FEBS Letters 574 (2004) 156–160 FEBS 28765

Exon 6 of human Jagged-1 encodes an autonomously folding unit

Corrado Guarnaccia, Alessandro Pintar*, Sándor Pongor

International Centre for Genetic Engineering and Biotechnology (ICGEB), Protein Structure and Bioinformatics Group, AREA Science Park, Padriciano 99, I-34012 Trieste, Italy

Received 12 July 2004; revised 4 August 2004; accepted 5 August 2004

Available online 21 August 2004

Edited by Irmgard Sinning

Abstract Human Jagged-1 is predicted to contain 16 epidermal growth factor-like (EGF) repeats. The oxidative folding of EGF-2, despite the several conditions tested, systematically led to complex mixtures. A longer peptide spanning the C-terminal part of EGF-1 and the complete EGF-2 repeat, on the contrary, could be readily refolded. This peptide, which corresponds to the entire exon 6 of the Jagged-1 gene, thus represents an autonomously folding unit. We show that it is structured in solution, as suggested by circular dichroism and NMR spectroscopy, and displays an EGF-like disulfide bond topology, as determined by disulfide mapping.
© 2004 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Keywords: EGF repeat; Solid phase peptide synthesis; Oxidative folding; Disulfide topology; NOTCH signaling

1. Introduction

Jagged-1 (SwissProt: JAG1_HUMAN) is one of the ligands of Notch receptors. Binding of Jagged-1 to Notch receptors triggers a cascade of proteolytic cleavages [1] that eventually leads to the release of the intracellular part of the receptor from the membrane, its translocation to the nucleus, and the activation of transcription factors [2,3]. These signaling events play a key role in cell differentiation and morphogenesis [4],as proved by the association of malfunctions in the NOTCH signaling network with several genetically inherited diseases [5] and with some forms of cancer [6]. Several mutations in the Jagged-1 gene have been linked with Alagille syndrome [7], whereas a mutation located in exon 6 of Jagged-1 has been correlated with familial tetralogy of Fallot [8], a congenital heart malformation. Jagged-1 is a type I, membrane anchored, multi-domain protein. The extra-cellular part of Jagged-1 is predicted to be made of a DSL domain, followed by a series of 16 epidermal growth factor-like (EGF) repeats and a von Willebrand factor type C (VWC) domain [9]. While for most EGF repeats there is a univocal correspondence between domains and exons, this is not the case of EGF-2. In this report, we show that exon 6 of Jagged-1 encodes an autonomously folding unit encompassing the C-terminal part of EGF-1 and the entire EGF-2 repeat. We assigned the topology of disulfide bonds, and show that the refolded peptide, J1ex6, is structured, as determined by circular dichroism (CD) and nuclear magnetic resonance (NMR).

2. Materials and methods

2.1. Peptide synthesis

The 33 amino acid peptide corresponding to residues 263–295 of human Jagged-1 (Swiss-Prot: JAG1_HUMAN), egf-2, and the 44 amino acid peptide corresponding to exon 6 of Jagged-1 (residues 252–295), J1ex6 (Fig. 1), were synthesized on solid phase (Fmoc/*t*-Bu chemistry). The synthesis was automatically performed with a PS3 Protein Technology synthesizer on a 0.05 mmol scale. Cysteines were manually added as *N*-α-Fmoc-S-trityl-L-cysteine pentafluorophenyl ester in order to avoid cysteine racemization [10]. After cleavage from the resin, the peptides were precipitated with diethylether, washed and freeze-dried. The reduced peptides were purified by RP-HPLC on a Zorbax 300SB-C18 column (Agilent).

2.2. Oxidative folding

Fractions from RP-HPLC were immediately diluted 10 times in the degassed refolding buffer (0.25 M TRIS–HCl, 2 mM EDTA, and GSH/GSSH 1:1 (w/w), pH 8) and refolded for 18 h. The folding reaction was monitored by acid quenching of reaction mixture aliquots with TFA and LC–MS analysis on an API 150 EX single quadrupole mass spectrometer (Applied Biosystems). J1ex6 was purified by RP-HPLC on a Zorbax 300SB-C18 column and freeze-dried.

