Analytical Biochemistry 385 (2009) 278–285

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Thermus thermophilus-derived protein tags that aid in preparation of insoluble viral proteins

Naoyuki Kondo [a,b], Akio Ebihara [c,1], Heng Ru [b], Seiki Kuramitsu [c,d], Aikichi Iwamoto [a,e], Zihe Rao [b,f], Zene Matsuda [a,b,*]

[a] *Research Center for Asian Infectious Diseases, Institute of Medical Science, University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan*
[b] *China–Japan Joint Laboratory of Structural Virology and Immunology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China*
[c] *RIKEN SPring-8 Center, Harima Institute, Hyogo 679-5148, Japan*
[d] *Department of Biological Sciences, Graduate School of Science, Osaka University, Osaka 560-0043, Japan*
[e] *Division of Infectious Diseases, Advanced Clinical Research Center, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan*
[f] *National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China*

ARTICLE INFO

Article history:
Received 1 September 2008
Available online 19 November 2008

Keywords:
Thermus thermophilus
HIV
Protein tag
Solubilization
Vpr
Membrane-spanning domain

ABSTRACT

The expression and solubilization of insoluble proteins have been facilitated by the introduction of protein tags. In our analyses of viral protein R (Vpr) of human immunodeficiency virus 1 (HIV-1), however, several conventional tag proteins enhanced its expression but failed to solubilize it. Therefore, we decided to explore whether proteins derived from *Thermus thermophilus* HB8 (*T. th.*), a highly heat-stable bacterium, could be used as tag proteins to enhance the solubilization of Vpr. Based on the data accumulated during the recent structural genomics project of *T. th.*, we selected 15 *T. th.* proteins with high expression levels and solubilities. From this group, we identified a *T. th.* tag protein that expressed Vpr in a soluble form. Furthermore, two *T. th.* tag proteins, including the identified one, were found to solubilize the extremely insoluble membrane-spanning domain of the envelope protein of HIV-1. When green fluorescent protein (GFP) was used as a passenger protein of *T. th.* tags, the brightness and stability of GFP were similar to those of untagged GFP, suggesting that the *T. th.* tags do not negatively affect the function of the passenger protein. Thus, data of structural genomics can be applied to generate a customized versatile protein tag for protein analyses.

© 2008 Elsevier Inc. All rights reserved.

The structures of various proteins have been determined during recent years, owing mainly to the progress of structural genomics (SG)[2] research. This information not only enriches our knowledge of certain aspects of basic protein science, such as protein-folding patterns and specific protein–protein interactions, but also provides practical contributions for structure-oriented drug design against pathogens [1–5]. The structures of some proteins remain unknown, however, because the proteins have unfavorable properties that hinder attempts to analyze their structures. Such properties include low expression level, low solubility, difficulty of purification, poor crystallizability, and instability of formed crystals. Adaptor proteins that interact with multiple other proteins, or membrane proteins with highly hydrophobic segments, often possess these properties.

The use of certain tag proteins can overcome some of these problems. A small peptide tag such as a polyhistidine tag, for example, can facilitate both the detection and purification of a target protein [6,7]. Some larger peptide tags (or protein tags) can likewise improve detection and purification while also enhancing the expression level and solubility of a protein of interest. Indeed, maltose binding protein (MBP) [8], glutathione *S*-transferase (GST) [9], thioredoxin (TRX) [10], and N utilization substance A (NusA) [11] have been widely used, and several expression vectors are commercially available. Several groups have compared the efficiency of conventional tag proteins for expression [12–15]. Although each of these studies used different expression vectors and passenger proteins, it seems reasonable to conclude that one should try several tag proteins for a particular passenger protein because the potency of a tag can be affected by multiple factors,

* Corresponding author. Address: Research Center for Asian Infectious Diseases, Institute of Medical Science, University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. Fax: +81 3 6409 2208.
E-mail address: zmatsuda@ims.u-tokyo.ac.jp (Z. Matsuda).
[1] Present address: Laboratory of Applied Biochemistry, Faculty of Applied Biological Sciences, Gifu University, Gifu 501-1193, Japan.

[2] *Abbreviations used:* SG, structural genomics; MBP, maltose binding protein; GST, glutathione *S*-transferase; TRX, thioredoxin; NusA, N utilization substance A; HIV, human immunodeficiency virus; SIV, simian immunodeficiency virus; AIDS, acquired immunodeficiency syndrome; Vpr, virus protein R; MSD, membrane-spanning domain; *T. th., Thermus thermophilus;* GFP, green fluorescent protein; Vpx, viral protein X; PCR, polymerase chain reaction; LB, Luria–Bertani; IPTG, isopropylthiogalactoside; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; PBS, phosphate-buffered saline; GdmHCl, guanidium hydrochloride; CD, circular dichroism; UV, ultraviolet; FACS, fluorescence-activated cell sorting.

0003-2697/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2008.10.050

including the specific properties of the passenger proteins as well as the linker between the tag and passenger proteins.

In many cases, tag proteins serve as molecular chaperons during the synthesis and folding of the passenger protein [12], only to be removed after production and purification. For an extremely insoluble passenger protein, a tag protein is retained so as to facilitate solubilization. If a proper tag protein is selected, the passenger protein's function is well maintained even with a fused tag form [16]. Moreover, the remaining tag proteins, if their structures are known, can serve as models of molecular replacement for phase determination [16,17]. Lysozyme [18], MBP [19], and GST [20] have been successfully used for this purpose.

In our laboratory, we are studying the several proteins of primate lentiviruses, human immunodeficiency virus (HIV), and simian immunodeficiency virus (SIV) that are linked with acquired immunodeficiency syndrome (AIDS). One of them is virus protein R (Vpr), and another is gp41, a subunit of envelope protein. Vpr is a highly insoluble 14-kDa accessory gene product of HIV-1. It works as a pleiotropic adaptor protein in the life cycle of HIV-1 [21–26]. The gp41 subunit plays a critical role for membrane fusion and contains highly hydrophobic membrane-spanning domain (MSD). The membrane-spanning domain of gp41 (gp41-MSD) anchors the HIV-1 envelope protein to lipid bilayers [27]. During our study of these insoluble proteins, we tried to improve solubilization by using several conventional tag proteins but had limited success. Therefore, we tried to identify new tag proteins that would facilitate the production and structure determination of proteins of interest. We sought candidate tag proteins from *Thermus thermophilus* (*T. th.*), one of the well-characterized model organisms of the SG project [28,29]. Because *T. th.* survives at 85 °C and its proteins are heat-stable [30], the structures of more than 400 of approximately 2000 known gene products have been determined during the past 5 years [28]. Drawing on information that has accumulated from SG research on *T. th.* proteins, we tested proteins with high expression and solubility characteristics and known high-resolution three-dimensional structures for their utility as candidate tag proteins.

