ACADEMIC PRESS

Biochemical and Biophysical Research Communications 293 (2002) 537–541

BBRC

www.academicpress.com

Silent mutations affect in vivo protein folding in *Escherichia coli*

Patricia Cortazzo,[a,1] Carlos Cerveñansky,[b] Mónica Marín,[a,*] Claude Reiss,[c] Ricardo Ehrlich,[a] and Atilio Deana[a,2]

[a] *Sección Bioquímica, Facultad de Ciencias, Iguá 4225, Montevideo 11400, Uruguay*

[b] *Instituto de Investigaciones Biológicas Clemente Estable – Facultad de Ciencias, Unidad de Bioquímica Analítica, Av. Italia 3318, Montevideo 11400, Uruguay*

[c] *Centre de Génétique Moléculaire, CNRS, F-91198 Gif-sur-Yvette cedex, France*

Received 25 March 2002

Abstract

As an approach to investigate the molecular mechanism of in vivo protein folding and the role of translation kinetics on specific folding pathways, we made codon substitutions in the EgFABP1 (*Echinococcus granulosus* fatty acid binding protein1) gene that replaced five minor codons with their synonymous major ones. The altered region corresponds to a turn between two short alpha helices. One of the silent mutations of EgFABP1 markedly decreased the solubility of the protein when expressed in *Escherichia coli*. Expression of this protein also caused strong activation of a reporter gene designed to detect misfolded proteins, suggesting that the turn region seems to have special translation kinetic requirements that ensure proper folding of the protein. Our results highlight the importance of codon usage in the in vivo protein folding.

Keywords: EgFABP1; Codon usage; Protein folding; Translation rate

The mechanism of in vivo protein folding remains one of the most intriguing problems to be elucidated in molecular biology. For many proteins the amino acid sequence contains the necessary information to determine its native tertiary structure. For those proteins, the native structure could be obtained in vitro relatively easily after denaturation. But the in vivo folding pathways are supposed to be affected by a number of factors not present in the simplified in vitro assays, such as physico-chemical conditions of the cellular environment, transient interactions with other proteins such as chaperones and translation rate [1].

Chaperones are thought to prevent newly synthesised proteins from misfolding and aggregating, impeding undesired hydrophobic interactions, and allowing alternative folding pathways [2]. Furthermore, since in vivo and in vitro folding is a co-translational process [3–7], the ribosome environment and the translation rate could also affect the nascent peptide folding process or the interaction with chaperones [8,9].

Concerning this last point, an interesting correlation was observed between synonymous codon usage and protein secondary structures. A tendency for the presence of rare codons in turns, loops, and domain linkers was observed [10,11]. It was thus suggested that translation kinetics implies a step-by-step synthesis process that determines partial peptide folding events and that could be necessary to ensure the correct folding of defined portions of the protein [12]. Slowly translated regions could represent separated folding events. Finally, it was observed that alterations in translation kinetics could influence the in vivo protein folding pathway [8,13,14].

To explore the role of translation rate on protein folding, we studied the effect of synonymous codon substitutions on the in vivo aggregation of EgFABP1 (*Echinococcus granulosus* Fatty Acid Binding Protein1) protein expressed in *E. coli*. EgFABP1 belongs to a family of low molecular weight proteins (approximately

* Corresponding author. Fax: +598-2-525-8617.

E-mail addresses: cortazzo@fcien.edu.uy (P. Cortazzo), lpp@iibce.edu.uy (C. Cerveñansky), marin@fcien.edu.uy (M. Marín), claudereiss@wanadoo.fr (C. Reiss), ehrlich@fcien.edu.uy (R. Ehrlich), deana@saturn.med.nyu.edu (A. Deana).

[1] Present address: Laboratorio de Inmunobiología, Departamento de Bioquímica, Facultad de Medicina, Uruguay.

[2] Present address: Skirball Institute of Biomolecular Medicine, NYU School of Medicine, New York 10016, USA. Fax: +01-212-263-8951.

0006-291X/02/$ - see front matter
PII: S0006-291X(02)00226-7

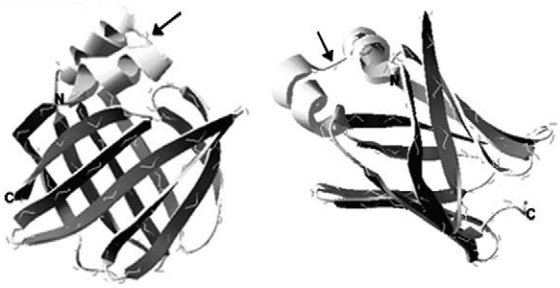

Fig. 1. EgFABP1 tridimensional model. Both structures correspond to a 90° rotation view. The arrows show the turn region between the two short alpha helices (amino acids 22–26), where rare codons were substituted by major ones.

