Table 1 (*continued*)

No. of Arabidopsis genes	PlantP funciton[a]	PlantP symbol[a]	PlantP no.[a]	PlantP family[a]	Relative luminescence signal[b]	Autophosphorylation[c]
at5g24430	CDPK-related protein kinase isoform 4	CRK4	21984	Family 4.2.1 – calcium dependent protein kinase	11.7	Yes
at4g00720	GSK3/shaggy-like protein kinase tetha	ASK-tetha	21794	Family 4.5.4 – GSK3/shaggy like protein kinase family	11.5	Yes
at4g26890	Putative NPK1-related protein kinase	–	21878	Family 4.4.1 – unknown function kinase	11.1	nd
at1g12970	–	–	-	–	11.1	–
at1g09020	Activator subunit of SNF1-related protein kinase SNF4	–	10475	–	10.9	–
at5g26750	GSK3/shaggy-like protein kinase ASK-alpha	ASK-alpha	21309	Family 4.5.4 – GSK3/shaggy like protein kinase family	10.6	Yes
at5g04510	3-Phosphoinositide-dependent kinase-1 PDK1	–	21950	Family 4.2.6 – IRE/NPH/PI dependent/S6 kinase	10.5	Yes
at3g51850	Calcium-dependent Protein Kinase, isoform 13	CPK13	21791	Family 4.2.1 – calcium dependent protein kinase	10.5	Yes
at2g40120	Putative protein kinase	–	21546	Family 4.5.8 – unknown function kinase	10.4	nd
at4g24400	SNF1-related protein kinase, subfamily 3	SnRK3.13	21869	Family 4.2.4 – SNF1 related protein kinase (SnRK)	10.3	nd
at1g01540	Hypothetical protein	–	21038	Family 1.6.3 – receptor like cytoplasmic kinase V	10.3	Yes
at1g73500	Putative protein kinase	–	21164	Family 4.1.3 – MAP2K	10.2	Yes
at1g20930	Putative cdc2-like protein kinase	–	21078	Family 4.5.2 – CDC2 like kinase family	10.0	Yes
at3g05840	GSK3/shaggy-like protein kinase, ASK-gamma	ASK-gamma	21676	Family 4.5.4 – GSK3/shaggy like protein kinase family	10.0	Yes
at3g61160	GSK3/shaggy-like protein kinase beta	ASK-beta	21787	Family 4.5.4 – GSK3/shaggy like protein kinase family	10.0	Yes

Abbreviations: Yes, detected; nd, not detected; –, not analyzed.
[a] Their classification were according to PlantP database (http://plantsp.sdsc.edu/).
[b] This works.
[c] Autoradiography of *in vitro* kinase assays (Sawasaki et al., 2004).

related with cell growth (Bai et al., 2009). Furthermore, *Brassica napus* BuPERK (a putative orthologous gene of *Arabidopsis thaliana* At-PERK1) gene expression is induced by wound and fungal pathogens stress (Silva and Goring, 2002). Thus, both of the signaling pathway of CPK and PERK be induced by the stress and ABA, and may have cross-talk via a Ca^{2+}-signaling pathway, or regulation by direct trans-phosphorylation.

The reversible protein phosphorylation on Ser, Thr, and Tyr is a key post-translational modification in eukaryotes with stunning regulatory and signalling potential. We clarified the Ser/Thr autophosphorylating activity of 759 PKs that the biochemical characterization was indistinct. In addition, the autophosphorylation of PKs biologically suggested that it be the important. This study will be a basis for understanding the function of PKs in phosphorylation network for future research. The function of the Tyr phosphorylation has been largely neglected because a mammalian protein Tyr kinase homolog was not found in plants. By using this method, may be able to identify protein Tyr kinase and phosphorylation network in plants.

4. Conclusion

In this study, we found that the wheat cell-free system was an excellent expression system to produce recombinant plant PKs efficiently and to carry out *in vitro* phosphorylation assays without purification, the interference of endogenous PK and phosphatase activity. Using a full-length cDNA library of *A. thaliana* genes encoding PKs, we demonstrated that the wheat germ cell-free system is capable of profiling the autophosphorylation activity of 759 PKs, and also found 67 highly active PKs. The annotation analysis revealed that some of these PKs may be involved in phosphosignaling pathways such as signal transduction, stress response, and the regulation of cell division. Information from this study may shed light on many unknown plant PKs.

5. Experimental

5.1. General

Details of the following procedures were either described or cited previously (Ogasawara et al., 1999; Madin et al., 2000; Sawasaki et al., 2002a,b): isolation of wheat germs and preparation of the extract, generation of DNA template by "split-primer" PCR, synthesis of mRNA, *in vitro* protein synthesis, and the estimation of protein quantity synthesized by densitometric scanning of the Coomassie brilliant blue (CBB)-stained band and autoradiography.

5.2. Construction of DNA templates for transcription

The unique primer for the "split-primer" PCR for each of the 768 cDNAs from the RAFL clones was designed according to the sequence in the database (Seki et al., 2002). The DNA templates for transcription were constructed by "split-primer" PCR technique described in previous reports. (Sawasaki et al., 2002a; Sawasaki et al., 2007) The first PCR product was amplified with 10 nM of each of the following primers: a gene specific primer, 5′-CCACC CACCACCACCAATGnnnnnnnnnnnnnnnnnnn (n denotes the coding region of the target gene), and AODA2303 (5′-GTCAGACC CCGTAGAAAAGA) or AODS (5′-TTTCTACGGGGTCTGACGCT). The second PCR products for protein synthesis were constructed with 100 nM SPu 5′-GCGTAGCATTTAGGTGACACT, 1 nM deSP6E02bls-S1 (5′-GGTGACACTATAGAACTCACCTATCTCTCTACACAAAAC-ATTTC CCTACATACAACTTTCAACTTCCTATTATGGGCCTGAACGACATCTTCG AGGCCCAGAAGATCGAGTGGCACGAACTCCACCCACCACCACCAATG) and 100 nM AODA2303 or AODS. By this "split-primer" PCR, the bls was fused onto the N-terminals of all the genes for protein biotinylation (Sawasaki et al., 2008).

5.3. Cell-free protein synthesis

In vitro transcription and cell-free protein synthesis were performed as described (Sawasaki et al., 2005). Transcript was made from each of the DNA templates mentioned above using the SP6 RNA polymerase. The synthetic mRNAs were then precipitated with ethanol and collected by centrifugation using a Hitachi R10H rotor. Each mRNA (usually 30–35 μg) was washed and transferred into a translation mixture. The translation reaction was performed in the bilayer mode (Sawasaki et al., 2002a) with slight modifications. The translation mixture that formed the bottom layer consisted of 60 A260 units of the wheat germ extract (CellFree Sciences, Yokohama, Japan) and 2 μg creatine kinase (Roche Diagnostics K.K., Tokyo, Japan) in 25 μl of SUB-AMIX® (CellFree Sciences). The SUB-AMIX® contained (final concentrations) 30 mM Hepes/KOH at pH 8.0, 1.2 mM ATP, 0.25 mM GTP, 16 mM creatine phosphate, 4 mM DTT, 0.4 mM spermidine, 0.3 mM each of the 20 amino acids, 2.7 mM magnesium acetate, and 100 mM potassium acetate. SUB-AMIX® (125 μl) was placed on the top of the translation mixture, forming the upper layer. After incubation at 16 °C for 18 h, the synthesized proteins were confirmed by SDS–PAGE. For biotin labeling, 1 μl of crude biotin ligase (BirA) produced by the wheat cell-free expression system was added to the bottom layer, and 0.5 μM (final concentration) of d-biotin (Nacalai Tesque, Inc., Kyoto, Japan) was added to both upper and bottom layers, as described previously (Sawasaki et al., 2008).

5.4. Analysis of autophosphorylation activity by luminescence method

In vitro autophosphorylation assays were carried out in a total volume of 15 μl consisting of 50 mM Tris–HCl (pH 7.6), 100 mM potassium acetate, 10 mM $MgCl_2$, 1 mM DTT, 66 μM ATP, 2 μl biotinylated or non-biotinylated (as control) PK at 30 °C for 1 h in a 384-well Optiplate (PerkinElmer Life and Analytical Sciences, Boston, MA, USA). In accordance with the AlphaScreen IgG (ProteinA) detection kit (PerkinElmer Life and Analytical Sciences, Boston, MA, USA) instruction manual, 10 μl of detection mixture containing 50 mM Tris–HCl pH 7.6, 100 mM potassium acetate, 10 mM $MgCl_2$, 0.1 mM DTT, 5 μg/ml anti-phospho Ser/Thr antibody (Upstate Biotechnology, Lake Placid, NY, USA), 1 mg/ml BSA, 0.1 μl streptavidin-coated donor beads and 0.1 μl anti-IgG acceptor beads were added to each well of the 384 Optiplate followed by incubation at 23 °C for 1 h. Luminescence was analyzed by the AlphaScreen detection program. All data are the average of two independent experiments, and the background was controlled for each using the relevant non-biotinylated PK. The relative luminescence signal of biotinylated PK was calculated by normalizing each signal against that of non-biotinylated PK.

5.5. Functional characterization

As a further annotation of the 67 high luminescent signal PKs, we used the gene ontology system (Ashburner et al., 2000). Comparison of frequencies with Arabidopsis within the Biological Processes and Cellular Component gene ontology categories was done using the whole *Arabidopsis* genome annotation tool on the TAIR website (http://www.arabidopsis.org/) and DAGViz (Yano et al., 2009).

Acknowledgements

This work was partially supported by the Special Coordination Funds for Promoting Science and Technology by the Ministry of Education, Culture, Sports, Science and Technology, Japan (T.S. and Y.E.). We thank Michael Andy Goren for proofreading this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.phytochem.2011.02.029.

References

Afzal, A.J., Wood, A.J., Lightfoot, D.A., 2008. Plant receptor-like serine threonine kinases: roles in signaling and plant defense. Mol. Plant Microbe Interact. 21, 507–517.

Anil, V.S., Harmon, A.C., Rao, K.S., 2000. Spatio-temporal accumulation and activity of calcium-dependent protein kinases during embryogenesis, seed development, and germination in sandalwood. Plant Physiol. 122, 1035–1043.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat. Genet. 25, 25–29.

Bai, L., Zhang, G., Zhou, Y., Zhang, Z., Wang, W., Du, Y., Wu, Z., Song, C.P., 2009. Plasma membrane-associated proline-rich extensin-like receptor kinase 4, a novel regulator of Ca signalling, is required for abscisic acid responses in *Arabidopsis thaliana*. Plant J. 60, 314–327.

Belkhadir, Y., Chory, J., 2006. Brassinosteroid signaling: a paradigm for steroid hormone signaling from the cell surface. Science 314, 1410–1411.

Benschop, J.J., Mohammed, S., O'Flaherty, M., Heck, A.J., Slijper, M., Menke, F.L., 2007. Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. Mol. Cell Proteomics 6, 1198–1214.

Berardini, T.Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., Mueller, L.A., Yoon, J., Doyle, A., Lander, G., Moseyko, N., Yoo, D., Xu, I., Zoeckler, B., Montoya, M., Miller, N., Weems, D., Rhee, S.Y., 2004. Functional annotation of the *Arabidopsis* genome using controlled vocabularies. Plant Physiol. 135, 745–755.

Bianchi, M.W., Guivarc'h, D., Thomas, M., Woodgett, J.R., Kreis, M., 1994. Arabidopsis homologs of the shaggy and GSK-3 protein kinases: molecular cloning and functional expression in *Escherichia coli*. Mol. Gen. Genet. 242, 337–345.

Binder, B.M., Harper, J.F., Sussman, M.R., 1994. Characterization of an Arabidopsis calmodulin-like domain protein kinase purified from *Escherichia coli* using an affinity sandwich technique. Biochemistry 33, 2033–2041.

Botella, J.R., Arteca, J.M., Somodevilla, M., Arteca, R.N., 1996. Calcium-dependent protein kinase gene expression in response to physical and chemical stimuli in mung bean (*Vigna radiata*). Plant Mol. Biol. 30, 1129–1137.

Bush, D.S., 1995. Calcium regulation in plant cells and its role in signaling. Annu. Rev. Plant Physiol. Plant Mol. Biol. 46, 95–122.

Chaudhuri, S., Seal, A., Gupta, M.D., 1999. Autophosphorylation-dependent activation of a calcium-dependent protein kinase from groundnut. Plant Physiol. 120, 859–866.

Chehab, E.W., Patharkar, O.R., Hegeman, A.D., Taybi, T., Cushman, J.C., 2004. Autophosphorylation and subcellular localization dynamics of a salt- and water deficit-induced calcium-dependent protein kinase from ice plant. Plant Physiol. 135, 1430–1446.

Chico, J.M., Raíces, M., Téllez-Iñón, M.T., Ulloa, R.M., 2002. A calcium-dependent protein kinase is systemically induced upon wounding in tomato plants. Plant Physiol. 128, 256–270.

Clark, S.E., 2001. Cell signalling at the shoot meristem. Nat. Rev. Mol. Cell Biol. 2, 276–284.

Cross, D.A., Alessi, D.R., Cohen, P., Andjelkovich, M., Hemmings, B.A., 1995. Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. Nature 378, 785–789.

Endo, Y., Sawasaki, T., 2004. High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system. J. Struct. Funct. Genomics 5, 45–57.

Endo, Y., Sawasaki, T., 2006. Cell-free expression systems for eukaryotic protein production. Curr. Opin. Biotechnol. 17, 373–380.

Fan, Y.X., Wong, L., Deb, T.B., Johnson, G.R., 2004. Ligand regulates epidermal growth factor receptor kinase specificity: activation increases preference for GAB1 and SHC versus autophosphorylation sites. J. Biol. Chem. 279, 38143–38150.

Frylinck, L., Dubery, I.A., 1998. Protein kinase activities in ripening mango *Mangifera indica* L., fruit tissue. III. Purification and characterisation of a calcium-regulated protein kinase. Biochim. Biophys. Acta 1387, 342–354.

Gribskov, M., Fana, F., Harper, J., Hope, D.A., Harmon, A.C., Smith, D.W., Tax, F.E., Zhang, G., 2001. PlantsP: a functional genomics database for plant phosphorylation. Nucleic Acids Res. 29, 111–113.

Harmon, A.C., Putnam-Evans, C., Cormier, M.J., 1987. A calcium-dependent but calmodulin-independent protein kinase from soybean. Plant Physiol. 83, 830–837.

Haystead, T.A., Dent, P., Wu, J., Haystead, C.M., Sturgill, T.W., 1992. Ordered phosphorylation of p42mapk by MAP kinase. FEBS Lett. 306, 17–22.

He, J.X., Gendron, J.M., Yang, Y., Li, J., Wang, Z.Y., 2002. The GSK3-like kinase BIN2 phosphorylates and destabilizes BZR1, a positive regulator of the brassinosteroid signaling pathway in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 99, 10185–10190.