We tested 15 candidate tag proteins on Vpr of HIV-1 and identified a *T. th.* tag protein that could express a soluble form of Vpr. This tag and another tag were also able to help expression of a gp41-MSD. Analyses using green fluorescent protein (GFP) as a fusion passenger protein suggested that these *T. th.* tags did not significantly affect the biophysical properties of the passenger proteins. Our results suggest that the information accumulated during the SG project, as exemplified here for *T. th.*, can be a versatile resource for the identification of customized tag proteins that may facilitate biophysical and structural analyses of highly insoluble proteins.

Materials and methods

Selection of candidate tag proteins from T. th

We used the accumulated data from expression, purification, and crystallization trials in the Whole Cell Project of *T. th.* [29,31] to select candidate proteins. From the more than 400 proteins with known structures, we selected proteins that showed good expression and solubility. In the Whole Cell Project of *T. th.*, we arbitrarily defined seven classes to indicate a protein's expression level and solubility fitness: (i) no expression, (ii) low expression and low solubility, (iii) high expression and low solubility, (iv) high expression and half solubility, (v) low expression and high solubility, (vi) high expression and half solubility, and (vii) high expression and high solubility. We selected proteins that fell in the high expression and high solubility category. Among these selected proteins, we further narrowed the field by choosing candidates whose resolution of the determined structure is higher than 2 Å.

Construction of expression plasmids

All of the proteins were expressed in *Escherichia coli* BL21 Star (DE3) (Invitrogen) using the same plasmid derived from pET-47b (Novagen). pET-47b contains a T7 promoter, a six polyhistidine tag-coding sequence, and an HRV 3C protease recognition sequence upstream of the multiple cloning site. A tag protein was added at either the N or C terminus of a passenger protein (Fig. 1). To add the tag at the N terminus, we used a modified pET-47b plasmid called pET-47md that had been produced by cloning an oligonucleotide cassette with *Asc*I, *Pac*I, *Bam*HI, and *Bsr*GI sites, in that order, between the *Kpn*I and *Avr*II sites of pET-47b. To fuse a candidate tag protein at the C terminus of a passenger protein, pET-47b was used. The pET-47b vector has *Bam*HI, *Bsr*GI, *Asc*I, and *Pac*I sites, in that order, in the multiple cloning site.

The genes for tag proteins were amplified by polymerase chain reaction (PCR) using primers that contained *Asc*I and *Pac*I sites at the 5′ and 3′ termini of the genes while also using KOD Plus (Toyobo), Pfu Turbo (Stratagene), and Ex Taq (Takara) DNA polymerases. The template DNAs we used included 15 *T. th.* expression plasmids for *T. th.* genes [31,32] (RIKEN Bioresource Center DNA bank), pMALp2e (New England Biolabs) for *mbp,* pGEX5x-1 (GE Healthcare) for *gst,* pET-50b (Novagen) for *nusA,* pSP65HXB2gpt for *vpr,* and the synthetic *vpx* (viral protein X) gene whose sequence is based on that of SIVmac251 and optimized for yeast codon use. Modified GFPs, which were used as passenger proteins, were also amplified by PCR using primers with *Bam*HI and *Bsr*GI sites at the 5′ and 3′ termini, respectively [33]. The MSD portion of gp41 [27] was synthesized as a complementary pair of oligode-

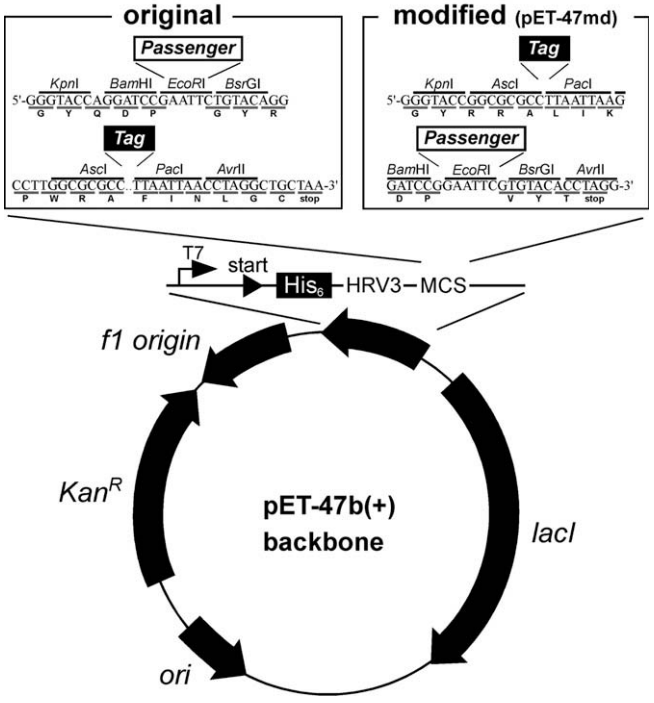

Fig. 1. Schematic representation of plasmids used in this study. The upper panel shows the details of the multiple cloning site of pET-47b and pET-47md. The lower panel shows the structure of the backbone. T7, T7 promoter; His_6, polyhistidine tag; HRV3, region coding amino acid sequences recognized by HRV 3C protease; MCS, multiple cloning site; Tag, genes of tag proteins; Passenger, genes of passenger proteins.

oxyribonucleotides with *Bam*HI and *Bsr*GI sites at the 5′ and 3′ termini, respectively. All amplicons were first cloned into pCR4blunt–TOPO vectors using the TOPO Cloning Kit (Invitrogen), and their sequences were verified before being cloning into pET-47b or pET-47md. After cloning the individual passenger gene between the *Bam*HI and *Bsr*GI sites of pET-47b or pET-47md, the respective tag genes were cloned using the *Asc*I and *Pac*I sites. As a control, we constructed expression vectors containing tag genes but without passenger genes; this was accomplished by cloning the tag genes to pET-47md with the *Asc*I and *Pac*I sites. We verified all sequences of the constructed expression vectors.

Analysis of protein expression on a small-scale

BL21 Star (DE3) (Invitrogen) was transformed by the constructed expression plasmids and cultured on Luria–Bertani (LB) plates containing 50 μg/ml kanamycin. Fresh colonies were picked and inoculated into 2 ml of LB medium containing 50 μg/ml kanamycin. All *E. coli* cells were grown at 37 °C throughout the experiments. When the cells had reached a log phase, isopropylthiogalactoside (IPTG) was added in a final concentration of 1 mM. After a 5-h growth period, the cells were harvested by centrifuging at 15,000*g* for 10 min at 4 °C using an MX-301 centrifuge (Tomy). We extracted the expressed proteins using 300 μl of BugBuster (Novagen) that contained 0.3 μl of Benzonase (Novagen). The total lysate and supernatant was subjected to sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) using a 5% to 20% SDS gradient gel (DRC). We confirmed the expressed protein with the peptide mass fingerprinting method using AXIMA–CFR Plus (Shimadzu).

Measurement of fluorescent activity of GFP

BL21 Star (DE3) cells that had been transformed using the GFP expression vectors (GFP alone or tag-fused GFP at the N terminus) were grown for 5 h in LB medium containing 1 mM IPTG and then harvested by centrifugation (15,000*g*, 4 °C). The cell pellets were suspended in 0.7 ml of phosphate-buffered saline (PBS), and the optical densities at 600 nm were determined. The fluorescence of the suspended cells was analyzed using a FACSCalibur flow cytometer (BD Biosciences). We normalized the measured mean fluorescence signal by dividing it by the obtained optical densities at 600 nm (OD_{600}).