Table 1
Amino acid sequence and codon subtitutions in EgFABP1 variants

Variant	Sequence						
	F_{27}	E_{21}	R_{22}	L_{23}	G_{24}	V_{25}	D_{26}
Wild type	GAA	CGC	CTT	GGG	GTG	GAT	TTC
1	GAA	CG**T**	CTT	GGG	GTG	GAT	TTC
2	GAA	CG**T**	CTT	GG**T**	GTG	GAT	TTC
3	GAA	CG**T**	CTT	GGG	GT**T**	GAT	TTC
4	GAA	CG**T**	CTT	GGG	GTG	GA**C**	TTC
5	GAA	CG**T**	CT**G**	GGG	GTG	GAT	TTC
6	GAA	CGC	CT**G**	GG**T**	GT**T**	GAT	TTC
7	GAA	CG**T**	CT**G**	GG**T**	GT**T**	GA**C**	TTC

The nucleotides changed in synonymous codon substitutions in EgFABP1 variants 1–7 are shown in bold.

14 kDa) involved in the transport of hydrophobic ligands, mainly long chain fatty acids [15,16]. The EgFABP1 structure consists of a barrel comprised of 10 anti-parallel β-strands organised in two orthogonal sheets and a short helix-turn-helix motif in the N-terminal end (Fig. 1) [17].

To examine the effect of translation rate on EgFABP1 folding, a short region of five codons located between the two alpha helices was gradually modified and the resulting changes in protein conformation were determined by monitoring the changes in protein aggregation. An in vivo protein-folding reporter was constructed to study misfolding or aggregation states and employed to analyse the expression of EgFABP1 variants. Our results show that synonymous codons usage can affect the conformational state of the proteins and suggest that codon specific translation rate may influence the in vivo protein folding.

Materials and methods

Bacterial strains and plasmids. Escherichia coli K12 strain XL1 Blue (Stratagene) was used as a host for expression vectors carrying EgFABP1 variants. *E. coli* K12 strain RYC1000 {F^-, *ara*D139, Δ*lac*169, *rbs*7, *rps*L, *rel*A, *thi*A, *gyr*A, *rec*A56} was used for β-galactosidase assays. pPC0 is a high copy plasmid derived from pTTQ18 [18] and contains an EgFABP1′ gene that differs from the original EgFABP1 reported sequence (GenBank accession number X65947) because of the presence of N, S, and M residues at positions 2, 3, and 18, respectively. Plasmids pPC1–pPC7 contain different silent variants of the EgFABP1′ gene, located between codons 22 and 26 (see Table 1). The low copy plasmid placZPGroESL was derived from placZ290 [19] by cloning of the *Eco*RI–*Cla*I fragment containing the *E. coli groESL* promoter from pRSΔTGroESL (a pRS414 [20] derivative that has the *groESL* promoter and the first two codons of *groEL* in frame with the *lacZ* coding sequence from codon 9).

Oligonucleotide site-directed mutagenesis. Oligonucleotide site-directed mutagenesis was performed in a two-step PCR as described previously [21] with minor modifications. In the first step the T7 primer and the one containing modifications (5′ ATC ATG GAA CG T/C CT G/T GG G/T GT G/T GA C/T TTC GTC ACT CG 3′) were included in the PCR mixture. The initial amplification was performed with the following parameters: 93 °C for 1 min, 45 °C for 1 min, and 72 °C for 1 min for 35 cycles. The amplified product from this reaction served as the mega-primer along with the T3 primer for the secondary amplification with the following parameters: 93 °C for 1 min, 45 °C for 10 min, and 72 °C for 1 min for 35 cycles. The amplified product of the secondary PCR was digested with *Eco*RI and *Hin*dIII restriction enzymes and cloned into the corresponding sites in plasmid pTTQ18. This method is expected to produce 31 different codon variants of EgFABP1. The fidelity of cloning was confirmed by DNA sequencing.