2.3. CD spectroscopy

Samples for CD spectroscopy were prepared by dissolving the lyophilized peptide in water and adjusting the pH to 6. Peptide concentration (35 μM) was determined by UV adsorbance at 280 nm using calculated ε values of 20 110 and 19 630 M^{-1} cm^{-1} for the oxidized and reduced forms, respectively. Reduction of disulfide bonds was achieved by adding TCEP (1 M) to give a final concentration of 4 mM. CD spectra were recorded on a Jasco J-810 spectropolarimeter in the range 190–250 nm and 250–350 nm using 0.1×1 cm quartz cuvettes, respectively. For each spectrum, five scans were acquired at 10 nm min^{-1}, and the mean residue ellipticity (MRE, deg cm^2 $dmol^{-1}$ $residue^{-1}$) was calculated from the baseline-corrected spectrum. A quantitative estimation of secondary structure content was carried out using SELCON3, CONTINLL, CDSSTR, and K2D, all run from the DichroWeb server [11].

2.4. NMR spectroscopy

Samples for NMR spectroscopy were prepared by dissolving the lyophilized material in H_2O/D_2O (90/10, v/v) for a final sample concentration of ~0.5 mM and adjusting the pH to ~4.5 with 0.1 N NaOH. Spectra were recorded on a Bruker Avance DRX 800 operating at a ^1H frequency of 800.13 MHz and equipped with a triple

* Corresponding author. Fax: +39-040-226555.
E-mail address: pintar@icgeb.org (A. Pintar).

Abbreviations: CD, circular dichroism; DSS, 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt; GSH/GSSH, reduced/oxidized glutathione; Fmoc, 9-fluorenylmethoxycarbonyl; LC–MS, liquid chromatography coupled with mass spectrometry; NMR, nuclear magnetic resonance; RP-HPLC, reverse phase high performance liquid chromatography; TCEP, Tris(2-carboxyethyl)phosphine; TFA, trifluoroacetic acid

0014-5793/$22.00 © 2004 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.febslet.2004.08.022

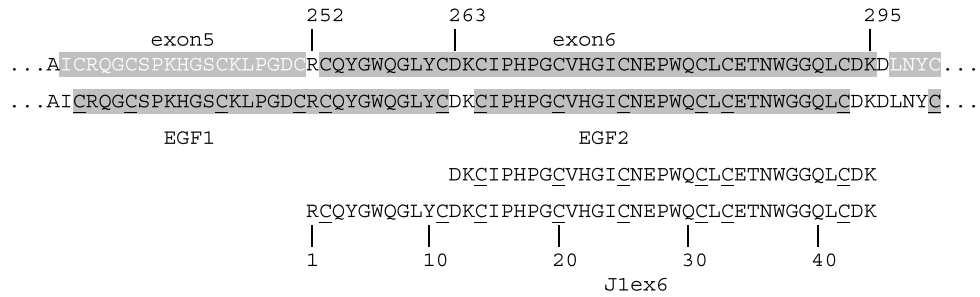

Fig. 1. Amino acid sequence (region of interest) of Jagged-1. Exons are alternatively shown in white and black over a gray background (upper line), EGF repeats are highlighted in gray (lower line); cysteines are underlined. The egf-2 and the J1ex6 peptides are also shown.

resonance, *z*-axis gradient cryo-probe. Solvent signal suppression was achieved using a WATERGATE pulse scheme [12]. 2D NOESY was recorded using a 150 ms mixing time. Data were transformed using X-WinNMR (Bruker). Chemical shifts were referenced to DSS.