Analysis of stability of tagged protein

A 20-μl aliquot of the supernatant of tag-free or tag-fused GFP-expressing cell lysate prepared as described above was added to a 500-μl buffer solution containing 20 mM Tris–HCl (pH 8.0) and increasing concentrations (1, 2, 3, 4, and 5 M) of guanidium hydrochloride (GdmHCl). After incubating the preparations for 1 day at room temperature, we measured the emission spectra from 500 to 600 nm using an F-4500 fluorescence spectrophotometer (Hitachi) set to an excitation wavelength of 470 nm.

Large-scale expression and purification

BL21 Star (DE3) cells transformed with individual expression plasmids were grown in 1 L of LB medium containing 50 μg/ml kanamycin. When OD_{600} reached 0.5 to 0.7, IPTG was added to produce a final concentration of 1 mM so as to induce the expression of target proteins. After a 5-h growth period, cells were pelleted down and frozen at −30 °C. To extract the protein, the cell pellets were resuspended in buffer A (20 mM NaPi [pH 7.4], 0.5 M NaCl, and 20 mM imidazole). The suspended cells were placed on ice and subjected to ultrasonication for 1 h. The disrupted cell extract was centrifuged at 5000*g* for 30 min at 4 °C to remove insoluble materials. The supernatants were loaded onto Ni Sepharose 6 Fast Flow columns (GE Healthcare) and eluted with a second buffer (20 mM NaPi [pH 7.4], 0.5 M NaCl, and 0.5 M imidazole). HRV 3 C protease (Novagen) was added to the eluted fractions, and the preparations were incubated at 4 °C for more than 16 h to remove the polyhistidine tags. After exchanging the second buffer for buffer A using VivaSpin (Sartorius), the solutions again were loaded onto Ni Sepharose 6 Fast Flow columns. The flow-through fraction was collected and concentrated to 2 to 5 ml in a buffer containing 50 mM Tris–HCl (pH 8.0) and 0.5 M NaCl. The concentrated fraction was subjected to Superdex 200 size exclusion chromatography using an AKTA Purifier (GE Healthcare). We collected the highest peaks and verified their content using SDS–PAGE.

Secondary structure analyses

Purified proteins were placed in buffer (20 mM NaPi [pH 7.0] and 0.5 M NaCl), and their circular dichroism (CD) in the far-UV (ultraviolet) region of 200 to 250 nm was measured at 25 °C using a Pistar-180 Spectrometer (Applied Photophysics).

Results

Expression of Vpr protein with conventional tag fusion

To obtain soluble Vpr protein for future analysis, we tried to express Vpr in *E. coli*. When Vpr was expressed without any tag, it did not express very well. Even when it was expressed, it did not come to a soluble fraction (Fig. 2A). This is consistent with previous studies [26,34]. The conventional tag genes *mbp, gst,* and *nusA* were introduced upstream or downstream of the HIV *vpr* gene in the expression plasmid (Fig. 1). Among the conventional tag proteins, GST and NusA facilitated expression of fusion protein when added at the N or C terminus, whereas MBP failed to achieve a high expression level (Fig. 2B). Unfortunately, all of the tagged Vpr proteins that were expressed were observed in the insoluble fraction (Fig. 2B).

Selection of candidate T. th. tag protein

To seek other effective tags for Vpr protein expression and solubilization, we decided to introduce proteins from *T. th.*; the SG project of *T. th.* [28] has provided many data sets, including those for protein expression, purification, and crystallization [29], that aided in our selection of viable candidates for this effort. We selected 59 candidates from more than 400 proteins based on their expression level and the solubility (see Materials and methods for details). From these 59, we selected 31 candidates with resolutions of structures higher than 2 Å. By coincidence, the molecular weight of each of the 31 candidates was less than 40 kDa. This may reflect the fact that smaller molecular weight proteins are easier to express and crystallize [35].

We further narrowed the candidate pool by selecting 14 proteins for which information on small-scale protein preparation was available. To these 14 candidates, we added a 57.9-kDa protein, the largest in the first-pass group of 59 proteins. We included this protein so as to have a larger molecular weight tag protein that could serve as a favorable model of molecular replacement for larger passenger proteins. The complete list of 15 candidate proteins appears in Table 1. As a group, there seemed to be no common functional or structural correlations, probably because we selected the candidates solely for their practical properties of expression and solubility and on their structural analyses.