Culture conditions and total protein analysis. E. coli cells containing the plasmids pPC1, pPC2, pPC3, pPC4, pPC5, pPC6, and pPC7 (carrying codon variants of EgFABP1) were grown in Luria Broth (LB) supplemented with 100 μg/ml ampicillin. EgFABP1 expression was induced for 3 h by addition of 2 mM IPTG to cell cultures in exponential phase ($A_{600} \approx 0.6$). For protein analysis, 1 ml culture samples were centrifuged and resuspended in SDS–PAGE sample lysis buffer. Total protein was quantified by the Bradford method [22]. Ten micrograms protein was resolved in a 15% SDS–PAGE. The protein level was estimated using Coomassie blue stained gels using the image analysis software ScionImage (www.scioncorp.com).

Analysis of EgFABP1 in soluble–insoluble fractions. E. coli cells from 1 ml IPTG induced cultures were centrifugated and the pellets were resuspended in 250 μl PBS. The bacterial cells were then lysed by sonication (3 × 30 s). Soluble and insoluble fractions were separated by centrifugation at 12,000*g* for 10 min. Proteins in the supernatant were quantified by the Bradford method [22]. To work with equivalent volumes, the insoluble fraction was resuspended in 250 μl PBS and solubilized in SDS–PAGE sample buffer. Ten micrograms protein of the soluble fractions and an equal volume of the insoluble fractions for each EgFABP1′ variant were resolved by SDS–PAGE. Gels were stained with Coomassie blue or transferred to a Hybond C (Amersham) membrane for Western blot analysis [23]. The anti-EgFABP1 primary antibody (provided by Dr. A. Esteves, Facultad de Ciencias, Uruguay) was detected by using an alkaline phosphatase-conjugated anti-rabbit antibody (Sigma), NBT/BCIP system. The relative concentrations of soluble/insoluble EgFABP1 were estimated in Coomassie blue stained gels and by Western blot as described above.

β-Galactosidase activity. E. coli K12 RYC1000 was co-transformed with plasmids containing EgFABP1′ variants (pPC1–pPC7) and the placZGroESL constructed for this purpose (described above). Cells were grown in minimal medium A supplemented with casamino acids (1%) glucose (20 mM), and thiamine (0.0005%) and containing ampicillin (100 μg/ml) and tetracycline (10 μg/ml) at 37 °C. Induction of EgFABP1 variants was carried out by adding 2 mM IPTG in cultures in exponential phase ($A_{600} \approx 0.6$). After 3 h of induction, β-galactosidase activity was measured as described previously [24].

Proteolytic digestion of EgFABP1 and peptide analysis by HPLC. Protein EgFABP1 variants 6 and 7 were PAGE purified from 350 μg cell extract by PAGE. In-gel proteolytic digestion (using trypsin or Endo-GluC) was performed by cutting out the band of interest from Coomassie stained gels. The resulting peptides were analysed by HPLC

using a VydacC18 column eluted with a standard gradient of acetonitrile in water with TFA (details in legend to Fig. 3). The protocol was mainly as described previously [25].

Results

Expression of the EgFABP1 variants

In an attempt to elucidate the effect of synonymous codon substitutions on the in vivo protein folding in *E. coli*, we introduced a series of silent mutations in *EgFABP1′* a variant of the *EgFABP1* coding sequence. By oligonucleotide site-directed mutagenesis, five rare codons located between codons 22 and 26 were replaced by synonymous frequent ones, obtaining seven codon variants (numbered from 1 to 7) of *EgFABP1′* (Table 1).

The expression levels of the seven variants of EgFABP1 (pPC1–pPC7) in *E. coli* were compared by electrophoresis of total protein extracts in 15% SDS–PAGE. Table 2 shows the levels of EgFABP1 with respect to the total protein content of extracts. Our data indicate that the substitution of codons in this region did not affect the EgFABP1 variant protein levels, which are all relatively high.

A silently mutated protein exhibits an altered solubility

The major question addressed in this study was whether the synonymous codon substitutions could affect the in vivo protein folding. As an approach to detect changes in protein conformation, we first studied changes in protein aggregation by analysing the soluble–insoluble fractions using SDS–PAGE and Western blotting. All of the variants, with the exception of variant 6, showed a similar distribution of EgFABP in the soluble–insoluble fractions. Variant 6 showed an accumulation of approximately 30% in the insoluble fraction and a concomitant reduction in the soluble fraction (Fig. 2 and Table 2).