Hegeman, A.D., Rodriguez, M., Han, B.W., Uno, Y., Phillips Jr., G.N., Hrabak, E.M., Cushman, J.C., Harper, J.F., Harmon, A.C., Sussman, M.R., 2006. A phyloproteomic

characterization of *in vitro* autophosphorylation in calcium-dependent protein kinases. Proteomics 6, 3649–3664.
Hematy, K., Hofte, H., 2008. Novel receptor kinases involved in growth regulation. Curr. Opin. Plant Biol. 11, 321–328.
Huse, M., Kuriyan, J., 2002. The conformational plasticity of protein kinases. Cell 109, 275–282.
Ikeda, Y., Koizumi, N., Kusano, T., Sano, H., 2000. Specific binding of a 14–3-3 protein to autophosphorylated WPK4, an SNF1-related wheat protein kinase, and to WPK4-phosphorylated nitrate reductase. J. Biol. Chem. 275, 31695–31700.
Jonak, C., Heberle-Bors, E., Hirt, H., 1995. Inflorescence-specific expression of AtK-1, a novel Arabidopsis thaliana homologue of shaggy/glycogen synthase kinase-3. Plant Mol. Biol. 27, 217–221.
Kanchiswamy, C.N., Takahashi, H., Quadro, S., Maffei, M.E., Bossi, S., Bertea, C., Zebelo, S.A., Muroi, A., Ishihama, N., Yoshioka, H., Boland, W., Takabayashi, J., Endo, Y., Sawasaki, T., Arimura, G., 2010. Regulation of *Arabidopsis* defense responses against *Spodoptera littoralis* by CPK-mediated calcium signaling. BMC Plant Biol. 10, 97.
Kim, T.W., Guan, S., Sun, Y., Deng, Z., Tang, W., Shang, J.X., Sun, Y., Burlingame, A.L., Wang, Z.Y., 2009. Brassinosteroid signal transduction from cell-surface receptor kinases to nuclear transcription factors. Nat. Cell Biol. 11, 1254–1260.
Knight, H., Knight, M.R., 2001. Abiotic stress signaling pathways: specificity and cross-talk. Trends Plant Sci. 6, 262–267.
Kurland, C.G., 1982. Translational accuracy *in vitro*. Cell 28, 201–202.
Li, J., Nam, K.H., 2002. Regulation of brassinosteroid signaling by a GSK3/SHAGGY-like kinase. Science 295, 1299–1301.
Li, J., Wen, J., Lease, K.A., Doke, J.T., Tax, F.E., Walker, J.C., 2002. BAK1, an *Arabidopsis* LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. Cell 110, 213–222.
Madin, K., Sawasaki, T., Ogasawara, T., Endo, Y., 2000. A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. Proc. Natl. Acad. Sci. USA 97, 559–564.
Matsuoka, K., Komori, H., Nose, M., Endo, Y., Sawasaki, T., 2010. Simple screening method for autoantigen proteins using the N-terminal biotinylated protein library produced by wheat cell-free synthesis. J. Proteome Res. 9, 4264–4273.
Merkle, D., Douglas, P., Moorhead, G.B., Leonenko, Z., Yu, Y., Cramb, D., Bazett-Jones, D.P., Lees-Miller, S.P., 2002. The DNA-dependent protein kinase interacts with DNA to form a protein–DNA complex that is disrupted by phosphorylation. Biochemistry 41, 12706–12714.
Monroy, A.F., Dhindsa, R.S., 1995. Low-temperature signal transduction: induction of cold acclimation-specific genes of alfalfa by calcium at 25 °C. Plant Cell 7, 321–333.
Nam, K.H., Li, J., 2002. BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. Cell 110, 203–212.
Nayler, O., Schnorrer, F., Stamm, S., Ullrich, A., 1998. The cellular localization of the murine serine/arginine-rich protein kinase CLK2 is regulated by serine 141 autophosphorylation. J. Biol. Chem. 273, 34341–34348.
Netzer, W.J., Hartl, F.U., 1997. Recombination of protein domains facilitated by co-translational folding in eukaryotes. Nature 388, 343–349.
Nishihama, R., Soyano, T., Ishikawa, M., Araki, S., Tanaka, H., Asada, T., Irie, K., Ito, M., Terada, M., Banno, H., Yamazaki, Y., Machida, Y., 2002. Expansion of the cell plate in plant cytokinesis requires a kinesin-like protein/MAPKKK complex. Cell 109, 87–99.
Nozawa, A., Matsubara, Y., Tanaka, Y., Takahashi, H., Akagi, T., Seki, M., Shinozaki, K., Endo, Y., Sawasaki, T., 2009. Construction of a protein library of Arabidopsis transcription factors using a wheat cell-free protein production system and its application for DNA binding analysis. Biosci. Biotechnol. Biochem. 73, 1661–1664.
Ogasawara, T., Sawasaki, T., Morishita, R., Ozawa, A., Madin, K., Endo, Y., 1999. A new class of enzyme acting on damaged ribosomes: ribosomal RNA apurinic site specific lyase found in wheat germ. EMBO J. 18, 6522–6531.
Oh, M.H., Wang, X., Kota, U., Goshe, M.B., Clouse, S.D., Huber, S.C., 2009. Tyrosine phosphorylation of the BRI1 receptor kinase emerges as a component of brassinosteroid signaling in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 106, 658–663.
Parniske, M., 2008. Arbuscular mycorrhiza: the mother of plant root endosymbioses. Nat. Rev. Microbiol. 6, 763–775.
Patharkar, O.R., Cushman, J.C., 2000. A stress-induced calcium-dependent protein kinase from *Mesembryanthemum crystallinum* phosphorylates a two-component pseudo-response regulator. Plant Cell 24, 679–691.
Pavlov, M.Y., Ehrenberg, M., 1996. Rate of translation of natural mRNAs in an optimized *in vitro* system. Arch. Biochem. Biophys. 328, 9–16.
Pimienta, G., Pascual, J., 2007. Canonical and alternative MAPK signaling. Cell Cycle 6, 2628–2632.
Roberts, D.M., Harmon, A.C., 1992. Calcium modulated proteins: targets of intracellular calcium signals in higher plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 43, 375–414.
Roux, P.P., Blenis, J., 2004. ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions. Microbiol. Mol. Biol. Rev. 68, 320–344.
Ryo, A., Tsurutani, N., Ohba, K., Kimura, R., Komano, J., Nishi, M., Soeda, H., Hattori, S., Perrem, K., Yamamoto, M., Chiba, J., Mimaya, J., Yoshimura, K., Matsushita, S., Honda, Yoshimura, A., Sawasaki, T., Aoki, I., Morikawa, Y., Yamamoto, N., 2008. SOCS1 is an inducible host factor during HIV-1 infection and regulates the intracellular trafficking and stability of HIV-1 Gag. Proc. Natl. Acad. Sci. USA 105, 294–299.
Saijo, Y., Hata, S., Kyozuka, J., Shimanoto, K., Izui, K., 2000. Over-expression of a single Ca^{2+}-dependent protein kinase confers both cold and salt/drought tolerance on rice plants. Plant J. 23, 319–327.
Sawasaki, T., Ogasawara, T., Morishita, R., Endo, Y., 2002a. A cell-free protein synthesis system for high-throughput proteomics. Proc. Natl. Acad. Sci. USA 99, 14652–14657.
Sawasaki, T., Hasegawa, Y., Tsuchimochi, M., Kamura, N., Ogasawara, T., Endo, Y., 2002b. A bilayer cell-free protein synthesis system for high-throughput screening of gene products. FEBS Lett. 514, 102–105.
Sawasaki, T., Hasegawa, Y., Morishita, R., Seki, M., Shinozaki, K., Endo, Y., 2004. Genome-scale, biochemical annotation method based on the wheat germ cell-free protein synthesis system. Phytochemistry 65, 1549–1555.
Sawasaki, T., Gouda, M.D., Kawasaki, T., Tsuboi, T., Tozawa, Y., Takai, K., Endo, Y., 2005. The wheat germ cell-free expression system: methods for high-throughput materialization of genetic information. Methods Mol. Biol. 310, 131–144.
Sawasaki, T., Morishita, R., Gouda, M.D., Endo, Y., 2007. Methods for high-throughput materialization of genetic information based on wheat germ cell-free expression system. Methods Mol. Biol. 375, 95–106.
Sawasaki, T., Kamura, N., Matsunaga, S., Saeki, M., Tsuchimochi, M., Morishita, R., Endo, Y., 2008. *Arabidopsis* HY5 protein functions as a DNA-binding tag for purification and functional immobilization of proteins on agarose/DNA microplate. FEBS Lett. 582, 221–228.
Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T., Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M., Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Shinozaki, K., 2002. Functional annotation of a full-length *Arabidopsis* cDNA collection. Science 296, 141–145.
Shiu, S.H., Bleecker, A.B., 2001. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. Proc. Natl. Acad. Sci. USA 99, 10763–10768.
Silva, N.F., Goring, D.R., 2002. The proline-rich, extensin-like receptor kinase-1 (PERK1) gene is rapidly induced by wounding. Plant Mol. Biol. 50, 667–685.
Stahl, Y., Simon, R., 2005. Plant stem cell niches. Int. J. Dev. Biol. 49, 479–489.
Sugiyama, N., Nakagami, H., Mochida, K., Daudi, A., Tomita, M., Shirasu, K., Ishihama, Y., 2008. Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in Arabidopsis. Mol. Syst. Biol. 4, 193.
Tadokoro, D., Shimizu, S., Takahama, S.K., Hayashi, S., Endo, y., Sawasaki, T., 2010. Characterization of a caspase-3-substrate kinome using an N- and C-terminally tagged protein kinase library produced by a cell-free system. Cell Death Dis. 1, e89.
Takahashi, H., Nozawa, A., Seki, M., Shinozaki, K., Endo, Y., Sawasaki, T., 2009. A simple and high-sensitivity method for analysis of ubiquitination and polyubiquitination based on wheat cell-free protein synthesis. BMC Plant Biol. 9, 39.
Takayama, S., Isogai, A., 2005. Self-incompatibility in plants. Annu. Rev. Plant Biol. 56, 467–489.
Trewavas, A., 1999. Le calcium, C'est la vie: calcium makes waves. Plant Physiol. 120, 1–6.
Urao, T., Katagiri, T., Mizoguchi, T., Yamaguchi-Shinozaki, K., Hayashida, N., Shinozaki, K., 1994. Two genes that encode Ca^{2+}-dependent protein kinases are induced by drought and high-salt stresses in *Arabidopsis thaliana*. Mol. Gen. Genet. 244, 331–340.
van Bentem, S., Anrather, D., Dohnal, I., Roitinger, E., Csaszar, E., Joore, J., Buijnink, J., Carreri, A., Forzani, C., Lorkovic, Z.J., Barta, A., Lecourieux, D., Verhounig, A., Jonak, C., Hirt, H., 2008. Site-specific phosphorylation profiling of Arabidopsis proteins by mass spectrometry and peptide chip analysis. J. Proteome Res. 7, 2458–2470.
Wang, Q.M., Fiol, C.J., DePaoli-Roach, A.A., Roach, P.J., 1994. Glycogen synthase kinase-3 beta is a dual specificity kinase differentially regulated by tyrosine and serine/threonine phosphorylation. J. Biol. Chem. 269, 14566–14574.
Wang, X., Li, X., Meisenhelder, J., Hunter, T., Yoshida, S., Asami, T., Chory, J., 2005. Autoregulation and homodimerization are involved in the activation of the plant steroid receptor BRI1. Dev. Cell 8, 855–865.
Wu, J., Rossomando, A.J., Her, J.H., Del Vecchio, R., Weber, M.J., Sturgill, T.W., 1991. Autophosphorylation *in vitro* of recombinant 42-kilodalton mitogen-activated protein kinase on tyrosine. Proc. Natl. Acad. Sci. USA 88, 9508–9512.
Xu, J., Tian, Y.S., Peng, R.H., Xiong, A.S., Zhu, B., Jin, X.F., Gao, F., Fu, X.Y., Hou, X.L., Yao, Q.H., 2010. AtCPK6, a functionally redundant and positive regulator involved in salt/drought stress tolerance in Arabidopsis. Planta 231, 1251–1260.
Yano, K., Aoki, K., Suzuki, H., Shibata, D., 2009. DAGViz: a directed acyclic graph browser that supports analysis of Gene Ontology annotation. Plant Biotechnol. 26, 9–13.
Yoon, G.M., Cho, H.S., Ha, H.J., Liu, J.R., Lee, H.P., 1999. Characterization of NtCDPK1, a calcium-dependent protein kinase gene in *Nicotiana tabacum*, and the activity of its encoded protein. Plant Mol. Biol. 39, 991–1001.

Parasitology International 60 (2011) 313–320

Contents lists available at ScienceDirect

Parasitology International

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

PEXEL-independent trafficking of *Plasmodium falciparum* SURFIN$_{4.2}$ to the parasite-infected red blood cell and Maurer's clefts

Jean Semé Fils Alexandre [a,b], Kazuhide Yahata [a], Satoru Kawai [c], Motomi Torii [d], Osamu Kaneko [a,*]

[a] Department of Protozoology, Institute of Tropical Medicine (NEKKEN) and the Global Center of Excellence Program, Nagasaki University, Sakamoto, Nagasaki 852-8523, Japan
[b] Centro Nacional de Control de Enfermedades Tropicales, Santo Domingo, Dominican Republic
[c] Center for Tropical Medicine and Parasitology, Dokkyo Medical University, Tochigi 321-0293, Japan
[d] Department of Molecular Parasitology, Ehime University Graduate School of Medicine, Shitsukawa, Toon, Ehime 791-0925, Japan

ARTICLE INFO

Article history:
Received 22 April 2011
Received in revised form 9 May 2011
Accepted 10 May 2011
Available online 17 May 2011

Keywords:
Malaria
Maurer's clefts
Plasmodium falciparum
Protein trafficking
SURFIN

ABSTRACT

SURFIN$_{4.2}$ is a parasite-infected red blood cell (iRBC) surface associated protein of *Plasmodium falciparum*. To analyze the region responsible for the intracellular trafficking of SURFIN$_{4.2}$ to the iRBC and Maurer's clefts, a panel of transgenic parasite lines expressing recombinant SURFIN$_{4.2}$ fused with green fluorescent protein was generated and evaluated for their localization. We found that the cytoplasmic region containing a tryptophan rich (WR) domain is not necessary for trafficking, whereas the transmembrane (TM) region was. Two PEXEL-like sequences were shown not to be responsible for the trafficking of SURFIN$_{4.2}$, demonstrating that the protein is trafficked in a PEXEL-independent manner. N-terminal replacement, deletion of the cysteine-rich domain or the variable region also did not prevent the protein from localizing at the iRBC or Maurer's clefts. A recombinant SURFIN$_{4.2}$ protein possessing 50 amino acids upstream of the TM region, TM region itself and a part of the cytoplasmic region was shown to be trafficked into the iRBC and Maurer's clefts, suggesting that there are no essential trafficking motifs in the SURFIN$_{4.2}$ extracellular region. A mini-SURFIN$_{4.2}$ protein containing WR domain was shown by Western blotting to be more abundantly detected in a Triton X-100-insoluble fraction, compared to the one without WR domain. We suggest that the cytoplasmic region containing the WR may be responsible for their difference in solubility.

1. Introduction

During its asexual replication in the human host, *Plasmodium falciparum*, the apicomplexan parasite responsible for malaria, dramatically remodels the infected red blood cell (iRBC) [1]. This process involves the generation of a parasitophorous vacuole (PV) in which parasites reside and replicate, the transportation of parasite proteins into the iRBC across the PV membrane (PVM), the generation of parasite-derived membranous structures in the cytoplasm of the host RBC called Maurer's clefts that play a major role as protein-sorting points, and the formation of knobs on the iRBC surface [2–4]. Some of the most severe malaria pathologies caused by *P. falciparum*, such as cerebral and placental malaria, are specifically linked to the adherence of the iRBCs to capillary vessels (cytoadhesion) and to uninfected RBCs (rosseting). *P. falciparum* erythrocyte membrane protein 1 (*Pf*EMP1), a parasite protein transported to the surface of the iRBC has previously been shown to mediate these phenomena [5–7]. Understanding the molecular mechanisms and pathways by which parasite-proteins such as *Pf*EMP1, are trafficked to the cytosol and thence to the surface of the iRBC is, therefore, critical for a clear insight into the pathogenesis of *P. falciparum* malaria.

RBCs lack a protein secretory apparatus so the parasite must establish de novo secretion machinery within the host cell cytoplasm in order to transport its own proteins into the iRBC across the PVM. The mechanisms that enable such trafficking are incompletely understood; however, many *P. falciparum* proteins destined for export into the iRBC contain both an N-terminal hydrophobic signal sequence and a short conserved pentameric host cell-targeting motif (RxLxE/Q/D) termed the *Plasmodium* export element (PEXEL) or the vacuolar transport signal (VTS) [8,9]. The N-terminal signal sequence is required for the proteins to enter the constitutive secretory pathway *via* the endoplasmic reticulum (ER) [10,11], where the PEXEL/VTS motif is cleaved by plasmepsin V, an ER residing aspartic protease [12,13], and the newly formed N-terminus (xE/Q/D) allows translocation into the iRBC cytosol by a PVM residing translocon called the "*Plasmodium* translocon of exported proteins" (PTEX) complex [14]. Our understanding of the mechanisms behind the transport of proteins within the iRBC remains vague, but some tentative explanations

Abbreviations: aa, amino acid(s); CRD, cysteine-rich domain; ER, endoplasmic reticulum; GFP, green fluorescence protein; IFA, indirect immunofluorescence assay; iRBC, infected red blood cell; PBS, phosphate buffered saline; PEXEL, *Plasmodium* export element; PNEP, PEXEL negative exported protein; PVM, parasitophorous vacuole membrane; TM, transmembrane; Tx, Triton-X 100; Var, variable region; WR, tryptophan-rich.

* Corresponding author. Tel.: +81 95 819 7838; fax: +81 95 819 7805.
E-mail address: okaneko@nagasaki-u.ac.jp (O. Kaneko).

1383-5769/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.parint.2011.05.003

have been raised; transport may be mediated through vesicles, through complex membrane networks, non-lipid enclosed protein aggregates, or lipid enclosed structures such as Maurer's clefts [4,15].

A few hundred proteins are known to contain a PEXEL/VTS motif, defining a large *Plasmodium* exportome [9,16]. However, several parasite proteins that are transported to the iRBC lack both an N-terminal signal sequence and a PEXEL/VTS motif, and are termed "PEXEL negative exported proteins" (PNEPs) [17]. The most well characterized PNEPs are skeleton-binding protein 1 (SBP1) [18], the membrane associated histidine-rich protein 1 (MAHRP1) [19] and the ring-exported proteins 1 (REX1) and 2 (REX2) [17,20,21]. Most of the reported PNEPs are believed to be trafficked to the iRBC *via* the classical secretory pathway, involving initial transport to the ER, but no shared signal or transport-related sequence has been identified to date for subsequent transport to the iRBC. SBP1, MAHRP1 and REX2 lack a signal peptide but contain a transmembrane (TM) region which, along with sequences at their N-terminal region, has been implicated in protein transport [22–24]; however, the hydrophobic N-terminal region of the REX1 protein has been shown to be the only region required for transport of this particular protein [25].