2.5. Disulfide mapping

All proteolysis reactions were carried out at 37 °C for 18–48 h in sodium acetate buffer (50 mM, pH 5.2 or pH 6.0) containing 5 mM $CaCl_2$, and fragments fractionated and analyzed by LC–MS. In the first reaction, the purified J1ex6 peptide (60 μg) was dissolved in 90 μl of buffer and trypsin (3 μg) was added. After completion, the reaction mixture was further incubated for 48 h at 37 °C in the presence of thermolysin (6 μg). Fragments were isolated by RP-HPLC and subjected to a further proteolysis in the presence of proteinase K (1/30, w/w) for 48 h. Digestion of the intact J1ex6 with proteinase K was carried out in the same conditions.

3. Results and discussion

3.1. Synthesis and oxidative folding

The peptide corresponding to the second EGF repeat of Jagged-1 (egf-2, residues 263–295, Fig. 1) was prepared by solid phase chemical synthesis. The peptide was purified by RP-HPLC and refolded in the presence of a redox couple, the oxidative folding reaction being monitored by acid trapping followed by LC–MS. We tested different redox couples (GSH/GSSH, cysteine/cystine) at different molar ratios (1:1 to 10:1), different pH values (sodium borate, pH 9; ammonium acetate, pH 7–8; TRIS buffers in the pH range 7.0–8.5) in the presence or absence of calcium ions, but we always obtained a complex mixture of three-disulfide species (Fig. 2A). From the analysis of the Jagged-1 gene structure, we remarked that exon 6, which is encoding EGF-2, is actually encoding also the C-terminal region of the preceding putative EGF (Fig. 1). We thus extended the peptide synthesis to residue 252, comprising part of the EGF-1 repeat. The corresponding peptide (residues 252–295, J1ex6) was purified and refolded in the presence of GSH/GSSH as redox couple, pH 7.5, to readily yield a single product (Fig. 2B), which was purified to homogeneity by RP-HPLC. J1ex6, therefore, corresponds to the product of exon 6 of the Jagged-1 gene and behaves as an autonomously folding unit.

3.2. CD and NMR spectroscopy

To verify if the refolded, purified J1ex6 is structured in solution, CD and NMR spectra were recorded. The CD spectrum in the far-UV region (Fig. 3A) suggests the presence of both α and β elements. The positive band at 222 nm corresponds to a pronounced shoulder in the UV spectrum and can be tentatively assigned to tryptophans transitions [13], whereas the negative band at 234 nm is often observed in small disulfide-rich polypeptides, where it can be either positive or negative [14]. The spectrum changes drastically upon reduction of disulfide bonds by TCEP, and shows the typical features normally displayed by denatured proteins, like the negative band slightly below 200 nm and the broad negative band around 220 nm (Fig. 3A). Evident changes can also be observed in the near-UV CD spectrum (Fig. 3B). The broad shoulder around 280 nm, which displays the fine structure typical of tryptophans transitions, is reduced in intensity and the CD activity around 260 nm corresponding to disulfide bond transitions is close to zero. A quantitative estimation of secondary structure [11] did not yield satisfactory results. However, this is not surprising. Determination of secondary structure from CD spectra in small disulfide-rich

