to the fibrillation propensity (see Figures S4 and S5). The chiral 2DUV *xxxy* signals in contrast are much richer and show a high sensitivity to the protein geometry and fibrillation propensity. This is illustrated by the simulated spectra of RNase A segments R1-F, R2-F, R3-F, R4-F, R7-NF, and R8-NF shown in Figure 2 A. The F spectra show more elaborate peak patterns, than the NF. The spectra of segments R7-NF and R8-NF contain only four peaks, while the other F segments have at least six. Simulated 2DUV *xxxy* spectra of R5, R6, and R8 with the F and NF shuffled sequences are shown in Figure 2 B. We denote the energy difference between the shuffled sequences with lower and higher energy by ΔE. Again, the spectra of the F group contain more peaks than of the NF. In the NF segments, the signals are mostly distributed close to diagonal line. Many additional small crosspeaks are observed in the *xxxy* spectra of F segments. Since crosspeaks are induced by couplings between chromophores, this also suggests the strong RRI in the F group.

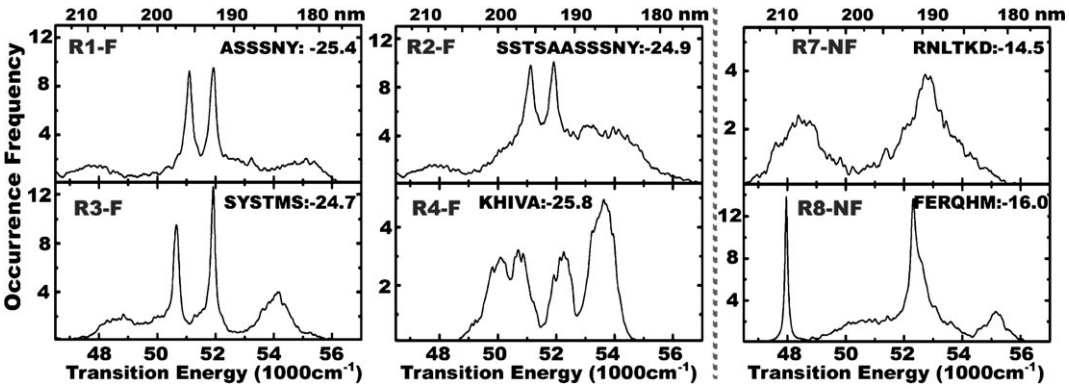

Figure 1. Histograms of the $\pi \rightarrow \pi^*$ transition energies of eight RNase A segments as indicated. The residue sequences and Rosetta energies (in kcal mol^{-1}) are given in each panel.