The recently identified *P. falciparum* surface-associated interspersed gene (*surf*) family encodes high molecular mass proteins, and one of these, $SURFIN_{4.2}$, has been shown to be co-transported along with *Pf*EMP1 and RIFIN to the iRBC surface [26]. The N-terminal of $SURFIN_{4.2}$, predicted to be extracellular, contains a moderately conserved cysteine-rich putative globular domain (CRD) preceding a variable segment (Var) followed by a putative TM region. The C-terminal region of $SURFIN_{4.2}$ contains three tryptophan-rich (WR) domains which are highly conserved among SURFIN protein members, intersected by stretches of higher variability. The protein does not appear to contain either a hydrophobic signal peptide sequence or a classical PEXEL motif. Although there are two PEXEL-like sequences located at the N-terminal segment amino acid positions (aa) 24–30 (R_KI_FE) and aa 118–122 (R_TL_ED), they may not be true signals because the first does not completely agree with the consensus PEXEL motif, and the second was located in the putative globular domain CRD. This suggests that $SURFIN_{4.2}$ may be transported *via* a PEXEL-independent pathway. Using a serial deletion approach, we have attempted to identify the regions involved in *P. falciparum* $SURFIN_{4.2}$ transport into the iRBC. We generated a panel of transgenic parasite lines expressing green fluorescent protein (GFP)-tagged recombinant $SURFIN_{4.2}$ and show that the protein is trafficked as a PNEP. We show that the TM region, the only predicted hydrophobic region of the protein, is necessary for entry into the parasite's secretory pathway and for subsequent trafficking into the iRBC. These findings confirm the importance of hydrophobic regions for the trafficking of PNEPs.

2. Materials and methods

2.1. Plasmid construction

A panel of plasmids that were used to make final *P. falciparum* transfection constructs were prepared based on the Multisite Gateway® system (Invitrogen, Carlsbad, CA) [27]. DNA fragments containing attB1 and attB2 sites were inserted into pUC19, resulting the pB12 plasmid. DNA fragments encoding aa 1–419 of $SURFIN_{4.2}$ (CRD and a part of Var), a triple HA tag, and aa 734–764 of $SURFIN_{4.2}$ (TM) were ligated into pB12 to make the pB12-$SURF_{4.2}$CRD-Var1-HA-TM plasmid. DNA fragments encoding aa 765–1347, 1320–1728 or 1712–2380 of $SURFIN_{4.2}$ (first, second or third WR domain) were ligated into pB12-$SURF_{4.2}$CRD-Var1-HA-TM, resulting in the pB12-$SURF_{4.2}$CRD-Var1-HA-TM-WR1, pB12-$SURF_{4.2}$CRD-Var1-HA-TM-WR2 or pB12-$SURF_{4.2}$CRD-Var1-HA-TM-WR3, respectively. DNA fragments encoding 50 amino acids at aa 684–734 adjacent to TM region, a triple HA tag, TM, and WR1 regions were ligated into pB12, resulting in the pB12-$SURF_{4.2}$VarC-HA-TM-WR1. All DNA fragments were amplified from *P. falciparum* 3D7 line parasites using KOD Plus DNA polymerase (Toyobo). pB12-$SURF_{4.2}$CRD-Var1-TM was further modified by site-directed mutagenesis using oligonucleotides with desired modifications as follows: To abolish the PEXEL-like sequences at aa 25–29 ($R_{25K}I_{27F}E_{29}$) or aa 118–122 ($R_{118T}L_{120E}D_{122}$), these sequences were replaced to $A_{25K}A_{27F}A_{29}$ or $A_{118T}A_{120E}A_{122}$, respectively, yielding pB12-$SURF_{4.2}$CRD-Var1-HA-TM-Pexel-1mut or pB12-$SURF_{4.2}$CRD-Var1-HA-TM-Pexel-2mut. To assess the N-terminal sequence for the trafficking, the N-terminal region at aa 1–42 (M_1LFVVELDSRLEKSADKRISVERFRKIFEIYVEDKLEELKRS$_{42}$) of $SURFIN_{4.2}$ was replaced with the N-terminal region at aa 1–15 (M_1DVHVNQLKNISPID$_{15}$) of *P. falciparum* adenylosuccinate lyase (ASL, PFB0295w), a *P. falciparum* enzyme not considered to be transported to the iRBC, to yield pB12-$SURF_{4.2}$CRD-Var1-HA-TM-RepN. pB12-$SURF_{4.2}$-Var1-HA-TM and pB12-$SURF_{4.2}$CRD-HA-TM were generated from pB12-$SURF_{4.2}$CRD-Var1-TM by removing a region encoding aa 46–196 containing the CRD and a region encoding aa 198–419 containing the Var1 region, respectively. These pB12-based plasmids were subjected to a BP recombination reaction with pDONR™221 (Invitrogen) according to the manufacturer's instructions, resulting in the corresponding pENT12 Gateway Entry vectors. These pENT12-based plasmids were then subjected to a Gateway MultiSite LR recombination reaction with other Entry vectors, *Pf*CRT 5′-pENTR4/1 (as a promoter component) and GFPm2-pENTR2/3 (as a tag sequence), and a Destination vector, pCHDR-3/4 (kind gifts from D. G McFadden), according to the manufacturer's instruction [27]. Initially, we used the promoter region of $SURFIN_{4.2}$, however, the signal was very weak, thus we decided to use CRT (chloroquine resistance transporter; MAL7P1.27) promoter, which has been used to study *P. falciparum* protein trafficking to the iRBC cytosol and Maurer's clefts, to overexpress recombinant proteins for the visualization of their clear location. Previous transcriptome data indicated that the promoter activity of CRT was stronger than that of $SURFIN_{4.2}$ and that CRT was mainly transcribed at schizont, ring, and early trophozoite stages, slightly longer than $SURFIN_{4.2}$ which was mainly transcribed at schizont and early ring stages. [28,29]. Obtained plasmids were verified by their restriction enzyme digestion pattern and sequencing. Schematic structures of the recombinant proteins expressed from the episomal form of the plasmids in the transfected *P. falciparum* are shown in Fig. 1. Detailed information may be found in Supplementary material.

2.2. Parasite culture and transfection

The *P. falciparum* MS822 line was used in this study. This line was isolated in Mae Sot, Thailand in 1988, maintained in vitro for less than 3 months, and kept at the Institute of Tropical Medicine, Nagasaki University [30]. Parasites were cultured in RPMI-1640 medium containing 5% heat-inactivated pooled type AB human serum and 0.25% Albumax II (Invitrogen), 200 mM hypoxanthine (SIGMA, St. Louis, MO), 10 μg/mL gentamycin (Invitrogen) and human RBC (type O) at 2% hematocrit. Human RBCs and plasma were obtained from the Nagasaki Red Cross Blood Center. Serum was produced from the acid–citrate–dextrose-containing plasma by removing the clot that had formed after adding calcium. *P. falciparum* transfection was performed essentially as previously described [31]. Briefly, RBCs were resuspended in 400 μL of incomplete Cytomix (120 mM KCl, 0.15 mM $CaCl_2$, 2 mM EGTA, 5 mM $MgCl_2$, 10 mM K_2HPO_4/KH_2PO_4, and 25 mM Hepes) containing 100 μg of plasmid DNA. Electroporations were performed in 2 mm cuvette using a Gene Pulser Xcell Electroporation System (Bio-Rad, Hercules, CA) with a condition of 320 mΩ, 0.32 kV and 975 μF. Observed time constants were 15–30 ms. Parasites were synchronized to ring stage by 5% sorbitol treatment, then 40 h later, mature trophozoites/schizonts-iRBCs were resuspended with the plasmids-preloaded RBCs (final 0.2% parasitemia). At day 3 post transfection, 5 nM of the anti-folate drug WR99210 (kind gift from Dr. D. Jacobus) was supplied to the culture medium, and was maintained until drug-resistant parasites reappeared. Resistant parasites

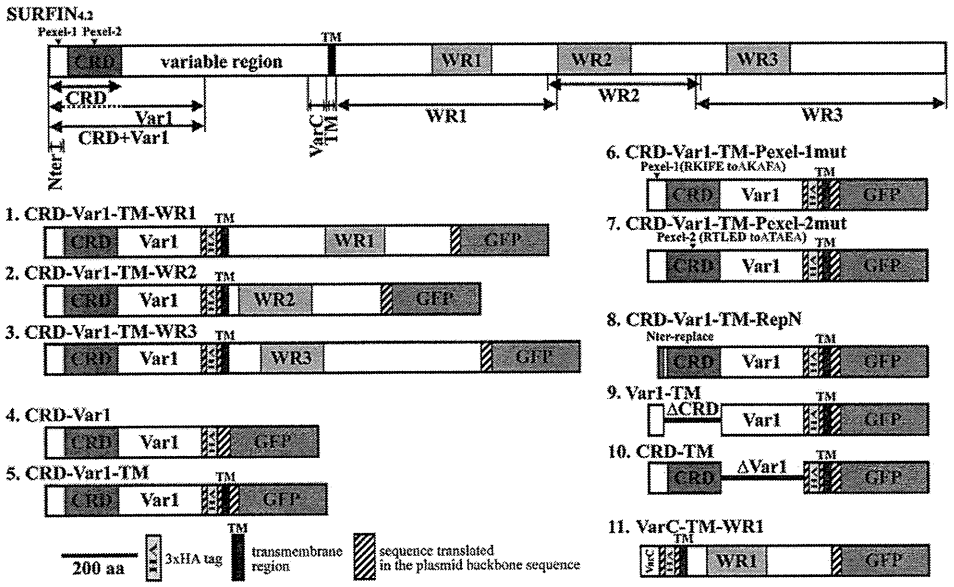

Fig. 1. Schematic structure of the recombinant proteins expressed in the transfected *P. falciparum* lines. CRD, cysteine-rich domain; VarC, C-terminal of the variable region; HA, triple hemagglutinin-tag; Nter, N-terminal; TM, transmembrane; Var1, variable region 1; and WR, tryptophan-rich. ΔCRD or ΔVar1 indicates that CRD or Var1 was deleted from the protein.

were usually detected before the 30th day of culture in the presence of the drug and were subsequently maintained in culture containing 25 nM WR99210.

2.3. Fluorescence live imaging and indirect immunofluorescence assay (IFA)

For fluorescence live imaging, 10 μL of parasite culture were incubated with 1 μg/mL of Hoechst 33342 (Moleculer Probe) for 30 min at 37 °C and placed on a glass slide for observation. Parasites expressing GFPm2 were visualized using a fluorescence microscope (Eclipse 80i; Nikon, Japan) and a digital camera (VB-7010; Keyence, Japan) equipped with 100×oil immersion lens. For IFA, thin smears of *P. falciparum* iRBCs were prepared on glass slides, fixed with 4% paraformaldehyde/0.075% glutaraldehyde in PBS at room temperature for 5 min, rinsed with 50 mM glycine in PBS, and blocked with PBS containing 3% BSA (SIGMA) for 30 min [32]. For single staining, the smears were reacted with rabbit anti-GFP polyclonal antibody (ab6556; Abcam, Cambridge, UK), followed by Alexa-Fluor 488-conjugated secondary anti-rabbit IgG antibody (Invitrogen). For double staining, rabbit anti-GFP antibody and rabbit anti-SBP1 antibody (a kind gift from Dr. T. Tsuboi) were labeled with Alexa-Fluor 488 and-594, respectively, using Zenon® Rabbit IgG labeling kit (Invitrogen). The smears were then incubated with rabbit anti-GFP antibody (1:500) and rabbit anti-SBP1 antiserum (1:1000) in PBS containing 3% BSA for 1 h at 37 °C. Parasite nuclear staining was carried out by adding 4′, 6-diamidino-2-phenylindole (DAPI; Invitrogen; final 1 μg/mL). Stained parasites were mounted with ProLong® Gold antifade reagent (Invitrogen) and visualized as described above. Some images were analyzed by using ImageJ software (1.44p; http://rsbweb.nih.gov/ij/).

2.4. Extraction of parasite proteins and Western blot analysis

Mature trophozoite and schizonts-iRBCs were collected by centrifugation on a 40/70% Percoll–sorbitol gradient. The enriched parasite fractions ($2–4\times10^8$ parasites) were subjected to protein extraction during which process the water-soluble fraction was collected following a freeze–thaw procedure in PBS containing a mixture of protease inhibitors (PI; cOmplete; Roche, Basel, Switzerland) repeated three times. The pellets were washed twice with PBS-PI, and proteins further extracted in PBS-PI containing 1% Triton X-100 (Tx; Calbiochem) for 30 min on ice. The insoluble materials were washed twice with PBS-PI-Tx, then proteins were further extracted by incubation with PBS-PI containing 2% SDS (Nacalai, Japan) for 30 min at room temperature.

Parasite extracts were subjected to electrophoresis on 5–20% SDS-polyacrylamide gradient mini gels (ATTO, Japan) under reducing conditions. The protein bands were transferred from gels to PVDF membranes (Millipore, Billerica, MA). The membranes were then probed with rabbit anti-GFP polyclonal antibody (1:500; ab6556; Abcam), for 1 h at room temperature followed by a secondary incubation with the horseradish peroxidase-conjugated goat anti-rabbit IgG antibodies (Promega) at a concentration of 1:25,000. Purified mouse anti-glycophorin A antibody (CD235a; BD Pharmingen) was used to detect glycophorin A as a positive control. Bands were visualized with Immobilon™ Western Chemiluminescent HRP Substrate (Millipore) and detected using a chemiluminescence detection system (LAS-4000EPUVmini; Fujifilm, Japan). The relative molecular sizes of the proteins were calculated based in reference to the molecular size standards (Precision Plus Dual Color Standards; Bio-Rad).

3. Results

3.1. Successful generation of mini-SURFIN$_{4.2}$ proteins

The production of *P. falciparum* transfectants expressing very large proteins such as SURFIN$_{4.2}$ (predicted molecular weight of 286 kDa in 3D7 parasite line) is technically challenging. Therefore, in order to investigate which region is responsible for the trafficking of SURFIN$_{4.2}$ into the iRBCs, and thence Maurer's clefts, we first attempted to generate and evaluate mini-SURFIN$_{4.2}$ proteins containing regions that are conserved among SURFIN members; N-terminal 419 amino acids containing N-terminal segment (aa 1–50) and CRD (aa 51–195), which is conserved among *P. falciparum* SURFIN members, and the N-terminal side of the variable region (Var1; aa 196–419), which is relatively conserved between SURFIN$_{4.2}$ and SURFIN$_{4.1}$ (identity 19.5% and similarity 28.7%), TM, and either one of the WR domains (WR1,

WR2, or WR3). All mini-SURFIN$_{4.2}$ proteins, SURFIN$_{4.2}$CRD-Var1-TM-WR1, -WR2, or -WR3, were tagged with GFP for visualization (Fig. 1). Although the extracellular region was tagged with triple HA, anti-HA mouse monoclonal antibody (4B2; Wako, Japan) could not detect any signals by IFA and Western blot analysis with unknown reason, thus HA tag was not used in this study. Live imaging using a fluorescence microscope produced only very weak fluorescent signals, and so the precise location of the GFPm2-tagged protein was impossible to assess. By double staining IFA, all mini-SURFIN$_{4.2}$ proteins were observed in the iRBC cytoplasm as punctate dots that co-localized with the Maurer's cleft protein SBP1 [18], indicating that all 3 mini-SURFIN$_{4.2}$ proteins were transported to the Maurer's clefts (Fig. 2). In addition to the Maurer's cleft localized signal, diffuse fluorescence was also observed in the parasite, as well as in the iRBC. As all of the mini-SURFIN$_{4.2}$ proteins were able to traverse the PVM and reach the iRBC cytosol and the Maurer's clefts, we selected SURFIN$_{4.2}$CRD-Var1-TM-WR1 for further evaluation.

3.2. The TM region, but not the WR domain is essential for the SURFIN$_{4.2}$ trafficking to the iRBC and Maurer's clefts

In order to evaluate the importance of the TM region and WR domain in SURFIN$_{4.2}$ trafficking, we generated parasites expressing only SURFIN$_{4.2}$CRD-Var1 (TM and WR were removed) or SURFIN$_{4.2}$ CRD-Var1-TM (WR was removed). By live imaging, weak, but detectable GFP signals were observed for these transfectants, and we found that SURFIN$_{4.2}$CRD-Var1 was exclusively located in the parasite cytoplasm, as would be expected for a protein lacking a signal sequence for transport to the ER. We speculate that the protein was located within the parasite cytosol (Fig. 3A). In contrast, SURFIN$_{4.2}$ CRD-Var1-TM was observed in a punctate dot pattern in the iRBC, in addition to diffuse fluorescence in the parasite cytoplasm. By double staining IFA, SURFIN$_{4.2}$CRD-Var1 was observed only in the parasite, confirming the live imaging results, whereas SURFIN$_{4.2}$CRD-Var1-TM colocalized with SBP1, indicating a Maurer's cleft localization (Fig. 3B). Thus, the IFA data indicate that the SURFIN$_{4.2}$ cytoplasmic region containing WR domain is not required for trafficking to the iRBC or Maurer's clefts, but the TM region is essential.

A diffused fluorescence pattern in the iRBC, as observed for the mini-SURFIN$_{4.2}$ proteins, appeared to be reduced for SURFIN$_{4.2}$CRD-Var1-TM with the double staining IFA images. Because the single staining with Alexa-Fluor 488-conjugated secondary antibody gave clearer images than the double staining using the Zenon antibody-labeling kit, we used two representative single staining images to measure and compare the signal intensity of the recombinant proteins in the iRBC cytosol for SURFIN$_{4.2}$CRD-Var1-TM and SURFIN$_{4.2}$CRD-Var1-TM-WR1. After subtracting background signals, signal intensities in the iRBC cytosol for SURFIN$_{4.2}$CRD-Var1-TM were 13 to 22 units (Fig. 4B; CRD-Var1-TM #1 and #2), whereas those for SURFIN$_{4.2}$CRD-Var1-TM-WR were 66 and 69 units (Fig. 4B; CRD-Var1-TM-WR1 #1 and #2, respectively). This indicates that the fluorescence signal in the iRBC cytosol is weaker for SURFIN$_{4.2}$CRD-Var1-TM than SURFIN$_{4.2}$ CRD-Var1-TM-WR1 and suggests that the SURFIN$_{4.2}$CRD-Var1-TM is less abundant in iRBC cytosol than SURFIN$_{4.2}$CRD-Var1-TM-WR1.

In order to evaluate their solubility, parasite proteins were sequentially extracted by a repeated-freeze thaw procedure (FT; water-soluble fraction protein), followed by Tx extraction (Tx; membrane bound protein), and SDS extraction (SDS; Tx-insoluble fraction) and were detected with rabbit anti-GFP antibody. About 105-kDa bands were detected for SURFIN$_{4.2}$CRD-Var1 and SURFIN$_{4.2}$CRD-Var1-TM and a 230-kDa band for SURFIN$_{4.2}$CRD-Var1-TM-WR1 by Western blot. Expected band sizes were 83, 86, and 158 kDa, respectively (Fig. 5A). Although the band sizes detected by Western blot are much larger than the expected size, this is not an uncommon observation for *P. falciparum*-derived proteins which have a deviated amino acid composition due to a highly A/T-rich genome (76.3% in the exon) [33]. In addition to the target protein bands, a ~60-kDa band was observed for all fractions (Fig. 5A), but this band was also observed in the extract from the wild-type non-transfected MS822 parasite, and so was not derived from the recombinant proteins expressed in the transfected parasite lines. Although the identity of this band is unclear, it is likely derived from parasites, because this band was not

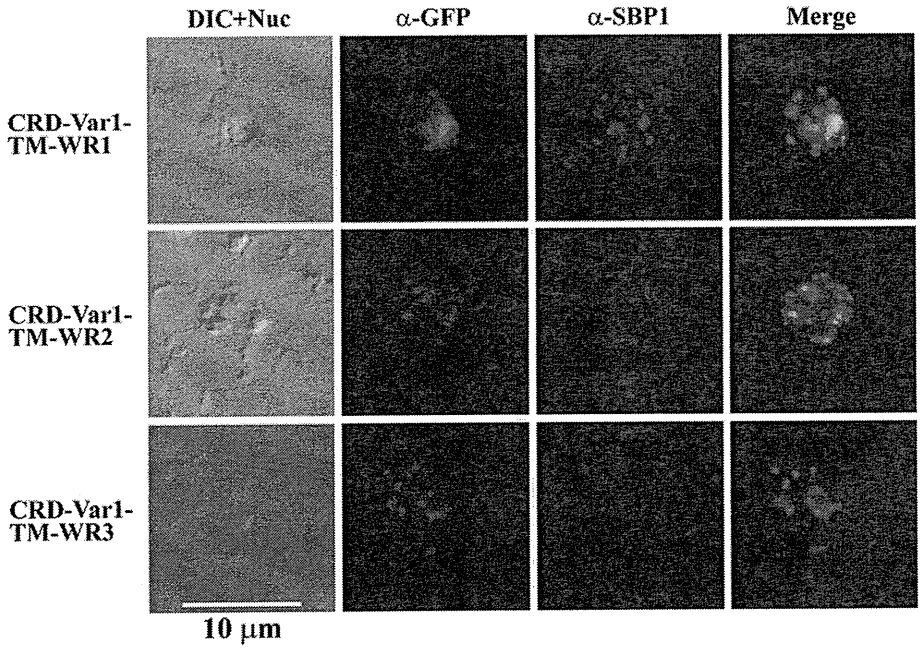

Fig. 2. Indirect immunofluorescence assay of three mini-SURFIN$_{4.2}$ proteins. Double staining IFA for 3 mini-SURFIN$_{4.2}$-expressing transfectants is shown. α-GFP and α-SBP1 indicate anti-GFP (mini-SURFIN$_{4.2}$) and anti-SBP1 (Maurer's cleft protein). Negative controls using normal rabbit antibody did not produce detectable signals (not shown). CRD, cysteine-rich domain; TM, transmembrane; Var1, variable region 1; and WR, tryptophan-rich.

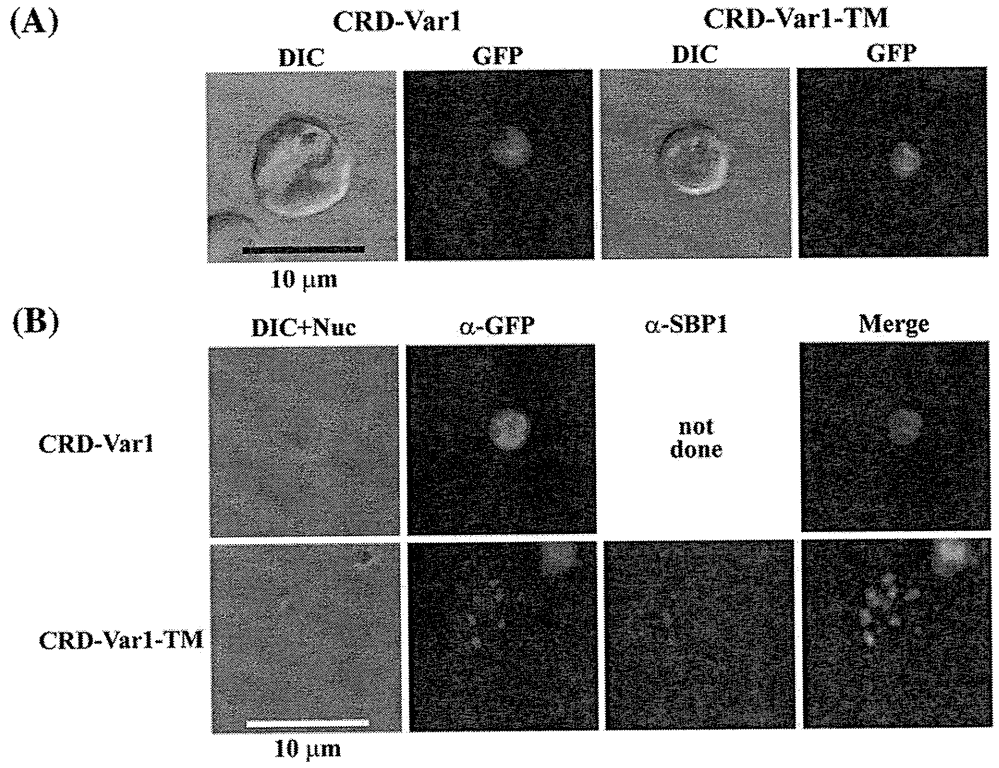

Fig. 3. Indirect immunofluorescence assay of $SURFIN_{4.2}$CRD-Var1 and $SURFIN_{4.2}$CRD-Var1-TM proteins. (A) Live imaging of GFP-expressing parasites. (B) Single staining for $SURFIN_{4.2}$CRD-Var1 and double staining for $SURFIN_{4.2}$CRD-Var1-TM with anti-GFP (α-GFP, green) and anti-SBP1 (α-SBP1, red). Nuclei were stained with DAPI. Negative control using normal rabbit antibody did not produce detectable signals (not shown). CRD, cysteine-rich domain; TM, transmembrane; and Var1, variable region 1.

observed in the extract from the parasite-uninfected RBC (Fig. 5B). Positive glycophorin A bands for the extracts from both the wild-type parasite and the uninfected RBC indicated that the protein extraction from the uninfected RBC was successful. It should be noted that the rabbit anti-GFP antibodies did not show any signal when wild-type parasites were subjected to IFA. We found that $SURFIN_{4.2}$CRD-Var1 was exclusively detected in the soluble FT fraction, indicating that this protein was in soluble form, which is consistent with the observation of its localization in the parasite's cytoplasm (Fig. 3). $SURFIN_{4.2}$CRD-Var1-TM was detected in the Tx-soluble fraction more abundantly than in the Tx-insoluble SDS fraction. Conversely, $SURFIN_{4.2}$CRD-Var1-TM-WR1 was detected in the SDS fraction more abundantly than in the Tx fraction. Thus both proteins appeared to be associated with membrane structures, and the cytoplasmic region containing the WR1 may be responsible for their difference in solubility.

3.3. $SURFIN_{4.2}$ is trafficked to the iRBC cytosol in a PEXEL-independent manner

$SURFIN_{4.2}$ contains two PEXEL-like sequences, one was termed Pexel 1 in this study and was located at the N-terminal region, spanning aa 25–29 (R_KI_FE), for which the 3rd position was isoleucine instead of leucine in the authentic PEXEL motif. The other was termed Pexel 2, and was located in the CRD at aa 118–122 (R_TL_ED). In order to evaluate their involvement in the transport of the protein into the iRBC, we generated two parasite lines expressing $SURF_{4.2}$CRD-Var1-TM-Pexel-1mut or $SURF_{4.2}$CRD-Var1-TM-Pexel-2mut, for which the conserved residues of the PEXEL-like sequence were replaced by alanine (A_KA_FA or A_TA_EA, respectively). Double staining IFA revealed that $SURFIN_{4.2}$CRD-Var1-TM-Pexel-1mut and -2mut both showed a punctate dot pattern in the iRBC that colocalized with the Maurer's cleft protein SBP1 along with parasite localized fluorescence (Fig. 6A). Thus, there was no appreciable difference between the trafficking of these proteins and that of the original $SURFIN_{4.2}$CRD-Var1-TM recombinant protein. These observations suggest that the PEXEL-like sequences of $SURFIN_{4.2}$ play no evident function in the transport of $SURFIN_{4.2}$ to the iRBC cytosol and Maurer's clefts. Thus, $SURFIN_{4.2}$ is being trafficked as a PNEP.

3.4. Removal of N-terminal 42 amino acid segment, CRD, or Var1 did not prevent the $SURFIN_{4.2}$ trafficking to the iRBC cytosol and Maurer's clefts

To further evaluate the importance of the different regions of the $SURFIN_{4.2}$ extracellular region in the trafficking of the protein to the iRBC, we generated three following parasite lines: two lines expressing $SURFIN_{4.2}$Var1-TM or $SURFIN_{4.2}$CRD-TM, in which the CRD or Var1 region were deleted from $SURFIN_{4.2}$CRD-Var1-TM and the one line expressing $SURFIN_{4.2}$CRD-Var1-TM-RepN, in which the N-terminal first 42 amino acids of $SURFIN_{4.2}$CRD-Var1-TM was replaced by the N-terminal first 15 amino acids of *P. falciparum* adenylosuccinate lyase (*Pf*ASL), an enzyme involved in the purine metabolism in the cell cytosol and is not considered to be transported to the iRBC [34]. Double staining IFA revealed that $SURFIN_{4.2}$Var1-TM and $SURFIN_{4.2}$CRD-TM colocalized with SBP1 in a punctate dot pattern in the iRBC and was also present in the parasite cytoplasm. No difference was observed between these two lines and the line expressing $SURFIN_{4.2}$CRD-Var1-TM (Fig. 6B). A more diffused iRBC localization with less obvious dot pattern formation was observed with the line expressing $SURF_{4.2}$CRD-Var1-TM-RepN compared to that expressing $SURFIN_{4.2}$CRD-Var1-TM. Nonetheless, the transport of the $SURF_{4.2}$

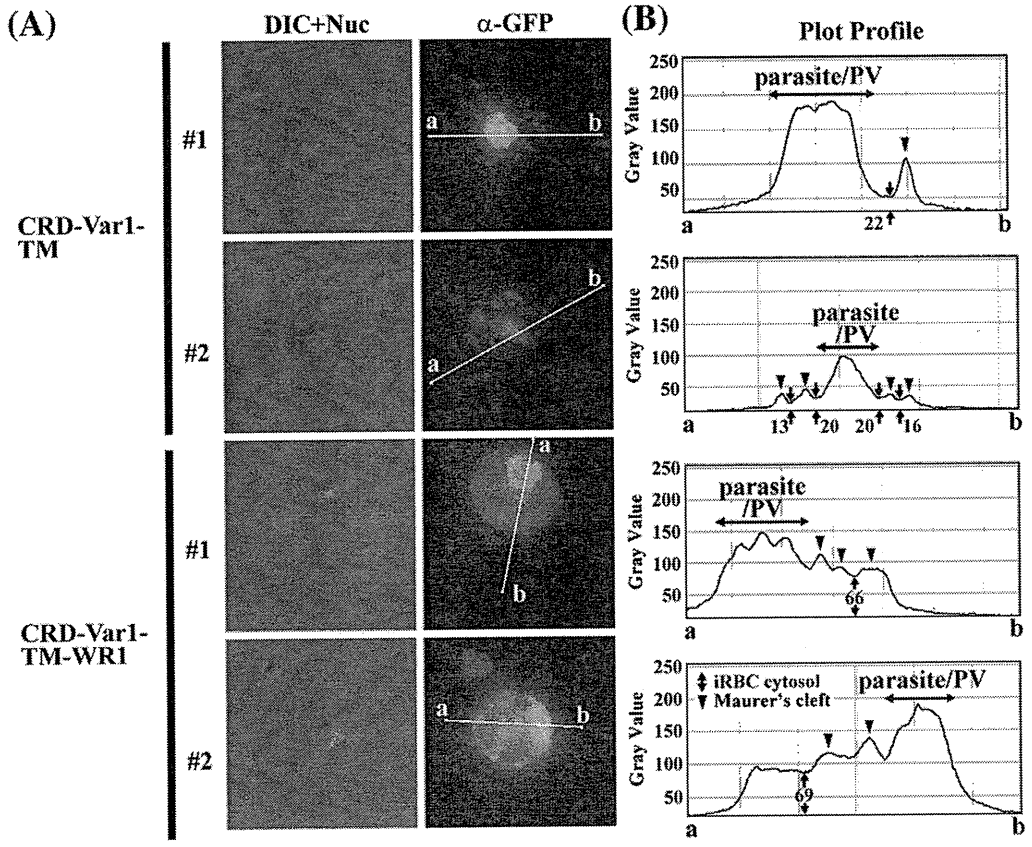

Fig. 4. Comparison of the signal intensity between SURFIN$_{4.2}$CRD-Var1-TM and SURFIN$_{4.2}$CRD-Var1-TM-WR1 proteins in the parasite-infected red blood cell. (A) Indirect immunofluorescence assay with anti-GFP antibody was performed for both parasites expressing SURFIN$_{4.2}$CRD-Var1-TM and SURFIN$_{4.2}$CRD-Var1-TM-WR1 at the same time. The fluorescence signal intensities were measured from point (a) to (b). (B) Plot profiles were made by using ImageJ software for the regions shown in panel A. Parasite/PV indicates parasite cytosol or parasitophorous vacuole. Signal intensities are shown by gray level pixel intensity values.

CRD-Var1-TM-RepN protein was not completely abrogated and signals, although faint, still colocalized with the Maurer's cleft SBP1. Thus, any of the N-terminal segment (aa 1–42), the CRD (aa 46–196), or the variable region (aa 198–733) at the extracellular region of SURFIN$_{4.2}$ do not carry a specific motif necessary for protein transport to the iRBC.

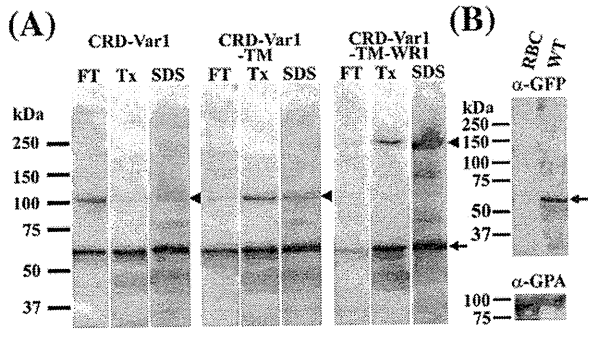

Fig. 5. Western blot of SURFIN$_{4.2}$CRD, SURFIN$_{4.2}$CRD-TM, and SURFIN$_{4.2}$CRD-TM-WR1 proteins subjected to different extraction procedures. (A) Parasite proteins were sequentially extracted by repeated-freeze thaw procedure (FT; soluble fraction protein), followed by Triton X-100 extraction (Tx; membrane bound protein), and SDS extraction (SDS, Tx-insoluble fraction) and detected with rabbit anti-GFP antibody. (B) Triton X-100 extracts of parasite-uninfected red blood cell (RBC) and the wild-type MS822 parasite line (WT) were subjected for Western blot analysis with rabbit anti-GFP (α-GFP) or mouse anti-glycophorin A (α-GPA). The arrow marks a ~60-kDa band also observed in wild-type parasites, but not in uninfected RBC. Arrowheads indicate expressed recombinant proteins, SURFIN$_{4.2}$CRD, SURFIN$_{4.2}$CRD-TM, and SURFIN$_{4.2}$CRD-TM-WR1.

To confirm these findings, we truncated the entire external domain from the mini-SURFIN$_{4.2}$ protein, and added 50 amino acids (SSGQVRRSGGQGSETYIVGTSQSGFHKNEVIPSIKDKSGKTQIVSNEKGG) preceding the TM region in order to support the integrality of the TM region for membrane insertion, to generate a parasite expressing SURFIN$_{4.2}$VarC-TM-WR1, thus this protein contains 50 amino acids derived from SURFIN$_{4.2}$ followed by a triple HA tag as an extracellular region. The recombinant protein was transported to the iRBC and observed in a punctate dot pattern in the iRBC cytosol, colocalizing with SBP1 (Fig. 7). This indicates that the extracellular region of SURFIN$_{4.2}$ is not required for the trafficking of the protein to the iRBC.

4. Discussion

In this study, we generated GFPm2-fused mini-SURFIN$_{4.2}$ proteins that, following their transfection into a *P. falciparum* parasite line, was observed to be trafficked into the iRBC and Maurer's clefts. Using this system, we then attempted to identify the specific region of the protein responsible for the iRBC and/or Maurer's cleft localization. We found that the TM region, but not the cytoplasmic region containing WR domain was essential for protein transport. We consider it likely that the TM region is responsible for initiating the trafficking of the protein into the ER. Two PEXEL-like sequences were found not to be essential for the movement of the protein into the iRBC and Maurer's clefts, indicating that SURFIN$_{4.2}$ trafficking is PEXEL-independent. N-terminal replacement, deletion of the CRD or Var region did not

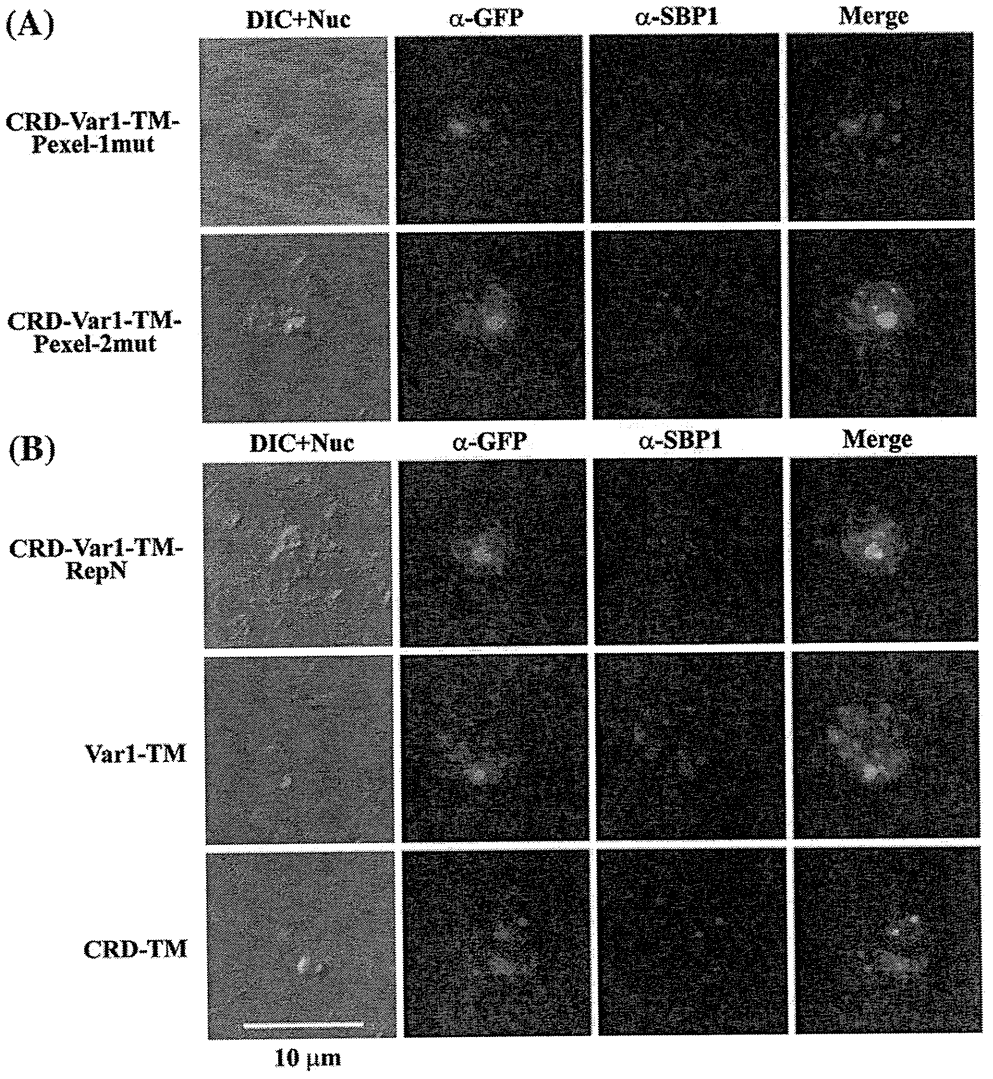

Fig. 6. Indirect immunofluorescence assay of modified $SURFIN_{4.2}$CRD-Var1-TM proteins. (A) Double staining for $SURFIN_{4.2}$CRD-Var1-TM-Pexel-1mut and CRD-Var1-TM-Pexel-2mut, and for (B) CRD-Var1-TM-RepN, Var1-TM, and CRD-TM proteins with anti-GFP (α-GFP, green) and anti-SBP1 (α-SBP1, red). Nuclei were stained with DAPI. Negative control using normal rabbit antibody did not produce detectable signals (not shown). CRD, cysteine-rich domain; RepN, N-terminus replacement; TM, transmembrane; and Var1, variable region 1.

prevent iRBC and Maurer's cleft localization, suggesting that no trafficking motif exists in these regions.

By sequential extraction of recombinant $SURFIN_{4.2}$ proteins, we found that mini-$SURFIN_{4.2}$ with an intact WR domain showed more resistance to Triton-X 100 extraction than a similar protein in which WR domain had been removed. As endogenous $SURFIN_{4.2}$ was insoluble in Triton-X 100 but soluble in SDS [26], we suggest that the cytoplasmic region, probably the WR domain, contributes to this difference. Insolubility with a neutral detergent such as Triton-X 100, was also reported for *Pf*EMP1 [35]. A large degree of sequence similarity was shown between the cytoplasmic WR domain of $SURFIN_{4.2}$, *Pf*EMP1, and another iRBC protein *Pf*332 [26]. The cytoplasmic regions of both *Pf*EMP1 and *Pf*332 are known to bind to RBC actin, the former also binding to spectrin [36,37]. Therefore, we

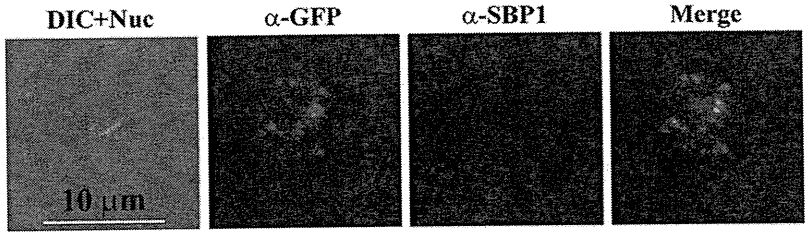

Fig. 7. Indirect immunofluorescence assay of $SURFIN_{4.2}$VarC-TM-WR1 protein. Double staining with anti-GFP (α-GFP, green) and anti-SBP1 (α-SBP1, red). Nuclei were stained with DAPI. Negative control using normal rabbit antibody did not produce detectable signals (not shown).

suggest that the WR domain of $SURFIN_{4.2}$ also associates with the RBC cytoskeleton, although further evaluation is required to assess this hypothesis.

Similar to most of the PNEPs reported so far, the $SURFIN_{4.2}$ TM region was found to be essential for protein trafficking. *Pf*SBP1 [11], MAHRP1 [19] and REX2 [24] share this feature, with their TM regions known to play important roles in protein transport. However, the N-terminal sequence of these PNEPs was also found to be essential for correct protein trafficking. In addition to the TM region, *Pf*SBP1 was shown to require the N-terminal segment at aa 16–26, which contain highly negative net charge residues (DEPTQLQDAVP) for transport into the iRBC [11]. This may also be the case for MAHRP1, as the N-terminal 50 amino acids of this protein, which is acidic, along with *Pf*SBP1 TM region is able to transport protein into the iRBC [11]. Conversely, REX2 appears to contain region resembling a PEXEL motif after cleavage in the ER at aa 5–10 ($L_5xE_7hhS_{10}$; h indicates hydrophobic residues), for which only the glutamate residue at aa 7 was found to be critical for trafficking [24]. In the mini-$SURFIN_{4.2}$ proteins we expressed in this study, none of the regions from the $SURFIN_{4.2}$ extracellular regions shown to be indispensable for trafficking to the iRBC, thus the trafficking of $SURFIN_{4.2}$ appears not depend on specific sorting signals, nor potential escorter proteins, but other factors in addition to the TM region.

Acknowledgments

We thank R. Culleton for critical reading and helpful comments. We are grateful to G. McFadden for *Pf*CRT 5′-pENTR4/1, GFP-pENTR2/3, and pCHDR-3/4, D. Jacobus for WR99210, T. Tsuboi for anti-*Pf*SBP1 rabbit serum, N. Iyoku for her expertise, and I. Sekine, head of the Nagasaki Red Cross Blood Center for human RBC and plasma. This work was supported in part by Grants-in-Aids for Scientific Research 20406009 and the Global COE Program, Nagasaki University, from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, and Daiichi-Sankyo Foundation of Life Science (to O.K.). J.A. is a recipient of MEXT PhD scholarship, Japan. All experiments conducted in this study were approved by the committee for Living Modified Organisms in Nagasaki University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.parint.2011.05.003.

References

[1] Tilley L, Sougrat R, Lithgow T, Hanssen E. The twists and turns of Maurer's cleft trafficking in *P. falciparum*-infected erythrocytes. Traffic 2008;9:187–97.
[2] Przyborski JM, Wickert H, Krohne G, Lanzer M. Maurer's clefts — a novel secretory organelle? Mol Biochem Parasitol 2003;132:17–26.
[3] Lanzer M, Wickert H, Krohne G, Vincensini L, Braun-Breton C. Maurer's clefts: a novel multi-functional organelle in the cytoplasm of *Plasmodium falciparum*-infected erythrocytes. Int J Parasitol 2006;36:23–36.
[4] Wickham ME, Rug M, Ralph SA, Klonis N, McFadden GI, Tilley L, et al. Trafficking and assembly of the cytoadherence complex in *Plasmodium falciparum*-infected human erythrocytes. EMBO J 2001;20:5636–49.
[5] Baruch DI. Adhesive receptors on malaria-parasitized red cells. Baillière's Best Pract Res Clin Haematol 1999;12:747–61.
[6] Chen Q, Schlichtherle M, Wahlgren M. Molecular aspects of severe malaria. Clin Microbiol Rev 2000;13:439–50.
[7] Newbold C, Craig A, Kyes S, Rowe A, Fernandez-Reyes D, Fagan T. Cytoadherence, pathogenesis and the infected red cell surface in *Plasmodium falciparum*. Int J Parasitol 1999;29:927–37.
[8] Hiller NL, Bhattacharjee S, van Ooij C, Liolios K, Harrison T, Lopez-Estraño C, et al. A host-targeting signal in virulence proteins reveals a secretome in malarial infection. Science 2004;306:1934–7.
[9] Marti M, Good RT, Rug M, Knuepfer E, Cowman AF. Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science 2004;306:1930–3.
[10] Marti M, Baum J, Rug M, Tilley L, Cowman AF. Signal-mediated export of proteins from the malaria parasite to the host erythrocyte. J Cell Biol 2005;171:587–92.
[11] Saridaki T, Fröhlich KS, Braun-Breton C, Lanzer M. Export of *Pf*SBP1 to the *Plasmodium falciparum* Maurer's clefts. Traffic 2009;10:137–52.
[12] Boddey JA, Hodder AN, Günther S, Gilson PR, Patsiouras H, Kapp EA, et al. An aspartyl protease directs malaria effector proteins to the host cell. Nature 2010;463:627–31.
[13] Russo I, Babbitt S, Muralidharan V, Butler T, Oksman A, Goldberg DE. Plasmepsin V licenses *Plasmodium* proteins for export into the host erythrocyte. Nature 2010;463:632–6.
[14] de Koning-Ward TF, Gilson PR, Boddey JA, Rug M, Smith BJ, Papenfuss AT, et al. A newly discovered protein export machine in malaria parasites. Nature 2009;459:945–9.
[15] Melcher M, Muhle RA, Henrich PP, Kraemer SM, Avril M, Vigan-Womas I, et al. Identification of a role for the *Pf*EMP1 semi-conserved head structure in protein trafficking to the surface of *Plasmodium falciparum* infected red blood cells. Cell Microbiol 2010;12:1446–62.
[16] van Ooij C, Tamez P, Bhattacharjee S, Hiller NL, Harrison T, Liolios K, et al. The malaria secretome: from algorithms to essential function in blood stage infection. PLoS Pathog 2008;4:e1000084.
[17] Spielmann T, Hawthorne PL, Dixon MW, Hannemann M, Klotz K, Kemp DJ, et al. A cluster of ring stage-specific genes linked to a locus implicated in cytoadherence in *Plasmodium falciparum* codes for PEXEL-negative and PEXEL-positive proteins exported into the host cell. Mol Biol Cell 2006;17:3613–24.
[18] Blisnick T, Morales Betoulle ME, Barale JC, Uzureau P, Berry L, Desroses S, et al. *Pf*sbp1, a Maurer's cleft *Plasmodium falciparum* protein, is associated with the erythrocyte skeleton. Mol Biochem Parasitol 2000;111:107–21.
[19] Spycher C, Klonis N, Spielmann T, Kump E, Steiger S, Tilley L, et al. MAHRP-1, a novel *Plasmodium falciparum* histidine-rich protein, binds ferriprotoporphyrin IX and localizes to the Maurer's clefts. J Biol Chem 2003;278:35373–83.
[20] Hawthorne PL, Trenholme KR, Skinner-Adams TS, Spielmann T, Fischer K, Dixon MW, et al. A novel *Plasmodium falciparum* ring stage protein, REX, is located in Maurer's clefts. Mol Biochem Parasitol 2004;136:181–9.
[21] Hanssen E, Hawthorne P, Dixon MW, Trenholme KR, McMillan PJ, Spielmann T, et al. Targeted mutagenesis of the ring-exported protein-1 of *Plasmodium falciparum* disrupts the architecture of Maurer's cleft organelles. Mol Microbiol 2008;69:938–53.
[22] Saridaki T, Sánchez CP, Pfahler J, Lanzer M. A conditional export system provides new insights into protein export in *Plasmodium falciparum*-infected erythrocytes. Cell Microbiol 2008;10:2483–95.
[23] Spielmann T, Gilberger TW. Protein export in malaria parasites: do multiple export motifs add up to multiple export pathways? Trends Parasitol 2009;26:6–10.
[24] Haase S, Herrmann S, Grüring C, Heiber A, Jansen PW, Langer C, et al. Sequence requirements for the export of the *Plasmodium falciparum* Maurer's clefts protein REX2. Mol Microbiol 2009;71:1003–17.
[25] Dixon MW, Hawthorne PL, Spielmann T, Anderson KL, Trenholme KR, Gardiner DL. Targeting of the ring exported protein 1 to the Maurer's clefts is mediated by a two-phase process. Traffic 2008;9:1316–26.
[26] Winter G, Kawai S, Haeggström M, Kaneko O, von Euler A, Kawazu S, et al. SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes. J Exp Med 2005;201:1853–63.
[27] van Dooren GG, Marti M, Tonkin CJ, Stimmler LM, Cowman AF, McFadden GI. Development of the endoplasmic reticulum, mitochondrion and apicoplast during the asexual life cycle of *Plasmodium falciparum*. Mol Microbiol 2005;57:405–19.
[28] Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, et al. Discovery of gene function by expression profiling of the malaria parasite life cycle. Science 2003;301:1503–8.
[29] Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, et al. PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res 2009;37:D539–43.
[30] Nakazawa S, Culleton R, Maeno Y. In vivo and in vitro gametocyte production of *Plasmodium falciparum* isolates from Northern Thailand. Int J Parasitol 2011;41:317–23.
[31] Deitsch K, Driskill C, Wellems T. Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes. Nucleic Acids Res 2001;29:850–3.
[32] Tonkin CJ, van Dooren GG, Spurck TP, Struck NS, Good RT, Handman E, et al. Localization of organellar proteins in *Plasmodium falciparum* using a novel set of transfection vectors and a new immunofluorescence fixation method. Mol Biochem Parasitol 2004;137:13–21.
[33] Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 2002;419:498–511.
[34] Marshall VM, Coppel RL. Characterisation of the gene encoding adenylosuccinate lyase of *Plasmodium falciparum*. Mol Biochem Parasitol 1997;88:237–41.
[35] Aley SB, Sherwood JA, Howard RJ. Knob-positive and knob-negative *Plasmodium falciparum* differ in expression of a strain-specific malarial antigen on the surface of infected erythrocytes. J Exp Med 1984;160:1585–90.
[36] Oh SS, Voigt S, Fisher D, Yi SJ, LeRoy PJ, Derick LH, et al. *Plasmodium falciparum* erythrocyte membrane protein 1 is anchored to the actin-spectrin junction and knob-associated histidine-rich protein in the erythrocyte skeleton. Mol Biochem Parasitol 2000;108:237–47.
[37] Waller KL, Stubberfield LM, Dubljevic V, Buckingham DW, Mohandas N, Coppel RL, et al. Interaction of the exported malaria protein *Pf*332 with the red blood cell membrane skeleton. Biochim Biophys Acta 2010;1798:861–71.

Parasitology International 61 (2012) 317–323

Contents lists available at SciVerse ScienceDirect

Parasitology International

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

Stable allele frequency distribution of the polymorphic region of SURFIN$_{4.2}$ in *Plasmodium falciparum* isolates from Thailand

Morakot Kaewthamasorn [a,e], Kazuhide Yahata [a], Jean Semé Fils Alexandre [a,f], Phonepadith Xangsayarath [a], Shusuke Nakazawa [a], Motomi Torii [b], Jetsumon Sattabongkot [c,1], Rachanee Udomsangpetch [d], Osamu Kaneko [a,*]

[a] Department of Protozoology, Institute of Tropical Medicine (NEKKEN) and the Global Center of Excellence Program, Nagasaki University, Sakamoto, Nagasaki 852-8523, Japan
[b] Department of Molecular Parasitology, Ehime University Graduate School of Medicine, Shitsukawa, Toon, Ehime 791-0295, Japan
[c] Department of Entomology, Armed Forces Research Institute of Medical Sciences, Bangkok 10400, Thailand
[d] Department of Pathobiology, Faculty of Science, Mahidol University, Bangkok 10400, Thailand
[e] Parasitology Unit, Department of Pathology, Faculty of Veterinary Science, Chulalongkorn University, Bangkok 10330, Thailand
[f] Centro Nacional de Control de Enfermedades Tropicales, Santo Domingo, República Dominicana

ARTICLE INFO

Article history:
Received 5 December 2011
Received in revised form 17 December 2011
Accepted 20 December 2011
Available online 27 December 2011

Keywords:
Plasmodium falciparum
SURFIN
Positive diversifying selection
Allele frequency distribution

ABSTRACT

Plasmodium falciparum SURFIN$_{4.2}$ (PFD1160w) is a polymorphic protein expressed on the surface of parasite-infected erythrocytes. Such molecules are expected to be under strong host immune pressure, thus we analyzed the nucleotide diversity of the N-terminal extracellular region of SURFIN$_{4.2}$ using *P. falciparum* isolates obtained from a malaria hypoendemic area of Thailand. The extracellular region of SURFIN$_{4.2}$ was divided into four regions based on the amino acid sequence conservation among SURFIN members and the level of polymorphism among SURFIN$_{4.2}$ sequences; N-terminal segment (Nter), a cysteine-rich domain (CRD), a variable region 1 (Var1), and a variable region 2 (Var2). Comparison between synonymous and non-synonymous substitutions, Tajima's *D* test, and Fu and Li's *D** and *F** tests detected signatures of positive selection on Var2 and to a lesser extent Var1, suggesting that these regions were likely under host immune pressure. Strong linkage disequilibrium was detected for nucleotide pairs separated by a distance of more than 1.5 kb, and 7 alleles among 19 alleles detected in 1988–1989 still circulated 14 years later, suggesting low recombination of the analyzed *surf*$_{4.2}$ sequence region in Thailand. The allele frequency distribution of polymorphic areas in Var2 did not differ between two groups collected in different time points, suggesting the allele frequency distribution of this region was stable for 14 years. The observed allele frequency distribution of SURFIN$_{4.2}$ Var2 may be fixed in Thai *P. falciparum* population as similar to the observation for *P. falciparum* merozoite surface protein 1, for which a stable allele frequency distribution was reported.

1. Introduction

Malaria is a serious life-threatening disease caused by *Plasmodium* protozoan parasite. The estimated number of malaria case was 225 million and over 780,000 people died of malaria in 2009, majority of them were children under 5 years old living in Africa [1]. The erythrocytic cycle of *Plasmodium falciparum* is characterized by complex interactions between parasite-derived surface exposed antigens, host-cell receptors and immune defense proteins. Parasite adhesins expressed on the surface of malaria-infected erythrocytes provide adhesive properties and at the same time allow for host immune evasion by virtue of antigenic variation and antigenic polymorphism [2]. Recently, a new multigene family called *surf* was identified, consisting of at least 10 members in the genome of *P. falciparum* 3D7 parasite line and infected erythrocyte surface localization was shown for one of the member, SURFIN$_{4.2}$ (gene ID: PFD1160w, chr 4) [3]. All SURFIN members are predicted to be type 1 transmembrane proteins and share a similar protein makeup, containing a conserved cysteine-rich domain (CRD) in the N-terminal of their extracellular region, and tryptophan-rich domains (WRD) in the C-terminal predicted intracellular region. An orthologous gene, *Pv*STP1 has been identified in *Plasmodium vivax*. The CRD of SURFIN/*Pv*STP members shares homology with proteins encoded by the super multigene family *pir*, which are

* Corresponding author at: Department of Protozoology, Institute of Tropical Medicine (NEKKEN), Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan. Tel.: +81 95 819 7838; fax: +81 95 819 7805.
E-mail addresses: morakot.k@chula.ac.th, kotscmi@hotmail.com (M. Kaewthamasorn), kyahata@nagasaki-u.ac.jp (K. Yahata), semefils@yahoo.es (J.S.F. Alexandre), xangsayalathdith@yahoo.com (P. Xangsayarath), nakazawa@nagasaki-u.ac.jp (S. Nakazawa), torii@m.ehime-u.ac.jp (M. Torii), tmjetsumon@mahidol.ac.th (J. Sattabongkot), scrud@mahidol.ac.th (R. Udomsangpetch), okaneko@nagasaki-u.ac.jp (O. Kaneko).
[1] Present address: Mahidol Vivax Research Center, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand.

1383-5769/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.parint.2011.12.003

expressed on surface of infected erythrocytes and are found in numerous malaria parasite species including *Plasmodium yoelii*, *Plasmodium berghei*, *Plasmodium chabaudi*, *Plasmodium knowlesi*, and *P. vivax* [4]. The intracellular WRD of SURFIN/*Pv*STP shares homology with the intracellular region of the other parasite-encoded erythrocyte surface ligands, such as the *P. falciparum Pf*EMP-1 family and *P. knowlesi* surface variant antigen SICA*var* family. Thus, it is possible that the *pir* gene products have evolved from the extracellular region of SURFIN/*Pv*STP type proteins, and, similarly, *Pf*EMP-1 and SICA*var* proteins may be derived from the extracellular region [3]. $SURFIN_{4.2}$ is co-transported to Maurer's clefts with *Pf*EMP-1 and RIFIN, and is expected to be translocated to the erythrocyte membrane for surface exposure. $SURFIN_{4.2}$ also accumulates in the parasitophorous vacuole, and can be detected in an amorphous cap at the newly released merozoite apex, suggesting that it may have a role in the erythrocyte invasion [3]. Such molecules are expected to be under a strong host immune pressure, and indeed the extracellular region is polymorphic between 3D7 and FCR3 parasite lines [3]. Recently, a population genetics-based analysis using Kenyan *P. falciparum* isolates detected positive diversifying selection and linkage disequilibrium with a distance of ~1.5 kb [5] on $surf_{4.2}$ exon 1 encoding the extracellular region. To further understand the host-pathogen interaction through $SURFIN_{4.2}$ and design an universal intervention strategy targeting this molecule, it is important to understand the nucleotide diversity in different geographic areas. Thus, we obtained sequences encoding the extracellular region of $SURFIN_{4.2}$ from Thai *P. falciparum* isolates collected in a malaria hypoendemic area and evaluated allelic diversity, whether or not there was positive selection, the degree of linkage disequilibrium, and temporal changes in allele frequency distribution in a malaria hypoendemic area.

2. Materials and methods

2.1. Parasite DNA isolation

P. falciparum parasites were obtained from Thailand in three different periods. Thirty samples (1988/1989 group) were collected from November 1988 to January 1989 in Mae Sod, which lies in the north west of Thailand near the Thai-Myanmar border. They were adapted to culture, cryopreserved and kept at the Department of Protozoology, Institute of Tropical Medicine (NEKKEN), Nagasaki University as previously described [6,7]. The parasites were thawed and maintained *in vitro* essentially as described previously [8]. Human erythrocytes and plasma used for culture were obtained from the Nagasaki Red Cross Blood Center. Parasites were harvested when parasitemia reached about 2%, and parasite genomic DNA (gDNA) was extracted using DNAzol BD (Invitrogen). Following elution, gDNA was stored at −30 °C. Twenty-eight blood samples were collected onto filter papers (Whatman 31 ETChr) in 2003 (n=21) and 2005 (n=7) after the approval by the Ethical Review Committee of Mahidol University. Genomic DNA was extracted using QIAamp DNA Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's instruction. Filter papers were separately and carefully handled in a clean bench to avoid DNA contamination among the samples.

2.2. Polymerase chain reaction (PCR) amplification and sequencing

A DNA fragment of $surf_{4.2}$ gene encoding the extracellular region (2314 bp, nucleotide positions (nt) -22 - 2292 after 3D7 sequence) was primarily amplified with forward primer F0 (ATATTTTCCCATTTTTGATAATATG) and reverse primer R2-2 (CTTATTAAATACCAAAAACATAAAAAG). The amplification was performed in a 20 μL reaction mixture containing 200 μM dNTPs, 1× KOD -Plus- buffer, 2 mM $MgSO_4$, 500 nM of each primer and 0.4 units of KOD -Plus- DNA polymerase (Toyobo, Japan) using a GeneAmp 9700 PCR thermocycler (Applied Biosystems, Foster City, CA). Negative control was always set using distilled water as a template solution. Thermal cycling profile contained an initial denaturation at 94 °C for 2 min; 40 cycles of 92 °C for 15 s, 54 °C for 20 s, 68 °C for 3 min; and final-extension at 68 °C for 5 min. PCR products were then resolved by electrophoresis on a 1.0% agarose gel (Takara, Japan). After ethidium bromide staining, the PCR products were visualized by UV transillumination. Primary PCR products were then diluted with distilled water and adjusted the DNA concentration to approximately 50–100 ng/μL to serve as a template for the nested PCR amplification. For the samples, especially those obtained from filter papers, producing undetectable or very faint target bands after primary PCR amplification, we 100-fold diluted the primary PCR product with distilled water and used for the nested PCR amplification with the same condition for the primary PCR amplification. Two sets of primers were used; F7 (CTTTTTGTTGTTGAGCTCGACAGC) and R2 (CCTGATCTGTCGAATAAATAGC) for the 5' side 1201 bp (nt 4–1204 after 3D7 sequence) and F3 (ATTGAAGTTGATTGTGCTGAAG) and R8 (TATCCCTTTTGAAAAAATCCCTC) for the 3' side part 1158 bp (nt 1060–2217 after 3D7 sequence).

PCR products were treated with ExoSAP-IT (USB Corporation) and directly sequenced from both directions with ABI PRISM® BigDye™ Terminator ver1.1 (Applied Biosystems, Foster City, CA) using an ABI 3730 DNA analyzer (Applied Biosystems) according to the manufacturer's instructions. The primers used for the sequencing are summarized in Table S1. Samples showing dual peaks, suggestive of a mixed infection, were sequenced after cloning the full length of extracellular region into pGEM-T Easy® plasmid (Promega, Madison, WI). We employed sequences supported by at least three independent plasmid clones. After sequencing reaction, sequencing mixture was subjected to ethanol precipitation to remove the fluorescent reaction mixture. All sequences were validated by at least two independently PCR-amplified DNA fragments to avoid potential error during the PCR amplification and sequencing with special care to the singletons (substitutions appearing only once among the sequences).

We analyzed four putatively neutral loci to serve as controls for allele frequency distribution. Single nucleotide polymorphisms (SNP) determined were nt 267 and 1008 of adenylosuccinate lyase (PFB0295w, chr 2), nt 165 and 319 of aspartate aminotransferase (PFB0200c, chr 2), nt 1989 of pyruvate kinase (PF10_0363, chr 10), and nt 819 and 969 of actin II (PF14_0124, chr 14). These genes were selected because they were located on different chromosomes to those harboring genes associated with drug resistance (dihydrofolate reductase-thymidylate synthase, chr 4; multidrug resistance protein, chr 5; chloroquine resistant transporter, chr 7; and dihydropteroate synthetase, chr 8), to avoid a potential hitchhiking effect by the drug pressure towards these genes. We amplified DNA region surrounding selected SNPs using primers listed in Table S2. PCR conditions and sequencing reactions were similar to those described for $surf_{4.2}$ with slight modifications. We adjusted the annealing temperature to 56 °C, reduced the extension time to 45 s~1 min depending on the amplicon, and sequenced directly.

2.3. Data analysis

Nucleotide diversity (π) and its standard error (SE) were computed with the Jukes and Cantor method using MEGA 4.0 software [9]. The mean numbers of synonymous substitutions per synonymous sites (d_S) and nonsynonymous substitutions per nonsynonymous sites (d_N) and their standard errors were computed using the Nei and Gojobori method [9,10] with the Jukes and Cantor correction, implemented in MEGA 4.0. The statistical difference between d_N and d_S was tested using a one-tailed Z-test with 500 bootstrap pseudosamples in MEGA 4.0. A value of d_N significantly higher than d_S at the 95% confidence level was taken as evidence for positive selection. Nucleotide diversity was plotted by a sliding window method (90 bases with a step size of 3 bases) using DnaSP 5.0 [11].

Departures from the predictions of the neutral model of molecular evolution were tested by a set of neutrality tests based on measures of allele frequencies or heterozygosity within species using Tajima's *D*,

Fu and Li's D^* and F^* parameters using DnaSP 5.0, under the assumption that the size of population was stably maintained larger than effective population size. Tajima's D statistic relies on the difference between an average pairwise nucleotide diversity (π) and an estimated nucleotide diversity under neutrality (θ) derived from the number of segregating sites (S) [12]. Fu and Li's tests rely on the differences between estimates of θ based on the number of singletons and that based on S (D^* index) or π (F^* index) [13]. Minimum number of recombination events was evaluated according to Hudson and Kaplan (1985) using DnaSP 5.0 [14,15]. Linkage analysis was performed by comparing a variance obtained from the distribution of the number of loci at which each pair of allele was different and an expected variance obtained by a Monte Carlo simulation (100,000 iterations) and a standardized I_A (I_A^S), a function of the recombination rate with zero value indicating a linkage equilibrium, were calculated using LIAN3.5 [16]. Linkage analysis was also performed to obtain |D'| and r^2 values, indices of the linkage disequilibrium, using DnaSP 5.0 [17,18].

3. Results

3.1. Polymorphism of the $surf_{4.2}$ gene

Single contiguous nucleotide sequences (2166 bp from nt 28 to 2193) encoding the extracellular region of $SURFIN_{4.2}$ were obtained from *P. falciparum* Thai field isolates collected at three different time points; 44 sequences (19 alleles) in 1988–1989 (collectively termed as 1988/1989 group), 23 sequences (14 alleles) in 2003 and 7 sequences (6 alleles) in 2005 (totally 74 sequences containing 28 alleles). The sequence names are found in the legend of Fig. 3. A total of 255 polymorphic nucleotide sites and no insertion/deletion were observed in the overall samples sequenced with an average pairwise nucleotide diversity of 0.043. In order to identify the area(s) accumulating polymorphism, we divided the extracellular region of $SURFIN_{4.2}$ into three regions based on amino acid sequence conservation among SURFIN members; N-terminal segment (Nter; amino acid positions (aa) 1–50, nt 1–150), CRD (aa 51–195, nt 151–585), and a variable region (aa 196–739, nt 586–2217). Then, the obtained nucleotide sequence was divided based on this definition, thus Nter was assessed based on the sequence from nt 28 to nt 150 and variable region from nt 586 to nt 2193. Although the polymorphic sites were distributed across the entire sequence, 225 of them were located in the variable region (14.0% of 1608 bp), while Nter had 2 (1.6% of 123 bp) and CRD had 28 polymorphic sites (6.4% of 435 bp). This was evident in sliding window plot of nucleotide diversity (Fig. 1).

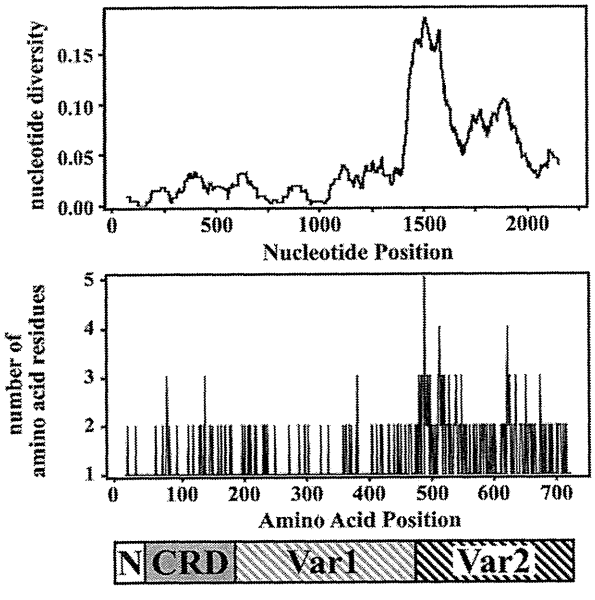

Fig. 1. Sliding window plot of nucleotide diversity and amino acid polymorphism of $SURFIN_{4.2}$ extracellular region in *Plasmodium falciparum* Thai isolates. Nucleotide diversity is plotted with a window length of 90 bp and step size of 3 bp (top) and the number of the amino acid residues at each amino acid position is plotted (middle). To visualize the location sites showing high diversity, a scheme of the extracellular region of $SURFIN_{4.2}$ is shown (bottom), which is divided into 4 parts; N-terminal (Nter), cysteine-rich domain (CRD), and variable regions (Var1 and Var2). A total of 74 sequences from Thai isolates are used. Nucleotide and amino acid positions are after the 3D7 line sequences.

These trends were extended to the amino acid level. Among 188 polymorphic sites, 2 were in Nter, 23 were in CRD, and 163 were in the variable region. In order to pinpoint which regions had accumulated the most polymorphisms, we plotted the location of the substitution and the number of amino acids observed at each site (Fig. 1). We noticed that more polymorphism accumulated towards the C-terminal side of the variable region. We divided $SURFIN_{4.2}$ variable region into two sub-regions for detailed analysis; the N-terminal Var1 region from aa 196 to 482 and C-terminal Var2 region from aa 483 to 739. The amount of polymorphic amino acid sites in CRD (23/145 = 15.9%) and Var1 (48/287 = 16.7%) was comparable and most of the polymorphism were dimorphic. Whereas a larger number of polymorphic sites were observed in Var2 (115/249 = 46.2%) with four different amino acids at aa 504 and 613, and five amino acids at aa 489.

3.2. Positive diversifying selection on $surf_{4.2}$

Because of the observed high polymorphism, we then evaluated signatures of a positive diversifying selection on $surf_{4.2}$ in Thai isolates by comparing synonymous and non-synonymous substitutions using all 74 sequences (Table 1). A significant excess of non-synonymous substitutions over synonymous substitutions was detected when the entire sequence was evaluated ($p = 0.0004$). The same analysis performed against the Var2 region also detected a significant excess of non-synonymous substitutions over synonymous substitutions ($p = 0.005$). Although excess of non-synonymous substitutions over synonymous substitutions was observed for Nter, CRD, and Var1 regions, these regions were excluded from this analysis, because the numbers of synonymous differences (Sd) were too low (0, 0.537, and 2.639, respectively) for an optimal analysis. The results suggest that positive selection is acting on the extracellular region of $SURFIN_{4.2}$, at least in the Var2 region.

We further evaluated signatures of selection by population genetics-based approaches; Tajima's D, Fu and Li's D^*, Fu and Li's F^* tests for 1988/1989 group (Table 2). Samples obtained from the other periods were excluded due to the low sample number. Significant positive values of Tajima's D were not detected for the entire extracellular region or four sub-regions. However, sliding window plot analysis depicted significant positive D values in Var1 and Var2 regions, suggesting the action of positive diversifying selection on these regions (Fig. 2). A significant positive value of Fu and Li's D^* (1.66) and F^* (2.02) were detected for the entire extracellular region ($p < 0.02$), respectively. When the four sub-regions were separately assessed, significant deviations greater than zero were detected for Var2 ($D^* = 1.77$ and $F^* = 2.13$; $p < 0.02$) and Var1 ($F^* = 1.84$; $p < 0.05$). Sliding window plot analysis of Fu and Li's D^* revealed significant positive deviation of greater than zero in the Var2 region. Sliding window plot analysis of Fu and Li's F^* revealed significant positive deviation of greater than zero in the Var1 and Var2 regions, for which the value of the Var2 region was higher than that of Var1 region. These results were consistent with the analysis based on the sub-regions. Collectively, comparison between non-synonymous and synonymous substitutions, Tajima's D test, and Fu and Li's D^* and F^* tests all detected the signature of positive diversifying selection on the Var2 region at the 98% confidence level and Tajima's D

Table 1
Nucleotide diversity of *Plasmodium falciparum surf4.2* from Thai isolates (n = 74).

Region (position)	Number of sites (base)	k (SE)	*Nd* (SE)	*N* (SE)	*Sd* (SE)	*S* (SE)	π (SE)	d_N (SE)	d_S (SE)	d_N/d_S	*p* ($d_N > d_S$)
Extracellular (28–2193)	2166	92.027 (5.484)	78.825 (5.350)	1711.288 (7.900)	13.203 (2.192)	454.712 (8.182)	0.0440 (0.0027)	0.0479 (0.0035)	0.0298 (0.0049)	1.61	0.0004
Nter (28–150)	123	0.870 (0.570)	0.870 (0.589)	98.833 (1.741)	0 (0)	24.167 (1.692)	0.0071 (0.0047)	0.0089 (0.0060)	0 (0)	∞	
CRD (151–585)	435	7.284 (1.599)	6.747 (1.451)	357.214 (3.469)	0.537 (0.368)	77.786 (3.443)	0.0170 (0.0035)	0.0191 0.0043)	0.0070 (0.0050)	2.73	
Var1 (586–1446)	861	20.563 (2.598)	17.924 (2.613)	673.883 (5.348)	2.639 (1.030)	187.117 (5.457)	0.0244 (0.0032)	0.0272 (0.0045)	0.0143 (0.0058)	1.90	
Var2 (1447–2193)	747	63.311 (4.419)	53.284 (4.273)	581.358 (4.701)	10.026 (1.829)	165.642 (4.227)	0.0917 (0.0071)	0.0998 0.0088)	0.0641 (0.0118)	1.56	<0.005

Extracellular, extracellular region; Nter, N-terminal segment; CRD, cysteine-rich domain; Var1, variable region 1; Var2, variable region 2. sites, sites nucleotide analyzed. k, the average number of nucleotide differences; *N* and *S*, average numbers of nonsynonymous and synonymous sites; π, pairwise nucleotide diversity; d_N, number of nonsynonymous substitutions over number of nonsynonymous sites; d_S, number of synonymous substitutions over number of synonymous sites; SE, standard error computed using the Nei-Gojobori method with Jukes-Cantor correction. SE was estimated using the bootstrap method with 500 replication. The numbers of synonymous (*Sd*) and nonsynonymous (*Nd*) differences were calculated by Nei-Gojobori method. *p* value indicates the statistical difference between d_N and d_S, tested using a one-tail Z-test with 500 bootstrap pseudosamples implemented in MEGA4.0.

and Fu and Li's F^* tests detected that on Var1 region at the 98% confidence level.

3.3. Linkage disequilibrium and recombination of *surf4.2*

To further assess the polymorphic nature of SURFIN$_{4.2}$, polymorphic amino acid sites were aligned and allele numbers were assigned (Fig. 3). We noticed that aa 483–539 were clustered and could be divided into three patterns; allele 1, 8, and 19 patterns. This pattern appeared to be extended further 3' side at the amino acid positions 577–586 and 599–602 where dimorphic substitutions were seen in the parasites possessing allele 3 pattern at aa 483–539. Thus, we evaluated linkage disequilibrium (LD) on this gene by calculating a standardized I_A (I_A^S) and found that I_A^S of 1988/1989 group sequences was 0.1459 ($p<0.00001$), indicating significant LD of *surf4.2* sequences. Significant |D'| and r^2 values seen in the right upper corner of the panels in Fig. 4 indicated LD between sites over a long distance, for example, significant values could be seen for the distance more than 1.5 kb for all plots (Fig. 4). Previous report on the Kenyan isolates detected a strong LD between nt 76 in Nter and nt sites located in nt 1473–1870 in Var2 (~1.5 kb apart; $p<0.001$ after Bonferroni correction) and an epistatic relationship between these sites was proposed [5]. Our analysis also detected LD between nt 76 and nt sites located in nt 1352–1955 for 1988/1989 group ($p<0.001$) after Bonferroni correction, which was consistent to the previous observation. Because more than 5 sequences belonged to the alleles 1, 4 or 19, which likely contributed to the observed LD, only one sequence from each allele were used to assess if LD was still detected or not. An obtained I_A^S value was 0.1119 ($p<0.00001$) for 1988/1989 (n = 19), further supporting the LD on *surf4.2* in Thai isolates.

Because LD was observed for the sites with the long distance, we evaluated the recombination events on *surf4.2* using sequences originated from Thailand. To this end, 35 minimum recombination events were detected throughout the entire sequence obtained for 1988/1989 (detected recombination between nonsynonymous substitutions were plotted in Fig. 3), except some areas in Var2 where multiple amino acids were clustered, such as aa 483–539. This indicates that recombination could occur on this gene except highly polymorphic areas, which would prevent efficient recombination between too diversified sequences. The recombination and 4th and 5th amino acids at the aa 483–539 occurred within allele 8 pattern, suggesting that this pattern contributed to the complexity of *surf4.2* polymorphism more than remaining 2 patterns.

3.4. Frequency distribution of the polymorphism of *surf4.2*

When a parasite population possesses more than one allele of an antigen-encoding gene at one time point, the most common allele may be targeted by immune selection pressure, and as a result, relatively rare alleles may expand in the population. Under such a hypothesis, a fluctuation of allele-frequency distribution may be observed. Thus we evaluated if the allele-frequency distribution of *surf4.2* in 2003 was changed from 1988/1989. We selected four areas in Var2 region where more than four amino acids were clustered for this analysis (Fig. 3), because positive selection was detected on this region. As a control to evaluate the temporal change of the allele-frequency distribution, we obtained the information of nucleotide polymorphisms of four gene loci (nt 165 and 319 of PFB0200c, aspartate aminotransferase; nt 267 and 1008 of PFB0295w, adenylosuccinate lyase; nt 1989 of PF10_0363, pyruvate kinase; and nt 819 and 969 of PF14_0124, actin II) that were expected to be neutral to immune or drug selection pressure (Fig. S1). Because two *surf4.2* sequences were detected from MS802, MS803, MS819, MS820, MS824, MS829, AQ1097, and AQ1105, and one or two of the 4 loci showed mixed peaks for MS826 and MS818, sequences obtained from these parasites were excluded from the analysis. Allele frequency distribution of four putatively neutral loci were not significantly different between 1988/1989 and 2003 (Table S3), suggesting that at least there were no obvious change in the population structure that could be detected with these loci. We

Table 2
Test of neutrality for *Plasmodium falciparum surf4.2* from Thai isolates collected in 1988–1989 (n = 44).

Region	Nucleotide position	Number of sites (base)	Two variants				More than two variants			Tajima's	Fu and Li's	
			η	*S*	(Singleton)	(Not singleton)		π	θ	*D*	*D**	*F**
Extracellular	28–2193	(2166)	262	247	11	222	14	0.041	0.028	1.75	1.66**	2.02**
Nter	28–150	(123)	2	2	0	2	0	0.007	0.004	1.36	0.76	1.09
CRD	151–585	(435)	24	24	3	21	0	0.015	0.013	0.72	0.81	0.93
Var1	586–1446	(861)	55	55	4	51	0	0.023	0.015	1.89	1.35	1.84*
Var2	1447–2193	(747)	181	166	4	148	14	0.083	0.056	1.80	1.77**	2.13**

Extracellular, extracellular region; Nter, N-terminal segment; CRD, cysteine-rich domain; Var1, variable region 1; Var2, variable region 2. sites, nucleotide sites analyzed; η, the total number of mutations; *S*, number of segregating sites; π, observed nucleotide diversity; θ, the expected nucleotide diversity under neutrality derived from *S*. * indicates $p<0.05$ and ** indicates $p<0.02$. Sequence number is after 3D7 line sequence.

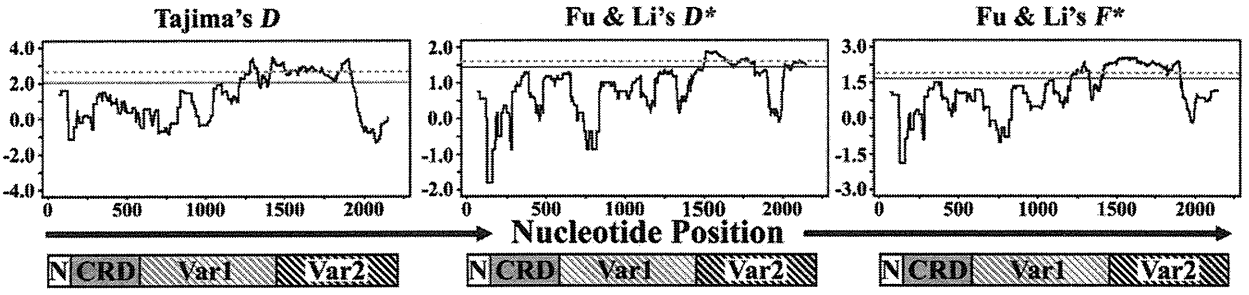

Fig. 2. Sliding window plots of Tajima's test (D) and Fu and Li's tests (D^* and F^*) for *Plasmodium falciparum* $surf_{4.2}$ sequence encoding extracellular region in Thai isolates. Forty-four samples for 1988/1989 group are used. Sites above the solid and dashed lines are significantly departed from neutrality (two-tailed; $P<0.05$ and $P<0.02$, respectively), indicating diversifying selection. Nucleotide numbers are after the 3D7 line sequence. Window length is 90 bp, and step size is 3 bp.

found also there were no statistically significant difference for the allele frequency distribution of $surf_{4.2}$ between two groups (Fig. 5), thus no obvious fluctuation of allele-frequency distribution were detected. This suggested that the frequency distribution of at least these four areas in the Var2 region were stable or did not change significantly in 14 years.

4. Discussion

In this study, we analyzed the extracellular domain-encoding nucleotide sequence of *P. falciparum* $surf_{4.2}$ for diversity, the signature of positive selection, the degree of linkage disequilibrium (LD), and temporal changes in allele frequency distribution in a *P. falciparum* population from Thailand. We found that the $surf_{4.2}$ gene of Thai isolates was highly polymorphic, particularly at the C-terminal side of the variable region (Var2 region), which is situated just before a predicted transmembrane region. Multiple tests detected the signature of positive selection on the Var2 region and, to a lesser extent, on the Var1 region, suggesting that they were potentially to be under host immune pressure. Similar experiments conducted for Kenyan isolates also showed similar tendency; positive selection toward the C-terminal side of extracellular region [5]. Although a significant departure from neutrality was not detected on the CRD, the nucleotide diversity of this region ($\pi=0.0170$) is comparable to the known malaria vaccine candidate antigen *ama-1* ($\pi=0.0166$), for which the action of the positive selection was previously detected [19]. Because sliding window analysis for Tajima's D, Fu and Li's D^* and F^* all showed negative values on some areas of CRD region (Fig. 2), we consider that a functional constraint likely prevented an extensive diversification of this region. Of note is that similar negative values were observed in the N-terminal side of Var1 region and this can be also seen in the analysis of Kenyan *P. falciparum* isolates [5]. The region between the CRD and transmembrane region of SURFIN members is highly diversified, and we expected that the variable (Var1 + Var2) region of $SURFIN_{4.2}$ has less functional

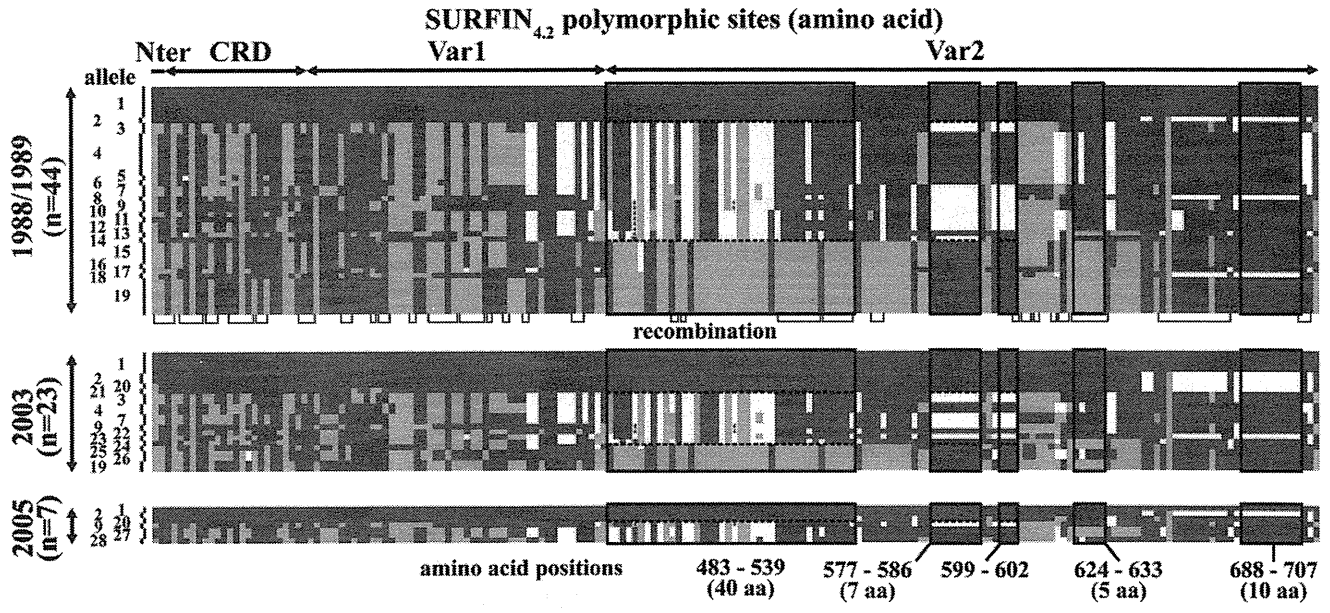

Fig. 3. Polymorphic amino acid sites of $SURFIN_{4.2}$ extracellular region among 74 Thai isolates. Only polymorphic amino acid sites were selected and aligned, then allelic type numbers were given to each sequence as follows; allele 1 (MS843, MS840, MS826, MS821, MS814A1, MS805, AA1329, AQ1132, TMPF44, AQ1125, and AQ1423), 2 (MS804, AQ1097a, AQ1099, and PA020), 3 (MS820b, MS819a, AQ1098, and AQ1142), 4 (MS844, MS946, MS818, MS816, MS815, MS813, MS808, MS824b, AQ1133, and AQ1097b), 5 (MS802b), 6 (MS811), 7 (MS809, MS833, TMPF18, and TMPF09), 8 (MS842), 9 (MS817, MS837, AQ1105b, and PA009), 10 (MS803a), 11 (MS827), 12 (MS807 and MS948), 13 (MS802a), 14 (MS829a), 15 (MS820a, MS819b, MS812, and MS825), 16 (MS828), 17 (MS947), 18 (MS806), 19 (MS831, MS830, MS829b, MS824a, MS803b, MS810, MS838, AQ1126, and AQ1129), 20 (TMPF34), 21 (AQ1139), 22 (TMPF11), 23 (AQ1127), 24 (AQ1105a, Q2D015, and TMPF338), 25 (AQ1130), 26 (AQ1101 and TMPF15), and 27 (PA021), 28 (AQ1459). MS843 sequence is used as a reference and shown with green color, because this sequence is most dominant (11/74; 14.9%). Amino acids different from MS843 are shown with pink or yellow for dimorphic sites, pink and yellow for trimorphic sites. Fourth or fifth substitutions are shown with cyan or blue colors and black dots. Singleton amino acid substitutions are shown with red color. Amino acid positions 483–539 are clustered and can be divided into 3 patterns; allele 1, 8, and 19 pattern (boxed and separated by dashed lines). These trends can be seen at the amino acid positions 577–586 and 599–602 (boxed and separated by dashed lines). Thin lines under the scheme for 1988/1989 group connect the sites for which recombination events are detected. Note that the recombination between amino acid positions 483–539 are only detected within the allele 8 group and 4th and 5th amino acids are seen in this group.

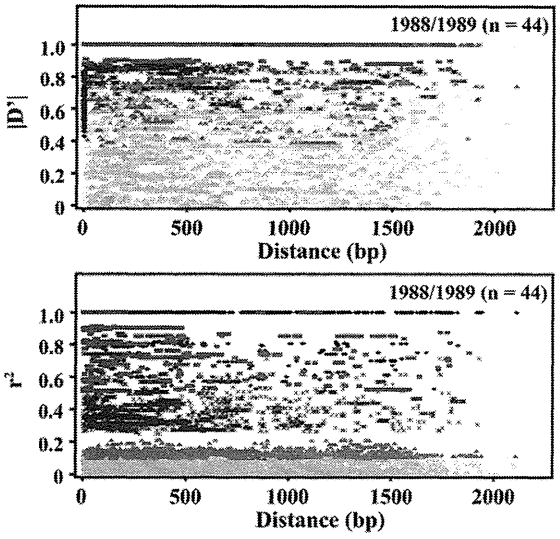

Fig. 4. Linkage disequilibrium (LD) of $surf_{4.2}$ region encoding extracellular region. LD is evaluated for 1988/1989 group using only informative sites after excluding sites with the frequency of the rare allele less than 10% and sites segregating for more than two nucleotides (number of polymorphic sites analyzed was 213). |D'| and r^2 were obtained by computing with DnaSP 5.0. Red filled square, $p<0.001$ after Bonferroni correction (more conservative); asterisks, $p<0.001$; black filled triangle, $0.001 \leq p<0.01$; dark gray filled triangle, $0.01 \leq p<0.05$; light gray open triangle, not significant. Correlation coefficients for |D'| and r^2 were −0.227 and −0.258, respectively.

constraint, thus the observed high diversity of C-terminal side Var1 and Var2 regions with the signature of positive selection is consistent with our expectation, whereas the negative values of Tajima's *D*, Fu and Li's D^* and F^* in N-terminal side of Var1 region is unexpected. Although Var1 region is not conserved among SURFIN members, this region might gain a functional role or structural importance during a shape-up process of current $SURFIN_{4.2}$ structure.

In contrast to the Kenyan *P. falciparum* $surf_{4.2}$ sequence, which showed 68 distinct allele in a total of 69 sequences, we found less alleles in Thai isolates (28 distinct alleles in a total of 74 sequences). Seven alleles (allele 1, 2, 3, 4, 7, 9, and 19) among 19 alleles detected in 1988/1989 were present in 2003/2005 sample collection. As recombination events were detected in both the Thai and the Kenyan sequences [5], we consider that the following factors contributed to this observation: 1) self-fertilization is dominant and frequency of the recombination of $surf_{4.2}$ gene locus is low, 2) epistatic relation between two sites with long distance (>1.5 kb), suggested by LD analysis in both the Thai and Kenyan sequences, suppresses the emergence of recombinant types of particular combinations, 3) high nucleotide diversity, especially in Var1 and Var2 regions may prevent efficient recombination. In the African population where recombination is more frequent than Asia, LD is less than Thai population. Thus in Thailand, even though recombination may occur, the effect of the epistatic relation and/or high diversity dominates the frequency of the recombination, and as a result same $surf_{4.2}$ alleles may circulate in the same population for long time.

Amino acid substitutions with an intermediate frequency are favored and expected to be selected under a frequency-dependent selection (balancing selection) and the selection pressure against $surf_{4.2}$ is likely to be host immunity. The frequency distribution of selected areas in Var2 region did not change for 14 years, suggesting that the allele frequency distribution of this region was stable. To explain this observation, we consider following 3 scenarios: 1) there is no pressure to maintain particular allele-frequency distribution of $SURFIN_{4.2}$ Var2 region, and the current distribution was formed by chance (neutral); 2) The cycle of fluctuation is longer than 14 years so that current data set is not sufficient to detect signature of fluctuation, if any (adapted with a fluctuation); and 3) the observed allele frequency of $SURFIN_{4.2}$ Var2 region has already adapted to the local environment (e.g., human immunity and/or human genetic background) in Thai *P. falciparum* population, and stably maintained (adapted without a fluctuation). Unless the structure of Thai *P. falciparum* population is dramatically changed, for example, by a genetic drift effect due to a small population size or a migration of new alleles, it is difficult to assess the first scenario. A stable allele frequency distribution of four putative neutral loci for 14 years also failed to rule out the first scenario, however, because of the positive diversifying selections were detected on this region, we consider scenarios 2 or 3 are more likely. To assess the possibility of second scenario, further study is required with more samples

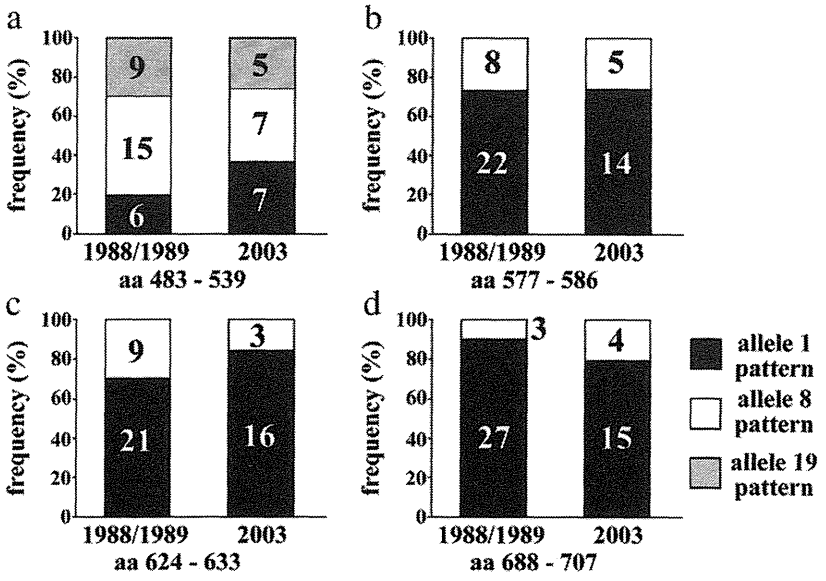

Fig. 5. Stable allele frequency distribution of the selected polymorphic areas of *Plasmodium falciparum* $SURFIN_{4.2}$ Var2 region in Thai isolates between 1988/1989 and 2003. Only sequences for which allele of the putatively neutral loci were determined are used (30 samples for 1988/1989 group and 19 samples for 2003 group). Allele frequencies are calculated for the regions where more than 4 amino acids are clustered (a, amino acid positions (aa) 483–539; b, aa 577–586; c, aa 624–633; d, and aa 688–707). Classification is detailed in Fig. 2. Allele frequency distributions are not significantly different between two groups (a, $p=0.42$; b, $p=0.98$; c, $p=0.32$; d, $p=0.41$). Chi-square test was used for (a) and (b), and Fisher's exact test was used for (c) and (d) (two-tail).

from different time points. The third scenario is appealing, because such temporally stable allele frequency distribution was also seen in most of the region of *P. falciparum* merozoite surface protein 1 (MSP1), a vaccine candidate antigen under diversifying selection, for 7 years in the Gambia [20] and for 10 years in Tanzania [21]. Allele frequency distribution of block 2 (most diversified region) of MSP1 was shown to be stable among *P. falciparum* populations in the different geographic areas and antibodies against this region were strongly associated with the protection against *P. falciparum* infection [22]. Spatially and temporally stable serotype frequency distribution can be observed for *Streptococcus pneumoniae*, a causative agent of the respiratory infection, however, several reports support that human immunity is unlikely to be the selection pressure [reviewed in 23]. Thus, although we consider that the selection pressure of the $SURFIN_{4.2}$ diversity is likely human immunity by observing its extensive diversity, it is formally possible that the other factor, such as host genetic background, may be responsible. Further studies are required if Var2 region of $SURFIN_{4.2}$ could be a target of the protective immunity and observed stable allele frequency distribution are maintained by allele-specific immunity, in addition to the identification of its biological role.

Acknowledgements

We thank S. Miyashita for her expertise. We are grateful to I. Sekine, head of the Nagasaki Red Cross Blood Center for human erythrocyte and plasma. This work was supported in part by Grants-in-Aids for Scientific Research 19590428 (to OK) and the Global COE Program, Nagasaki University (to OK) from the Ministry of Education, Culture, Sports, Science and Technology, Japan. The nucleotide sequence data reported in this paper are available in the GenBank™/EMBL/DDBJ databases under the accession numbers: AB679835–AB679908.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.parint.2011.12.003.

References

[1] World Health Organization. World Malaria Report 2010. Geneva, Switzerland: World Health Organization; 2010.
[2] Ferreira MU, da Silva Nunes M, Wunderlich G. Antigenic diversity and immune evasion by malaria parasites. Clinical and Diagnostic Laboratory Immunology 2004;11:987–95.
[3] Winter G, Satoru K, Haeggstrom M, Kaneko O, von Euler A, Kawazu S, et al. SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes. The Journal of Experimental Medicine 2005;201:1853–63.
[4] Janssen CS, Barrett MP, Turner CM, Phillips RS. A large gene family for putative variant antigens shared by human and rodent malaria parasites. Proceedings Biological Sciences / The Royal Society 2002;269:431–6.
[5] Ochola LI, Tetteh KK, Stewart LB, Riitho V, Marsh K, Conway DJ. Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in *Plasmodium falciparum*. Molecular Biology and Evolution 2010;27:2344–51.
[6] Nakazawa S, Culleton R, Maeno Y. In vivo and in vitro gametocyte production of *Plasmodium falciparum* isolates from Northern Thailand. International Journal for Parasitology 2011;41:317–23.
[7] Alexandre JSF, Kaewthamasorn M, Yahata K, Nakazawa S, Kaneko O. Positive selection on the *Plasmodium falciparum clag2* gene encoding a component of the erythrocyte-binding rhoptry protein complex. Tropical Medicine and Health 2011;39:77–82.
[8] Trager W, Jensen JB. Human malaria parasites in continuous culture. Science 1976;193:673–5.
[9] Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Molecular Biology and Evolution 2007;24:1596–9.
[10] Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 1986;3:418–26.
[11] Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 2009;25:1451–2.
[12] Tajima F. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 1993;135:599–607.
[13] Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics 1993;133:693–709.
[14] Maynard Smith J, Smith NH, Dowson CG, Spratt BG. How clonal are bacteria? Proceedings of the National Academy of Sciences of the United States of America 1993;90:4384–8.
[15] Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 1985;111:147–64.
[16] Haubold B, Hudson RR. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage Analysis. Bioinformatics 2000;16:847–8.
[17] Lewontin RC. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. Genetics 1964;49:49–67.
[18] Hill WG, Robertson A. Linkage disequilibrium in finite populations. Theoretical and Applied Genetics 1968;38:226–31.
[19] Polley SD, Conway DJ. Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. Genetics 2001;158:1505–12.
[20] Conway DJ, Greenwood BM, McBride JS. Longitudinal study of *Plasmodium falciparum* polymorphic antigens in a malaria-endemic population. Infection and Immunity 1992;60:1122–7.
[21] Tanabe K, Sakihama N, Rooth I, Björkman A, Färnert A. High frequency of recombination-driven allelic diversity and temporal variation of *Plasmodium falciparum msp1* in Tanzania. The American Journal of Tropical Medicine and Hygiene 2007;76:1037–45.
[22] Conway DJ, Cavanagh DR, Tanabe K, Roper C, Mikes ZS, Sakihama N, et al. A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. Nature Medicine 2000;6:689–92.
[23] Lipsitch M, O'Hagan JJ. Patterns of antigenic diversity and the mechanisms that maintain them. Journal of the Royal Society, Interface 2007;22:787–802.