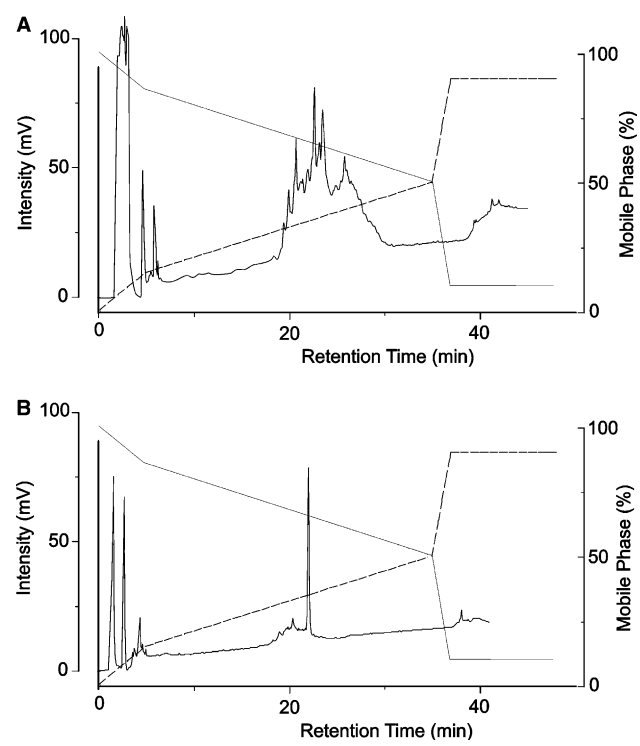

Fig. 2. Oxidative folding. RP-HPLC (UV detection at 214 nm) of the folding mixture for egf-2 (A) and J1ex6 (B) after 18 h refolding. Refolding conditions were identical: GSH/GSSH (1/1, mol/mol), TRIS buffer (0.25 M, pH 8.0), EDTA (2 mM); solvent A: 0.1% TFA in water (—); solvent B: 0.1% TFA in acetonitrile (- - -).

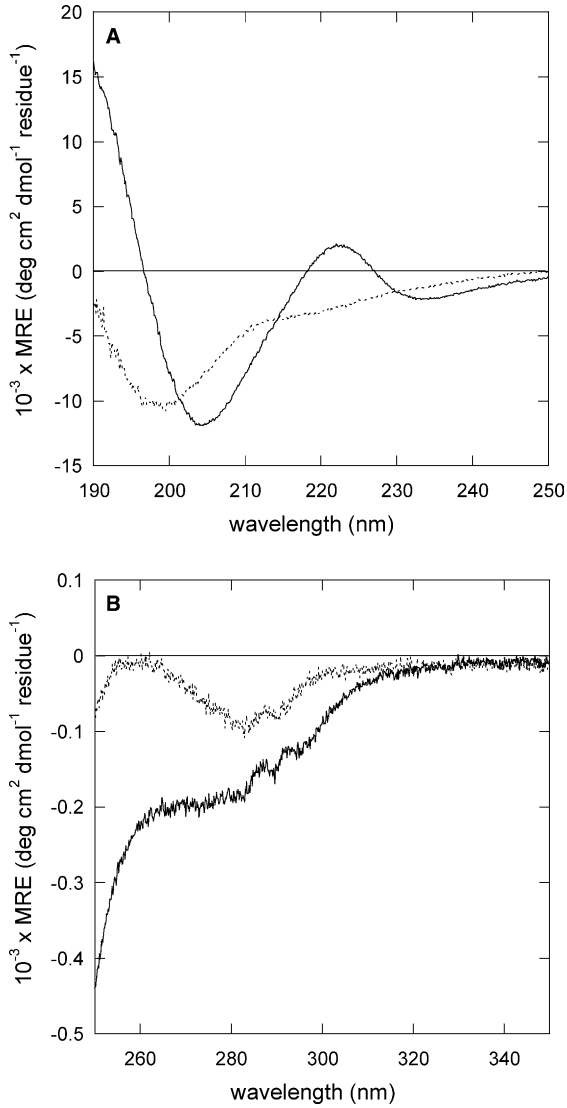

Fig. 3. CD spectroscopy. CD spectra in the far-UV (A) and near-UV (B) of oxidized (—) and reduced (···) J1ex6.

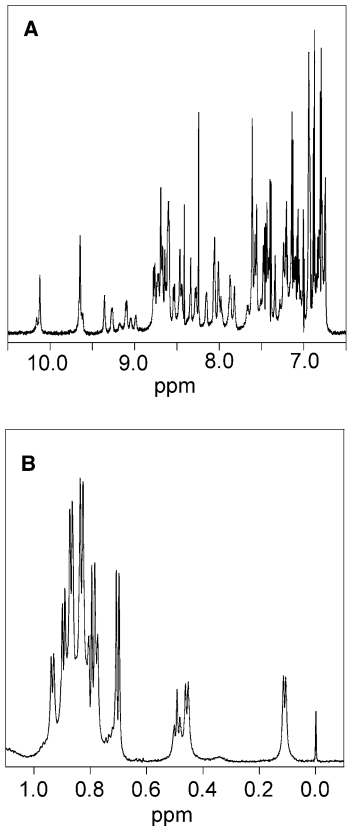

Fig. 4. NMR spectroscopy. ^{1}H NMR of J1ex6 in H_2O/D_2O (90/10 v/v). (A) Aromatics/backbone amide region; (B) methyl groups region.