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

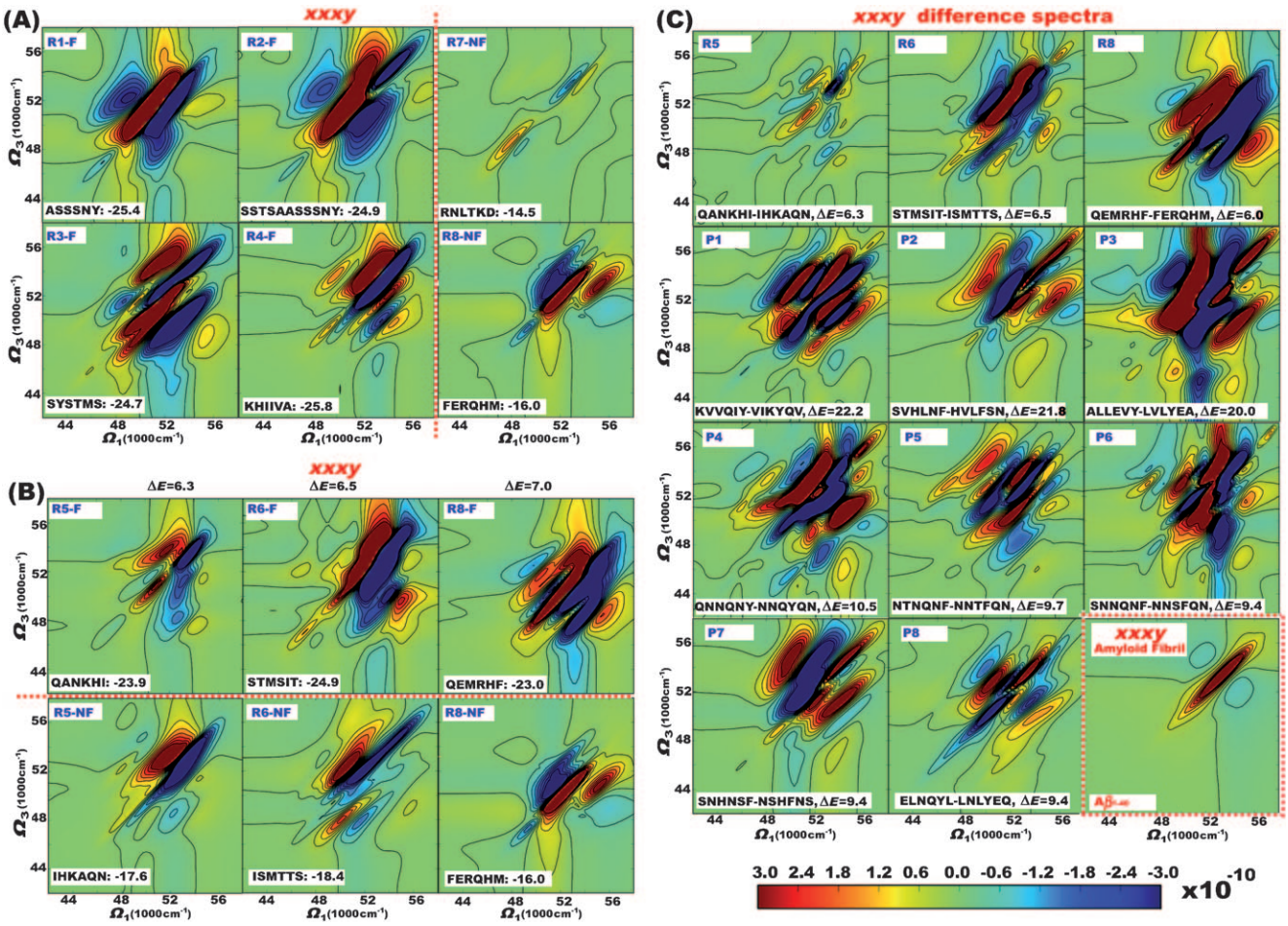

Figure 2. A) Simulated chiral 2DUV *xxxy* spectra for the same systems of Figure 1. B) Simulated chiral 2DUV *xxxy* spectra for the RNase A segments R5-F, R6-F, R8-F (top) and the corresponding shuffled sequences R5-NF, R6-NF, R8-NF (bottom). C) 2DUV *xxxy* difference spectra (lower energy minus higher energy) between the shuffled sequences with lower energies (and fibril forming) and higher energies for three RNaseA segments (R5, R6, R8) and eight other protein segments (P1 to P8). The residue sequences and energy differences (in kcal mol^{-1}) are given at the bottom of each spectra graph. Bottom right panel: 2DUV *xxxy* spectroscopy of a typical amyloid fibril Aβ_{1-40}. 2DUV magnitude scale bar is plotted at the right bottom.

The fibril formation propensity threshold of −23 kcal mol^{-1} for the Rosetta energy had also been tested and validated in thirteen wild-type protein segments. These segments were further shuffled to get sequences with the lowest and highest Rosetta energies.[12] We have simulated the 2DUV spectra of twelve of these protein segments (labeled P1 to P12) whose crystal structures are available in the protein data bank. The shuffled sequences with the lowest and highest Rosetta energies are labelled as P*xx*-L and P*xx*-H. Simulated spectra of P*xx*-L and P*xx*-H (in Figures S6, S7 and S8) help validate the correlation between the Rosetta energy and 2DUV spectral features. We found that an increase of Rosetta energy upon sequence shuffling is always accompanied by a decrease of the number of 2DUV *xxxy* peaks. These reflect the reduction of RRI which is correlated with the ability to form fibrils. These spectroscopic features are consistent with the transition diversity illustrated in Figures S6, S7 and S8, where in all cases the protein sequence with the lowest Rosetta energy shows much broader distributions with a larger number of peaks than the shuffled sequence with the highest Rosetta energy.