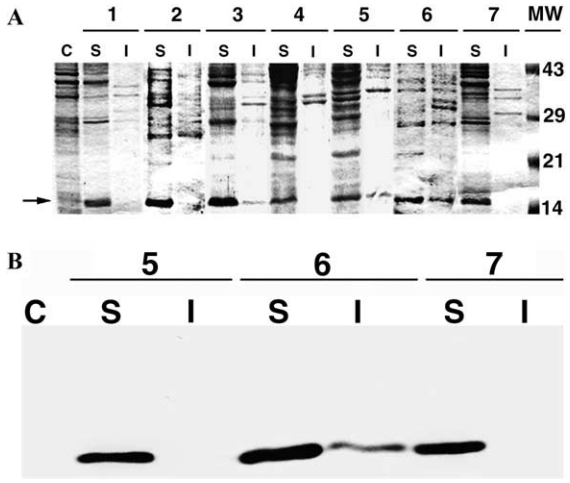

Fig. 2. Analysis of EgFABP1 solubility by SDS–PAGE and Western blot. The presence of EgFABP1 was analysed in soluble (S) and insoluble (I) protein fractions. Numbers 1–7 correspond to the different variants. The control (C) is total cell proteins before induction, all other lanes are after 3 h of induction with 2 mM IPTG. (A) Electrophoresis in 15% SDS–PAGE; (B) Western blot.

Table 2
Protein expression level, solubility, and induction of the stress promoter by EgFABP1 variants

Variant	Total EgFABP1 (%)[a]	Soluble EgFABP1 (%)[b]	β-Galactosidase activity (Miller units)[c]
1	30 ± 5	93 ± 3	31.2 ± 8.7
2	26 ± 4	94 ± 4	34.1 ± 7.9
3	25 ± 7	95 ± 6	32.5 ± 9.5
4	30 ± 5	83 ± 4	51.8 ± 18.8
5	26 ± 4	95 ± 5	35.6 ± 16.9
6	25 ± 3	52 ± 7	201.6 ± 26.0
7	33 ± 6	88 ± 4	32.8 ± 7.1

[a] EgFABP1 protein level in the total cell extract from Coomassie stained gels.

[b] Amount of EgFABP1 found in the soluble fraction calculated from Coomassie stained gels.

[c] The induction of the *groESL* promoter was determined using the in vivo protein folding reporter by measuring β-galactosidase activity (P*groESL–lacZ* fusion) following the expression of EgFABP1 variants.

The expression of a silently mutated protein induces the stress response

As misfolded or unfolded proteins are known to activate the heat shock response in the cell [26], their presence in vivo can be detected by the activity of the heat shock promoters. To assess whether differences in folding states among EgFABP1 synonymous variants could be correlated with the observed differences in protein solubility, we constructed the plasmid placZ-GroESL which places *lacZ* gene under the control of the *groESL* promoter.

The ability of the construct placZPGroESL to detect stress conditions was verified by measuring β-galactosidase activity in cells after different stress conditions such as heat shock and exposure to ethanol or methanol (data not shown).

To determine whether the expression of EgFABP1 synonymous variants was able to activate the stress response, *E. coli* strain K12 RYC1000 carrying placZ-GroESL was transformed with plasmids pPC1–pPC7. β-Galactosidase activity was measured for each of these strains. Interestingly, a clear increase (up to 6-fold) of β-galactosidase activity was observed for the strain transformed with pPC6 compared to the other strains (Table 2).

To preclude the possibility that the differences in protein solubility and induction of β-galactosidase activity were caused by unexpected changes in EgFABP1 amino acid sequence, the peptide profiles of EgFABP1

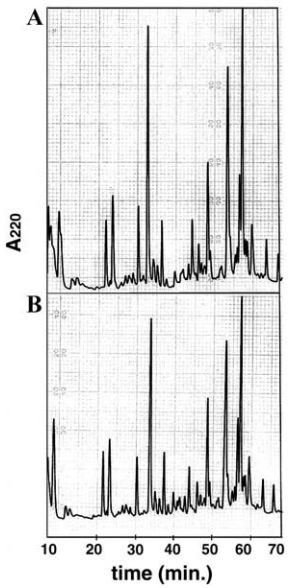

Fig. 3. HPLC of tryptic peptides from EgFABP1 variant 6 (panel A) and variant 7 (panel B). After in-gel proteolytic digestion, peptides were extracted with 60% acetonitrile/1% TFA in water and analysed by HPLC using a Vydac C18 column (15 cm × 2.3 mm). The elution gradient was: 0–2 min B: 2%, 2–62 min: B from 2–32%, 62–82 min: B from 32–60%, 82–100 min: B from 60–90% (solution A: 0.1% TFA in water and solution B: 0.07% TFA in acetonitrile), flow 0.15 mi/min, UV detection with 0.1 absorbance units full scale at 220 nm.

variants 6 and 7 obtained with both trypsin or Endo-GluC were compared and found to be identical (Fig. 3).