proteins is often hampered by the contribution of aromatic residues to the CD spectrum in the far-UV region [13,15]. This is especially the case of J1ex6, which has an unusually high content in aromatic residues (three Trp and two Tyr). Overall, CD suggests that J1ex6 is structured, as further evidenced by the difference between the CD of the oxidized and reduced forms.

These results were confirmed by ^{1}H NMR. A significant dispersion in the backbone NH chemical shifts can be observed in 1D spectra recorded in H_2O (Fig. 4A), as well as an unusual upfield shift in the resonances of some methyl groups (Fig. 4B), which can be ascribed to ring current effects arising from aromatic residues. The presence of several NOEs between aromatic side chains and methyl groups is a further evidence of structural compactness (data not shown). Because calcium ions are known to bind a subset of EGF domains [16], we recorded 1D NMR spectra also in the presence of 20 mM Ca^{2+} but could not observe any difference in the 1D spectrum (data not shown). Although binding of calcium cannot be ruled out at this point, calcium ions do not seem to be required for the structural integrity of J1ex6, nor do they induce any conformational change in J1ex6.

3.3. Disulfide mapping

To establish if exon 6 encodes a novel cysteine-rich domain or rather an unusual EGF-like repeat, the topology of disulfide bonds was determined through a series of sequential proteolysis reactions tailored on the amino acid sequence of J1ex6. Starting from the hypothesis of an EGF-like disulfide topology, the first reaction was carried out using trypsin, which cleaved J1ex6 in two fragments, 1–13 and 14–44 (Table 1). The first disulfide was then assigned as C2–C11. From the thermolysin digestion, three different fragments corresponding to the central region of J1ex6 could be identified. From the results of this proteolysis, it was possible to rule out a bead-like arrangement of the two central disulfide bonds. At the same time, the last disulfide was tentatively assigned as C33–C42. This assignment was independently confirmed by the proteolysis of the intact J1ex6 with proteinase K. In order to distinguish between the two possibilities for the central disulfides (C14–C25, C20–C31 or C14–C31, C20–C25), the fragment 14–20/24–31 was purified by RP-HPLC and digested with proteinase K. The central region of J1ex6 turned out to be very resistant to proteolysis despite the broad specificity of proteinase K. The identified fragments 14/24–29 and 15–20/30–31, however, were sufficient to assign the central disulfides as C14–C25 and C20–C31 (Table 1). J1ex6 has therefore the disulfide

Table 1
Disulfide mapping

Enzyme	Fragment	Sequence number	Mass (calc.)	Mass (found)	Disulfides
Trypsin	R**C**QYGWQGLY**C**DK	1–13	1617.8	1617.7 1618.3 (+1) 809.5 (+2)	2–11
	CIPHPG**C**VHGI**C**NEPWQ**C**L**C**ET-NWGGQL**C**DK	14–44	3435.9	3435.91 1719.0 (+2) 1146.0 (+3)	–
Thermolysin[a]	**C**IPHPG**C**VHGI**C**NEPWQ**C**	14–31	1989.3	1988.8 995.3 (+2) 664.0 (+3)	14–25, 20–31 or 14–31, 20–25
	CIPHPG**C**/I**C**NEPWQ**C**[b]	14–20/24–31	1714.0	1714.4 1715.5 (+1) 858.3 (+2) 572.5 (+3)	
	C/IPHPG**C**/I**C**NEPWQ**C**	14/15–20/24–31	1732.0	1731.8 1732.3 (+1) 866.8 (+2) 578.3 (+3)	
Proteinase K	L**C**ETNWGGQL**C**DK	32–44	1464.6	1465.0 1466.0 (+1) 733.5 (+2)	33–42
	C/I**C**NEPW[c]	14/24–29	880.0	879.6 880.5 (+1) 440.8 (+2)	14–25
	IPHPG**C**/Q**C**[c]	15–20/30–31	870.0	869.8 870.5 (+1) 436.0 (+2)	20–31