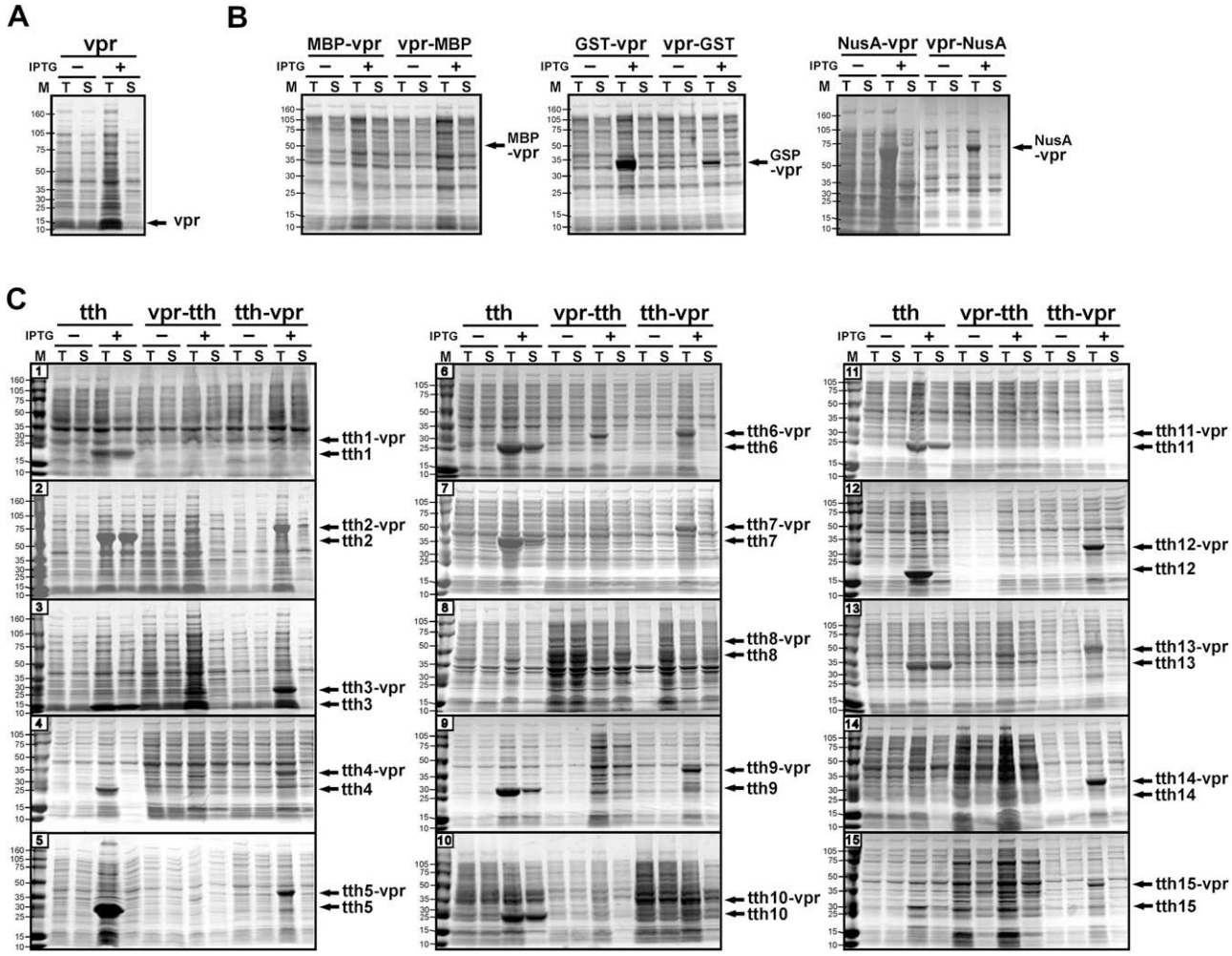

Fig. 2. Expression profile of tag protein-fused Vpr with (+) or without (−) induction by IPTG. Based on calculated molecular weights, the positions of expected proteins are marked by arrows along the right edge of each gel. (A) Expression profiles of tag-free Vpr. (B) Expression profiles of MBP-fused Vpr, GST-fused Vpr, and NusA-fused Vpr. (C) Expression profiles of individual *T. th.* protein candidates alone (tth) and when fused with Vpr at the C terminus (vpr–tth) or at the N terminus (tth–vpr). The number indicated in the box in the upper left corner of each gel corresponds to the number used for the candidate tth gene listed in Table 1. Gel columns indicate molecular weight markers (in kDa) (M), total lysate (T), and supernatant (S).

Table 1
Selected candidates of protein tags from *T. th.*

Name in this study	Locus tag in genome	Functional feature	M_W (kDa)	PDB ID	Resolution (Å)
tth1	TTHA0132	Conserved hypothetical protein	16.9	2dx6	1.78
tth2	TTHA0271	Molecular chaperon, GroEL	57.9	1wf4	2.80
tth3	TTHA0281	UPF0150 family protein	9.8	2dsy	1.90
tth4	TTHA0341	Molybdopterin biosynthesis enzyme MoaB	17.8	2is8	1.64
tth5	TTHA1053	Conserved hypothetical protein	20.7	2cve	1.60
tth6	TTHA1091	Conserved hypothetical protein	17.8	1vgg	1.75
tth7	TTHA1199	Ornithine carbamoyltransferase	33.3	2ef0	2.00
tth8	TTHA1275	C subunit of $V_1 - V_0$ ATPase	36.0	1v9m	1.85
tth9	TTHA1359	Transcriptional regulator	22.4	2zcw	1.50
tth10	TTHA1671	Adenylate kinase	20.7	3cm0	1.80
tth11	TTHA1699	Putative transediting enzyme of tRNA-Syn	16.6	2cx5	1.90
tth12	TTHA1713	Conserved hypothetical protein	14.4	2cu5	1.84
tth13	TTHA1838	Atypical ABC-ATPase, SufC	27.6	2d2f	1.90
tth14	TTHA1897	GidA-related protein	25.9	2cul	1.65
tth15	TTHB192	CRISPR-associated protein	23.8	1wj9	1.90

Expression of Vpr protein with T. th. protein fusion

Because the expression vectors used in this study (pET-47b and pET-47md) (Fig. 1) have the polyhistidine and protease recognition sequences that are absent from the vectors used in the *T. th.* SG project (pET-3a and pET-11a) [28,29], we began our investigation by evaluating the selected *T. th.* proteins in the pET-47b backbone. Our results showed that although all *T. th.* proteins were expressed well, 7 of the 15 *T. th.* proteins (Tth4, Tth5, Tth7, Tth8, Tth12, Tth14, and Tth15) failed to solubilize (Fig. 2C, tth). Because the general

conditions for the expression, such as growth medium, growth temperature (37 °C), and IPTG concentration, were identical to those of the original *T. th.* SG project, the differences in the used expression vectors, especially the presence of a polyhistidine tag and the protease recognition site in the pET-47md backbone, seem to have affected the solubility of the *T. th.* proteins.

Then we tested the expression level and the solubility of *T. th.* tag-fused Vpr. Except for tags Tth1, Tth8, and Tth11, all *T. th.* tags facilitate the expression level of the fusion protein when the *T. th.* proteins had been added at the N terminus of Vpr (Fig. 2C, tth–vpr). However, in the case of C terminus addition, only the Tth6-fused protein among the *T. th.* tag-fused Vpr proteins was expressed well (Fig. 2C, vpr–tth).We found that the Tth2 protein enhanced the solubility of Vpr when Tth2 was added to the N terminus of Vpr (Fig. 2C).

Efficacy of the tag proteins on other insoluble proteins

We investigated the efficacy of *T. th.* tag proteins on other insoluble proteins from primate lentiviruses. Based on the results of *T. th.* alone or *T. th.* in a Vpr fusion form (Fig. 2), seven *T. th.* tags (Tth2, Tth3, Tth6, Tth9, Tth10, Tth11, and Tth13) were introduced to the N terminus of two lentiviral proteins. Three conventional tags were also introduced to this position. One of these proteins, gp41-MSD, is a highly hydrophobic 2.2-kDa peptide that anchors the envelope protein to lipid bilayers [27]. The synthetic peptide of gp41-MSD remains insoluble under the condition of 100% dimethyl sulfoxide (unpublished data). Another protein, Vpx, is a virion-associated 12.9-kDa protein of SIV that plays an important role in nuclear transport of the incoming preintegration complex [26]. Vpr and Vpx are evolutionarily related to each other and have approximately 15% of their primary structures in common. Among the conventional tags, MBP and NusA enhanced the expression level of tag-fused gp41-MSD (Fig. 3A), but the expressed proteins could not be found in the soluble fractions (Fig. 3A). Six *T. th.* tags (Tth2, Tth6, Tth9, Tth10, Tth11, and Tth13) enhanced expression levels when fused at the N terminus (Fig. 3B). Both the Tth2 and Tth10 proteins were partially successful at solubilizing gp41-MSD (Fig. 3B). In the case of Vpx, expression levels were enhanced when fused with NusA, Tth2, Tth3, Tth6, Tth9, Tth10, Tth11, and Tth13, although none of these fusion proteins could be solubilized (data not shown). These results suggest that *T. th.*-tagged proteins can enhance the expression level of insoluble proteins; however, the efficacy depends heavily on the passenger proteins used.