Discussion

An important number of works that has focused in the study of folding mechanisms either in in vitro translation systems as well as in vivo leads to the conclusion that the protein folding is a co-translational event [3–7]. More recently, it has been observed that some amino acid substitutions in AcP (human muscle acylphosphatase) tend to favour aggregation over correct folding demonstrating that both processes compete with each other [27]. However, there is scarce information about the influence of translation rate on protein folding pathways or on protein aggregation rates.

In *E. coli*, different aspects of the translation kinetics and gene expression have been studied. In this organism, tRNAs levels have been determined in different culture conditions and a correlation between the frequency of codon usage, the tRNAs contents, and gene expression levels has been established [28,29]. Furthermore, translation kinetics controls the distribution of ribosomes on the mRNA and may influence the stability of mRNA by obscuring or exposing putative RNase cleavage sites [30]. However, the effects of translation kinetics on protein structure are poorly understood.

Previous work showed the influence of synonymous codons on protein folding in various linker regions of multi-domain multi-functional proteins or on relatively large proteins with different folding domains [8,31]. In this work we explored the effect of changes in synonymous codons on a very small turn between two short alpha helices. This region does not correspond to any interface between different folding domains in the Eg-FABP1 protein.

We chose to modify the codon usage in a region of EgFABP1 that corresponds to the turn (residues 22–26) between two alpha helices. These five consecutive rare codons were gradually substituted for synonymous frequent ones, R22 being the last residue from the first alpha helix. This region was selected because rare codons were frequently observed in turns, loops, and links [31]. We found that one of the synonymous variants (variant 6) exhibited a clear change in solubility accumulating in the insoluble protein fraction. Moreover, when the same EgFABP1′ variant was expressed with an in vivo protein-folding probe we detected the presence of misfolded proteins that induce the stress response. It is worth mentioning that in addition to the plasmid DNA resequencing of the variant 6 after being performed the expression experiments, we also verified that the amino acid sequence was not modified due to any translation error that could occur during high level expression conditions. We also analysed the effect of modifying the codon usage of other regions further downstream (residues 38–42), but no aggregation effect was observed (data not shown).

We hypothesise that accelerating the translation rate of this region is expected to perturb the initial folding steps of the already synthesised part of the protein, a process that could take place inside the polypeptide exit tunnel of the ribosome [32]. A close view of the structure of the polypeptide tunnel shows that the length is about 100 Å long and relatively straight and the diameter is 15–20 Å wide [32]. A tunnel with these proportions can accommodate more than 60 residues in an alpha helix conformation. We assume that, in case of variant 6, alterations in translational kinetics of the turn could cause misfolding of the first part of the protein. This perturbation in folding could lead to a disruption of the correct folding of the rest of the protein, which is primarily composed of two large beta sheets, that presumably folds outside the ribosome. The exposure of hydrophobic regions in variant 6, due to the inability of the protein to fold correctly, might result in the observed aggregation.

The differential behaviour of variant 7, which includes in addition to the substitutions present in variant 6, changes in positions 22 and 26, is intriguing. Why variant 7 did not show an aggregation effect like variant 6 is still unanswered. According to the hypothesis proposed initially, one might have expected a gradual ag-

gregation effect from variant 1 to variant 7, but interestingly, only the variant 6 showed clear aggregation behaviour. According to the procedure followed in this work, conformational changes can be detected through changes in protein solubility or by the induction of the stress response. It should be mentioned that two differentially folded proteins could exhibit similar solubility properties. Furthermore, the threshold for induction of the stress response by misfolded polypeptides has not been determined. Therefore, we cannot exclude that conformational variations have occurred in the other silent variants. Moreover it cannot be excluded that the different variants show a different proteolytic sensitivity avoiding aggregation or impeding the triggering of the stress response.

Finally, we conclude that the in vivo aggregation observed was due to a synonymous codon substitution. This behaviour could be related to subtle changes in translation rates of this region and could have dramatic effects by modifying the alternative folding pathways.