Proteolytic enzyme, sequence (cysteines in bold) and residue number of the fragments identified by LC–MS, calculated mass (Da) of the fragment, experimental mass (Da) of the fragment obtained by deconvolution of the observed multiple charge spectra, and disulfide bond assignments; (a) the thermolysin digestion was carried out in one pot after completion of the trypsin digestion; the fragment (b) was purified by RP-HPLC and subjected to proteinase K digestion, to yield fragments (c). See Section 2 for experimental details. In fragments made of two or more peptide chains joined by disulfide bonds, different chains are separated by a slash (/).

topology expected for an EGF-like repeat, with an additional disulfide bond in its N-terminal extension.

In conclusion, while rather unexpectedly we were not able to refold EGF-2 despite the variety of conditions used, J1ex6, which corresponds to the entire exon 6, behaves as an autonomously folding unit and is structured in solution.

In Jagged–1, 17 exons encode 16 putative EGF repeats. Of these, 14 are predicted with a relatively high confidence (*E*-values between 1.9×10^{-2} and 2.4×10^{-12}) [9]. On the contrary, EGF-1 and −2 are predicted with an *E*-value of 16 and 24, respectively. A univocal correspondence between EGF repeats and exons occurs in 12 out of 16 cases, but not for EGF-2. EGF-2 is encoded by exon 6, which also encodes the C-terminal part of EGF-1 (Fig. 1). In our hands, this part turned out to be required for the folding of EGF-2 to occur and the correct disulfide bonds to be formed. In other words, the C-terminal part of EGF-1 seems to work as an intra-molecular chaperone in the folding of EGF-2. A similar effect was observed in the folding of bovine pancreatic trypsin inhibitor, where a short, 13 residue pro-region containing a single cysteine facilitates the oxidative folding of this small, three-disulfide protein [17]. A role for pro-domains and pro-sequences in assisting folding of a number of proteins has been described [18,19].

In 1978, Blake had already proposed that exonic regions of DNA correspond to "folded protein units" [20]. The debate on the introns-early and introns-late theory on the origin of genes [21] later drifted towards the correspondence between exons and units of protein structure. This correspondence was strongly criticized by Doolittle [22] and recent evidence suggests that at least for ancient proteins, only the position of phase zero introns is correlated with boundaries of domains, defined as compact structural units by geometrical means [23]. Introns of phase 1 and 2 might have been added later and are not related to structural features [23]. In the case reported here, (i) the predicted sequence domain boundaries do not seem to coincide with the boundaries of the autonomously folding unit and (ii) this apparent folding unit is limited by non-zero phase introns (phase 2 and phase 1). Although general conclusions cannot be drawn at this stage, it is tempting to speculate that while exons encode autonomously folding units, or folding domains, as originally suggested [20], in some instances these might not necessarily coincide with structural domains. Because protein folding in the living cell is a process that has to be both highly productive and reproducible, evolutionary pressure might have operated at the exon/intron organization level of eukaryotic genes not only to select for optimized function but also in order to arrange multi-domain proteins in optimized folding units. Indeed, selection based on function optimization can work only on "foldable" proteins [20]. Interestingly, the average length of a typical EGF repeat (35–40 amino acids) is very close to the most probable value of the exon length found in eukaryotic genes [24]. This might have also been the length of early folding units [25]. The annotation of several eukaryotic genomes, together with the steadily increasing number of protein structures available, is leading to the development of dedicated tools for the study of exon position within three-dimensional structures [26]. Although the relationship between exons and folding units has been addressed [27], more data on the folding, rather than on the structure, of individual

exon-encoded polypeptides should provide insights into this possible association.

Acknowledgements: We acknowledge the support of the NMR National Facility (CERM) in Sesto Fiorentino, Italy. We thank Doriano Lamba (CNR) for use of the circular dichroism spectropolarimeter and Massimo Lucci (CERM) for assistance with NMR data acquisition.

References

[1] Weinmaster, G. (2000) Curr. Opin. Genet. Dev. 10, 363–369.
[2] Allman, D., Punt, J.A., Izon, D.J., Aster, J.C. and Pear, W.S. (2002) Cell 109, S1–S11.
[3] Iso, T., Kedes, L. and Hamamori, Y. (2003) J. Cell. Physiol. 194, 237–255.
[4] Artavanis-Tsakonas, S., Rand, M.D. and Lake, R.J. (1999) Science 284, 770–776.
[5] Gridley, T. (2003) Hum. Mol. Genet. 12, R9–R13.
[6] Radtke, F. and Raj, K. (2003) Nat. Rev. Cancer 3, 756–767.
[7] Spinner, N.B., Colliton, R.P., Crosnier, C., Krantz, I.D., Hadchouel, M. and Meunier-Rotival, M. (2001) Hum. Mutat. 17, 18–33.
[8] Eldadah, Z.A., Hamosh, A., Biery, N.J., Montgomery, R.A., Duke, M., Elkins, R. and Dietz, H.C. (2001) Hum. Mol. Genet. 10, 163–169.
[9] Letunic, I., Copley, R.R., Schmidt, S., Ciccarelli, F.D., Doerks, T., Schultz, J., Ponting, C.P. and Bork, P. (2004) Nucleic Acids Res. 32, D142–D144.
[10] Han, Y., Albericio, F. and Barany, G. (1997) J. Org. Chem. 62, 4307–4312.
[11] Lobley, A., Whitmore, L. and Wallace, B.A. (2002) Bioinformatics 18, 211–212.
[12] Piotto, M., Saudek, V. and Sklenar, V. (1992) J. Biomol. NMR 2, 661–665.
[13] Woody, R.W. (1994) Eur. Biophys. J. 23, 253–262.
[14] Prestrelski, S.J., Arakawa, T., Wu, C.S., O'Neal, K.D., Westcott, K.R. and Narhi, L.O. (1992) J. Biol. Chem. 267, 319–322.
[15] Sreerama, N., Manning, M.C., Powers, M.E., Zhang, J.X., Goldenberg, D.P. and Woody, R.W. (1999) Biochemistry 38, 10814–10822.
[16] Saha, S., Boyd, J., Werner, J.M., Knott, V., Handford, P.A., Campbell, I.D. and Downing, A.K. (2001) Structure (Cambridge) 9, 451–456.
[17] Weissman, J.S. and Kim, P.S. (1992) Cell 71, 841–851.
[18] Bryan, P.N. (2002) Chem. Rev. 102, 4805–4816.
[19] Eder, J. and Fersht, A.R. (1995) Mol. Microbiol. 16, 609–614.
[20] Blake, C.C.F. (1978) Nature 273, 267.
[21] Roy, S.W. (2003) Genetica 118, 251–266.
[22] Stoltzfus, A., Spencer, D.F., Zuker, M., Logsdon Jr., J.M. and Doolittle, W.F. (1994) Science 265, 202–207.
[23] de Souza, S.J., Long, M., Klein, R.J., Roy, S., Lin, S. and Gilbert, W. (1998) Proc. Natl. Acad. Sci. USA 95, 5094–5099.
[24] Sakharkar, M., Passetti, F., de Souza, J.E., Long, M. and de Souza, S.J. (2002) Nucleic Acids Res. 30, 191–194.
[25] Blake, C.C.F. (1983) Nature 306, 535–537.
[26] Vivek, G., Tan, T.W. and Ranganathan, S. (2003) Bioinformatics 19, 159–160.
[27] Panchenko, A.R., Luthey-Schulten, Z. and Wolynes, P.G. (1996) Proc. Natl. Acad. Sci. USA 93, 2008–2013.