Effect of tag proteins on the function of passenger proteins

It is of interest to know whether fusion of a tag may affect any functions of the fused passenger protein. We attempted to test what effect, if any, the *T. th.* tag-fused Vpr had on the cell cycle, but our efforts at transduction of the purified proteins into the mammalian cells were not too successful. In the case of gp41-MSD, there is no good biological assay because it is a portion of a protein. Therefore, we used GFP as a surrogate reporter for Tth2, Tth6, Tth9, Tth10, Tth11, and Tth13. Most of the tag proteins, such as Tth2, Tth3, Tth6, and Tth9, enhanced expression to levels comparable to those achieved with the conventional tags GST and NusA (Fig. 4A). Moreover, Tth2, Tth10, and Tth13 proteins enhanced the solubility of fused GFPs. The fluorescence-activated cell sorting (FACS) analyses of the fluorescence of the fused GFPs showed that the fluorescent intensity of MBP, GST, Tth11, and Tth13 fused proteins was less than half the intensity of intact GFP (Fig. 4B). As shown in Fig. 4A, both the expression level and solubility of these fusion proteins were similar to those of intact GFP; however, the fusion with these tags reduced the activity of GFP. In the case of Tth2, Tth3, Tth6, Tth9, and Tth10 proteins, the fusion affected the fluorescence of GFP only slightly.

We next examined the stability of the tagged GFPs under denaturing conditions by treating the proteins with GdmHCl (Fig. 4C and D). For the conventional tag proteins, the denaturation profile of GST-fused GFP was nearly identical to that of intact GFP, although for MBP and NusA the normalized fluorescence intensities of the tag-fused proteins were slightly lower than the intensity of intact GFP (Fig. 4C). On the other hand, all denaturation profiles of *T. th.* tag-fused proteins, except for Tth13, were similar to the profile of tag-free GFP (Fig. 4D). These results suggest that the *T. th.* tag proteins do not destabilize the GFP passenger protein. We also observed the endurance of the tag-fused proteins to heat treatment (see Supplemental Fig. 1 in supplementary material), suggesting the possibility of a simple purification protocol employing a heat treatment.

Purification and characterization of tag-fused proteins

To verify whether a large amount of tag protein can be obtained using a *T. th.* tag-mediated system, large-scale expression and purification were performed. Tth2- or Tth10-fused protein was used because these two tags could efficiently solubilize insoluble protein. After induction, the proteins were purified in two steps: using nickel–nitrilotriacetate (Ni–NTA) columns and through size exclusion (Fig. 5A). Despite the good expression, the Tth2–Vpr protein was harder to purify, probably owing to the residual insoluble properties derived from Vpr. Except for the Tth2–Vpr protein, at least 1 mg of pure protein was obtained from a 1-L culture. Meanwhile, because a high concentration of salt improved the solubility of tag-fused proteins, 0.5 M NaCl was added at all of the purification and analysis steps. CD spectra of purified proteins were measured. CD of the Tth2 protein showed the typical pattern of an α-helix-containing protein, which has negative maxima at 209 and 222 nm (Fig. 5B), consistent with the results of crystal structure of Tth2 protein (PDB ID: 1WF4). Other proteins fused with Tth2 proteins, such as Tth2–GFP, Tth2–Vpr, and Tth2–gp41-MSD, also

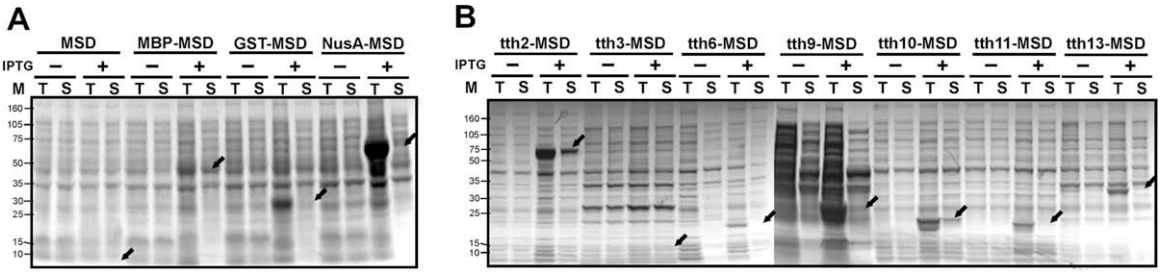

Fig. 3. Efficacy of tag protein for the expression of insoluble protein, gp41-MSD (MSD) from HIV-1. (A) Expression profiles of MSD alone and when fused with each of three conventional tags: MBP, GST, and NusA. (B) Expression profiles of tth-fused MSD. Each arrow indicates the position of the expected molecular weight of the fusion protein. Abbreviations are as defined in Fig. 2.