The differences of 2DUV spectra induced by sequence shuffling reflect variations of RRI. We have subtracted the 2DUV *xxxy* spectra of the sequences with lower Rosetta energies (and fibrils observed) from those with higher energies for three RNase A segments (R5, R6, R8) and eight other protein segments (P1 to P8). The resulting difference spectra displayed in Figure 2C show different patterns for various protein segments. However, we found a notable common feature in all cases: a positive (red) diagonal peak at around 54000–58000 cm^{-1} which is sometimes accompanied by two negative (blue) side bands. This feature is also common in the 2DUV *xxxy* spectra of a series of Aβ_{1-40} amyloid fibrils of various size.[24] One example of Aβ_{1-40} amyloid fibrils is shown at the bottom right panel of Figure 2C. In all cases shuffling the protein sequence to decrease the Rosetta energy and enhance the ability to form amyloid-like fibrils, results in additional positive diagonal signals at 54000–58000 cm^{-1} in the 2DUV *xxxy* spectra.

In summary, the present computational QM/MM investigation of the fibrillation propensity of protein segments reveals a Rosetta energy/fibrillation propensity relation in

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

proteins. We found that protein segments with high fibrillation propensity and lower Rosetta energy have more diversified transition energy distributions than those with low fibrillation propensity and higher Rosetta energy. Decrease of the Rosetta energy implies increasing RRI. Fluctuations of the QM Hamiltonians of the system are then amplified, increasing the splittings of the electronic transition energies, and broadening their distribution.

UV spectroscopy of proteins contains information about RRI and transition diversification, which is related to the fibrillation propensity. Due to high spectral congestion, 1DUV techniques such as absorption and CD do not provide clear signatures of the fibrillation propensity. 2DUV signals are much richer and better resolved. Simulated 2DUV spectra of six RNase A segments and twelve other protein segments and their shuffled sequences show that 2DUV *xxxy* signals are very sensitive to variations of RRI. We further demonstrated that chiral 2DUV (*xxxy*) spectra of protein segments with high fibrillation propensity contain additional peaks to those with low fibrillation propensity. 2DUV *xxxy* signals directly represent the electronic transition diversity of proteins, and provide a convenient tool for identifying the fibrillation propensity of protein sequences.

2DUV difference spectra can help understand the variations of intermolecular interactions in proteins. 2DUV *xxxy* spectroscopy differences of shuffled protein sequences share a common spectral feature: a positive diagonal peak at around 54000–58000 cm^{-1}, which also dominates the signals of typical amyloid $A\beta_{1-40}$ fibrils. In order to form fibrils, protein sequences with high Rosetta energy should be shuffled. An additional positive diagonal peak at around 54000–58000 cm^{-1} in their 2DUV *xxxy* spectra can serve as a indicator of the ability of protein sequences to form amyloid-like fibrils.

Received: August 13, 2010
Published online: November 12, 2010

Keywords: amyloid β-peptides · fibrillation propensity · protein folding · 2DUV spectroscopy