Acknowledgments

This work was supported by the CONICYT (Consejo Nacional de Investigación Científica y Técnica, Uruguay, Project 1034), CSIC (Comisión Sectorial de Investigación Científica, Universidad de la República Oriental del Uruguay), and by ECOS France–Uruguay. The authors would like to thank Dr. Alexis Diwa for a critical reading of the manuscript.

References

[1] J.L. Johnson, E.A. Craig, Cell 90 (1997) 201–204.
[2] F.U. Hartl, Nature 381 (1996) 571–579.
[3] A.N. Fedorov, T.O. Baldwin, J. Biol. Chem. 272 (1997) 32715–32718.
[4] A.N. Fedorov, T.O. Baldwin, Proc. Natl. Acad. Sci. USA 92 (1995) 1227–1231.
[5] E.V. Makayed, V.A. Kolb, A.S. Spirin, FEBS Lett. 378 (1996) 166–170.
[6] A.A. Komar, A. Kommer, I.A. Krasheninnikov, A.S. Spirin, J. Biol. Chem. 272 (1997) 10646–10651.
[7] V.A. Kolb, E.V. Makeyev, A.S. Spirin, J. Biol. Chem. 275 (2000) 16597–16601.
[8] A.A. Komar, T. Lesnik, C. Reiss, FEBS Lett. 462 (1999) 387–391.
[9] A.A. Komar, T. Lesnik, C. Cullin, E. Guillemet, R. Ehrlich, C. Reiss, FEBS Lett. 415 (1997) 6–10.
[10] M. Oresic, D. Shalloway, J. Mol. Biol. 281 (1998) 31–48.
[11] T.A. Thanaraj, P. Argos, Protein Sci. 5 (1996) 1973–1983.
[12] J. Frydman, Annu. Rev. Biochem. 70 (2001) 603–647.
[13] I.G. Ivanov, A.A. Saraffova, M.G. Abouhaidar, Int. J. Biochem. Cell Biol. 29 (1997) 659–666.
[14] T. Crombie, J.P. Boyle, J.R. Coggins, A.J. Brown, Eur. J. Biochem. 226 (1994) 657–664.
[15] A. Esteves, L. Joseph, M. Paulino, R. Ehrlich, Int. J. Parasitol. 27 (1997) 1013–1023.
[16] A. Esteves, B. Dallagiovanna, R. Ehrlich, Mol. Biochem. Parasitol. 58 (1993) 215–222.
[17] M. Paulino, A. Esteves, M. Vega, G. Tabares, R. Ehrlich, O. Tapia, J. Comput. Aided Mol. Des. 12 (1998) 351–360.
[18] M.J. Stark, Gene 51 (1987) 255–267.
[19] J.W. Gober, L. Shapiro, Mol. Biol. Cell 3 (1992) 913–926.
[20] R.W. Simons, F. Houman, N. Kleckner, Gene 53 (1987) 85–96.
[21] O. Landt, H.P. Grunert, U. Hahn, Gene 96 (1990) 125–128.
[22] M.M. Bradford, Anal. Biochem. 72 (1976) 248–254.
[23] J. Sambrook, E.F. Fritisch, T. Maniatis, Molecular Cloning: A Laboratory Manual, second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
[24] J.H. Miller, A Short Course in Bacterial Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1992, pp. 71–74.
[25] Coligan, J.E., Dunn, B.M., Ploegh, H.L., Speicher, D.W., Wingfield, P.T., 1995. Current Protocols in Protein Science, vol. 2.
[26] B. Bukau, Mol. Microbiol. 9 (1993) 671–680.
[27] F. Chiti, N. Taddei, F. Baroni, C. Capanni, M. Stefani, G. Ramponi, C.M. Dobson, Nat. Struct. Biol. 9 (2002) 137–143.
[28] O.G. Berg, C.G. Kurland, J. Mol. Biol. 270 (1997) 544–550.
[29] H. Dong, L. Nilsson, C.G. Kurland, J. Mol. Biol. 260 (1996) 649–663.
[30] A. Deana, R. Ehrlich, C. Reiss, Nucleic Acids Res. 26 (1998) 4778–4782.
[31] I.J. Purvis, A.J. Bettany, T.C. Santiago, J.R. Coggins, K. Duncan, R. Eason, A.J. Brown, J. Mol. Biol. 193 (1987) 413–417.
[32] P. Nissen, J. Hansen, N. Ban, P.B. Moore, T.A. Steitz, Science 289 (2000) 920–930.