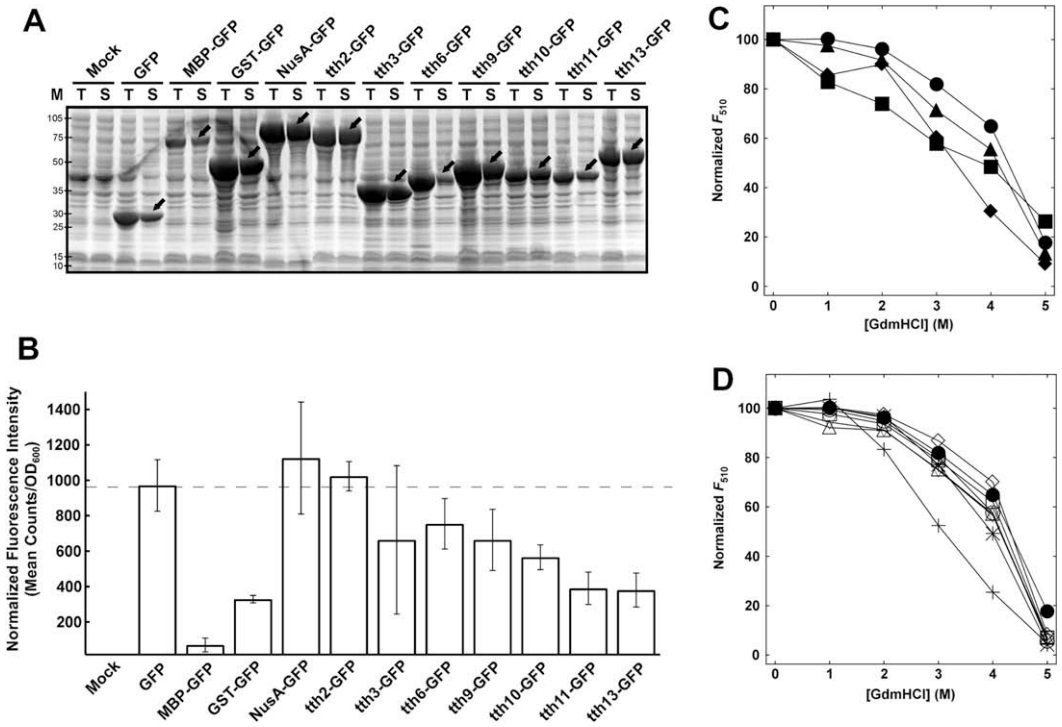

Fig. 4. Properties of tag-fused GFP. (A) Expression profile of tag-fused GFP. All cell extracts were prepared after 5 h exposure to 1 mM IPTG. Based on the calculated molecular weights, the positions of expected proteins are marked by arrows. Abbreviations are as defined in Fig. 2. (B) Fluorescence intensity of tag-fused GFP-expressing *E. coli* cells. The profiles of expression levels of GFP or tag-fused GFPs are the same as those in panel A. The measured mean fluorescence signal was normalized by dividing by the optical density of the culture at 600 nm. The standard deviations were calculated based on the results obtained in more than five independent experiments. (C,D) Residual intensity at 510 nm of proteins fused with conventional tags (C) or tth tags (D) and denatured by GdmHCl. Fluorescence intensities are expressed using nondenatured GFP as 100%. ●, tag-free GFP; ■, MBP-fused GFP; ▲, GST-fused GFP; ♦, NusA-fused GFP; ○, tth2-fused GFP; □, tth3-fused GFP; ◇, tth6-fused GFP; △, tth9-fused GFP; ▽, tth10-fused GFP; *, tth11-fused GFP; +, tth13-fused GFP.

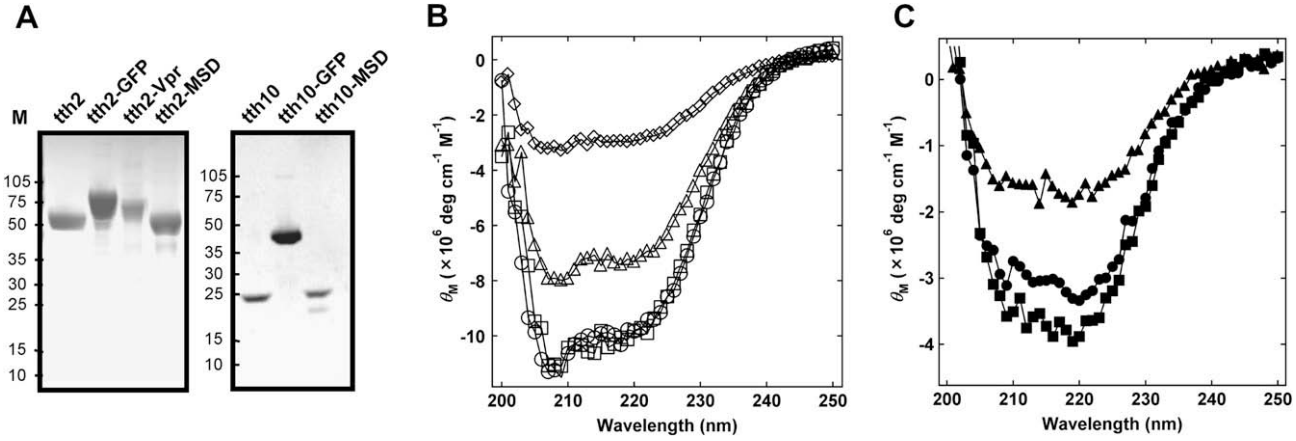

Fig. 5. Purification and secondary structure analyses of tag-fused proteins. (A) SDS–PAGE profile of purified proteins with MSD indicating gp41-MSD. (B) CD spectra of tth2-fused proteins. ○, tth2 alone; □, tth2-fused GFP; △, tth2-fused Vpr; ◇, tth2-fused MSD. (C) CD spectra of tth10-fused proteins. ●, tth10 alone; ■, tth10-fused GFP; ▲, tth10-fused MSD.

have similar negative maxima, suggesting that a considerable number of α-helical structures were present in the tag-fused proteins (Fig. 5B). The intensities of CD spectra for Tth2 and Tth2–GFP proteins were nearly identical, whereas those of Tth2–Vpr and Tth2–gp41-MSD proteins were lower. As for Tth10 and Tth10-fused proteins, the patterns of negative maxima were similar but the intensity of Tth10-fused gp41-MSD was reduced compared with that of Tth10 protein alone. Although it is difficult to explain the decrease in the intensities, the observed decreases may suggest that fusion-driven alterations to the proteins' structures occur.

Discussion

In this study, we used the information on properties of the *T. th.* proteins accumulated during structural genomics studies to produce tag proteins that could facilitate the high production of target proteins. We applied two major criteria for the selection of the candidate tag proteins: high expression level and high solubility. The additional criterion, a high-resolution structure, was employed because this characteristic can facilitate phase determination during X-ray diffraction; the fused tag can be used as a template for molecular replacement. The 15 proteins that

met these criteria were diverse in their functions and structures (Table 1).

We examined the versatility of the selected tag proteins by fusing them to the highly insoluble adaptor protein Vpr. The candidate *T. th.* tag proteins showed themselves to be more effective when placed at the N terminus of Vpr (Fig. 2). The efficacies of the enhancements to protein expression proved to be nearly comparable to—or sometimes better than—those achieved with conventional tag proteins. Thus, the proteins derived from *T. th.* can be used as alternative tag proteins for efficient protein expression. Furthermore, we successfully identified one *T. th.* tag protein, Tth2, that could express Vpr in a soluble form. In addition to Tth2, the Tth10 protein could solubilize the extremely insoluble gp41-MSD (Figs. 2 and 3) more effectively than MBP, GST, or NusA.

When we tested Vpx as the fusion partner, the *T. th.* tag-fused proteins were expressed better than tag-free proteins, although the tagged proteins did not solubilize (data not shown). This apparent difference in solubility could arise from many factors other than the difference of the passenger proteins. One possible factor is the influence of the sequence that links the tag and the passenger protein. Indeed, we observed that some *T. th.* tags became insoluble when fused to a polyhistidine tag in our pET-47 vector (Fig. 2C). Others have observed similar effects of the linker sequence on expression and, possibly, solubility [12–14]. The linker sequence also plays an important role in the crystallization of a tag-fused protein [17]. Therefore, for successful analyses, it seems to be essential that both the fusion tag and linker sequence be carefully selected.

The high crystallization characteristics of *T. th.* proteins may confer additional advantage to this tag system. For this reason, we intended to retain the *T. th.* tag proteins after the purification not only to keep the proteins soluble but also to facilitate their crystallization. Therefore, we estimated the effects that the tag proteins had on the passenger proteins. Our analyses of tag-fused GFPs showed that 7 of the 15 candidate *T. th.* tags did not diminish either the function or the stability of GFP (Fig. 4). Even the considerable effect that the Tth2 protein had on the function of the passenger protein compared favorably with that produced with GST. In contrast, MBP had a considerable negative effect on the function of GFP; it also decreased the expression level of GFP itself. This effect has been reported previously [36]. Our results suggest that *T. th.* tags have similar or milder effects on the function of passenger proteins compared with the effects produced with the conventional tags MBP, GST, and NusA.

We focused mainly on the practical properties of *T. th.* proteins rather than on their biological functions when selecting our candidate tag proteins. Therefore, the two successful candidates identified for Vpr and gp41-MSD in this study have unrelated biochemical properties; Tth2 protein is the molecular chaperon GroEL, and Tth10 protein is the phosphotransferase adenylate kinase (Table 1). Each is known to undergo a conformational change on ligand binding [37–40], a characteristic suggesting that these proteins have flexible structures. In Fig. 5, the molar ellipticities of the fusion proteins, especially those fused with insoluble proteins, were not cumulative against the ellipticity of *T. th.* tag alone. The decrease in the intensities of the *T. th.* tag-fused insoluble proteins, such as Tth2–Vpr, Tth2–gp41-MSD, and Tth10–gp41-MSD, may reflect the structure alternation of the *T. th.* tag proteins for solubilization. There seems to be no severe structural alteration in the passenger protein; the fluorescent intensities of GFP and Tth2- or Tth10-fused GFP were nearly identical (Fig. 4).

In the case of Tth2 protein, the chaperon activity of GroEL may mediate the process of protein-folding and result in the efficient solubilization of the passenger protein [41]. Indeed, coexpression of GroEL/GroES for the production of target proteins in *E. coli* is widely used for the production of correctly folded proteins [42–46]. It has been reported, however, that GroEL from *T. th.* does not have chaperon activity alone [47]. If Tth2 protein acts as a molecular chaperon, endogenous GroES in *E. coli* may function in a coordinate manner during protein synthesis. At this point, we cannot explain the exact mechanism of Tth2-induced solubilization.

In this study, we have used the systematically accumulated data of the structural genomics project of *T. th.* to develop tag proteins useful for protein expression and solubilization. The properties used for the selection of candidate tags, such as high expression level and high solubility, are beneficial but do not necessarily guarantee that the protein will be a successful solubilizing partner. Although different proteins may need different fusion tags to achieve maximum efficacy of expression and solubility, it is noteworthy that the highly hydrophobic gp41-MSD could be solubilized with the use of *T. th.* tag. It is estimated that approximately 30% of all proteins are membrane proteins, and structural analysis of these proteins can be difficult because these proteins are often insoluble. The approach shown here may facilitate their structural analysis. Our preliminary trials of crystallization of *T. th.* tag-fused Vpr and gp41-MSD yielded small crystals. We are interested in determining whether the attached tag protein can facilitate phase determination during X-ray crystallographic analyses, as we have proposed. Moreover, investigation of this type must be useful because the greater the number of effective solubility-enhancing tags we can identify, the better are the chances of understanding the mechanism of solubility enhancement.

Acknowledgments

This work was supported in part by the Program of Founding Research Centers for Emerging and Reemerging Infectious Diseases of the Ministry of Education, Culture, Sports, Science, and Technology (MEXT). We are grateful to the staff members of the Structural Biology Core Facility at the Institute of Biophysics, Chinese Academy of Sciences, for technical assistance. We thank Zhensheng Xie of Proteomic Platform at the Institute of Biophysics, Chinese Academy of Sciences, for identification of proteins with mass spectrometry. We thank Xinyu Wang for assistance with the Pistar instrument. We thank Kunito Yoshiike and Mark Bartlam for a critical reading of the manuscript. We also thank A. M. Menting, an editorial consultant, for preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ab.2008.10.050.

References

[1] S.G. Dahl, I. Sylte, Molecular modeling of drug targets: The past, the present, and the future, Basic Clin. Pharmacol. Toxicol. 96 (2005) 151–155.
[2] K. Klumpp, T. Mirzadegan, Recent progress in the design of small molecule inhibitors of HIV RNase H, Curr. Pharm. Des. 12 (2006) 1909–1922.
[3] P. Beltrao, C. Kiel, L. Serrano, Structures in systems biology, Curr. Opin. Struct. Biol. 17 (2007) 378–384.
[4] M. Grabowski, A. Joachimiak, Z. Otwinowski, W. Minor, Structural genomics: keeping up with expanding knowledge of the protein universe, Curr. Opin. Struct. Biol. 17 (2007) 347–353.
[5] J. Weigelt, L.D. McBroom-Cerajewski, M. Schapira, Y. Zhao, C.H. Arrowmsmith, Structural genomics and drug discovery: all in the family, Curr. Opin. Chem. Biol. 12 (2008) 32–39.
[6] M.C. Smith, T.C. Furman, T.D. Ingolia, C. Pidgeon, Chelating peptide-immobilized metal ion affinity chromatography: a new concept in affinity chromatography for recombinant proteins, J. Biol. Chem. 263 (1988) 7211–7215.
[7] K. Terpe, Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems, Appl. Microbiol. Biotechnol. 60 (2003) 523–533.

[8] D. Sachdev, J.M. Chirgwin, Fusions to maltose-binding protein: control of folding and solubility in protein purification, Methods Enzymol. 326 (2000) 312–321.
[9] D.B. Smith, Generating fusions to glutathione *S*-transferase for protein studies, Methods Enzymol. 326 (2000) 254–270.
[10] E.R. LaVallie, E.A. DiBlasio, S. Kovacic, K.L. Grant, P.F. Schendel, J.M. McCoy, A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm, Bio/Technology 11 (1993) 187–193.
[11] G.D. Davis, C. Elisee, D.M. Newham, R.G. Harrison, New fusion protein systems designed to give soluble expression in *Escherichia coli*, Biotechnol. Bioeng. 65 (1999) 382–388.
[12] R.B. Kapust, D.S. Waugh, *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused, Protein Sci. 8 (1999) 1668–1674.
[13] L. Niiranen, S. Espelid, C.R. Karlsen, M. Mustonen, S.M. Paulsen, P. Heikinheimo, N.P. Willassen, Comparative expression study to increase the solubility of cold adapted Vibrio proteins in *Escherichia coli*, Protein Expr. Purif. 52 (2007) 210–218.
[14] J.G. Marblestone, S.C. Edavettal, Y. Lim, P. Lim, X. Zuo, T.R. Butt, Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO, Protein Sci. 15 (2006) 182–189.
[15] S. Nallamsetty, D.S. Waugh, Solubility-enhancing protein MBP and NusA play a passive role in the folding of their fusion partners, Protein Expr. Purif. 45 (2006) 175–182.
[16] G.G. Privé, G.E. Verner, C. Weitzman, K.H. Zen, D. Eisenberg, H.R. Kaback, Fusion proteins as tools for crystallization: the lactose permease from *Escherichia coli*, Acta Crystallogr. D 50 (1994) 375–379.
[17] D.R. Smyth, M.K. Mrozkiewicz, W.J. Mcgrath, P. Listwan, B. Kobe, Crystal structures of fusion proteins with large-affinity tags, Protein Sci. 12 (2003) 1313–1322.
[18] J.P. Donahue, H. Patel, W.F. Anderson, J. Hawiger, Three-dimensional structure of the platelet integrin recognition segment of the fibrinogen gamma chain obtained by carrier protein-driven crystallization, Proc. Natl. Acad. Sci. USA 91 (1994) 12178–12182.
[19] B. Kobe, R.J. Center, B.E. Kemp, P. Poumbourios, Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins, Proc. Natl. Acad. Sci. USA 96 (1999) 4319–4324.
[20] Y. Zhan, X. Song, G.W. Zhou, Structural analysis of regulatory protein domains using GST-fusion proteins, Gene 281 (2001) 1–9.
[21] E. Le Rouzic, N. Belaïdouni, E. Estrabaud, M. Morel, J.C. Rain, C. Transy, F. Margottin-Goguet, HIV-1 Vpr function is mediated by interaction with the damage specific DNA-binding protein DDB1, Cell Cycle 6 (2007) 182–188.
[22] B. Schrofelbauer, Y. Hakata, N.R. Landau, HIV-1 Vpr function is mediated by interaction with the damage specific DNA-binding protein DDB1, Proc. Natl. Acad. Sci. USA 104 (2007) 4130–4135.
[23] K. Hrecka, M. Gierszewska, S. Srivastava, L. Kozaczkiewicz, S.K. Swanson, L. Florensm, M.P. Washburn, J. Skowronski, Lentiviral Vpr usurps Cul4-DDB1[VprBP] E3 ubiquitin ligase to modulate cell cycle, Proc. Natl. Acad. Sci. USA 104 (2007) 11778–11783.
[24] X. Wen, K.M. Duus, T.D. Friedrich, C.M. de Noronha, The HIV-1 protein Vpr acts to promote G2 cell cycle arrest by engaging a DDB1 and Cullin4A-containing ubiquitin ligase complex using VprBP/DCAF1 as an adaptor, J. Biol. Chem. 282 (2007) 27046–27057.
[25] L. Tan, E. Ehrlich, X.F. Yu, DDB1 and Cul4A are required for human immunodeficiency virus type 1 Vpr-induced G2 arrest, J. Virol. 81 (2007) 10822–10830.
[26] E. Le Rouzic, S. Benichou, The Vpr protein from HIV-1: distinct roles along the viral life cycle, Retrovirology 2 (2005) 11.
[27] K. Miyauchi, J. Komano, Y. Yokomaku, W. Sugiura, N. Yamamoto, Z. Matsuda, Role of the specific amino acid sequence of the membrane-spanning domain of human immunodeficiency virus type 1 in membrane fusion, J. Virol. 79 (2005) 4720–4729.
[28] Structural–Biological Whole Cell Project. Available from: <http://www.thermus.org>.
[29] H. Iino, H. Naitow, Y. Nakamura, N. Nakagawa, Y. Agari, M. Kanagawa, A. Ebihara, A. Shikai, M. Sugahara, M. Miyano, N. Kamiya, S. Yokoyama, K. Hirotsu, S. Kuramitsu, Crystallization screening test for the Whole-Cell Project on *Thermus thermophilus* HB8, Acta Crystallogr. F 64 (2008) 487–491.
[30] T. Oshima, K. Imahori, Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa, Int. J. Syst. Bacteriol. 24 (1974) 102–112.
[31] S. Yokoyama, H. Hirota, T. Kigawa, T. Yabuki, M. Shirouzu, T. Terada, Y. Ito, Y. Matsuo, Y. Kuroda, Y. Nishimura, Y. Kyogoku, K. Miki, R. Masui, S. Kuramitsu, Structural genomics projects in Japan, Nat. Struct. Biol. 7 (2000) 943–945.
[32] RIKEN Bioresource Center DNA Bank, *Thermus thermophilus*. Available from: <http://www.brc.riken.jp/lab/dna/en/thermus_en.html>.
[33] S. Cabantous, T.C. Terwilliger, G.S. Waldo, Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein, Nat. Biotechnol. 23 (2005) 102–107.
[34] P. Henklein, K. Bruns, M.P. Sherman, U. Tessmer, K. Licha, J. Kopp, C.M. de Noronha, W.C. Greene, V. Wray, U. Schubert, Functional and structural characterization of synthetic HIV-1 Vpr that transduces cells, localizes to the nucleus, and induces G2 cell cycle arrest, J. Biol. Chem. 275 (2000) 32016–32026.
[35] L. Slabinski, L. Jaroszewski, A.P. Rodrigues, L. Rychlewski, I.A. Wilson, S.A. Lesley, A. Godzik, The challenge of protein structure determination lessons from structural genomics, Protein Sci. 16 (2007) 247–282.
[36] A.H. Podmore, P.E. Reynolds, Purification and characterization of VanXYC, a D,D-dipeptidase/D,D-carboxypeptidase in vancomycin-resistant *Enterococcus gallinarum* BM4174, Eur. J. Biochem. 269 (2002) 2740–2746.
[37] M. Taniguchi, T. Yoshimi, K. Hongo, T. Mizobata, Y. Kawata, Stopped-flow fluorescence analysis of the conformational changes in the GroEL apical domain: relationships between movements in the apical domain and the quaternary structure of GroEL, J. Biol. Chem. 279 (2004) 16368–16376.
[38] M. Yokokawa, C. Wada, T. Ando, N. Sakai, A. Yagi, S.H. Yoshimura, K. Takeyasu, Fast-scanning atomic force microscopy reveals the ATP/ADP-dependent conformational changes of GroEL, EMBO J. 25 (2006) 4567–4576.
[39] K. Arora, C.L. Brooks III, Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism, Proc. Natl. Acad. Sci. USA 104 (2007) 18496–18501.
[40] P.C. Whitford, S. Gosavi, J.N. Onuchic, Conformational transitions in adenylate kinase: allosteric communication reduces misligation, J. Biol. Chem. 283 (2008) 2042–2048.
[41] A.L. Horwich, G.W. Farr, W.A. Fenton, GroEL–GroES-mediated protein folding, Chem. Rev. 106 (2006) 1917–1930.
[42] P. Goloubinoff, A.A. Gatenby, G.H. Lorimer, GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*, Nature 337 (1989) 44–47.
[43] S.C. Lee, P.O. Olins, Effect of overproduction of heat shock chaperones GroESL and DnaK on human procollagenase production in *Escherichia coli*, J. Biol. Chem. 267 (1992) 2849–2852.
[44] P. Caspers, M. Stieger, P. Burn, Overproduction of bacterial chaperones improves the solubility of recombinant protein tyrosine kinases in *Escherichia coli*, Cell. Mol. Biol. 40 (1994) 635–644.
[45] K.E. Amrein, B. Takacs, M. Stieger, J. Molnos, N.A. Flint, P. Burn, Purification and characterization of recombinant human p50csk protein–tyrosine kinase from an *Escherichia coli* expression system overproducing the bacterial chaperones GroES and GroEL, Proc. Natl. Acad. Sci. USA 92 (1995) 1048–1052.
[46] K. Nishihara, M. Kanemori, H. Yanagi, T. Yura, Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*, Appl. Environ. Microbiol. 66 (2000) 884–889.
[47] K. Amada, M. Yohda, M. Odaka, I. Endo, N. Ishii, H. Taguchi, M. Yoshida, Molecular cloning, expression, and characterization of chaperonin-60 and chaperonin-10 from a thermophilic bacterium, *Thermus thermophilus* HB8, J. Biochem. 118 (1995) 347–354.