[1] R. Tycko, *Curr. Opin. Struct. Biol.* **2004**, *14*, 96.
[2] B. Caughey, P. T. Lansbury, *Annu. Rev. Neurosci.* **2003**, *26*, 267.
[3] N. E. Lester-Coll, E. J. Riviera, S. J. Soscia, K. Doiron, J. R. Wands, S. M. de La Monte, *J. Alzheimer's Dis.* **2006**, *9*, 13.
[4] G. M. Rishton, *Nat. Chem. Biol.* **2008**, *4*, 159.
[5] B. Y. Feng, B. H. Toyama, H. Wille, D. W. Colby, S. R. Collins, B. C. H. May, S. B. Prusiner, J. Weissman, B. K. Shoichet, *Nat. Chem. Biol.* **2008**, *4*, 197.
[6] F. Chiti, C. M. Dobson, *Nat. Chem. Biol.* **2009**, *5*, 15.
[7] N. L. Fawzi, E. H. Yap, Y. Okabe, K. L. Kohlstedt, S. P. Brown, T. Head-Gordon, *Acc. Chem. Res.* **2008**, *41*, 1037.
[8] G. M. Shankar, S. Li, T. H. Mehta, A. G.-M. and. N. E. Shepardson, I. Smith, F. M. Brett, M. A. Farrell, M. J. Rowan, C. A. Lemere, C. M. Regan, D. M. Walsh, B. L. Sabatini, D. J. Selkoe, *Nat. Med.* **2008**, *14*, 837.
[9] M. Sunde, C. C. F. Blake, *Q. Rev. Biophys.* **1998**, *31*, 1.
[10] C. M. Dobson, *Nature* **2003**, *426*, 884.
[11] S. Maurer-Stroh, M. Debulpaep, N. Kuemmerer, M. L. de La Paz, I. C. Martins, J. Reumers, K. L. Morris, A. Copland, L. Serpell, L. Serrano, J. W. H. Schymkowitz, F. Rousseau, *Nat. Methods* **2010**, *7*, 237.
[12] L. Goldschmidt, P. K. Teng, R. Riek, D. Eisenberg, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 3487.
[13] M. J. Thompson, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4074.
[14] A. M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, *Nat. Biotechnol.* **2004**, *22*, 1302.
[15] G. G. Tartaglia, A. P. Pawar, S. Campioni, C. M. Dobson, F. Chiti, M. Vendruscolo, *J. Mol. Biol.* **2008**, *380*, 425.
[16] A. Trovato, F. Chiti, A. Maritan, F. Seno, *PLoS Comput. Biol.* **2006**, *2*, e170.
[17] H. S. Chung, Z. Ganim, K. C. Jones, A. Tokmakoff, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 14237.
[18] Y. S. Kim, L. Liu, P. H. Axelsen, R. M. Hochstrasser, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 7720.
[19] W. Zhuang, T. Hayashi, S. Mukamel, *Angew. Chem.* **2009**, *121*, 3804; *Angew. Chem. Int. Ed.* **2009**, *48*, 3750.
[20] S. Mukamel, Y. Tanimura, P. Hamm, *Acc. Chem. Res.* **2009**, *42*, 1207.
[21] S. H. Shim, R. Gupta, Y. L. Ling, D. B. Strasfeld, D. P. Raleigh, M. T. Zanni, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6614.
[22] D. B. Strasfeld, Y. L. Ling, S. H. Shim, M. T. Zanni, *J. Am. Chem. Soc.* **2008**, *130*, 6698.
[23] W. Zhuang, D. Abramavicius, D. V. Voronine, S. Mukamel, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 14233.
[24] J. Jiang, D. Abramavicius, C. Falvo, B. M. Bulheller, J. D. Hirst, S. Mukamel, *J. Phys. Chem. B* **2010**, *114*, 12150.
[25] M. T. Zanni, N. H. Ge, Y. S. Kim, R. M. Hochstrasser, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11265.
[26] W. Zhuang, N. G. Sgourakis, Z. Y. Li, A. E. Garcia, S. Mukamel, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15687.
[27] P. Corkum, S. D. Silvestri, K. A. Nelson, E. Riedle, R. W. Schoenlein, *Ultrafast Phenomena XVI*, Springer, Heidelberg, **2009**.
[28] A. A. Oskouei, O. Bram, A. Cannizzo, F. van Mourik, A. Tortschanoff, M. Chergui, *J. Mol. Liq.* **2008**, *141*, 118.
[29] A. A. Oskouei, O. Bram, A. Cannizzo, F. van Mourik, A. Tortschanoff, M. Chergui, *Chem. Phys.* **2008**, *350*, 104.
[30] C. Tseng, S. Matsika, T. C. Weinacht, *Opt. Express* **2009**, *17*, 18788.
[31] M. Beutler, M. Ghotbi, F. Noack, D. Brida, C. Manzoni, G. Cerullo, *Opt. Lett.* **2009**, *34*, 710.
[32] D. Abramavicius, J. Jiang, B. M. Bulheller, J. D. Hirst, S. Mukamel, *J. Am. Chem. Soc.* **2010**, *132*, 7769.
[33] B. Kuhlman, D. Baker, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10383.
[34] B. K. Y. Liu, *Nucleic Acids Res.* **2006**, *34*, W235.
[35] K. W. Kaufmann, G. H. Lemmon, S. L. DeLuca, J. H. Sheehan, J. Meiler, *Biochemistry* **2010**, *49*, 2987.
[36] J. Jiang, D. Abramavicius, B. M. Bulheller, J. D. Hirst, S. Mukamel, *J. Phys. Chem. B* **2010**, *114*.
[37] A. Mikhonin, S. Bykov, N. Myshakina, S. A. Asher, *J. Phys. Chem. B* **2006**, *110*, 1928.
[38] I. K. Lednev, V. A. Shashilov, M. Xu, *Current Science* **2009**, *97*, 180.

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim