

indicating that mutations of unknown genes (MODY-X) are responsible for 20–25% of MODY cases.¹⁹ The estimated prevalence of MODY-X is as many as 60–80% in Chinese,²⁰ Japanese^{21–23} and Korean families^{24,25} Owing to a different genetic background and absence of molecular genetic information regarding MODY in Thailand and Southeast Asia, we investigated the prevalence of mutations in six known genes responsible for MODY in Thai patients with MODY and early-onset type 2 diabetes.

Patients and methods

Patients and clinical data

Fifty-one unrelated probands with type 2 diabetes mellitus were recruited at the diabetic clinic, Siriraj Hospital, Bangkok, Thailand, according to the following criteria: (i) the proband and at least one first degree relative diagnosed with type 2 diabetes before age 35, (ii) two or more generations affected by diabetes, (iii) diabetes treatment with diet and/or oral agents for at least 2 years, (iv) no history of diabetic ketoacidosis (DKA), and (v) absence of anti-GAD antibody. Twenty-one of the probands were in concordance with strict MODY criteria.²² Non-diabetic subjects were 65 healthy staff members of the Department of Immunology and Department of Research and Development, Faculty of Medicine Siriraj Hospital. All of them had fasting plasma glucose (FPG) level < 5.6 mmol/l and had no history of diabetes in first-degree relatives. Approval for the study was granted by the Ethics Committee of the Faculty of Medicine Siriraj Hospital. All subjects were informed of the purpose and extent of the study before signing a consent form and enrollment into the study.

DNA was extracted from EDTA anti-coagulated peripheral blood by a standard phenol/chloroform method. Plasma glucose was determined by glucose oxidase method (Gluco-quant® Glucose/HK, Roche Diagnostics, Indianapolis, IN). Glycosylated haemoglobin (HbA1c) was measured by turbimetric inhibition immunoassay (Dimension® HA1C, Dade Behring Inc., DE). Total cholesterol and triglyceride were assayed by enzymatic calorimetric test (Cholesterol CHOD-PAP, Triglycerides GPO-PAP, Roche Diagnostics). HDL-cholesterol was determined by homogeneous enzymatic colourimetric test (HDL-C Plus 3rd Generation, Roche Diagnostics). LDL-cholesterol was calculated by Friedewald formula ($\text{LDL-cholesterol} = \text{total cholesterol} - \text{HDL-cholesterol} - \text{triglyceride}/5$) or measured directly when appropriate (LDL-C Plus second Generation, Roche Diagnostics). Anti-glutamic acid decarboxylase (GAD) antibody was measured by radioimmunoassay method (CIS Bio International ORIS Group Gif-Sur-Yvette Cedex, France). There was no interference with other autoantibodies such as insulin auto-antibodies (IAA). Insulin was determined by electrochemiluminescence assay (Insulin, Roche Diagnostics). C-peptide was measured by immunoradiometric assay (CIS Bio International ORIS Group Gif-Sur-Yvette Cedex, France).

Screening for sequence variations of six known genes responsible for MODY

All exons, flanking introns, 5' flanking and minimal promoter regions of *HNF-4α*, *GCK*, *HNF-1α*, *IPF-1*, *HNF-1β* and *NeuroD1* genes were amplified from a genomic DNA sample of each proband

by polymerase chain reaction (PCR). The PCR primers were designed by using Omega program version 2.0 (Molecular Ltd, Oxford, UK). Thermal cycling was carried out in a DNA Thermal Cycler 2400 (Perkin-Elmer, California, CA). The PCR products were analysed for sequence variation by single strand conformation polymorphism (SSCP) method using non-denaturing polyacrylamide gel electrophoresis and visualized by a standard silver staining protocol. The PCR product showing a mobility shift on the SSCP gel was subjected to direct sequencing by using ABI Prism BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, CA), as described in the manufacture's instructions. Fluorescent signals were detected with ABI Collection software in a computer connected to an automated sequencer. Nucleotide sequence was determined by Sequencer Navigator software and analysed by Chromas program version 1.4.4 (Conor McCarthy, Griffith University, Queensland, Australia).

Additional genotyping of identified sequence variations

Identified sequence variations which were present in at least one of 51 probands but absent in primary 15 non-diabetic controls were then subjected to genotyping in additional 50 non-diabetic controls by a PCR-restriction fragment length polymorphism (PCR-RFLP) method. If these variants did not create restriction sites for any restriction enzymes, new PCR primers containing introduced restriction sites were designed by a web-based program, dCAPS Finder 2.0 <<http://helix.wustl.edu/dcaps/dcaps.html>>, to analyse by RFLP technique. Variations that were not present in the non-diabetic controls were classified as potentially pathogenic mutations and were subjected to segregation analysis with diabetes in the affected families.

Statistical analysis

Comparisons of clinical characteristics of classic MODY probands and early-onset type 2 diabetic probands were analysed by using SPSS software (version 11.5; SPSS Inc., Chicago, IL). The significant level was determined at *P*-value of 0.05, by *t*-test or Mann-Whitney *U*-test.

Results

The clinical characteristics of the probands were summarized in Table 1. In contrast to patients with early-onset type 2 diabetes, the classic MODY patients were diagnosed at an earlier age. The diastolic blood pressure and serum cholesterol were significantly lower but the proportion of patients placed on insulin therapy was slightly higher and overall glycaemic control was poorer in the classic MODY probands compared to the patients with early-onset type 2 diabetes. The mean fasting C-peptide level of patients with early-onset type 2 diabetes was significantly higher than that of MODY.

Thirty-seven variations of six MODY genes were identified and sequence variants of *HNF-1α* gene were relatively more common among Thai patients studied (Table 2). Eleven of these variations were novel. Seventeen of these variations were not found in 30 chromosomes analysed in an initial group of non-diabetic controls, including *HNF-4α* R140Q and R312H; *GCK* -30 G > A, IVS +29 G > T and R327H; *HNF-1α* R203C, L281L, IVS5 +9 C > G, L459L,

Table 1. Clinical characteristics of patients classified as MODY and early-onset T2DM

	MODY	Early-onset T2DM	P value
Number of patients	21	30	
Age (years)	21.75 ± 9.66	36.88 ± 7.03	< 0.001*
Age at onset (years)	16.00 ± 5.06	31.25 ± 5.83	< 0.001*
Duration (years)	5.75 ± 7.47	5.75 ± 5.32	NS*
Biparental diabetes (n)	4	2	—
BMI (kg/m ²)	25.32 ± 6.27	27.54 ± 4.42	NS*
Waist (cm)	81.13 ± 10.65	89.31 ± 13.27	NS*
Waist : hip ratio	0.85 ± 0.08	0.89 ± 0.12	NS*
Systolic BP (mmHg)	116.25 ± 13.72	120.69 ± 9.21	NS*
Diastolic BP (mmHg)	70.42 ± 8.73	79.00 ± 9.24	0.020*
Retinopathy (n)	2	1	—
Nephropathy (n)	1	1	—
Neuropathy (n)	0	2	—
Macrovascular complication (n)	0	1	—
FPG (mmol/l)	11.26 ± 1.54	10.94 ± 0.92	NS*
HbA1c (%)	10.02 ± 4.03	8.21 ± 2.01	NS*
Total cholesterol (mmol/l)	4.54 ± 0.85	5.29 ± 0.64	0.019*
Triglyceride (mmol/l)	1.88 ± 0.44	2.76 ± 0.46	NS†
LDL (mmol/l)	2.80 ± 0.52	3.09 ± 0.4	NS*
HDL (mmol/l)	1.07 ± 0.09	1.18 ± 0.06	NS*
Fasting C-peptide (nmol/l)	0.49 ± 0.18	0.95 ± 0.43	0.028†

Note: No acanthosis nigricans was identified in any patients. Values are expressed as mean ± SD. Normal range for C-peptide assay 0.35–1.18 nmol/l. NS, not significant; FPG, fasting plasma glucose. Statistical method used: **t*-test, †Mann–Whitney *U*-test.

P475L, IVS7 +65 G > C, IVS7 +68 A > G, IVS8-19 G > A and G554fsX556; *HNF-1β* D82D; *NeuroD1* –303 G > A and A322N. These variations (excluding silent and intronic variants) were then subjected to genotyping in an additional 100 chromosomes of non-diabetic controls. Seven of them were not observed in the latter analysis and were classified as potentially pathogenic mutations (Table 3). No pathogenic mutation was found in *IPF-1* and *HNF-1β*.

HNF-1α R203C was previously reported in two families, one from Denmark²³ and another from Japan.²⁴ In our study this mutation transmitted from a grandmother, who has diabetes, to the proband (Fig. 1a). These data support segregation of the mutation with diabetes. However, the relatives of the proband who carried *HNF-4α* R312H were not available for this study. Thus, its role in the development of diabetes remains unclear.

Table 4 summarizes the clinical characteristics of patients carrying novel mutations of known MODY genes. There was no acanthosis nigricans and no biparental diabetes. None of them had chronic complications. Proband with novel *GCK* R327H was in the early-onset type 2 diabetes group and the BMI was obviously higher. However, the clinical characteristics were compatible with MODY2 phenotypes, including mild hyperglycaemia (FPG 7.5 mmol/l), and had good glycaemic control (HbA1c 6.7%) with diet control alone.^{11,25,26} The BMI of probands with novel *HNF-1α* and *NeuroD1* mutations were not obese by Asian standards except for the patient who carried *HNF-1α* P475L (BMI 24.49 kg/m²). The *HNF-1α* frameshift mutation (G554fsX556), segregated with diabetes in the family, transmitted from the mother to the proband (Fig. 1b).

However, relatives of the proband who carried *HNF-1α* P475L were not available for the study. Probands who carried *NeuroD1* –1972 G > A and A322N exhibited severe hyperglycaemia (FPG 15.04 and 17.76 mmol/l, respectively). However, *NeuroD1* A332N did not segregate completely with diabetes in the family (Fig. 1c) and relatives of the proband who carried *NeuroD1* –1972 G > A were not available for segregation analysis. Thus, the role of these sequence variations in glucose homeostasis remained to be explored.

Discussion

We have screened all six known genes responsible for MODY in 51 unrelated early-onset type 2 diabetic probands, 21 of them fitted into classic MODY criteria. By using PCR-SSCP analysis and direct sequencing, 37 sequence variations were identified and 7 of them were classified as potentially pathogenic mutations. Five of these mutations have not been previously reported. The average BMI of our MODY probands (25.32 ± 6.27 kg/m²) was modestly higher than the normal Asian standard. However, there is an overlap of BMI among MODY patients studied in Asia. Studies of Chinese MODY have indicated that the BMI of MODY3 patients was between 22 and 29 kg/m²^{20,21} while the BMI of MODY2 patients could be as high as 28 kg/m².²⁰ In the Korean population the average BMI of their patients being classified as MODY was approximately 22 kg/m².^{24,25} There is an epidemic of obesity in Asia and Thailand has a highest rate of obesity (6.8%), perhaps due to a greater level of adoption of western lifestyle.³² Furthermore, the prevalence of obesity and type

Table 2. Sequence variations identified in Thai subjects including 51 diabetic probands and 15 or 65 non-diabetic controls

Gene	Location	Nucleotide change	Amino acid change	Allele frequency	
<i>HNF-4α</i> (MODY1)	Promoter	−728 A > C	—	A 0.69	C 0.31
	Intron 1	IVS1-5 C > T	—	C 0.88	T 0.12
	Exon 4	12352 C > T	T139I	C 0.99	T 0.01
	Exon 4	12355 G > A†	R140Q	G 0.99	A 0.01
	Exon 8	22688 G > A	R312H	G 0.996	A 0.004
<i>GCK</i> (MODY2)	Promoter	−516 G > A	—	G 0.97	A 0.03
	Promoter	−680 A > G	—	A 0.96	G 0.04
	Intron 5	IVS5 +29 G > T	—	G 0.99	T 0.01
	Exon 8	42452 G > A†	R327H	G 0.996	A 0.004
	Intron 9	IVS9+8 C > T	—	C 0.58	T 0.42
<i>HNF-1α</i> (MODY3)	Exon 1	51 C > G	L17L	C 0.72	G 0.28
	Exon 1	79 A > C	I27L	A 0.73	C 0.27
	Intron 1	IVS1-42 G > A	—	G 0.68	A 0.32
	Intron 2	IVS2-51 T > A	—	T 0.58	A 0.42
	Exon 3	14832 C > T	R203C	C 0.996	T 0.004
	Exon 4	15523 C > T	L281L	C 0.99	T 0.01
	Exon 4	15546 G > C†	G288G	G 0.98	C 0.02
	Intron 5	IVS5 +9 C > G	—	C 0.99	G 0.01
	Intron 5	IVS5−42 G > T	—	G 0.86	T 0.14
	Exon 7	18771 C > T†	L459L	C 0.67	T 0.33
	Exon 7	18773 G > A	L459L	G 0.99	A 0.01
	Exon 7	18820 C > T†	P475L	C 0.99	T 0.01
	Exon 7	18856 G > A	S487N	G 0.67	A 0.33
	Intron 7	IVS7 +7 G > A	—	G 0.65	A 0.35
	Intron 7	IVS7 +65 G > C†	—	C 0.99	G 0.01
	Intron 7	IVS7 +68 A > G	—	A 0.99	G 0.01
	Intron 8	IVS8 −19 G > A†	—	G 0.99	A 0.01
	Exon 9	20750_20751ins AGTGAGTGAAGCCC†	G554fsX556	wt 0.996	ins 0.004
	Intron 9	IVS9−24 T > C	—	T 0.64	C 0.36
<i>IPF-1</i> (MODY4)	Promoter	−109_−110delG	—	G 0.71	delG 0.29
	Enhancer	−1867 G > A	—	G 0.64	A 0.36
<i>HNF-1β</i> (MODY5)	Exon 1	246 C > T†	D82D	C 0.99	T 0.01
	Intron 8	IVS8−22 C > T	—	C 0.96	T 0.04
<i>NeuroD1</i> (MODY6)	Promoter	−1631 A > G	—	A 0.81	G 0.19
	Promoter	−1972 G > A†	—	G 0.996	A 0.004
	Exon 2	133 G > A	A45T	G 0.84	A 0.16
	Exon 2	964 GC > AA†	A322N	GC 0.996	AA 0.004

Mutations which were not identified in 130 chromosomes of non-diabetic controls were classified as potentially pathogenic mutations, represented with bold characters. †Variations which were not previously reported.

2 diabetes in children and adolescents has also been increasing in Thailand.³³ It is possible that an interaction between genetic predisposition to diabetes and obesity occur in our MODY and early-onset type 2 diabetic patients. Unfortunately, we do not have any data regarding the patients' BMI before diagnosis of diabetes. There was no acanthosis nigricans, which is a feature of insulin resistance in all our patients being studied. Four of the 21 families conforming to strict MODY criteria have biparental diabetes. This scenario may be similar to the study by Xu *et al.*²¹ which reveals that these families may have early-onset type 2 diabetes and, through

chance, it also appeared as autosomal diabetes. The mean fasting C-peptide of our MODY probands was compatible with previous reports in other Asian MODY.^{20,21,24,25}

HNF-4α R312H is located in ligand-binding domain (LBD) of human HNF-4α. Arginine at this position is conserved across nine species. HNF-4α is a nuclear receptor that regulates many essential genes related to nutrient transport and metabolism.²⁷ Thus, alteration of amino acid at this position could potentially affect function of HNF-4α, such as DNA-binding, ligand-binding and interaction with co-regulatory molecules.²⁸

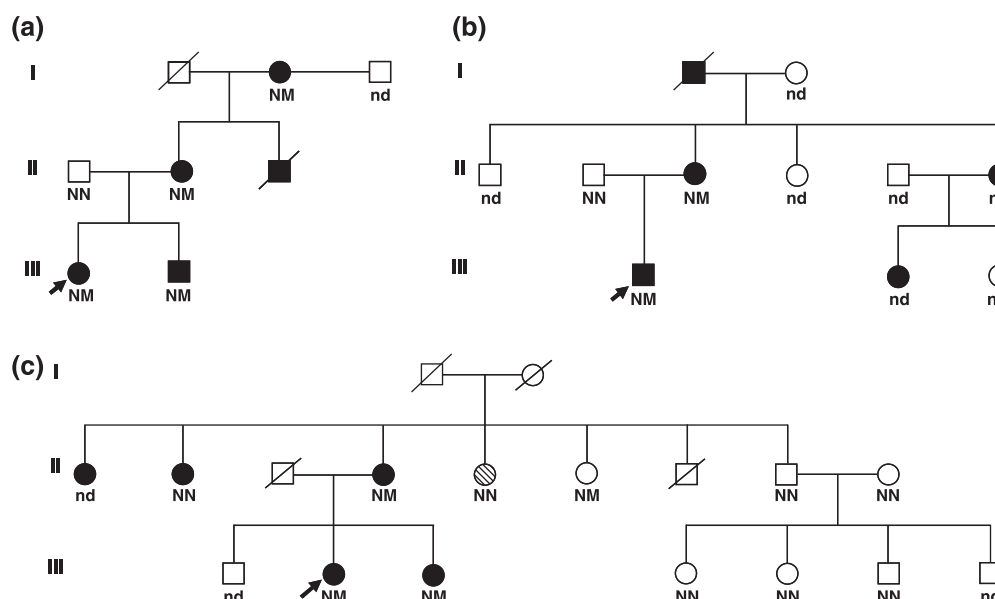


Fig. 1 Pedigrees of probands carried mutations of *HNF-1α* R203C (a), *HNF-1α* G554fsX556 (b), and *NeuroD1* A322N (c). 'N' refers to the normal allele and 'M' refers to a mutated allele. Symbols indicate the state of glucose tolerance: ○ and □, normal fasting glucose; ⊗, impaired glucose tolerance; ● and ■, diabetes; ○ and □ with 'nd', unknown.

Table 3. Possibly pathogenic mutations of four MODY genes observed in Thai patients

Gene	Location	Nucleotide change	Designation	Segregating with disease
<i>HNF-4α</i> (MODY1)	Exon 8	CGT > CAT	R312H‡	ND
<i>GCK</i> (MODY2)	Exon 8	CGC > CAC	R327H†‡	ND
<i>HNF-1α</i> (MODY3)	Exon 3	CGT > TGT	R203C‡	Yes
	Exon 7	CCG > CTG	P475L†‡	ND
	Exon 9	Insertion 14 nt	G554fsX556†	Yes
<i>NeuroD1</i> (MODY6)	Promoter	G > A	-1972 G > A†	ND
	Exon 2	GCT > AAT	A322N†‡	No

ND, not done. †Novel sequence variations. ‡Sequence variations alter amino acids that are conserved across species.

The arginine residue at position 203 of HNF-1α is conserved among species and located within DNA binding domain (DBD). Alteration from arginine to cysteine (HNF-1α R203C) should therefore affect *trans*-activation activity. Nevertheless, functional analysis of the mutant protein revealed biphasic activity in which a decreasing of activity was observed at a low concentration of transfected mutant gene, whereas an increase of function was demonstrated at a high level of transfected mutant gene.²⁴ In addition, the mutant protein showed weak but positive signal of nuclear localizing defects which could be involved, at least in part, in a biphasic effect on *trans*-activation activity. Thus, the precise mechanism of HNF-1α R203C in contributing to aetiology of MODY3 should be further elucidated.

Proline at position 475 and glutamine at position 554 of HNF-1α are conserved across several species. Substitution of proline by leucine at the position 475 (HNF-1α P475L) may affect the protein conformation. *HNF-1α* G554fsX556 occurred from a 14 nucleotide insertion at codon 554 and results in a frameshift alteration, giving

rise to a replacement of two amino acids (glycine and leucine by serine and glutamic acid) and an introduction of a stop codon (TGA) at the position 556. The mutant protein had a truncation of 76 amino acids at the C-terminus, compared with that of the wild-type HNF-1α protein. The effect of 76 amino acid deletion on intrinsic *trans*-activation of HNF-1α is currently under investigation by our group.

The *NeuroD1* -1972 G > A substitution occurred at a highly conserved promoter region of human and mouse genes which was identified as a critical region for basal transcriptional activation.²⁸ The *NeuroD1* A322N mutation causes an amino acid change in the region that associated with the co-activators CBP and p300. Alteration from alanine to asparagine might affect stability of the protein and might influence its association with co-activator proteins.

Although our sample size was rather small, such a low prevalence of mutation of known MODY genes in our patients is in concordance with the results from studies in other Asian populations. Enlarging the sample size in future studies should give more insight into

Table 4. Clinical characteristics of patients carrying novel mutations

	Subject† with GCK R327H	Subject† with HNF-1 α P475L	Subject‡ with HNF-1 α G554SfsX556	Subject‡ with NeuroD1 -1972 G > A	Subject‡ with NeuroD1 A322N
Age (years)	39	48	18	24	31
Age at onset (years)	39	31	12	20	14
Duration (years)	0	17	6	4	17
BMI (kg/m ²)	27.55	24.49	20.41	19.14	22.23
Waist (cm)	80	86	72	65	76
Waist : Hip ratio	0.83	0.89	0.77	0.76	0.83
Systolic BP (mmHg)	111	130	120	120	100
Diastolic BP (mmHg)	70	80	80	100	60
FPG (mmol/l)	7.49	6.94	9.27	15.04	17.76
HbA1c (%)	6.70	6.60	4.90	12.60	7.60
Fasting C-peptide (nmol/l)	NA	0.30	0.66	0.35	NA
Serum creatinine (μ mol/l)	61.88	106.08	61.88	53.04	88.40
Total Cholesterol (mmol/l)	5.98	5.10	3.55	5.83	8.62
Triglyceride (mmol/l)	0.90	0.76	0.43	1.33	3.60
LDL (mmol/l)	4.66	3.49	1.77	3.89	6.19
HDL (mmol/l)	1.14	1.26	1.58	0.78	0.78

NA, not available; FPG, fasting plasma glucose. Acanthosis nigricans was absent in all subjects. There was no biparental diabetes. Normal range for C-peptide assay 0.35–1.18 nmol/l. †early-onset type 2 diabetic proband. ‡MODY probands.

specific prevalence of MODY in Thailand. Another limitation of our study is the sensitivity of the PCR-SSCP method which was not determined in the present work. It is generally known, however, that the sensitivity of this method is approximately 80% or even greater if the size of PCR product falls within the length of 300 bp.^{34,35} Moreover, larger fragments have been successfully investigated by SSCP.³⁶ As most of our PCR fragments were shorter than 300 bp, we should have been able to detect the majority of sequence variations of the known MODY genes although there is a relatively small chance that we might have missed some variants.

As our early-onset type 2 diabetic patients had higher BMI and fasting C-peptide level, it is possible that insulin resistance may also play an important role in this group of patients, as was observed in Chinese²¹ and Caucasians.¹⁵ There is no prior study of molecular genetics of early-onset type 2 diabetes in Thailand. It would be interesting to explore this issue by investigating the roles of genes involved in β -cell function and insulin action in Thais with early-onset type 2 diabetes.

In summary, this is the first report on the prevalence of MODY subtypes in Thais, a Southeast Asian population. Despite the relatively small number of our patients, this data indicated the prevalence of genetic variability of known MODY genes in Thais. The possible pathogenic mutations of six known MODY genes account for only a small proportion of both classic MODY (19%) and early-onset type 2 diabetes (10%). We concluded that genetic variability of six known MODY genes is not a common cause of MODY and early-onset type 2 diabetes in Thais. Our results are consistent with that of the studies in Chinese,²⁰ Japanese²¹ and Korean.^{24,25} Thus, there are unidentified genes which await discovery in the majority of MODY patients in Asia. Identification of these novel genes will facilitate a better understanding of molecular mechanisms

underlying pathogenesis of MODY and type 2 diabetes and may lead to the development of more rational preventive and therapeutic strategies.

Acknowledgements

This work was financially supported by a Research Development Grant from the Faculty of Medicine Siriraj Hospital, Mahidol University, and Thailand Research Fund (TRF) (to NP and PY). PY is a TRF-Senior Research Scholar. WB and PJ are supported by the TRF-Royal Golden Jubilee PhD-Scholarship. Authors appreciate technical assistances given by Luksame Wattanamongkonsil, Kanjana Leejinda, Jatuporn Sujitjoon, Wiwit Tantibhedhyangkul, Thaniya Sricharunrat, Praneet Watanakejorn, Sirirat Ploybutr, Bangorn Chotisamer and Monchaya Tunlakit. Authors are grateful for valuable suggestions and comments from Dr Alessandro Doria M.D., Ph.D., Joslin Diabetes Center Harvard Medical School.

References

- 1 Fajans, S.S., Bell, G.I. & Polonsky, K.S. (2001) Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *New England Journal of Medicine*, **345**, 971–980.
- 2 Matschinsky, F.M. (2002) Regulation of pancreatic β -cell glucokinase: from basics to therapeutics. *Diabetes*, **51**(Suppl. 3), S394–S404.
- 3 Burke, C.V., Buettger, C.W., Davis, E.A. *et al.* (1999) Cell-biological assessment of human glucokinase mutants causing maturity-onset diabetes of the young type 2 (MODY-2) or glucokinase-linked hyperinsulinemia (GK-HI). *Biochemical Journal*, **342**, 345–352.
- 4 Byrne, M.M., Sturis, J., Clement, K. *et al.* (1994) Insulin secretory abnormalities in subjects with hyperglycemia due to glucokinase mutations. *Journal of Clinical Investigation*, **93**, 1120–1130.

- 5 Sturis, J., Kurland, I.J., Byrne, M.M. *et al.* (1994) Compensation in pancreatic β -cell function in subjects with glucokinase mutations. *Diabetes*, **43**, 718–723.
- 6 Frayling, T.M., Evans, J.C., Bulman, M.P. *et al.* (2001) β -cell genes and diabetes: molecular and clinical characterization of mutations in transcription factors. *Diabetes*, **50**(Suppl. 1), S94–S100.
- 7 Stoffel, M. & Duncan, S.A. (1997) The maturity-onset diabetes of the young (MODY1) transcription factor HNF4 α regulates expression of genes required for glucose transport and metabolism. *Proceedings of the National Academy of USA*, **94**, 13209–13214.
- 8 St-Onge, L., Wehr, R. & Gruss, P. (1999) Pancreas development and diabetes. *Current Opinion in Genetics and Development*, **9**, 295–300.
- 9 Clocquet, A.R., Egan, J.M., Stoffers, D.A. *et al.* (2000) Impaired insulin secretion and increased insulin sensitivity in familial maturity-onset diabetes of the young 4 (insulin promoter factor 1 gene). *Diabetes*, **49**, 1856–1864.
- 10 Froguel, P., Zouali, H., Vionnet, N. *et al.* (1993) Familial hyperglycemia due to mutations in glucokinase. Definition of a subtype of diabetes mellitus. *New England Journal of Medicine*, **328**, 697–702.
- 11 Velho, G., Blanche, H., Vaxillaire, M. *et al.* (1997) Identification of 14 new glucokinase mutations and description of the clinical profile of 42 MODY-2 families. *Diabetologia*, **40**, 217–224.
- 12 Beards, F., Frayling, T., Bulman, M. *et al.* (1998) Mutations in hepatocyte nuclear factor 1 β are not a common cause of maturity-onset diabetes of the young in the U.K. *Diabetes*, **47**, 1152–1154.
- 13 Lindner, T.H., Cockburn, B.N. & Bell, G.I. (1999) Molecular genetics of MODY in Germany. *Diabetologia*, **42**, 121–123.
- 14 Moises, R.S., Reis, A.F., Morel, V. *et al.* (2001) Prevalence of maturity-onset diabetes of the young mutations in Brazilian families with autosomal-dominant early-onset type 2 diabetes. *Diabetes Care*, **24**, 786–788.
- 15 Doria, A., Yang, Y., Malecki, M. *et al.* (1999) Phenotypic characteristics of early-onset autosomal-dominant type 2 diabetes unlinked to known maturity-onset diabetes of the young (MODY) genes. *Diabetes Care*, **22**, 253–261.
- 16 Costa, A., Bescos, M., Velho, G. *et al.* (2000) Genetic and clinical characterisation of maturity-onset diabetes of the young in Spanish families. *European Journal of Endocrinology*, **142**, 380–386.
- 17 Frayling, T.M., Bulman, M.P., Ellard, S. *et al.* (1997) Mutations in the hepatocyte nuclear factor-1 α gene are a common cause of maturity-onset diabetes of the young in the U.K. *Diabetes*, **46**, 720–725.
- 18 Massa, O., Meschi, F., Cuesta-Munoz, A. *et al.* (2001) High prevalence of glucokinase mutations in Italian children with MODY. Influence on glucose tolerance, first-phase insulin response, insulin sensitivity and BMI. *Diabetologia*, **44**, 898–905.
- 19 Chevre, J.C., Hani, E.H., Boutin, P. *et al.* (1998) Mutation screening in 18 Caucasian families suggest the existence of other MODY genes. *Diabetologia*, **41**, 1017–1023.
- 20 Ng, M.C., Cockburn, B.N., Lindner, T.H. *et al.* (1999) Molecular genetics of diabetes mellitus in Chinese subjects: identification of mutations in glucokinase and hepatocyte nuclear factor-1 α genes in patients with early-onset type 2 diabetes mellitus/MODY. *Diabetic Medicine*, **16**, 956–963.
- 21 Xu, J.U., Dan, Q.H., Chan, V. *et al.* (2005) Genetic and clinical characteristics of maturity-onset diabetes of the young in Chinese patients. *European Journal of Human Genetics*, **13**, 422–427.
- 22 Nishigori, H., Yamada, S., Kohama, T. *et al.* (1998) Mutations in the hepatocyte nuclear factor-1 α gene (MODY3) are not a major cause of early-onset non-insulin-dependent (type 2) diabetes mellitus in Japanese. *Journal of Human Genetics*, **43**, 107–110.
- 23 Tonooka, N., Tomura, H., Takahashi, Y. *et al.* (2002) High frequency of mutations in the HNF-1 α gene (TCF1) in non-obese patients with diabetes of youth in Japanese and identification of a case of digenic inheritance. *Diabetologia*, **45**, 1709–1712.
- 24 Hwang, J.S., Shin, C.H., Yang, S.W. *et al.* (2006) Genetic and clinical characteristics of Korean maturity-onset diabetes of the young (MODY) patients. *Diabetes Research and Clinical Practice*, **74**, 75–81.
- 25 Lee, H.J., Ahn, C.W., Kim, S.J. *et al.* (2001) Mutation in hepatocyte nuclear factor-1 α is not a common cause of MODY and early-onset type 2 diabetes in Korea. *Acta Diabetologica*, **38**, 123–127.
- 26 Frayling, T.M., Lindgren, C.M., Chevre, J.C. *et al.* (2003) A genome-wide scan in families with maturity-onset diabetes of the young: evidence for further genetic heterogeneity. *Diabetes*, **52**, 872–881.
- 27 Wang, H., Maechle, P., Antinozzi, P.A., Hagenfeldt, K.A. & Wollheim, C.B. (2000) Hepatocyte nuclear factor 4 α regulates the expression of pancreatic β -cell genes implicated in glucose metabolism and nutrient-induced insulin secretion. *Journal of Biological Chemistry*, **275**, 35953–35959.
- 28 Bogan, A.A., Dallas-Yang, Q., Ruse, M.D. Jr *et al.* (2000) Analysis of protein dimerization and ligand binding of orphan receptor HNF4 α . *Journal of Molecular Biology*, **302**, 831–851.
- 29 Johansen, A., Ek, J., Mortensen, H.B. *et al.* (2005) Half of clinically defined maturity-onset diabetes of the young patients in Denmark do not have mutations in HNF4A, GCK, and TCF1. *Journal of Clinical Endocrinology and Metabolism*, **90**, 4607–4614.
- 30 Yamada, S., Tomura, H., Nishigori, H. *et al.* (1999) Identification of mutations in the hepatocyte nuclear factor-1 α gene in Japanese subjects with early-onset NIDDM and functional analysis of the mutant proteins. *Diabetes*, **48**, 645–648.
- 31 Miyachi, T., Maruyama, H., Kitamura, T. *et al.* (1999) Structure and regulation of the human NeuroD (BETA2/BHF1) gene. *Brain Research Molecular Brain Research*, **69**, 223–231.
- 32 Yoon, K.H., Lee, J.H., Kim, J.W. *et al.* (2006) Epidemic obesity and type 2 diabetes in Asia. *Lancet*, **368**, 1681–1688.
- 33 Likitmaskul, S., Kiattisathavee, P., Chaichanwatanakul, K. *et al.* (2003) Increasing prevalence of type 2 diabetes mellitus in Thai children and adolescents associated with increasing prevalence of obesity. *Journal of Pediatric Endocrinology and Metabolism*, **16**, 71–77.
- 34 Hayashi, K. & Yandell, D.W. (1993) How sensitive is PCR-SSCP? *Human Mutation*, **2**, 338–346.
- 35 Hestekin, C.N. & Baron, A.E. (2006) The potential of electrophoretic mobility shift assays for clinical mutation detection. *Electrophoresis*, **27**, 3805–3815.
- 36 Jordanova, A., Kalaydjieva, L., Savov, A. *et al.* (1997) SSCP analysis: a blind sensitivity trial. *Human Mutation*, **10**, 65–70.



Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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



Functional defect of truncated hepatocyte nuclear factor-1 α (G554fsX556) associated with maturity-onset diabetes of the young

Suwattanee Kooptiwut^{a,*}, Jatuporn Sujitjoo^b, Nattachet Plengvidhya^{b,c}, Watip Boonyasrisawat^b, Nalinee Chongjaroen^b, Prapapron Jungtrakoon^b, Namoiy Semprasert^a, Hiroto Furuta^d, Kishio Nanjo^d, Napatawn Banchuin^b, Pa-thai Yenchitsomanus^{e,f}

^a Department of Physiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

^b Department of Immunology and Immunology Graduate Program, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

^c Division of Endocrinology and Metabolism, Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

^d The First Department, Wakayama Medical University, Japan

^e Division of Medical Molecular Biology, Medicine Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

^f Medical Biotechnology Unit, National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Bangkok, Thailand

ARTICLE INFO

Article history:

Received 5 March 2009

Available online 29 March 2009

Keywords:

Diabetes

Maturity-onset diabetes of the young

MODY

HNF-1 α

Frameshift mutation

Dual-luciferase assay

Thai

ABSTRACT

A novel frameshift mutation attributable to 14-nucleotide insertion in *hepatocyte nuclear factor-1 α* (HNF-1 α) encoding a truncated HNF-1 α (G554fsX556) with 76-amino acid deletion at its carboxyl terminus was identified in a Thai family with maturity-onset diabetes of the young (MODY). The wild-type and mutant HNF-1 α proteins were expressed by *in vitro* transcription and translation (TNT) assay and by transfection in HeLa cells. The wild-type and mutant HNF-1 α could similarly bind to human glucose-transporter 2 (GLUT2) promoter examined by electrophoretic mobility shift assay (EMSA). However, the transactivation activities of mutant HNF-1 α on human GLUT2 and rat *L-type pyruvate kinase* (L-PK) promoters in HeLa cells determined by luciferase reporter assay were reduced to approximately 55–60% of the wild-type protein. These results suggested that the functional defect of novel truncated HNF-1 α (G554fsX556) on the transactivation of its target-gene promoters would account for the β -cell dysfunction associated with the pathogenesis of MODY.

© 2009 Elsevier Inc. All rights reserved.

Introduction

Maturity-onset diabetes of the young (MODY) is a genetically heterogeneous monogenic form of diabetes characterized by an early onset (usually before 25 years), absence of ketosis, and an autosomal dominant pattern of inheritance [1–3]. Up to date, abnormalities of six known genes including *hepatocyte nuclear factor-4 α* (HNF-4 α), *glucokinase* (GCK), *HNF-1 α* , *insulin promoter factor-1* (IPF-1), *HNF-1 β* , and *neurogenic differentiation 1* (NeuroD 1), have been identified to be responsible for MODY subtypes 1–6, respectively [4]. The common subtype of MODY in Caucasian population is MODY3 which is caused by defects of HNF-1 α [5] while that in French population is MODY2 which results from defects of GCK [6]. It has been observed that MODY with unknown genetic defects (MODY X) in Asian populations, accounting for 60–80% of MODY cases [7,8], is more prevalent than that in Caucasian populations. Our previous work indicated that genetic mutations of the six known MODY genes may not be a major cause of MODY and

early-onset type 2 diabetes in Thais [9]. Nevertheless, mutations of HNF-1 α were detectable with a low frequency in Thai population [9].

HNF-1 α is a transcription factor that is expressed in liver, kidney, intestine, stomach, and pancreas [10–12]. HNF-1 α protein is composed of three functional domains: an N-terminal dimerization domain (residues 1–32), a DNA-binding domain (residues 150–280) with POU-like and homeodomain motif, and a C-terminal transactivation domain (residues 281–631). It regulates expression of several proteins including amylin [13], insulin [14], GLUT2 [15] and L-type pyruvate kinase (L-PK), 3-hydroxy-3-methylglutaryl coenzyme A reductase, mitochondrial 2-oxoglutarate dehydrogenase (OGDH) E1 subunit [16]. It also plays an important role in glucose metabolism and insulin secretion [17,18].

Recently, our group identified a novel frameshift mutation with 14-nucleotide insertion at codon 554 in exon 9 of HNF-1 α in a Thai family with MODY. This mutation resulted in a replacement of two amino acids (glycine and leucine by serine and glutamic acid) in the protein and an introduction of a stop codon (GTA) at the position 556 (G554fsX556). The encoded protein contained 555 amino acids with a truncation of C-terminus, compared to the normal

* Corresponding author. Fax: +66 2 4115009.

E-mail address: S_kooptiwut@hotmail.com (S. Kooptiwut).

full-length HNF-1 α protein. However, the functional effect of the truncated HNF-1 α to the β -cell function and pathogenesis of MODY is unknown. We therefore, investigated into binding and transactivation activities of the mutant HNF-1 α on two target-gene promoters in HeLa cells to understand its functional impact associated with the pathogenesis of MODY.

Materials and methods

Analysis of genes responsible for MODY. This project was approved by the Human Ethics Committee of the Faculty of Medicine Siriraj Hospital before its initiation. Fifty-one unrelated probands with MODY were recruited at the diabetic clinic, Siriraj Hospital, Thailand, according to the criteria as previously described [19].

Genomic DNAs were extracted from blood samples by a standard phenol/chloroform method. All exons, flanking introns, 5' flanking and minimal promoter regions of six known genes, *HNF-4 α* , *GCK*, *HNF-1 α* , *IPF-1*, *HNF-1 β* and *NeuroD1*, were analyzed by the polymerase chain reaction-single strand conformation polymorphism (PCR-SSCP). The PCR product showing a mobility shift on the SSCP gel was subjected to direct sequencing by using ABI Prism BigDye™ (Applied Biosystems, CA, USA), as described in the manufacturer's instructions.

Construction of a recombinant plasmid containing a novel frameshift mutation with 14-nucleotide insertion in HNF-1 α cDNA. A full-length, human wild-type (WT) HNF-1 α cDNA was cloned into pcDNA3.1 to create pcDNA3.1-HNF-1 α -WT. The 14-nt insertion at codon 554 of HNF-1 α (G554fsX556) was generated by using modified site-directed ligase-independent mutagenesis (SLIM) method described by Chiu et al. [20]. Briefly, two pairs of long-tailed (FL and RL) and short-tailed (Fs and Rs) primers were designed in order to insert 14 nucleotides at the codon 554 of the HNF-1 α cDNA construct. The two long-tailed primers contained overhanging 14-nucleotides that would be inserted into the HNF-1 α cDNA at their termini which were complementary to each other. The sequences of the two pairs of primers used for PCR were: FL 5'-AGTGAGTGAAGCCCGGGCTTCACACGCGGCATCT-3', Rs 5'-GGACTCACTGGAAGCTTCAGTGC-3', and Fs 5'-GGGCTTCACACGCGGCATCTCA G-3', RL 5'-GGGCTTCACTCACTGGACTCACTGGAAGCTTCA GT-3'; the underlined regions were the complementary 14 nucleotides to be inserted.

The linear PCR products from amplifications of the whole plasmid containing the 14-nt insert were obtained from two separate reactions using two different pairs of primers as described. The PCR products from two reactions were pooled together, denatured and re-annealed to allow formation of double stranded DNA. The re-annealed PCR products were transformed into competent *Escherichia coli* cells. The *E. coli* colonies were screened for the presence of the recombinant plasmid containing 14-nt insert within HNF-1 α cDNA, namely pcDNA3.1-HNF-1 α -G554fsX556. The recombinant plasmid was purified and its nucleotide sequence was examined by sequencing.

Construction of pGL3-Basic luciferase reporter vector containing human GLUT2 promoter or rat L-PK promoter. Human GLUT2 promoter region from nucleotides –1296 to +312 was isolated from human genomic DNA by PCR and cloned into pGL3-Basic luciferase reporter vector (Promega Corp., Madison, WI). Rat L-type pyruvate kinase (*L-PK*) gene was also regulated by HNF-1 α [21,22]. Thus, the rat *L-PK* promoter, between nucleotides –189 and +37 was amplified by PCR and cloned into pGL3-Basic luciferase reporter vector. Sequences of these constructs were verified by automated DNA sequencing.

In vitro expression of wild-type and mutant HNF-1 α . Wild-type HNF-1 α and mutant HNF-1 α (G554fsX556) proteins were synthesized by TNT assay (Promega) using pcDNA3.1-HNF-1 α -WT and

pcDNA3.1-HNF-1 α -G554fsX556 constructs. The proteins were separated on 10% SDS-PAGE and transferred onto a nitrocellulose membrane by electroblotting. The membranes were incubated with anti-HNF-1 α antibody (Santa Cruz Biotechnology), followed by horseradish peroxidase-conjugated anti-goat IgG (Santa Cruz Biotechnology), and the antibody binding was detected by using SuperSignal West Pico Chemiluminescence substrate (Pierce).

Expression of wild-type and mutant HNF-1 α in HeLa cells. HeLa cells were maintained in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal calf serum and transfected with 2 μ g each of pcDNA3.1, pcDNA3.1-HNF-1 α -WT, or pcDNA3.1-HNF-1 α -G554fsX556 construct using FUGENE® 6 transfection reagents (Roche Diagnostics). After 24 h, the transfected cells were lysed in M-PER reagent (Pierce) following the manufacturer's protocol. Then, 50 μ g of protein were subjected to 10% SDS-PAGE and transferred by electroblotting onto nitrocellulose membrane. The expressed proteins were detected by Western blot method using anti-HNF-1 α antibody as previously described.

Electrophoretic-mobility shift assay (EMSA). HeLa cells were transfected with either pcDNA3.1-HNF-1 α -WT or pcDNA3.1-HNF-1 α -G554fsX556 and cultured for 24 h. The nuclear proteins containing either wild-type or mutant HNF-1 α were extracted by using NE-PER extraction reagents (Pierce) as described in the manufacturer's protocol. The nuclear proteins were incubated with double-stranded oligonucleotide containing HNF-1 α -binding site from the human GLUT2 promoter sequences [23]; one strand of the oligonucleotides was labeled at its 5'-end with biotin (biotin-5'-tccttcTGCAATGCATACTAGGCctaggc-3'). The oligonucleotide-protein binding reaction was performed following the manufacturer's protocol. The oligonucleotide-protein complexes were separated on 5% non-denaturing polyacrylamide gel and subjected to electrophoresis in 0.5 $\times$ Tris-borate/EDTA. The complexes were transferred to the membrane, fixed, and detected by using LightShift Chemiluminescent EMSA kit (Pierce). To inhibit and confirm binding specificity, unlabeled double-stranded oligonucleotides was added into the binding reaction and then similarly processed. Electrophoretic mobility shift of the complexes, compared with the free biotinylated oligonucleotides that could be inhibited by unlabeled oligonucleotides, indicated the specific binding between the protein and biotinylated oligonucleotides.

Transactivation activities of wild-type and mutant HNF-1 α determined by luciferase reporter assay. HeLa cells were transfected with 500 ng of either pcDNA3.1-HNF-1 α -WT or pcDNA3.1-HNF-1 α -G554fsX556 using FUGENE® 6 transfection reagent (Roche Diagnostics) along with 100 ng of either pGL3-human GLUT2 or pGL3-rat *L-PK* promoter, and 10 ng of pRL-SV40 (to control transfection efficiency). After 24 h, the transactivation activities of the wild-type and mutant HNF-1 α proteins were measured by means of the Dual-Luciferase Reporter Assay System (Promega). For each plasmid construct, the experiments were repeated three times.

Statistical analysis. The data were analyzed by using SPSS software (version 11.5; SPSS Inc.), expressed as mean $\pm$ SD, and tested for their difference by unpaired *t*-test. The *p* value < 0.05 was considered significant.

Results and discussion

A novel frameshift mutation attributable to an insertion of 14 nucleotides in exon 9 of HNF-1 α introducing two amino-acid changes in the protein and creating an early stop codon at the position 556 (HNF-1 α -G554fsX556) was identified in a Thai family with MODY (Fig. 1). Thus, the mutant HNF-1 α protein was truncated by the deletion of 76 amino acids at its C-terminus. The mutation was observed in the male proband who had diabetes at the age of 12 years and also in his mother who had the disease

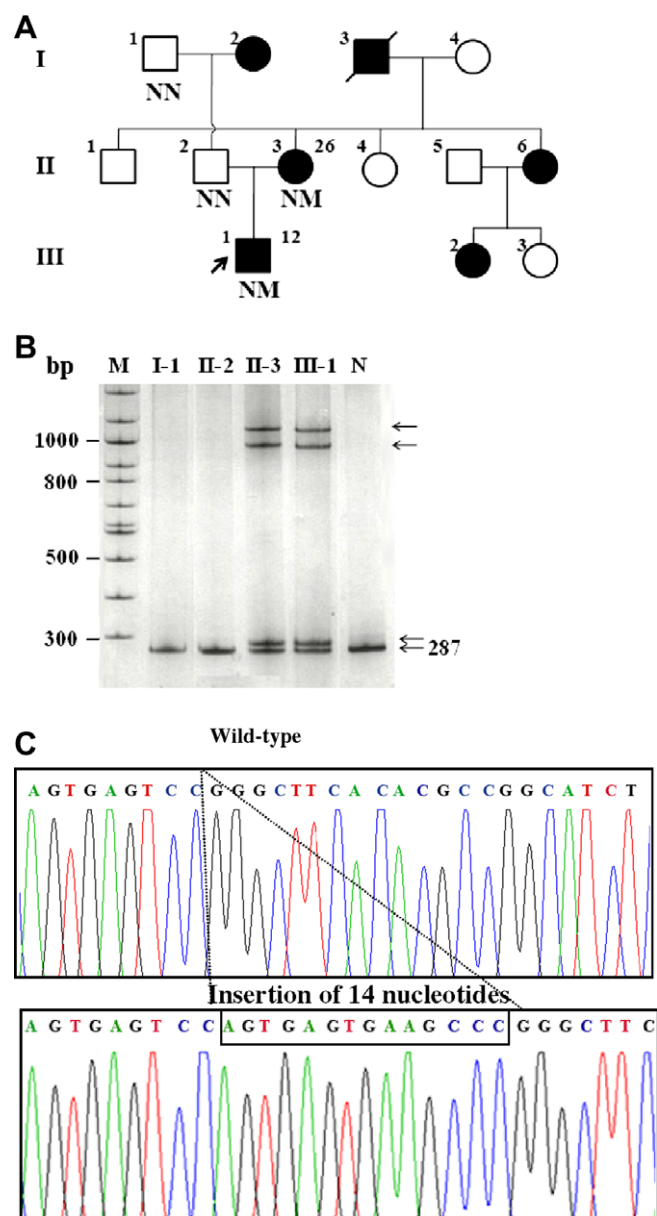


Fig. 1. (A) The pedigree of Thai family with MODY. *HNF-1α* G554fsX556 mutation was detected in the proband (III-1) and his affected mother (II-3) who had diabetes at the ages of 12 and 26 years, respectively. *Square* and *circular symbols* represent male and female members while *filled* and *open symbols* indicates diabetic and non-diabetic individuals, respectively. *Arrow* indicates the proband. *Roman numerals* on left side indicate the generation number and *Arabic numbers* on upper left of the symbols indicate individual numbers in the generation. Age at onset of diabetes is shown at the upper right of the symbol. The letters under the symbol indicate *HNF-1α* genotypes: N, normal; M, mutant. (B) The result of *HNF-1α* mutation screening (exon 9) by polymerase chain reaction-single strand conformation polymorphism (PCR-SSCP). Positive screening results were found in the PCR products from the mother (II-3) and proband (III-1) showing the mobility shift and double bands (heteroduplexes) of DNA. Lanes are indicated above the gel: M, ladder of DNA markers; I-1, II-2, II-3, and III-1 are the PCR products from the family members (the numbers are corresponding to those in the pedigree); N, normal. (C) Sequencing profiles of the exon 9 of *HNF-1α* in the mutation region: *top panel*, wild-type sequence; *lower panel*, mutant sequence with 14-nucleotide insertion indicated by box.

at the age of 26 years while his father who was healthy did not carry this mutation (Fig. 1). However, other members of the family were not available for the studies. This mutation was not observed in the screening of 214 non-diabetic subjects. Clinical findings of the proband, his mother and father were shown in Table 1.

Table 1
Clinical characteristics of the patient, affected mother, an healthy father.

	Proband	Mother	Father
Age (yrs)	18	50	52
Age at onset (yrs)	12	26	–
Duration (yrs)	6	24	–
BMI (kg/m ²)	20.41	22.51	18.34
Waist (cm)	79	81	77
Hip (cm)	95	93	89
Waist/Hip ratio	0.83	0.87	0.86
FPG (mmol/l)	13.20	11.71	6.33
Serum creatinine (μmol/L)	61.88	79.56	–
Total cholesterol (mmol/l)	3.55	4.82	–
Triglyceride (mmol/l)	0.43	0.723	–
LDL (mmol/l)	1.77	3.24	–
Treatment	OHA	OHA	–
Complications	–	Retinopathy mild neuropathy	–

Normal range of fasting plasma glucose < 5.6 mmol/l.
Normal range of serum creatinine: adult males, 70.72–123.76 mg/dl; adult females, 53.04–97.24 mg/dl.

To examine expression of the mutant *HNF-1α* (G554fsX556) protein and its functions in binding and transactivating target-gene promoters, we constructed the recombinant plasmid containing *HNF-1α* G554fsX556 cDNA for the studies. The protein expression was studied by TNT system and by HeLa cell transfection and detected by Western blot method. By both expression systems, the wild-type and mutant proteins were expressed to the predicted sizes of 68 and 59 kDa, respectively (Fig. 2A and B). In the transfected HeLa cells, we determined the expression of wild-type and mutant proteins in whole cell lysate in three independent experiments and found that the wild-type and mutant *HNF-1α* proteins had relatively similar quantities calculated as ratios with the densities of β-actin protein ($p = 0.219$) (Fig. 2B). When the cytoplasmic and nuclear compartments were separated, the wild-type and mutant proteins were found to be predominantly expressed in the nucleus (data not shown). This indicated that, similar to the wild-type *HNF-1α*, the mutant was stable and could migrate to localize in the nuclear compartment of the cells.

The binding abilities of the wild-type and mutant *HNF-1α* proteins expressed in HeLa cells to human *GLUT2* promoter containing *HNF-1α* binding site were determined by electrophoretic mobility shift assay (EMSA). The results showed that both wild-type and mutant proteins were able to bind and cause electrophoretic mobility shift of the biotinylated oligonucleotides (Fig. 2C), indicating that the binding ability of mutant protein to the *HNF-1α* binding site was still intact. In addition, their binding to the biotinylated oligonucleotides could completely be blocked by the addition of 200-fold excess of unlabeled oligonucleotide (Fig. 2C), confirming their binding specificity.

Since *HNF-1α* is a transcription factor that activates expression of *GLUT2* and *L-PK* genes [21,24], we examine whether the mutant *HNF-1α* protein, would exhibit any defect in activation of these two promoters. Human *GLUT2* or rat *L-PK* promoters cloned into pGL3-Basic luciferase reporter vector were co-transfected with either pcDNA3.1-*HNF-1α*-WT or pcDNA3.1-*HNF-1α*-G554fsX556 into HeLa cells, which has no endogenous *HNF-1α* [25]. The results showed that the wild-type *HNF-1α* could transactivate the human *GLUT2* and rat *L-PK* promoters, resulting in markedly increased luciferase activities. However, the mutant *HNF-1α* could not similarly transactivate these two promoters but only at the levels of 60% and 55% of that of the wild-type protein, respectively (Fig. 3A and B). The reduction of transactivation activity of the mutant protein is most likely due to the defect from the deletion of transactivation domain in its C-terminal region since similar amounts of wild-type and mutant proteins were expressed in the transfected HeLa cells as shown by the Western blot analysis

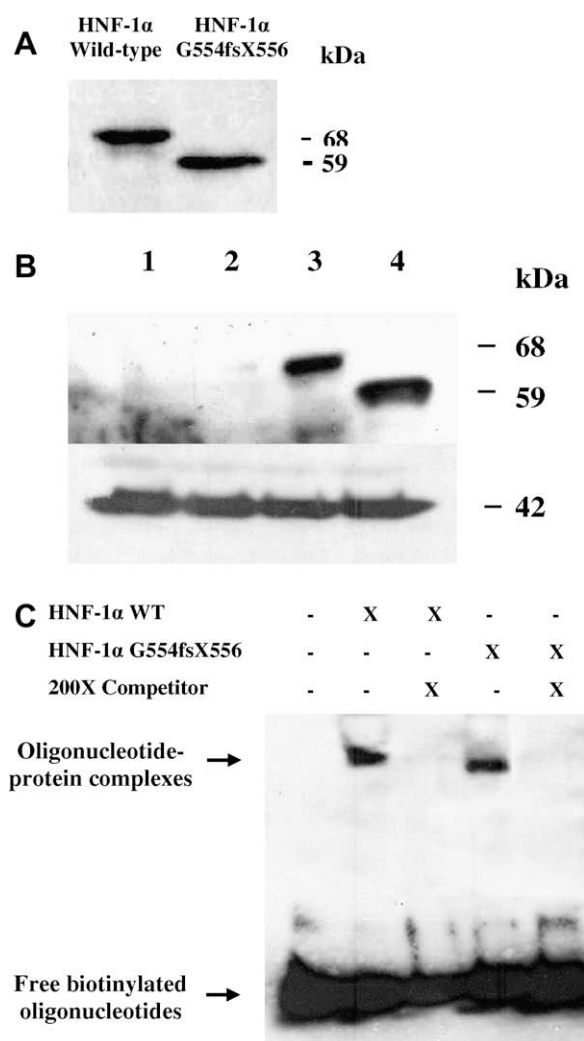


Fig. 2. (A) Western blot analysis of protein expression by TNT system. Equal amounts of proteins were loaded on SDS-PAGE, transferred onto nitrocellulose membrane, and detected by anti-HNF-1α antibody. The wild-type (HNF-1α WT) and mutant (HNF-1α G554fsX556) proteins were expressed at the molecular weights of 68 and 59 kDa, respectively. (B) Western blot analysis of protein expression in HeLa cells. Equal amounts of proteins were loaded on SDS-PAGE, transferred by onto nitrocellulose membrane, and detected by anti-HNF-1α antibody; lanes 1 and 2, empty vectors; lane 3, wild-type protein; lane 4, mutant protein. The wild-type and mutant proteins (in lanes 3 and 4) were expressed at the molecular weights of 68 and 59 kDa, respectively. Lower panel were β-actin protein with molecular weight of 42 kDa. (C) Electrophoretic mobility shift assay (EMSA) for analysis of the binding between wild-type or mutant HNF-1α protein to *GLUT2* promoter consisting of HNF-1α-binding sequence. The oligonucleotide-protein complexes were analyzed on 5% non-denaturing polyacrylamide gel. In the competitive assay, 200-fold excess of unlabeled double-stranded oligonucleotides (competitor) were added into the binding reaction. Both wild-type and mutant proteins (lanes 2 and 4) could bind to the biotinylated double-stranded oligonucleotides, which could be inhibited by the competitor (lanes 3 and 5).

(Fig. 2B) and both proteins could similarly attach to the HNF-1α binding site in the human *GLUT2* promoter (Fig. 2C).

The reduction in transcription activity of the mutant HNF-1α (G554fsX556) protein may down regulate the expression of its target genes such as *GLUT2* and *L-PK*. *GLUT2* is a major glucose transporter in pancreatic β-cells and plays an important role in glucose-induced insulin secretion [26]. Therefore, decrease in *GLUT2* expression may lead to the impairment of insulin secretion [27–29]. On the other hand, *L-PK* is an important enzyme in the glycolytic pathway in liver, kidney, small intestine and pancreatic β-cells [30]. The defect in *L-PK* expression may lead to defect in glucose metabolism that in-

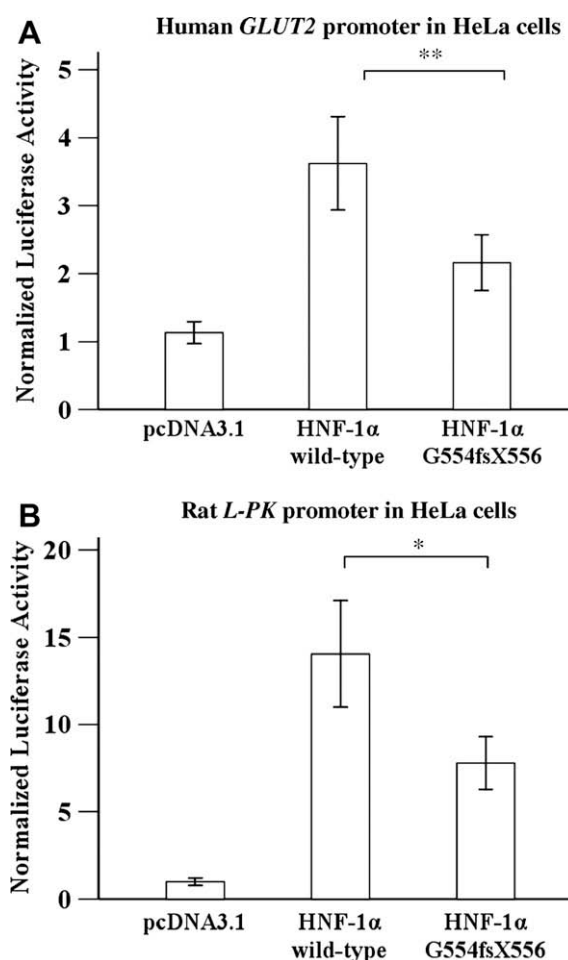


Fig. 3. (A) Transactivation activities of wild-type and mutant HNF-1α proteins on human *GLUT2* promoter in HeLa cells determined by dual-luciferase reporter assay. HeLa cells were transfected with pcDNA3.1-empty vector, pcDNA3.1-HNF-1α WT, or pcDNA3.1-HNF-1α G554fsX556 together with pGL3-human *GLUT2* vector and pRL-SV40. (B) Transactivation activities of wild-type and mutant HNF-1α proteins on rat *L-PK* promoter in HeLa cells determined by luciferase reporter assay. HeLa cells were transfected with pcDNA3.1-empty vector, pcDNA3.1-HNF-1α WT, or pcDNA3.1-HNF-1α G554fsX556 together with pGL3-rat *L-PK* vector and pRL-SV40.

volves in insulin secretion in pancreatic β-cells. Thus, the decrease in transactivation of the mutant HNF-1α (G554fsX556) protein on *GLUT2* and *L-PK* may result in the defect of insulin secretion and associated with MODY in the reported Thai family.

In conclusion, the HNF-1α G554fsX556 mutation identified in a Thai family with MODY should be pathogenic because although the mutant protein with truncation of the transactivation domain could be expressed and attach to the HNF-1α binding site, its transactivation activities on *GLUT2* and *L-PK* promoters was markedly decreased. This may explain the impairment of insulin secretion and insulin sensitivity leading to hyperglycemia in the patients with MODY caused by this pathogenic mutation.

Acknowledgments

This work was supported by Thailand Research fund (TRF), Siriraj Research Development Grant and Siriraj Chalearmprakit Fund (to SK), Siriraj Graduate Scholarship (to JS), Research Career Development Grant of Thailand Research Fund (TRF) (to NP), and TRF Senior Research Scholar Grant (to PY). We thank to Nuanghathai Sawasdee, and Malika Churintaraphan for technical assistances.

References

- [1] S. Fajans, G. Bell, D. Bowden, J. Halter, K. Polonsky, Maturity onset diabetes of the young (MODY), *Diabet. Med.* 13 (1996) S90–S95.
- [2] G. Velho, P. Froguel, Maturity-onset diabetes of the young (MODY), MODY genes and non-insulin-dependent diabetes mellitus, *Diabetes Metab.* 23 (Suppl 2) (1997) 34–37.
- [3] W.E. Winter, M. Nakamura, D.V. House, Monogenic diabetes mellitus in youth. The MODY syndromes, *Endocrinol. Metab. Clin. North Am.* 28 (1999) 765–785.
- [4] S.S. Fajans, G.I. Bell, K.S. Polonsky, Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young, *N. Engl. J. Med.* 345 (2001) 971–980.
- [5] F. Beards, T. Frayling, M. Bulman, Y. Horikawa, L. Allen, M. Appleton, G. Bell, S. Ellard, A. Hattersley, Mutations in hepatocyte nuclear factor 1beta are not a common cause of maturity-onset diabetes of the young in the UK, *Diabetes* 47 (1998) 1152–1154.
- [6] J. Hager, H. Blanché, F. Sun, N.V. Vaxillaire, W. Poller, D. Cohen, P. Czernichow, G. Velho, J.J. Robert, N. Cohen, P. Froguel, Six mutations in the glucokinase gene identified in MODY by using a nonradioactive sensitive screening technique, *Diabetes* 43 (1994) 730–733.
- [7] M.C. Ng, B.N. Cockburn, T.H. Lindner, V.T. Yeung, C.C. Chow, W.Y. So, J.K. Li, Y.M. Lo, Z.S. Lee, C.S. Cockram, J.A. Critchley, G.I. Bell, J.C. Chan, Molecular genetics of diabetes mellitus in Chinese subjects: identification of mutations in glucokinase and hepatocyte nuclear factor-1alpha genes in patients with early-onset type 2 diabetes mellitus/MODY, *Diabet. Med.* 16 (1999) 956–963.
- [8] H. Nishigori, S. Yamada, T. Kohama, T. Utsugi, H. Shimizu, T. Takeuchi, J. Takeda, Mutations in the hepatocyte nuclear factor-1 alpha gene (MODY3) are not a major cause of early-onset non-insulin-dependent (type 2) diabetes mellitus in Japanese, *J. Hum. Genet.* 43 (1998) 107–110.
- [9] N. Plengvidhya, W. Boonyasrisawat, N. Chongjaroen, P. Jungtrakoon, S. Sriussadaporn, S. Vannaseang, N. Banchuin, P. Yenchitsomanus, Mutations of maturity-onset diabetes of the young (MODY) genes in Thais with early-onset type 2 diabetes mellitus, *Clin Endocrinol (Oxf)*. (2008) [Epub ahead of print].
- [10] I. Bach, Z. Galcheva-Gargova, M.G. Mattei, D. Simon-Chazottes, J.L. Guenet, S. Cereghini, M. Yaniv, Cloning of human hepatic nuclear factor 1 (HNF1) and chromosomal localization of its gene in man and mouse, *Genomics* 8 (1990) 155–164.
- [11] C.J. Kuo, P.B. Conley, C.L. Hsieh, U. Francke, G.R. Crabtree, Molecular cloning, functional expression, and chromosomal localization of mouse hepatocyte nuclear factor 1, *Proc. Natl. Acad. Sci. USA* 87 (1990) 9838–9842.
- [12] J. Rey-Campos, T. Chouard, M. Yaniv, S. Cereghini, VHNF1 is a homeoprotein that activates transcription and forms heterodimers with HNF1, *EMBO J.* 10 (1991) 1445–1457.
- [13] L.M. Shepherd, S.C. Campbell, W.M. Macfarlane, Transcriptional regulation of the IAPP gene in pancreatic beta-cells, *Biochim. Biophys. Acta* 1681 (2004) 28–37.
- [14] K. Okita, Q. Yang, K. Yamagata, K.A. Hangenfeldt, J. Miyagawa, Y. Kajimoto, H. Nakajima, M. Namba, C.B. Wollheim, T. Hanafusa, Y. Matsuzawa, Human insulin gene is a target gene of hepatocyte nuclear factor-1alpha (HNF-1alpha) and HNF-1beta, *Biochem. Biophys. Res. Commun.* 263 (1999) 566–569.
- [15] J.Y. Cha, H. Kim, K.S. Kim, M.W. Hur, Y. Ahn, Identification of transacting factors responsible for the tissue-specific expression of human glucose transporter type 2 isoform gene. Cooperative role of hepatocyte nuclear factors 1alpha and 3beta, *J. Biol. Chem.* 275 (2000) 18358–18365.
- [16] H. Wang, P.A. Antinozzi, K.A. Hagenfeldt, P. Maechler, C.B. Wollheim, Molecular targets of a human HNF1 alpha mutation responsible for pancreatic beta-cell dysfunction, *EMBO J.* 19 (2000) 4257–4264.
- [17] H. Wang, K. Hagenfeldt-Johansson, L.A. Otten, B.R. Gauthier, P.L. Herrera, C.B. Wollheim, Experimental models of transcription factor-associated maturity-onset diabetes of the young, *Diabetes* 51 (Suppl 3) (2002) S333–S342.
- [18] K. Yamagata, T. Nammo, M. Moriwaki, A. Ihara, K. Iizuka, Q. Yang, T. Satoh, M. Li, R. Uenaka, K. Okita, H. Iwahashi, Q. Zhu, Y. Cao, A. Imagawa, Y. Tochino, T. Hanafusa, J. Miyagawa, Y. Matsuzawa, Overexpression of dominant-negative mutant hepatocyte nuclear factor-1 alpha in pancreatic beta-cells causes abnormal islet architecture with decreased expression of E-cadherin, reduced beta-cell proliferation, and diabetes, *Diabetes* 51 (2002) 114–123.
- [19] N. Plengvidhya, S. Kooptiwut, N. Songtawe, A. Doi, H. Furuta, M. Nishi, K. Nanjo, W. Tantibhedhyangkul, W. Boonyasrisawat, P.T. Yenchitsomanus, A. Doria, N. Banchuin, PAX4 mutations in Thais with maturity onset diabetes of the young, *J. Clin. Endocrinol. Metab.* 92 (2007) 2821–2826.
- [20] J. Chiu, P.E. March, R. Lee, D. Tillett, Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h, *Nucleic Acids Res.* 32 (2004) 174.
- [21] M. Parrizas, M.A. Maestro, S.F. Boj, A. Paniagua, R. Casamitjana, R. Gomis, F. Rivera, J. Ferrer, Hepatic nuclear factor 1-alpha directs nucleosomal hyperacetylation to its tissue-specific transcriptional targets, *Mol. Cell. Biol.* 21 (2001) 3234–3243.
- [22] K. Yamada, T. Noguchi, T. Matsuda, M. Takenaka, P. Monaci, A. Nicosia, T. Tanaka, Identification and characterization of hepatocyte-specific regulatory regions of the rat pyruvate kinase L gene. The synergistic effect of multiple elements, *J. Biol. Chem.* 265 (1990) 19885–19891.
- [23] N. Ban, Y. Yamada, Y. Someya, K. Miyawaki, Y. Ihara, M. Hosokawa, S. Toyokuni, K. Tsuda, Y. Seino, Hepatocyte nuclear factor-1alpha recruits the transcriptional co-activator p300 on the GLUT2 gene promoter, *Diabetes* 51 (2002) 1409–1418.
- [24] D.Q. Shih, S. Screenan, K.N. Munoz, L. Philipson, M. Pontoglio, M. Yaniv, K.S. Polonsky, M. Stoffel, Loss of HNF-1alpha function in mice leads to abnormal expression of genes involved in pancreatic islet development and metabolism, *Diabetes* 50 (2001) 2472–2480.
- [25] A. Nicosia, R. Tafi, P. Monaci, Trans-dominant inhibition of transcription activator LFB1, *Nucleic Acids Res.* 20 (1992) 5321–5328.
- [26] M.T. Guilleam, P. Dupraz, B. Thorens, Glucose uptake, utilization, and signaling in GLUT2-null islets, *Diabetes* 49 (2000) 1485–1491.
- [27] E. Chankiewitz, D. Peschke, L. Herberg, I. Bazwinsky, E. Mühlbauer, H. Brömme, E. Peschke, Did the gradual loss of GLUT2 cause a shift to diabetic disorders in the New Zealand obese mouse (NZO/HI)? *Exp. Clin. Endocrinol. Diabetes* 114 (2006) 262–269.
- [28] O. Laukkanen, J. Lindström, J. Eriksson, T. Valle, H. Hämäläinen, P. Ilanne-Parikka, S. Keinänen-Kiukaanniemi, J. Tuomilehto, M. Uusitupa, M. Laakso, F.D.P. Study, Polymorphisms in the SLC2A2 (GLUT2) gene are associated with the conversion from impaired glucose tolerance to type 2 diabetes: the Finnish Diabetes Prevention Study, *Diabetes* 54 (2005) 2256–2260.
- [29] S. Reddy, M. Young, C. Poole, J. Ross, Loss of glucose transporter-2 precedes insulin loss in the nonobese diabetic and the low-dose streptozotocin mouse models: a comparative immunohistochemical study by light and confocal microscopy, *Gen. Comp. Endocrinol.* 111 (1998) 9–19.
- [30] K. Yamada, T. Noguchi, Nutrient and hormonal regulation of pyruvate kinase gene expression, *Biochem. J.* 337 (1999) 1–11.



Interaction of dengue virus envelope protein with endoplasmic reticulum-resident chaperones facilitates dengue virus production

Thawornchai Limjindaporn^{a,b,c,d,*,1}, Wiyada Wongwiwat^{b,d,1}, Sansanee Noisakran^{d,g}, Chatchawan Srisawat^c, Janjuree Netsawang^{b,d}, Chunya Puttikhunt^{d,g}, Watchara Kasinrerker^{e,f}, Panisadee Avirutnan^{d,g}, Somchai Thiemmecha^d, Rungtawan Sriburi^h, Nopporn Sittisombut^{g,h}, Prida Malasit^{d,g}, Pa-thai Yenchitsomanus^{d,g}

^a Department of Anatomy, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

^b Department of Immunology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

^c Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

^d Office for Research and Development (Medical Molecular Biology Unit), Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

^e Department of Clinical Immunology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, Thailand

^f Biomedical Technology Research Center, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Bangkok, Thailand

^g Medical Biotechnology Unit, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Bangkok, Thailand

^h Department of Microbiology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

ARTICLE INFO

Article history:

Received 19 November 2008

Available online 25 December 2008

Keywords:

Dengue virus
Envelope protein
Endoplasmic reticulum
Chaperone
Replication

ABSTRACT

Dengue virus infection is an important mosquito-borne disease and a public health problem worldwide. A better understanding of interactions between human cellular host and dengue virus proteins will provide insight into dengue virus replication and cellular pathogenesis. The glycosylated envelope protein of dengue virus, DENV E, is processed in the endoplasmic reticulum of host cells and therefore reliant on host processing functions. The complement of host ER functions involved and nature of the interactions with DENV E has not been thoroughly investigated. By employing a yeast two-hybrid assay, we found that domain III of DENV E interacts with human immunoglobulin heavy chain binding protein (BiP). The relevance of this interaction was demonstrated by co-immunoprecipitation and co-localization of BiP and DENV E in dengue virus-infected cells. Using the same approach, association of DENV E with two other chaperones, calnexin and calreticulin was also observed. Knocking-down expression of BiP, calnexin, or calreticulin by siRNA significantly decreased the production of infectious dengue virions. These results indicate that the interaction of these three chaperones with DENV E plays an important role in virion production, likely facilitating proper folding and assembly of dengue proteins.

© 2008 Elsevier Inc. All rights reserved.

Introduction

Dengue virus (DENV) infection is one of the most important mosquito-borne viral diseases, which is endemic in many countries. Clinical severity ranges from febrile dengue fever (DF) to dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS). DENV belongs to the *Flaviviridae* family and contains a single positive-stranded RNA genome, encoding a single precursor polypeptide. Host and viral proteases cleave this polypeptide into three structural proteins (capsid, membrane and envelope) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) [1].

* Corresponding author. Address: Department of Anatomy, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand. Fax: +662 418 4793.

E-mail address: limjindaporn@yahoo.com (T. Limjindaporn).

¹ These authors contribute equally to this work.

Replication of flaviviruses occurs in association with the endoplasmic reticulum where virions assemble and bud into the lumen of the ER. Virus particles transit through the Golgi where they undergo maturation prior to being released by exocytosis [1]. DENV envelope protein (DENV E) is the major component of the virion surface contains two functionally important glycosylation sites, Asn-67 and Asn-153. DENV E lacking Asn-67 was able to infect cells but the production of new infectious particles was abolished. In addition, DENV E lacking Asn-153 showed reduced infectivity [2]. Although DENV E biosynthesis and assembly is thought to occur in the ER, limited information is available on the involvement of host ER chaperones in the folding and assembly of DENV E. We therefore performed a yeast two-hybrid assay to identify host proteins that interact with DENV E. One of the proteins identified in this screen was human immunoglobulin heavy chain binding protein (BiP), a member of the heat shock protein 70 family found in the ER lumen, the interaction between DENV E and BiP in

dengue virus-infected Vero cells was confirmed by co-immunoprecipitation and co-localization studies. Two additional ER-resident chaperones, calnexin and calreticulin, were similarly examined and both exhibited co-immunoprecipitation and co-localization with DENV E. Significantly, knocking down the expression of BiP, calnexin, or calreticulin by siRNA decreased the yield of infectious dengue virions, indicating that the role of these ER-resident chaperones in the folding and assembly of DENV E is essential for dengue virion production.

Materials and methods

Yeast two-hybrid screening. Two-hybrid screening was performed by the interaction mating method as described by Finley and Brent [3]. Domain III of DENV E, nucleotides 1819 to 2118 encoding amino acids 295 to 394, was amplified from pBluescript II KS (S1SP6-4497), a plasmid which contains the 5' portion of DENV serotype two strain 16681 [4] by a pair of nucleotide primers 5' AAG CCG GAA TTC AAA GGA ATG TCA TAC 3' for the forward direction and 5' GCC CGC GGA TCC CTA TTT CTT AAA CCA G 3' for the reverse direction. Domain III is an immunoglobulin-like domain and does not contain a transmembrane domain [5]. The PCR reaction was carried out in a GeneAmp PCR System 9700 (Applied Biosystems) consisting of 35 cycles of denaturation at 94 °C for 45 s, annealing at 55 °C for 45 s and extension at 72 °C for 2 min. Subsequently, the amplified DNA was cloned as an in frame fusion with the LexA DNA binding domain in the yeast expression vector pEG202, which contains a *HIS3* selectable marker. The resulting bait plasmid, pEG-E, was transformed into *S. cerevisiae* strain RFY 206 (MATa *his3A200 leu2-3 lys2A201 ura3-52 trp1A::hisG*) containing a Lexop-*lacZ* reporter plasmid, pSH18-34, under *URA3* selection. A galactose-inducible HeLa cell cDNA prey library was constructed in plasmid pJZ 4-5 containing a *TRP1* selectable marker and transformed into strain RFY 231 (MATa *his3 leu2::3-Lexop-LEU2 ura3 trp1 LYS2*) [6]. The bait strain was mated with the library strains and plated on galactose drop-out medium lacking histidine, tryptophan, uracil and leucine (gal/raf-u, -h, -w, -l) to select for diploids. The production of a DENV E binding protein by a prey plasmid was expected to activate the 3Lexop-*LEU2* reporter. Putative positive clones were patching to four indicator plates: (glu/-u, -h, -w, -l), (gal/raf-u, -h, -w, -l), (glu/X-Gal-u, -h, -w) and (gal/raf/X-Gal-u, -h, -w). Prey plasmids were rescued from clones exhibiting a galactose-inducible *Leu⁺ lacZ⁺* phenotype by transformation into a *Trp⁻ Escherichia coli* strain KC8 [3]. To verify the interaction, recovered prey plasmids were introduced into yeast strain RFY 231 along with the *lacZ* plasmid and bait plasmid and again tested on the indicator plates.

Co-immunoprecipitation. Sub-confluent monolayers of 1×10^7 Vero cells in a 100-mm dish were infected with DENV serotype two strain 16681 at an MOI of 1 for 48 h. Cells were washed twice with 5 ml of PBS and detached by incubating with 2.5 mM EDTA in PBS for 15 min. Detached cells were collected by centrifugation. The cell pellets were lysed with a buffer containing 150 mM NaCl, 20 mM Tris-HCl pH 7.4, 5 mM EDTA, 0.5% deoxycholate, 0.1% SDS, 1% Triton X-100 and protease inhibitors (complete, EDTA-free, Roche) and the lysate was incubated on ice for 20 min. Cell debris was removed by centrifugation at 13,000g for 10 min at 4 °C. Five microgram of goat anti-BiP antibody, goat anti-calnexin antibody, goat anti-calreticulin antibody or an isotype-matched control antibody (Santa Cruz Biotechnology) were added to lysates and incubated 4 h in the presence of Protein G Sepharose beads (Amersham Pharmacia Biosciences). Subsequently, the Protein G Sepharose beads were collected by centrifugation at 13,000g for 5 min and washed twice with washing buffer. Lastly, the bound proteins were eluted by boiling in SDS-PAGE loading buffer, separated by SDS-PAGE, and transferred to nitrocellulose membranes.

After blocking with 5% skim milk, the membranes were incubated with mouse anti-DENV E monoclonal antibody (3H5) [7,8] followed by probing with horseradish peroxidase (HRP)-conjugated rabbit anti-mouse antibody (Dako). The protein bands were detected using ECL reagent (Amersham Pharmacia Biosciences).

Co-localization. Vero cells were grown on cover slips and infected with DENV serotype two strain 16681 at an MOI of 1 for 48 h. Thereafter, the cells were washed, fixed with 4% paraformaldehyde in PBS for 20 min, and permeabilized with 0.2% Triton X-100 in PBS for 10 min at room temperature. After washing three times with 0.1% Triton X-100 in PBS, the cover slips were incubated for an hour with (i) mouse anti-DENV E antibody (3H5) and goat anti-BiP antibody (Santa Cruz Biotechnology) (ii) mouse anti-DENV E antibody (3H5) and rabbit anti-calnexin antibody (Santa Cruz Biotechnology) (iii) mouse anti-DENV E antibody (3H5) and goat anti-calreticulin antibody (Santa Cruz Biotechnology). After washing, the cover slips (i and iii) were incubated with both Alexa 488-conjugated donkey anti-goat antibody and Alexa 594-conjugated donkey anti-mouse antibody (Molecular Probes) as secondary antibodies at room temperature for an hour. Cover slip (ii) was incubated with both Alexa 488-conjugated donkey anti-mouse antibody and Cy3-conjugated donkey anti-rabbit antibody (Molecular Probes) as secondary antibodies. Florescent images were captured by a confocal microscope (model LSM 510 Meta, Carl Zeiss).

Knock-down experiments by siRNA and infection assays. The BiP siRNA (5' GCGGAACCTTCGATGTGTCTCTCT 3'), calnexin siRNA (5' ATA GAATGTGGTGGTGCCTATGTGA 3'), calreticulin siRNA (5' CCCGCTGG ATCGAATCCAAACACAA 3') were purchased from Invitrogen, USA and used to knock-down BiP, calnexin and calreticulin by transfection into Vero cells using Lipofectamine™ 2000 reagent (Invitrogen, USA). Transfection with irrelevant siRNA (Invitrogen, USA, Cat. No. 12935-300) was performed as a negative control. After 6 h of transfection, cells were fed with 10% FBS in MEM medium for 30 h. Samples were taken for mRNA and protein analysis using real-time PCR (Lightcycler RNA amplification kit, Roche) and Western blot analysis, respectively. Then, siRNA-transfected cells were infected with DENV serotype two strain 16681 at an MOI of 1 for 3 h. The infected siRNA-transfected cells were washed with PBS and fed with 2% FBS in MEM medium for 24 h. The culture supernatants were collected to measure the amount of DENV production by a focus forming unit (FFU) assay as previously described [9].

Results and discussion

DENV E interacts with BiP in a yeast two-hybrid system

To identify human proteins that interacted with DENV E, we screened over 10^7 clones from a HeLa cDNA library using DENV E as bait. Forty five putative positive clones were obtained. Sequence analysis of the inserts showed that three of the cDNA inserts were identical and encoded amino acids 467–655 of BiP. BiP, also known as glucose regulated protein (GRP78), is an isoform five of the heat shock protein 70, which functions as a molecular chaperone involving in folding and assembly of several cellular and viral membrane proteins [10–14]. The specificity of the interaction between DENV E and BiP is shown in Fig. 1A wherein cells containing the DENV E bait plasmid and BiP prey plasmid exhibited galactose-dependent leucine prototrophy and *lacZ* expression. As controls, cells containing the DENV C bait plasmid and BiP prey plasmid or cells containing the DENV NS5 bait plasmid and BiP prey plasmid did not exhibit galactose-dependent leucine prototrophy and *lacZ* expression.

BiP, in complex with other ER chaperones, facilitates the proper folding of proteins in the secretory system. It has two distinct functional regions. The amino-terminal region of BiP possesses an

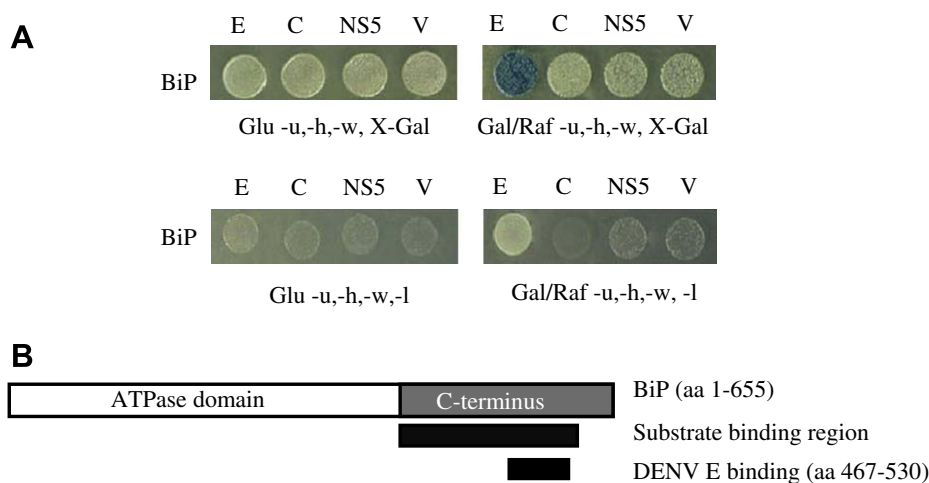


Fig. 1. DENV E-human BiP interaction in a yeast two-hybrid system. (A) Yeast strain RFY231 was co-transformed with a bait plasmid, a BiP prey plasmid and *lacZ* reporter plasmid. The bait plasmids used were pEG-E, expressing the LexA-DENV E fusion protein (E), pEG-C expressing the LexA-DENV C fusion protein (C), pEG-NS5 expressing the LexA-DENV NS5 fusion protein (NS5), and the empty bait plasmid pEG202, expressing the LexA fusion protein (V). A specific interaction was indicated by galactose-dependent β -galactosidase expression, as evidenced by blue colonies on the galactose containing X-Gal plate and white colonies on the leucine deficient plate. (B) The region of BiP that interacted with DENV E was located between the amino acid positions 467–530 in the substrate-binding domain of the carboxyl-terminal region of BiP. (For interpretation of color mentioned in this figure the reader is referred to the web version of the article.)

ATPase activity [15,16] whereas its carboxyl-terminal region contains an 18 kDa substrate-binding domain and a 10 kDa oligomerization domain [17,18]. The region of BiP that interacted with DENV E in yeast two-hybrid screening was further mapped by yeast two-hybrid system and located between amino acids 467–530 in the substrate-binding domain of the carboxyl-terminal region of BiP (Fig. 1B).

DENV E associates with BiP, calnexin and calreticulin in dengue virus-infected cells

In order to confirm DENV E-BiP interaction in mammalian cells, lysates from dengue virus-infected Vero cells were tested for co-immunoprecipitation of the proteins. As shown in Fig. 2A,

immunoprecipitation with anti-BiP antibody pulled down DENV E protein suggesting the association of DENV E and BiP in dengue virus-infected cells. Furthermore, co-localization of DENV E and BiP in the ER was evident when DENV E and BiP fluorescence images were superimposed (Fig. 2B). In hepatitis B virus (HBV), BiP interacts with the large surface protein (L) and plays a role in HBV morphogenesis by regulating proper folding of the L protein and assembly of the envelope protein [19]. The envelope protein of HIV type 1 has also been shown to interact with BiP [11]. Thus the interaction of DENV E and BiP may contribute to DENV morphogenesis by regulating the correct folding of DENV E.

Based on the immunofluorescence intensity of the images in Fig. 2, BiP appeared to be strongly induced by DENV infection. This is consistent with previous studies showing that DENV infection

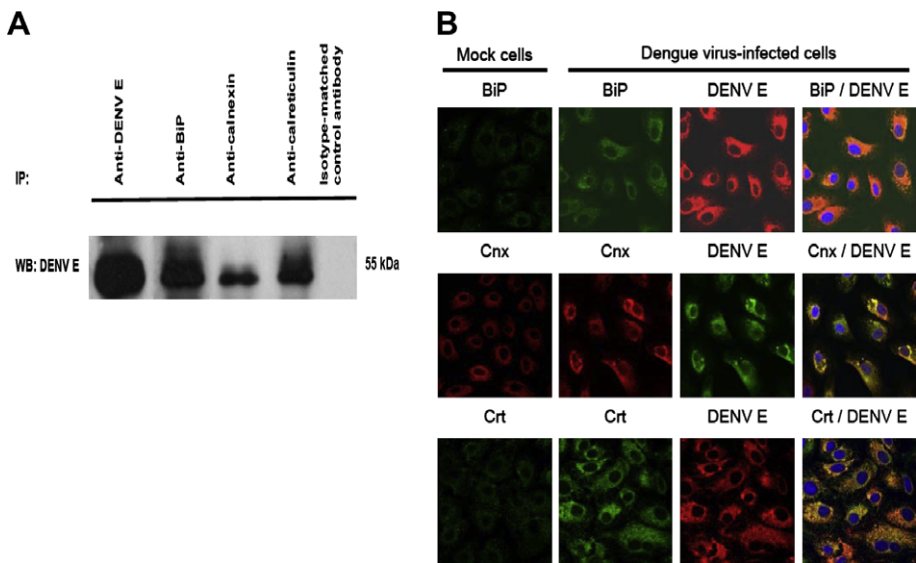


Fig. 2. Interaction of DENV E with human BiP, calnexin and calreticulin in dengue virus-infected cells. (A) Lysates of dengue virus-infected Vero cells were immunoprecipitated with mouse anti-DENV E antibody (lane 1), goat anti-BiP antibody (lane 2), goat anti-calnexin antibody (lane 3), goat anti-calreticulin antibody (lane 4) and goat isotype-matched control antibody (lane 5). Immune complexes were detected by Western blot analysis using anti-DENV E monoclonal antibody. (B) Dengue virus-infected Vero cells at 48 h after infection were subjected to double immunofluorescence staining for DENV E and chaperones (BiP, calnexin [Cnx] and calreticulin [Crt]) and observed for their co-localization under a laser-scanning confocal microscope. Mock-infected cells served as controls.

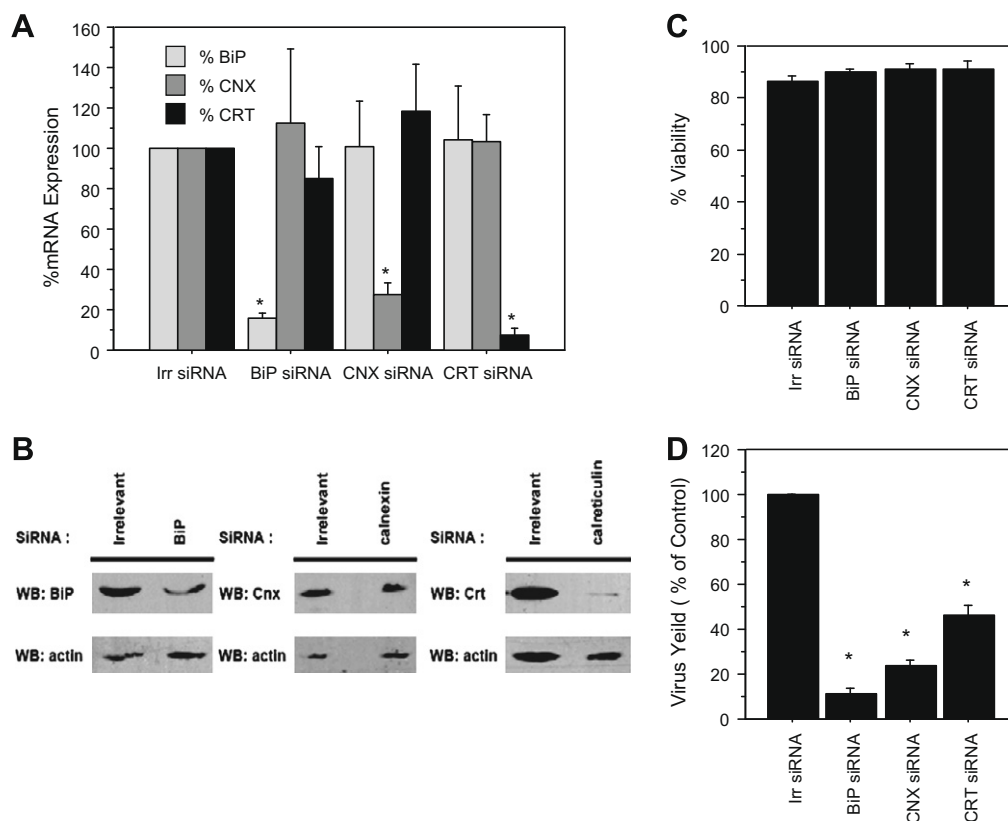


Fig. 3. Knocking-down the expression of ER chaperones by siRNA decreases the yield of infectious dengue virus production. Vero cells were transfected with siRNA against BiP, calnexin, calreticulin or an irrelevant siRNA (negative control) and harvested at 36 h post-transfection. The mRNA and proteins were measured by (A) real-time PCR and (B) Western blot analysis using polyclonal antibodies specific to BiP, calnexin and calreticulin, respectively. The decreased expression of BiP, calnexin and calreticulin was observed in the corresponding siRNA-transfected Vero cells relative to irrelevant siRNA-transfected control cells ($p \leq 0.05$). The viability of transfected cells with indicated siRNA was measured by trypan blue staining (C). Transfected cell lysates were then infected with dengue virus at an MOI of 1. At 24 h post-infection, the supernatants were collected to measure the amount of dengue virion production by focus forming unit (FFU) assay. The virus yield is expressed as a percentage of the yield obtained from cells transfected with irrelevant siRNA (D). The siRNAs to BiP, calnexin and calreticulin decreased the yield of viral progeny, relative to irrelevant siRNA-transfected control cells ($p \leq 0.05$). Results are derived from three independent experiments.

induces the unfolded protein response [20,21] and that BiP is induced by other flaviviruses [22,23].

Interestingly, cell surface associated BiP was previously identified as a putative dengue virus receptor in hepatic HepG2 cells [24] and our two-hybrid screen used domain III of DENV E which is thought to bind host cell receptors [5]. Therefore, we tested the ability of anti-BiP antibodies to block DENV infection. However, in our study using Vero cells both polyclonal antibody directed against BiP and monoclonal antibody directed against the carboxyl-terminal region of BiP, the binding site between DENV E and BiP, failed to inhibit DENV infection (data not shown), suggesting that an alternate or additional receptors are present on Vero cells.

BiP is one of multiple chaperone systems important for ER quality control [25]. Calnexin and calreticulin are lectin-like chaperones that interact with glycosylated proteins and are important for viral glycoprotein processing and maturation [26]. Since DENV E is a glycoprotein, we examined its interaction with calnexin and calreticulin. As shown in Fig. 2A, immunoprecipitation with either anti-calnexin antibody or anti-calreticulin antibody pulled down DENV E protein. Immunoprecipitation of dengue virus-infected cell lysates with an isotype-matched control antibody did not bring down DENV E, demonstrating the specificity of these interactions. As further evidence of interaction, dengue virus-infected Vero cells were examined for co-localizations of the proteins by immunofluorescence and confocal microscopy. Co-localization of DENV E and calnexin or calreticulin in the ER and peri-nuclear regions was clearly evident (Fig. 2B). The intensity of staining with anti-calnex-

in and anti-calreticulin was strongly enhanced following dengue virus infection (Fig. 2B) suggesting that, like BiP, expression of these two chaperones is induced.

Knocking-down the expression of BiP, calnexin and calreticulin by siRNA decreases dengue virus production

Defects in the mechanisms controlling proper protein folding and assembly mediated by ER chaperones affects morphogenesis and production of virions [2,27–29]. We asked whether the interaction of BiP, calnexin or calreticulin with DENV E influence dengue virion production. Transfection of Vero cells with siRNA against BiP, calnexin or calreticulin was performed and levels of the corresponding mRNA and proteins were measured by real-time PCR and Western blot analysis. The mRNA expression of BiP, calnexin and calreticulin was 15.57%, 27.69% and 7.4%, respectively, relative to irrelevant siRNA-transfected control cells (Fig. 3A). A corresponding decrease in protein expression of BiP, calnexin and calreticulin proteins was observed by Western blot analysis (Fig. 3B). The transfected cells were subsequently infected with DENV, and 24 h post-infection virions present in the supernatants were titrated by a focus forming unit assay (FFU). The reduced expression of BiP, calnexin or calreticulin decreased the yield of viral progeny a minimum of 50% (calreticulin) and as much as 90% (BiP) clearly demonstrating the involvement of BiP, calnexin and calreticulin in the production of DENV infectious particles (Fig. 3D). This is in agreement with the observation that BiP,

calnexin and calreticulin are essential for replication and infectivity of many human pathogenic viruses [28–30]. Chaperone proteins are only a part of the host machinery necessary for virus replication and assembly. Recently, silencing of ER-associated degradation (ERAD) components including DERL2, NSFL1C, UBE3A, UFDL1, SEC61G and SEC61A1 was shown to reduce dengue virus infection in the Hela cells [31].

In summary, this work is the first to demonstrate direct interactions between DENV E and the ER-resident chaperones including BiP, calnexin and calreticulin, and the involvement of these chaperones in the production of DENV infectious particles. Future studies will be directed toward elucidating the detailed molecular mechanism by which ER chaperones play a role in dengue virion assembly and cytoplasmic egress.

Acknowledgments

We thank Dr. Russell L Finley Jr., Wayne State University, for reagents of yeast two-hybrid system and his continuous support. We appreciate Dr. William A Fonzi, Georgetown University, for the helpful discussion. This work is financially supported by the Thailand Research Fund (TRF) to T.L., the TRF Royal Golden Jubilee (RGJ) Ph.D. Program to W.W. and J.N., and the TRF Senior Research Scholar Program to P.M. and P.Y.

References

- [1] T.J. Chambers, C.S. Hahn, R. Galler, C.M. Rice, Flavivirus genome organization, expression, and replication, *Annu. Rev. Microbiol.* 44 (1990) 649–688.
- [2] J.A. Mondotte, P.Y. Lozach, A. Amara, A.V. Gamarnik, Essential role of dengue virus envelope protein N glycosylation at asparagine-67 during viral propagation, *J. Virol.* 81 (2007) 7136–7148.
- [3] R. Brent, R.L. Finley Jr., Understanding gene and allele function with two-hybrid methods, *Annu. Rev. Genet.* 31 (1997) 663–704.
- [4] R. Sriburi, P. Keelapang, T. Duangchinda, S. Pruksakorn, N. Maneeekarn, P. Malasit, N. Sittisombut, Construction of infectious dengue 2 virus cDNA clones using high copy number plasmid, *J. Virol. Methods* 92 (2001) 71–82.
- [5] R.J. Kuhn, W. Zhang, M.G. Rossmann, S.V. Pletnev, J. Corver, E. Lenches, C.T. Jones, S. Mukhopadhyay, P.R. Chipman, E.G. Strauss, T.S. Baker, J.H. Strauss, Structure of dengue virus: implications for flavivirus organization, maturation, and fusion, *Cell* 108 (2002) 717–725.
- [6] J. Gyuris, E. Golemis, H. Chertkov, R. Brent, Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2, *Cell* 75 (1993) 791–803.
- [7] M.K. Gentry, E.A. Henchal, J.M. McCown, W.E. Brandt, J.M. Dalrymple, Identification of distinct antigenic determinants on dengue-2 virus using monoclonal antibodies, *Am. J. Trop. Med. Hyg.* 31 (1982) 548–555.
- [8] E.A. Henchal, J.M. McCown, D.S. Burke, M.C. Seguin, W.E. Brandt, Epitopic analysis of antigenic determinants on the surface of dengue-2 virions using monoclonal antibodies, *Am. J. Trop. Med. Hyg.* 34 (1985) 162–169.
- [9] N. Jirakanjanakit, T. Sanohsomneing, S. Yoksan, N. Bhamarapravati, The micro-focus reduction neutralization test for determining dengue and Japanese encephalitis neutralizing antibodies in volunteers vaccinated against dengue, *Trans. R. Soc. Trop. Med. Hyg.* 91 (1997) 614–617.
- [10] S.M. Hurtley, D.G. Bole, H. Hoover-Litty, A. Helenius, C.S. Copeland, Interactions of misfolded influenza virus hemagglutinin with binding protein (BiP), *J. Cell Biol.* 108 (1989) 2117–2126.
- [11] P.L. Earl, B. Moss, R.W. Doms, Folding, interaction with GRP78-BiP, assembly, and transport of the human immunodeficiency virus type 1 envelope protein, *J. Virol.* 65 (1991) 2047–2055.
- [12] B. Kahn-Perles, J. Salamerio, O. Jouans, Biogenesis of MHC class I antigens: involvement of multiple chaperone molecules, *Eur. J. Cell Biol.* 64 (1994) 176–185.
- [13] Y. Gaudin, Folding of rabies virus glycoprotein: epitope acquisition and interaction with endoplasmic reticulum chaperones, *J. Virol.* 71 (1997) 3742–3750.
- [14] A. Choukhi, S. Ung, C. Wychowski, J. Dubuisson, Involvement of endoplasmic reticulum chaperones in the folding of hepatitis C virus glycoproteins, *J. Virol.* 72 (1998) 3851–3858.
- [15] K.M. Flaherty, C. DeLuca-Flaherty, D.B. McKay, Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein, *Nature* 346 (1990) 623–628.
- [16] T.G. Chappell, B.B. Konforti, S.L. Schmid, J.E. Rothman, The ATPase core of a clathrin uncoating protein, *J. Biol. Chem.* 262 (1987) 746–751.
- [17] B. Gao, E. Eisenberg, L. Greene, Effect of constitutive 70-kDa heat shock protein polymerization on its interaction with protein substrate, *J. Biol. Chem.* 271 (1996) 16792–16797.
- [18] L. King, M. Chevalier, S.Y. Blond, Specificity of peptide-induced depolymerization of the recombinant carboxy-terminal fragment of BiP/GRP78, *Biochem. Biophys. Res. Commun.* 263 (1999) 181–186.
- [19] D.Y. Cho, G.H. Yang, C.J. Ryu, H.J. Hong, Molecular chaperone GRP78/BiP interacts with the large surface protein of hepatitis B virus in vitro and in vivo, *J. Virol.* 77 (2003) 2784–2788.
- [20] I. Umareddy, O. Pluquet, Q.Y. Wang, S.G. Vasudevan, E. Chevet, F. Gu, Dengue virus serotype infection specifies the activation of the unfolded protein response, *Virol. J.* 4 (2007) 91.
- [21] C.Y. Yu, Y.W. Hsu, C.L. Liao, Y.L. Lin, Flavivirus infection activates the XBP1 pathway of the unfolded protein response to cope with endoplasmic reticulum stress, *J. Virol.* 80 (2006) 11868–11880.
- [22] E. Liberman, Y.L. Fong, M.J. Selby, Q.L. Choo, L. Cousens, M. Houghton, T.S. Yen, Activation of the grp78 and grp94 promoters by hepatitis C virus E2 envelope protein, *J. Virol.* 73 (1999) 3718–3722.
- [23] H.L. Su, C.L. Liao, Y.L. Lin, Japanese encephalitis virus infection initiates endoplasmic reticulum stress and an unfolded protein response, *J. Virol.* 76 (2002) 4162–4171.
- [24] S. Jindadamrongwech, C. Thepparit, D.R. Smith, Identification of GRP 78 (BiP) as a liver cell expressed receptor element for dengue virus serotype 2, *Arch. Virol.* 149 (2004) 915–927.
- [25] M. Ni, A.S. Lee, ER chaperones in mammalian development and human diseases, *FEBS Lett.* 581 (2007) 3641–3651.
- [26] M. Pieren, C. Galli, A. Denzel, M. Molinari, The use of calnexin and calreticulin by cellular and viral glycoproteins, *J. Biol. Chem.* 280 (2005) 28265–28271.
- [27] M.P. Courageot, M.P. Frenkiel, C.D. Dos Santos, V. Deubel, P. Despres, Alpha-glucosidase inhibitors reduce dengue virus production by affecting the initial steps of virion morphogenesis in the endoplasmic reticulum, *J. Virol.* 74 (2000) 564–572.
- [28] N.J. Buchkovich, T.G. Maguire, Y. Yu, A.W. Paton, J.C. Paton, J.C. Alwine, Human cytomegalovirus specifically controls the levels of the endoplasmic reticulum chaperone BiP/GRP78, which is required for virion assembly, *J. Virol.* 82 (2008) 31–39.
- [29] L. Maruri-Avidal, S. Lopez, C.F. Arias, Endoplasmic reticulum chaperones are involved in the morphogenesis of rotavirus infectious particles, *J. Virol.* 82 (2008) 5368–5380.
- [30] A. Mehta, X. Lu, T.M. Block, B.S. Blumberg, R.A. Dwek, Hepatitis B virus (HBV) envelope glycoproteins vary drastically in their sensitivity to glycan processing: evidence that alteration of a single N-linked glycosylation site can regulate HBV secretion, *Proc. Natl. Acad. Sci. USA* 94 (1997) 1822–1827.
- [31] M.N. Krishnan, A. Ng, B. Sukumaran, F.D. Gilfoy, P.D. Uchil, H. Sultana, A.L. Brass, R. Adametz, M. Tsui, F. Qian, R.R. Montgomery, S. Lev, P.W. Mason, R.A. Koski, S.J. Elledge, R.J. Xavier, H. Agaisse, E. Fikrig, RNA interference screen for human genes associated with West Nile virus infection, *Nature* 455 (2008) 242–245.

Simple, Efficient, and Cost-Effective Multiplex Genotyping with Matrix Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry of Hemoglobin Beta Gene Mutations

Wanna Thongnoppakhun,* Surasak Jiemsup,[†] Suganya Yongkiettrakul,[†] Chompunut Kanjanakorn,* Chanin Limwongse,*[‡] Prapon Wilairat,[§] Anusorn Vanasant,[†] Nanyawan Rungroj,* and Pa-thai Yenchitsomanus*

From the Departments of Research and Development,* and Medicine,[‡] Faculty of Medicine Siriraj Hospital, and the Department of Biochemistry,[§] Faculty of Science, Mahidol University, Bangkok; and the National Center for Genetic Engineering and Biotechnology (BIOTEC),[†] Thailand Science Park, Klong Luang, Pathumthani, Thailand

A number of common mutations in the hemoglobin β (*HBB*) gene cause β -thalassemia, a monogenic disease with high prevalence in certain ethnic groups. As there are 30 *HBB* variants that cover more than 99.5% of *HBB* mutant alleles in the Thai population, an efficient and cost-effective screening method is required. Three panels of multiplex primer extensions, followed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry were developed. The first panel simultaneously detected 21 of the most common *HBB* mutations, while the second panel screened nine additional mutations, plus seven of the first panel for confirmation; the third panel was used to confirm three *HBB* mutations, yielding a 9-Da mass difference that could not be clearly distinguished by the previous two panels. The protocol was both standardized using 40 samples of known genotypes and subsequently validated in 162 blind samples with 27 different genotypes (including a normal control), comprising heterozygous, compound heterozygous, and homozygous β -thalassemia. Results were in complete agreement with those from the genotyping results, conducted using three different methods overall. The method developed here permitted the detection of mutations missed using a single genotyping procedure. The procedure should serve as the method of choice for *HBB* genotyping due to its accuracy, sensitivity, and cost-effectiveness, and can be applied to studies of other gene variants that are potential disease biomarkers. (*J Mol Diagn* 2009, 11:334–346; DOI: 10.2353/jmoldx.2009.080151)

To date, 739 point mutations in the hemoglobin, β (*HBB*) gene causing β -thalassemia (MIM# 141900) have been reported in HbVar: A Database of Human Hemoglobin Variants and Thalassemias (<http://globin.cse.psu.edu/globin/hbvar/menu.html>, accessed March 2009), but each ethnic group has a limited number of common mutations and a considerable number of rarer mutations.¹ The c.79G>A (also known as CD26G>A or Hb E) is the most frequent *HBB* variant in Southeast Asia including Thailand.² “Thai” generally refers to speakers of Thai (Tai) languages. The ethnic groups of Thailand comprise Thais (constituting 85% of the population) and Hill Peoples living primarily in the north, as well as other groups including the Chinese and minorities in the south.³ In the Thai population, approximately 40 *HBB* mutations have been identified,⁴ of which 30 variants account for more than 99.5% of all mutant *HBB* alleles (Table 1). Thus, an efficient, cost-effective, and highly accurate screening method is required for their detection.

Many simple methods for genotyping *HBB* mutations have been used, including restriction fragment length polymorphism analysis,⁵ reverse dot-blot hybridization,^{6–8} amplification refractory mutation system,⁹ single strand conformation polymorphism analysis,^{4,10} denaturing gradient gel electrophoresis (DGGE),^{11,12} and direct DNA sequencing.¹³ Recent advances in genotyping technologies have enabled high sample-throughput screening of several mutations in a large number of samples. Allele-specific arrayed primer extension has been designed for the simultaneous detection of 15 nondeletion α -globin gene defects and 23 β -globin gene muta-

Supported by the National Center for Genetic Engineering and Biotechnology, National Science and Technology for Development Agency (NSTDA), Thailand (BT-B-02-MM-SU-4904), Siriraj Chalermprakiat Fund (to W.T.) and Mahidol University Research Grant (to C.L.). P.Y. is a Senior Research Scholar of The Thailand Research Fund.

Accepted for publication March 6, 2009.

Address reprint requests to Wanna Thongnoppakhun, Division of Molecular Genetics, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkoknoi, Bangkok 10700, Thailand, E-mail: siwtn@mahidol.ac.th; or Pa-thai Yenchitsomanus, Division of Medical Molecular Biology, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkoknoi, Bangkok 10700, Thailand. E-mail: grpye@mahidol.ac.th.

Table 1. Thirty Common and Rare *HBB* Mutations of Patients with β -Thalassemia and Carriers from the Siriraj-Thalassemia Program Project, Faculty of Medicine Siriraj Hospital, Mahidol University during 2003–2004

Common <i>HBB</i> mutations (13)		<i>HBB</i> mutations causing abnormal Hb (10)		Rare <i>HBB</i> mutations (7)	
Common name	HGVS nomenclature	Common name	HGVS nomenclature	Common name	HGVS nomenclature
CD26G>A (Hb E)	c.79G>A*	CD147+AC (Hb Tak)	c.441_442insAC* ^{††}	CD43G>T	c.130G>T*
CD41/42-TTCT	c.124_127delTTCT* [†]	CD126T>G (Hb Dhonburi)	c.380T>G*	CD123/125 (–8 bp)	c.370_377delACC CCACC [†]
CD17A>T	c.52A>T* ^{††}	CD136G>A (Hb Hope)	c.410G>A*	–87C>A	c. –137C>A [†]
–28A>G	c. –78A>G*	CD6G>A (Hb C)	c.19G>A* [†]	CD15-T	c.46delT [†]
IVS2#654C>T	c.316–197C>T*	CD56G>A (Hb J-Bangkok)	c.170G>A*	CD8/9+G	c.27_28insG [†]
IVS1#5G>C	c.92 + 5G>C*	CD83G>A (Hb Pyrgos)	c.251G>A*	CD27/28+C	c.84_85insC [†]
CD19A>G	c.59A>G*	CD6A>C (Hb G Makassar)	c.20A>C* [†]	CD41-C	c.126delC* [†]
(Hb Malay)					
CD71/72 + A	c.216_217insA*	CD6A>T (Hb S)	c.20A>T* ^{††}		
IVS1#1G>T	c.92 + 1G>T [†]	CD121G>C (Hb D Punjab)	c.364G>C*		
–31A>G	c. –81A>G [†]	CD1T>C (Hb Raleigh)	c.5T>C [†]		
–30T>C	c. –80T>C*				
CD35C>A	c.108C>A [†]				
CD0T>G	c.2T>G*				

Each column is listed in order of decreasing frequency.

HGVS, Human Genome Variation Society.

**HBB* mutations detectable by Panel 1 Multiplex SBE.

[†]*HBB* mutations detectable by Panel 2 Multiplex SBE.

^{††}*HBB* mutations detectable by Panel 3 Multiplex VSET.

tions commonly found in Southeast Asian countries to overcome the need to use multiple reverse dot-blot analyses.¹⁴ Multiplex minisequencing also has been widely applied as a basic molecular technique, with subsequent detection using a variety of different platforms, including capillary electrophoresis,¹⁵ denaturing high performance liquid chromatography^{16,17} and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS).^{18,19}

MALDI-TOF MS has been developed as a genotyping tool based on the differences in mass of variant DNA sequences.²⁰ This technique provides highly accurate identification due to its ability in a single run to detect directly the absolute masses of multiple variant sites.²¹ MALDI-TOF MS in combination with multiplex minisequencing has proven to be a cost-effective and efficient procedure for high-throughput genotyping of a number of disease-causing genes or of single nucleotide polymorphisms (SNPs).^{19,22–24}

Nevertheless, bottlenecks in multiplex genotyping using MALDI-TOF MS include optimization of highly multiplex-primer extension (PE) reactions²⁵ and the need to completely remove contaminating salt adducts that can compromise spectral quality and reduce accuracy of mass assignments.^{25–27} For genotyping of *HBB*, there is the additional problem of the very close proximity and partial overlapping of the mutations to one another. Recently, a successful single analysis of the eight most common *HBB* mutations in Taiwanese population has been achieved by using eight parallel minisequencing reactions and pooling of the minisequencing reaction products for subsequent sequential desalting and multiplex MALDI-TOF analysis.¹⁹

To reduce analysis time and cost of *HBB* genotyping, this study aimed to maximize multiplexing in both PCR and PE steps, based on having well-designed primers and well-optimized reaction conditions that give best

yields for every possible allele in each multiplex reaction. We have developed an alternative approach for genotyping the 30 specific *HBB* mutations in the Thai population, which comprises tetraplex PCR to amplify four fragments spanning all 30 mutations, multiplex PE reaction of the PCR products, desalting with magnetic bead separation, and analysis of PE products by MALDI-TOF MS. Three separate panels of multiplex PE reactions were developed for 21 mutations, 16 variants including nine additional mutations and seven mutations identical to the first panel, and an optional third panel for confirmation of c.20A>T, c.52A>T, or c.441_442insAC (Hb Tak). Using this approach, the 30 *HBB* mutations were reliably and unambiguously detected and the technique was validated using a total of 162 randomly selected β -thalassemia samples previously genotyped by DGGE, restriction fragment length polymorphism, and/or direct sequencing techniques.

Materials and Methods

Collection of DNA Samples Carrying *HBB* Mutations

DNA samples were anonymously obtained from DNA bank of Siriraj-Thalassemia Program Project, Faculty of Medicine Siriraj Hospital, Mahidol University during 2005, with patients' informed consent at the time of blood collection conducted as part of a routine medical examination, and the process conforming to institutional and national ethical guidelines. Genomic DNA was isolated from whole blood using Puregene DNA isolation kit (Gentra Systems Inc., Minneapolis, MN). These samples have been analyzed previously in our laboratory for *HBB* genotypes using DGGE, restriction fragment length polymorphism, or direct DNA sequencing.

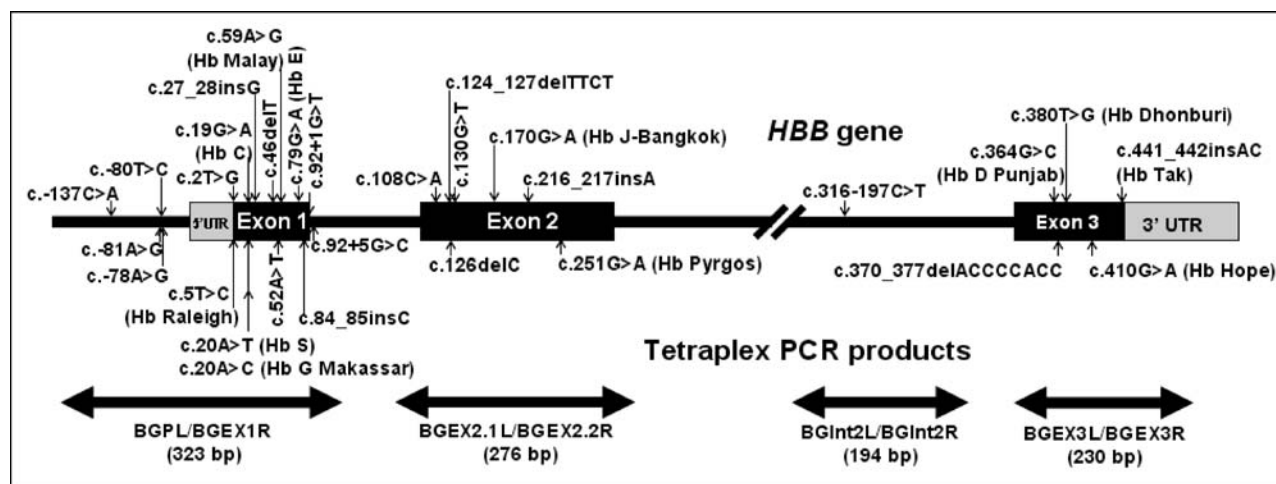


Figure 1. Schematic diagram of *HBB* gene showing distribution of 30 *HBB* mutations of interest and positions of tetraplex PCR products.

For method development of multiplex genotyping with MALDI-TOF MS, 40 DNA samples of known *HBB* genotypes from a normal control, 30 β -thalassemia carriers with 30 different *HBB* mutations (Table 1), either single heterozygotes or compound heterozygotes, as well as nine different homozygotes were tested. Subsequently, 162 DNA samples randomly selected from β -thalassemia patients (compound heterozygote or homozygote) and carriers were tested for method validation.

Multiplex PCR Amplification and Purification

Four pairs of PCR primers, BGPL/BGEX1R, BGEX2.1L/BGEX2.2R, BGInt2L/BGInt2R, and BGEX3L/BGEX3R,⁴ were used in a single PCR reaction to generate tetraplex-PCR amplicons of 323, 276, 194, and 230 bp, respectively, which encompass all of the 30 *HBB* mutations, as shown in Figure 1. A 25- μ L volume of multiplex PCR reaction comprised 1.5 mmol/L $MgCl_2$, 200 μ mol/L dNTPs, 0.5 U of Immolase DNA polymerase (Bioline USA Inc.), 100 ng of genomic DNA, and 200 nmol/L PCR primers, except for BGEX2.1L and BGEX2.2R, which required 600 nmol/L. Thermal cycling followed in a GeneAmp PCR System 2400 (PE Applied Biosystems, Foster City, CA) at 95°C for 7 minutes, followed by 35 cycles of 94°C for 30 seconds, 57°C for 30 seconds, and 72°C for 20 seconds, and an additional extension step at 72°C for 10 minutes.

Excess PCR primers and dNTPs in the reaction were removed by incubating 5 μ L of PCR mixture with 2 μ L of "ExoSAP-IT" mix (U.S.B. Corporation, Cleveland, OH) at 37°C for 30 minutes followed by 20 minutes at 80°C.

Primer Design for Multiplex Primer-Extension Reaction

To economize on time and cost, multiplex PE was designed to cover all 30 mutations using at most three reactions. PE reaction can be a single-base extension (SBE) (PinPoint assay)²⁸ or a very short extension (VSET).²¹ For SBE, the primer anneals immediately next

to the mutation site and is extended with a single nucleotide, using terminating dideoxynucleoside triphosphate (ddNTP), whereas in VSET, the primer is extended by only one base from one allele and by two bases from the other allele using three ddNTPs and a fourth nucleotide in the deoxy form. Due to a high capacity of multiplexing, SBE was performed in our study while VSET was used to distinguish clearly the 9-Da mass difference between extension product containing dideoxyadenosine monophosphate and that of dideoxythymidine monophosphate.

Primers used throughout this study were synthesized by BioService Unit, National Center for Genetic Engineering and Biotechnology, Bangkok, Thailand. Design of the extension primers was based on the positions of individual mutations, taking into account the molecular mass range of primers or products from both alleles that can be reliably and unequivocally differentiated from one another by MALDI-TOF MS (4000 to 9000 Da). The masses of the primers and their corresponding extension products were calculated using Mongo Oligo Mass Calculator v2.05 (<http://library.med.utah.edu/masspec/mongo.htm>, accessed November 2006) and adding the mass of a particular ddNTP to that of the primers. Primers were grouped into three panels for multiplexing as shown in Table 2.

Multiplex PE Reaction

Optimization of multiplex PE reactions was performed by testing with various combinations of extension primers until maximum multiplexing was obtained. Primer concentrations were adjusted to obtain balanced intensities of MALDI-TOF MS signals. Finally, three panels, which included 19, 14, and 3 primers per reaction for detection of 21, 16 (with 7 repeated tests), and 3 (as confirmation) *HBB* mutations, respectively, were obtained as Panel 1 and Panel 2 for multiplex SBE and Panel 3 for multiplex VSET assay (Table 2).

Multiplex SBE reaction (Panels 1 and 2) was performed in a 20- μ L volume containing 5 μ L of ExoSAP-IT treated

Table 2. Expected Molecular Mass of Extension Primers and Products of Multiplex Single-Base Extension (SBE) and Very Short Extension (VSET) Assay Employed for Detection of *HBB* Mutations

Primer No	Mutation site in <i>HBB</i> gene	Primer sequence	Primer strand	Primer conc. (nmol/L)	Primer mass (Da)	Normal product mass (Da)	Mutant product mass (Da)	Mass difference
Panel 1: Multiplex SBE-for detection of 21 <i>HBB</i> mutations (using 19 extension primers)								
1.1	c.19G>A (Hb C)	5'-GCATCTGACTCCT-3'	Sense	425	3885.6	+ddG = 4198.8	+ddA = 4182.8	-16
1.2	c.20A>T (Hb S)	5'-CCGGCAGACTTCTCC-3'	Antisense	200	4489.0	+ddT = 4777.2	+ddA = 4786.2	+9
	c.20A>C (Hb G Makassar)						+ddG = 4802.2	+25
1.3	c.216_217insA	5'-TGCTCGGTGCTTTA-3'	Sense	300	4550.0	+ddG = 4863.2	+ddA = 4847.2	-16
1.4	c.410G>A (Hb Hope)	5'-GAGGGCATTAGCCACA-3'	Antisense	150	4915.3	+ddC = 5188.5	+ddT = 5203.5	+15
1.5	c.79G>A (Hb E)	5'-CCACTGCCCCAGGCCT-3'	Antisense	100	5092.3	+ddC = 5365.6	+ddT = 5380.5	+15
1.6	c.52A>T	5'-GTACTGCCTGTTGGGC-3'	Sense	385	5258.4	+ddA = 5555.6	+ddT = 5546.6	-9
1.7	c.-78A>G	5'-ATGGCTCTGCCCTGACTT-3'	Antisense	225	5441.6	+ddT = 5729.8	+ddC = 5714.8	-15
1.8	c.170G>A (Hb J-Bangkok)	5'-CTCCTGATGCTGTATAGG-3'	Sense	175	5496.6	+ddG = 5809.8	+ddA = 5793.8	-16
1.9	c.59A>G (Hb Malay)	5'-ACCACCAACTTCATCCACG-3'	Antisense	100	5661.7	+ddT = 5949.9	+ddC = 5934.9	-15
1.10	c.-80T>C	5'-CATTAGCCAGGGCTGGGCA-3'	Sense	150	5853.8	+ddT = 6142.0	+ddC = 6127.0	-15
1.11	c.441_442insAC (Hb Tak)	5'-GCAAGAAAGCGAGCTTAGTG-3'	Antisense	100	6215.1	+ddA = 6512.3	+ddT = 6503.3	-9
1.12	c.380T>G (Hb Dhonburi)	5'-ACTTCTGATAGGCAGCCTGC-3'	Antisense	87.5	6397.2	+ddA = 6694.4	+ddC = 6670.4	-24
1.13	c.130G>T	5'-GGGACAGATCCCCAAAGGACT-3'	Antisense	150	6449.0	+ddC = 6722.2	+ddA = 6746.2	+24
1.14	c.2T>G	5'-ACTAGCAACCTCAAACAGACACCA-3'	Sense	150	7252.8	+ddT = 7541.0	+ddG = 7566.0	+25
1.15	c.251G>A (Hb Pyrgos)	5'-TGGGTACCTGGACAACCTCAAGG-3'	Sense	125	7362.8	+ddG = 7676.0	+ddA = 7660.0	-16
1.16	c.316-197C>T	5'-ACAGTGATAATTTCTGGGTTAAGG-3'	Sense	150	7446.9	+ddC = 7720.1	+ddT = 7735.1	+15
1.17	c.126delC	5'-GTTCCTGACCTTGGACCCAGAGTT-3'	Sense	350	7978.2	+ddC = 8251.4	+ddT = 8266.4	+15
	c.124_127delTTCT						+ddG = 8291.4	+40
1.18	c.364G>C (Hb D Punjab)	5'-CTCTCTCTCGCCCATCACTTTGGCAAA-3'	Sense	450	8395.5	+ddG = 8708.7	+ddC = 8668.7	-40
1.19	c.92 + 5G>C	5'-TCCTTAAACCTGTCTTGTAACCTTGATA-3'	Antisense	225	8488.6	+ddC = 8761.8	+ddG = 8801.8	+40
Panel 2: Multiplex SBE-for detection of 16 <i>HBB</i> mutations (using 14 extension primers)								
2.1	c.19G>A (Hb C)	5'-GCATCTGACTCCT-3'	Sense	425	3885.6	+ddG = 4198.8	+ddA = 4182.8	-16
2.2	c.52A>T	5'-ATCCACGTTTCACCT-3'	Antisense	385	4158.8	+ddT = 4447.0	+ddA = 4456.0	+9
2.3	c.20A>T (Hb S)	5'-CCGGCAGACTTCTCC-3'	Antisense	140	4489.0	+ddT = 4777.2	+ddA = 4786.2	+9
	c.20A>C (Hb G Makassar)						+ddG = 4802.2	+25
2.4	c.27_28insG	5'-GCTCCTGAGGAGAAG-3'	Sense	250	4642.1	+ddT = 4930.3	+ddG = 4955.3	+25
2.5	c.126delC	5'-CCTTGGACCCAGAGGTT-3'	Sense	125	5186.4	+ddC = 5460.0	+ddT = 5475.0	+15
	c.124_127delTTCT						+ddG = 5500.0	+40
2.6	c.108C>A	5'-AGGCTGCTGGTGGTCTA-3'	Sense	125	5257.5	+ddC = 5530.7	+ddA = 5554.7	+24
2.7	c.441_442insAC (Hb Tak)	5'-GAGAAAGCGAGCTTAGTG-3'	Antisense	75	5612.7	+ddA = 5909.9	+ddT = 5900.9	-9
2.8	c.46delT	5'-CATTGCCGTTACTGCCCTG-3'	Sense	125	5730.7	+ddT = 6018.9	+ddG = 6043.9	+25
2.9	c.-81A>G	5'-GGCAGGAGCCAGGCTGGGC-3'	Sense	125	6249.1	+ddA = 6546.3	+ddG = 6562.3	+16
2.10	c.84_85insC	5'-GATGAAGTTGGTGGTCAAGGCC-3'	Sense	150	6871.5	+ddT = 7159.7	+ddC = 7144.7	-15
2.11	c.-137C>A	5'-AGACCTCACCTGTGGAGCCACAC-3'	Sense	100	7267.8	+ddC = 7541.0	+ddA = 7565.0	+24

(Continued)

Table 2. (Continued)

Primer No	Mutation site in <i>HBB</i> gene	Primer sequence	Primer strand	Primer conc. (nmol/L)	Primer mass (Da)	Normal product mass (Da)	Mutant product mass (Da)	Mass difference
2.12	c.370_377del ACCCACCC	5'-ACTTTCCTGATAGGCAGCCTGC A CTG-3'	Antisense	150	7633.0	+ddG = 7946.2	+ddA = 7930.2	-16
2.13	c.92 + 1G>T	5'- CT CCTGCTCTTGTAACCTTGATAC CAA -3'	Antisense	125	7856.2	+ddC = 8129.4	+ddA = 8153.4	+24
2.14	c.5T>C (Hb Raleigh)	5'-ACTAGCAACCTCAAACAGACACCA TGG -3'	Sense	200	8215.4	+ddT = 8503.6	+ddC = 8488.6	-15
Panel 3: Multiplex VSET—for confirmation of three <i>HBB</i> mutations (using three extension primers)								
3.1	c.20A>T (Hb S)	5'- TGCA TCT GACTCCT G -3'	Sense	150	4519	+ddA = 4816.2	+dT + ddG = 5136.4	320.2
3.2	c.52A>T	5'- G TACTGCG G CTGTGGGGC-3'	Sense	375	5258.4	+ddA = 5555.6	+dT + ddA = 5859.8	303.8
3.3	c.441_442insAC (Hb Tak)	5'-GCAAGAAAGCGAGCTTAGTG-3'	Antisense	100	6215.1	+ddA = 6512.3	+dT + ddG = 6832.5	320.2

Bold bases are modifications for mass adjustment.

Underlined-bold bases are modifications for removal of self-complementary or hairpin formation.

Italic-bold bases are positions where there may be mismatches of bases due to the presence of other closely-located *HBB* mutations or SNPs.

PCR product, 75 to 450 nmol/L extension primers (Table 2), 50 μ mol/L ddNTP (Amersham Biosciences UK Ltd, Buckinghamshire, UK), 1 μ l of Thermo Sequenase buffer, and 1 U of Thermo Sequenase DNA polymerase (Amersham Biosciences UK Ltd). For multiplex VSET reaction (Panel 3), the 20- μ l reaction volume contained 5 μ l of ExoSAP-IT treated PCR product, 100 to 375 nmol/L of extension primers (Table 2), 50 μ mol/L each of ddATP, ddCTP, ddGTP, and dTTP (instead of ddTTP), 1 μ l of Thermo Sequenase buffer, and 1 U of Thermo Sequenase DNA polymerase.

Thermal cycling for all PE reactions was performed in a GeneAmp PCR System 2400, consisting of an initial denaturation at 96°C for 1 minute, followed by 90 cycles of touch-down protocol using the same denaturation step at 96°C for 15 seconds and extension step at 60°C for 60 seconds in every cycle, but the annealing temperature for 15 seconds was reduced from starting 60°C to 42°C in 2°C decrement every two cycles. Annealing temperature of 42°C was then maintained for the subsequent 72 cycles.

Sample Preparation and MALDI-TOF MS Analysis

PE products (20 μ l) were purified using 5 μ l of magnetic beads (GenoPure) according to manufacturer's instructions (Bruker Daltonics GmbH, Bremen, Germany) to obtain single-stranded oligonucleotide products.

For preparation of matrix solution, a saturated solution of 3-hydroxypicolinic acid in 50% acetonitrile was mixed with 50 mmol/L di-ammonium hydrogen citrate at a 1:1 ratio (v/v). A 0.5 μ l aliquot of matrix solution was spotted onto an AnchorChip target plate (Bruker-Daltonics) and dried at room temperature. Subsequently, a 0.5 μ l aliquot of purified PE sample was applied on top of the dried matrix and dried at room temperature. Mass spectra were recorded in a linear Bruker reflex V delayed extraction MALDI-TOF mass spectrometer (Bruker Daltonics), using

negative ion and linear mode. External calibration was performed using a standard oligonucleotide mixture (0.25 pmol/ μ L of Oligo 12 mer, average m/z = 3646.4; 1.25 pmol/ μ L of Oligo 20 mer, average m/z = 6118.0; and 5 pmol/ μ L Oligo 30 mer, average m/z = 9192.0). Data were analyzed using Flex Analysis software (Bruker Daltonics).

Genotype Determination

Genotype was manually assigned for each mutation based on mass difference between primer and extension peaks of the MALDI spectra by determining signal ratio of specific to nonspecific PE reaction products.

Results

Design of Extension Primer Sets

The extension primers were designed specifically for each of the 30 *HBB* mutations by dividing into two multiplex-SBE panels (Table 2). Very close proximity to each other in some of the 30 *HBB* mutations made it difficult to design extension primers that were interference-free. Base positions where there might be mismatches due to the presence of other closely located *HBB* mutations are shown in bold italics in Table 2.

Panel 1 multiplex SBE was designed to detect a total of 21 *HBB* mutations using 19 extension primers (two were shared by two variant sites). Detectable *HBB* mutations included ten common and two rare mutation types, as well as nine abnormal Hb variants (as classified in Table 1). Data obtained from Siriraj-Thalassemia Program Project, Faculty of Medicine Siriraj Hospital, Mahidol University during 2003–2004, have revealed that the eight most common *HBB* mutations account for about 90% of mutant alleles in the Thai population, namely, c.79G>A (Hb E), c.124_127delTTCT, c.52A>T, c.-78A>G, c.316–197C>T, c.92 + 5G>C, c.59A>G (Hb Malay),

and c.216_217insA (Table 1). Thus Panel 1 multiplex SBE was capable of detecting 95% of mutant *HBB* alleles in Thailand.

Panel 2 multiplex SBE, using 14 extension primers, was designed to detect 16 *HBB* mutations. Nine variants other than those in Panel 1 could be detected in Panel 2, including three common mutations, one abnormal Hb variant, and five rare mutations (as classified in Table 1). The three common variants (c.-81A>G, c.92 + 1G>T, and c.108C>A) were grouped into this panel to avoid incompatibility with their closely located mutations in Panel 1. Seven mutations in Panel 1 were also detected in Panel 2 as confirmatory tests, namely, those having 9-Da mass difference between wild-type and mutant alleles causing difficulty in interpretation (c.20A>T, c.52A>T, and c.441_442insAC); the smallest mass allele yielding poor signal (c.19G>A); the second most frequent variant in Thais (c.124_127delTTCT) and its extension-primer sharing variant (c.126delC); as well as the one (c.20A>C) coupled with c.20A>T.

In Panel 3, instead of multiplex SBE, multiplex VSET was established as an optional procedure for resolving ambiguous results obtained from the three *HBB* mutations, namely, c.20A>T, c.52A>T, and c.441_442insAC, having 9-Da mass differences found in the first two panels.

Optimization of the Three Multiplex PE Reactions

To evaluate the minisequencing performance of the extension primer for each *HBB* mutation before incorporating it into the multiplex SBE set, each primer was individually used in a single-plex SBE reaction and the result

was analyzed by MALDI-TOF MS. Using various different combinations of extension primers, successful set of 19-plex and 14-plex reaction was obtained for multiplex Panels 1 and 2 respectively. Optimal PE conditions were further tested by varying the amounts of extension primers to adjust the signal balance of MALDI-TOF MS peaks in the same multiplex set and by adjusting thermal profiles of PE reactions to obtain the highest yields of every possible extension products.

To reduce the possibility of primer interactions in the multiplex SBE panels, as well as interference of unextended primers in MALDI-TOF MS analysis, an optimized minimum concentration of each extension primer was used. However, this could still yield very high signal-to-noise ratio of mass spectra. An optimal thermal cycling profile was achieved by the use of a modified touchdown thermal-cycling minisequencing, with an initial annealing temperature at 60°C followed by gradual reduction to 42°C. With amplification of up to 90 cycles, more primers were consumed, resulting in higher product yields and lower peak intensities of the primers. As a result, more extension products could be clearly determined.

Genopure magnetic bead separation kit for purification of oligonucleotides was used to purify PE products. Good peak intensity of ionization, high signal-to-noise ratio, and negligible salt-adduct peaks were observed from the purified PE products.

Standardization of Multiplex SBE and VSET

Multiplex SBE, optimized for Panels 1 and 2, was used for genotyping of a normal control and heterozygotes (single and compound) carrying 30 *HBB* mutations (Figure 2 and Table 1). Mass differences between each pair of wild-

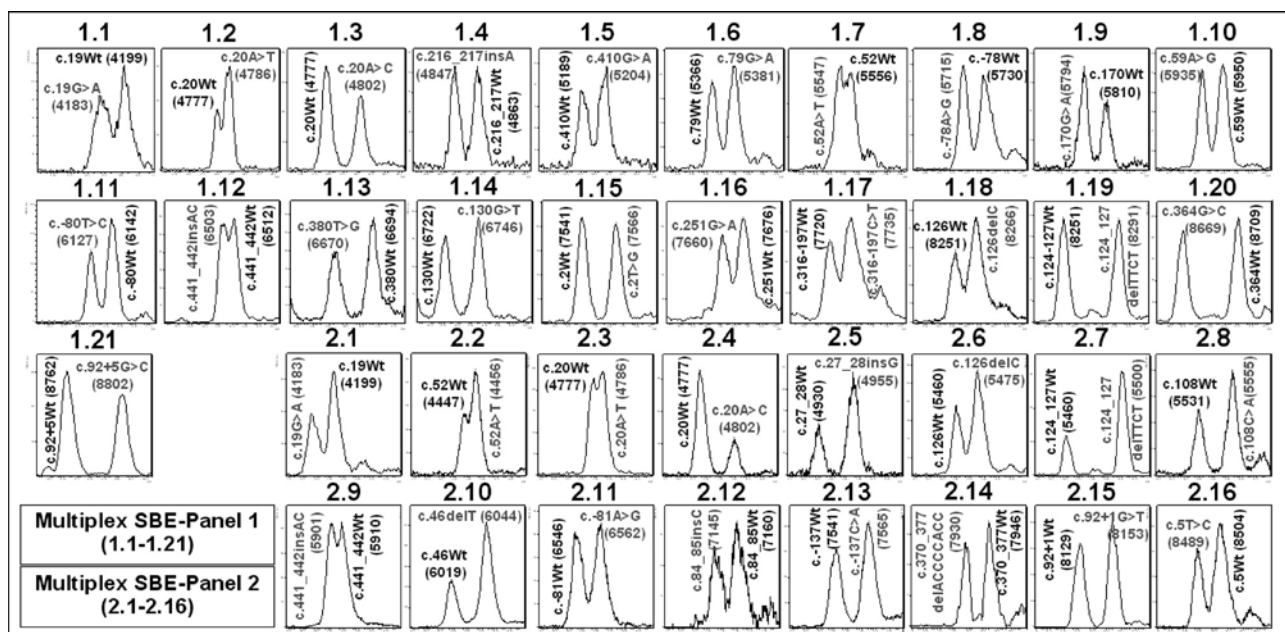


Figure 2. Mass spectra of *HBB* mutations detected by **Panel 1** and **Panel 2** multiplex SBE. Representative mass spectra obtained from 21 heterozygous *HBB* mutation sites detected by **Panel 1** multiplex SBE (1.1–1.21) and 16 by **Panel 2** (2.1–2.16) are shown. Genotype of the extension product is indicated above each peak and mass value is in parentheses. Wt = wild-type.

type and mutant allele of 9, 15, 16, 24, 25, and 40 Da could be clearly distinguished by MALDI-TOF MS, except for the 9-Da difference between A and T of c.20A>T, c.52A>T, and c.441_442insAC, which appeared as broad peaks (Figure 2: 1.2, 1.7, and 1.12). Mass spectra of the extended products plus residual primers in Panels 1 and 2 multiplex SBE of representative samples are shown in Figure 3, A–D. Inclusion in both multiplex SBE panels of the three mutations producing a 9-Da difference allowed confirmation of the results. In cases where there were high quality PCR and clean PE products, the three mutations could be unequivocally genotyped, and where inconclusive signals were obtained, Panel 3 multiplex VSET was performed to obtain the correct genotyping (Figure 4, A–D).

For discrimination of homozygotes from heterozygotes, we evaluated the performance of Panels 1 and 2 multiplex SBEs using nine (available) homozygous *HBB* mutations (Table 3). For Panel 1 multiplex SBE, eight homozygotes could be genotyped, giving the expected results. Five homozygotes, c.124_127delTTCT, c.52A>T, c.–81A>G, c.19G>A, and c.20A>T, could also be genotyped accurately by Panel 2 multiplex SBE. More-

over, Panel 3 multiplex VSET correctly identified homozygotes of c.52A>T and c.20A>T.

However, due to the crowded features of the 30 *HBB* mutations, interferences or interactions among various *HBB* alleles and extension primers are of concern. Observations that need to be taken into account for screening unknown samples with the three Panels are listed in Table 4. For example, in Panel 1, homozygous c.–81A>G (which is not within the detection scope of Panel 1) results in the loss of detection of c.–80T>C because c.–81A>G is located at the 3' end of an extension primer for c.–80T>C, leading to a strong mismatch between primer and template.

Homozygous c.79G>A (Hb E) was also evaluated in Panel 2 multiplex SBE. This mutation site is located at the seventh base from the 3' end of the c.84_85insC primer and has no effect on primer binding (Figure 5). Thus, similar results could be expected in Panel 1 multiplex SBE for the case of homozygous c.46delT and homozygous c.84_85insC when the effect on the binding of primer for c.52A>T and c.79G>A is considered respectively, as the mutation sites are at the sixth base from the 3' end of both primers. However, the presence of base

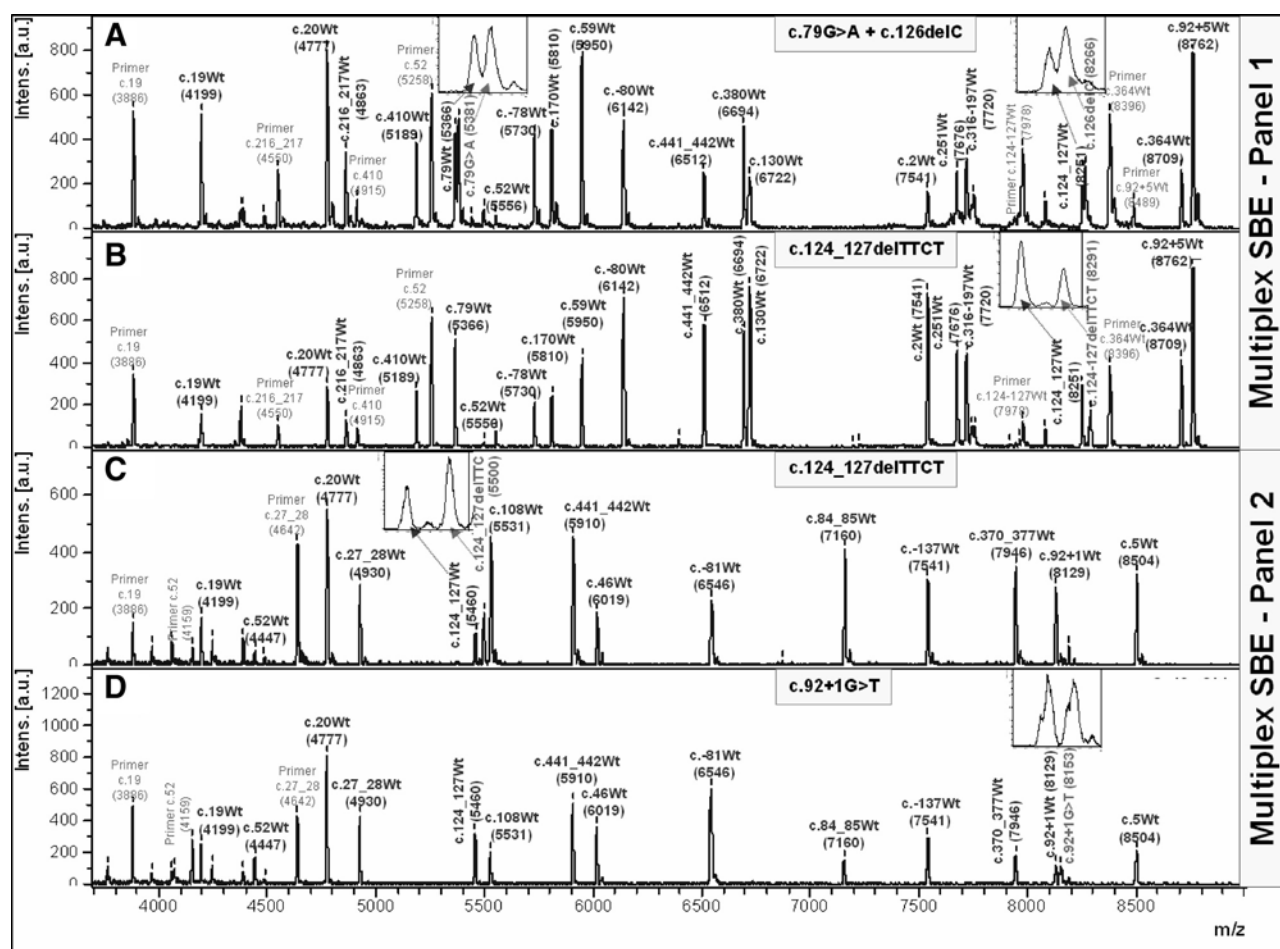


Figure 3. Mass spectra of 19-plex minisequencing reaction for genotyping of 21 *HBB* mutation sites in **Panel 1** multiplex SBE and 14-plex minisequencing reaction for 16 *HBB* mutations in **Panel 2** multiplex SBE. **Panel 1** multiplex SBE: (A) c.79G>A + c.126delC (compound heterozygous) and (B) c.124_127delTTCT heterozygous. **Panel 2** multiplex SBE: (C) c.124_127delTTCT heterozygous and (D) c.92 + 1G>T heterozygous. Genotype of extension product is indicated above each peak and mass value is in parentheses. Peaks labeled "Primer" indicate residual extension primers. Magnified views of heterozygote peaks are indicated by arrows and shown in insets.

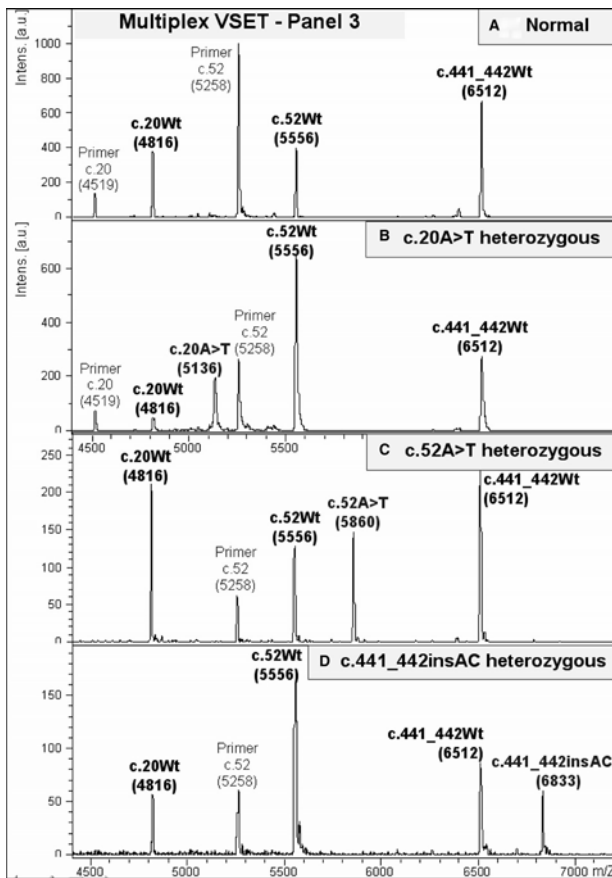


Figure 4. Mass spectra of 3-plex primer-extension reaction for genotyping in Panel 3 multiplex VSET: (A) normal, (B) c.20A>T heterozygote, (C) c.52A>T heterozygote, and (D) c.441_442insAC heterozygote. Genotype of extension product is indicated above each peak and mass value is in parentheses. Peaks labeled "Primer" indicate residual extension primers.

mismatch at the third position from 3' end of an extension primer may inhibit PE reaction, as in the case of homozygous c.2T>G detected by c.5T>C primer in Panel 2 multiplex SBE (Figure 5).

Primers used for Panel 3 multiplex VSET could detect c.20A>T, c.52A>T, and c.441_442insAC mutations. The primer for detecting c.20A>T could also pick up c.20A>C mutation. However, it was observed that heterozygous c.20A>C gave no expected signal of mass 4792 Da from the mutant C allele (as SBE, not VSET). This may have been caused by a preferential incorporation of ddA for the wild-type allele (mass = 4816.2 Da) over that

of the mutant and/or by poor ionization/desorption of oligonucleotide molecules of this mutant allele.

Validation of Multiplex SBE and VSET in Thalassemia Samples

One hundred and sixty-two randomly selected thalassemia subjects whose *HBB* mutations had previously been genotyped by DGGE, restriction fragment length polymorphism, and/or DNA sequencing were used to validate the technique. Samples consisted of 120 heterozygous healthy carriers having 15 different *HBB* mutations, and 42 patients (eight homozygotes and 34 compound heterozygotes) having two homozygous and 10 different compound heterozygous *HBB* mutations (Table 5). Every sample identified using a combination of Panels 1 + 2 (104 samples) or Panels 1 + 2 + 3 (29 samples) consistently demonstrated the expected genotype. Moreover, 10 samples determined by PE MALDI-TOF MS were shown to have been misidentified by conventional methods. For example, c.79G>A (Hb E) was frequently discovered in samples that previously had not been detected by DGGE.

Discussion

We have developed a multiplex-PE method for genotyping using MALDI-TOF MS, based on mass difference between wild-type and mutant alleles generated by an addition of a single base. Using optimized minimum primer concentrations, a touchdown thermal cycling protocol, and only two multiplex panels, the method enabled efficient genotyping of 30 commonly found *HBB* mutations specific to the Thai population. The procedure was validated by correctly genotyping 162 thalassemia samples previously identified using conventional techniques of DGGE, restriction fragment length polymorphism, and/or direct DNA sequencing.

Due to the close proximity of the 30 mutation sites, primers used in genotyping *HBB* mutations were chosen so as to avoid false priming, self annealing, and primer-dimer artifact, genotyping errors that could arise in negative controls. Presence of any additional SNP sites in the extension primer sequence itself is an important factor when genotyping homozygous samples carrying such SNP sites as shown in Table 4. Other critical parameters

Table 3. *HBB* Genotypes Tested

Status of subject with known <i>HBB</i> genotype				
Single heterozygous (trait)			Compound heterozygous	Homozygous
c.-137C>A	c.27_28insG	c.124_127delTTCT	c.126delC + c.79G>A	c.-81A>G
c.-81A>G	c.46delT	c.130G>T	c.251G>A + c.79G>A	c.-78A>G
c.-78A>G	c.52A>T	c.170G>A	c.410G>A + c.79G>A	c.19G>A
c.-80T>C	c.59A>G	c.216_217insA		c.20A>T
c.2T>G	c.79G>A	c.316-197C>T		c.52A>T
c.5T>C	c.84_85insC	c.364G>C		c.59A>G
c.19G>A	c.92 + 1G>T	c.370_377delACCCCACC		c.79G>A
c.20A>T	c.92 + 5G>C	c.380T>G		c.92 + 5G>C
c.20A>C	c.108C>A	c.441_442insAC		c.124_127delTTCT

Table 4. Allelic Loss Observed in Multiplex Panel

Multiplex panel tested	Homozygous DNA sample tested	Observed allelic loss (of either allele)	Cause	Solution
Panel 1	c.-81A>G (detectable by Panel 2)	c.-80T>C	Homozygous c.-81A>G causes a strong mismatch at the 3' end of an extension primer for c.-80T>C	DNA sequencing focusing on c.-80T>C
Panel 2	c.19G>A (Hb C) and c.20A>T (Hb S) (detectable by both Panels 1 and 2)	c.27_28insG	Homozygous c.19G>A and c.20A>T result in a mismatched site at the 9 th and 8 th base respectively from 3' end of primer for c.27_28insG whose length is only 14 bases. The primer is furthermore at a disadvantage in competition for the same template region with the other two for c.19 and c.20	DNA sequencing focusing on c.27_28insG or rechecking it by exclusion of primers for c.19G>A and c.20A>T from Panel 2
	c.92 + 5G>C (detectable by Panel 1)	c.92 + 1G>T	Homozygous c.92 + 5G>C has a base mismatch at the 4 th base from 3' end of a primer for c.92 + 1G>T	DNA sequencing focusing on c.92 + 1G>T
Panel 3	c.19G>A (Hb C) (detectable by both Panels 1 and 2)	c.20A>T (Hb S)	Homozygous c.19G>A has a mismatch at 3' end of extension primer for c.20A>T (only in this panel)	c.20A>T (Hb S) can be detected in Panel 1 and 2 without this interference (different primer is used)

determining success of multiplex-minisequencing protocols include high quantity and quality of amplicons, as well as high quality purified primers with appropriate sequences, mass ranges (about 4000 to 9000 Da) and

thermal profiles.²⁹ In this study, optimization of primer extension reaction was achieved by varying combinations of extension primers taking into account priority of *HBB*-mutation types and ensuring compatibility of all

Homozygous c.79G>A (Hb E) (This homozygote <i>does</i> exist.) c.84_85insC primer (Panel 2, 2.10)	c. 79 84 90 5' -GATGAAGTTGGTGGT A AGGCCCTGGGC-3' 3' -CTACTTCAACCACCA T TCGGGACCCG-5'	Mismatched bases does not affect SBE reactions.
	5' -GATGAAGTTGGTGGT A AGGCC-3' (Sense strand)	
Homozygous c.46delT (This homozygote <i>might not</i> exist.) c.52A>T primer (Panel 1, 1.6)	c. 46 52 5' -TTACTGCCCTG-GGGCAAGGTGAA-3' 3' -AATGACGGGAC- C CCCGTTCCACTT-5'	Mismatched bases likely not affect SBE reactions (prediction).
	5' -GTACTGCGCTG T GGGGC-3' (Sense strand)	
c.79G>A (Hb E) primer (Panel 1, 1.5) Homozygous c.84_85insC (This homozygote <i>might not</i> exist.)	3' -TCCGG G ACCCGTCCACC-5' (Anti-sense) 5' -AGTTGGTGGTGAGGC C TGGGCAGGTTG-3' 3' -TCAACCACCACTCCGG G ACCCGTCCAAC-5'	Mismatched bases likely not affect SBE reactions (prediction).
	c. 69 79 84 90	
Homozygous c.2T>G (This homozygote <i>might not</i> exist.) c.5T>C (Hb Raleigh) primer (Panel 2, 2.14)	c. 2 5 10 15 20 5' -ACTAGCAACCTCAAACAGACACCA G GGTGCATCTGACTCCTGA-3' 3' -TGATCGTTGGAGTTTGTCTGTGGT C CCACGTAGACTGAGGACT-5'	Mismatched bases likely inhibit SBE reactions (prediction).
	5' -ACTAGCAACCTCAAACAGACACCA T GG-3' (Sense strand)	

Figure 5. Mismatched bases generated by extension primers used in multiplex SBE reaction and by homozygous alleles of *HBB* mutations. Mismatched positions near the 3' end are likely to cause allelic loss in multiplex SBE reaction, whereas those at sixth or seventh position are not. Mismatched bases are indicated in bold and italics in box.

Table 5. Blinded Validation Analysis of Multiplex PE Panels Coupled with MALDI-TOF MS in 162 Randomly Selected Samples Carrying *HBB* Mutations Previously Genotyped by Conventional Methods

<i>HBB</i> mutation	Number of samples tested by PE panel that gave 100% concordant results compared with conventional methods				
	Panel 1 alone	Panel 2 Alone	Panel 3 alone	Panel 1 + 2 (in same sample)	Panel 1 + 2 + 3 (in same sample)
Homozygous mutation (2 types, 8 samples)					
c.-78A>G	2	1	—	1	—
c.79G>A	6	6	—	5	—
Compound heterozygous mutation (10 types, 34 samples)					
c.-81A>G/c.79G>A	1	2	—	1	—
c.-78A>G/c.59A>G	1	1	—	1	—
c.-78A>G/c.79G>A	2	2	—	2	—
c.52A>T/c.79G>A	5	10	8	5	5
c.59A>G/c.79G>A	1	1	—	1	—
c.92 + 1G>T/c.79G>A	2	4	—	2	—
c.124_127delTTCT/c.79G>A	4	8	—	4	—
c.364G>C/c.79G>A	1	2	—	1	—
c.441_442insAC/c.79G>A	1	2	2	1	1
c.216_217insA/c.441_442insAC	1	2	2	1	1
Heterozygous mutation (15 types, 120 samples)					
c.-81A>G	1	1	—	1	—
c.-78A>G	10	9	—	8	—
c.2T>G	2	1	—	1	—
c.19G>A	2	1	1	1	1
c.52A>T	26	27	36	19	19
c.59A>G	5	2	—	2	—
c.79G>A	3	1	—	1	—
c.92 + 1G>T	3	4	—	3	—
c.92 + 5G>C	3	1	—	1	—
c.124_127delTTCT	33	35	—	27	—
c.216_217insA	3	3	—	3	—
c.316-197C>T	8	7	—	7	—
c.370_377delACCCACCC	1	1	—	1	—
c.410G>A	2	2	—	2	—
c.441_442insAC	3	4	5	2	2
Total	132	140	54	104	29

primers in the set, as well as uniqueness of individual masses. The optimized minimum concentration of primers used in the minisequencing reactions yielded a high signal-to-noise ratio, resulting in easier identification of the extended products. As excess primers can dimerize to form false peaks in the mass spectrum and can compete for the ion current, thereby reducing detection sensitivity of MALDI-TOF MS for the DNA fragments of interest,³⁰ a modified touchdown minisequencing thermal-cycling program with a high cycle number of up to 90 cycles yielded sufficient products for subsequent MS detection, as previously demonstrated by Meyer et al.³¹

As shown in Figure 2, unequal signal intensities for heterozygous alleles are frequently observed in this study. Such uneven extension signals could be caused by differences in SBE efficiency due to preferential incorporation of one ddNTP over another, as well as by competitive ionization and desorption of oligonucleotide molecules in multiplex reactions. This phenomenon is acceptable unless the signals from both alleles are more than 10 times different and the low signal no longer satisfies the signal-to-noise criterion for genotype determination.³² Although low concentrations of primers were used, some primer signals still masked extension signals. Extension products of c.52A>T (either A or T) sometimes exhibited very low or no peak intensity. However, if a peak absence was observed in DNA samples whose

genotypes could be assigned by the other two mutant alleles (in cases of homozygote or compound heterozygote), it could be ignored. Alternatively, problematic samples should be specifically rechecked by improving every step including multiplex-PCR amplification, primer extension, and purification, as well as sample preparation and MALDI-TOF MS analysis. If a repeat test still failed in one panel, the remaining panel could provide a back-up analysis for this mutation (*viz.* c.52A>T). Thus, false-negative results were not obtained from our protocol, resulting in 100% sensitivity.

In general, DNA analysis by MALDI-TOF MS is considered more complicated than protein analysis. Larger DNA fragments can give less intense and broader signals due to DNA instability or fragmentation (mainly depurination), contributing to the effect of a more complex isotope distribution of the poor mass spectrum.²⁵ Higher sensitivity, resolution, and accuracy of the MALDI mass spectrum can only be achieved by using an appropriate TOF instrumental configuration.³³ Linear-mode MALDI-TOF MS analysis, although unable to resolve isotopic peaks of parent and fragment ions, was used in our study to reduce fragmentation and to enhance signals from DNA molecules of mass larger than 3000 Da. Discrimination of monoisotopic ions of 1-Da mass difference, as obtained from the reflectron mode, cannot be seen in the linear mode, but only average masses were shown as a

Table 6. Cost Comparison between MALDI-TOF Genotyping and Conventional Methods Used for *HBB* Genotyping in this Study

	MALDI-TOF genotyping			Conventional method		
	1 Panel	2 Panels	3 Panels	DGGE	RFLP	DNA Sequencing
Number of detectable mutations	21	30	30*	18	1 (Hb E)	11†
Percent total <i>HBB</i> alleles in Thai population	95%	99.5%	99.5*%	65%	30%	4.5%
Analysis time (h)	22	22	22	17	22	14
Reagent cost (\$)						
DNA preparation	5.00	5.00	5.00	5.00	5.00	5.00
PCR	1.25	1.25	1.25	7.50 (1.25 × 6)	1.25	7.50 (1.25 × 6)
Post-PCR	3.75	7.50	11.25	—	0.75	52.50
Analysis step	2.00	4.00	6.00	2.50	—	7.50
Total cost	12.00	17.80	23.50	15.00	7.00	72.50
Cost (\$) per mutation	0.57	0.59	0.78	0.83	7.00	6.60
	Average for 30 mutations = 0.69			Average for 30 mutations = 3.15		

*Optional for specific confirmation of c.20A>T (Hb S), c.52A>T, and c.441_442insAC.

†*HBB* mutations requiring DNA sequencing analysis are c.−137C>A, c.2T>G, c.5T>C (Hb Raleigh), c.19G>A (Hb C), c.20A>T (Hb S), c.20A>C (Hb G Makassar), c.46delT, C.251G>A (Hb Pyrgos), c.364G>C (Hb D Punjab), c.380T>G (Hb Dhonburi), and c.410G>A (Hb Hope).

continuous single peak of each SBE allele due to the aggregation of ions of the individual isotopes, thereby generating the overlapping mass spectral peaks. This can lead to a high background and may result in false positives in those cases with low signal-to-noise spectra. In fact, only a potential false-positive result was generated by a Na⁺ salt adduct (21.98 Da) in the product of c.46delT wild-type allele (6019 Da), giving rise to a 6041-Da peak close to 6044 Da of the mutant allele. However, mutation c.46delT is rare, and when it is present, the mutant peak was always about twice that of the wild-type peak (see Figure 2, 2.10). Nevertheless, a mass difference of 25 Da in the mutant case should allow discrimination from the 22-Da difference of the salt adduct.

The c.79G>A (Hb E) mutation is often missed when the DGGE technique is used, as heteroduplex DNA species formed have very similar mobility in the gel. However, the PE reaction coupled with MALDI-TOF can readily discriminate such alleles based on differences in mass of the extended products.

Any individual DNA sample requested for *HBB* genotyping (regardless of relevant clinical information) should be screened first by Panel 1 multiplex SBE, which covers 95% of mutant *HBB* alleles in the Thai population (cost/mutation of \$0.58). Panel 2 multiplex SBE, which detects an additional 4.5% of the mutations, should be performed on samples not identified by Panel 1. Panel 3 multiplex VSET is used only to confirm the results of three mutations, c.20A>T, c.52A>T, and c.441_442insAC, which constitute approximately 15% of Thai mutant *HBB* alleles. Even based on the use of up to three multiplex panels, unit cost of the protocol for detection of all 30 *HBB* mutations is \$0.78 per genotype determination (average cost \$0.69), while that using a combination of three conventional methods is \$3.15. The major cost of MALDI-TOF genotyping is that for Thermo Sequenase DNA polymerase used in PE reactions. Other contributions to the cost include the purification steps using ExoSAP-IT and Genopure kit, as well as MALDI-TOF MS analysis. To reduce reagent costs further, the reaction volume for primer ex-

tension can be reduced to 10 μ l, allowing the subsequent purification step to use only one half of the original protocol volume. A comparison of the costs involved in MALDI-TOF genotyping methods in this study and other conventional methods is summarized in Table 6. These estimates include only reagent costs. The costs of all plastic wares, consumables, and personnel were assumed to be similar for all of the methods. Nonetheless, the high levels of multiplexing and automation of MALDI-TOF genotyping likely should reduce personnel costs below those of other methods. Instruments can be amortized over a large number of tests for a rather long period of time, and thus the costs can be assumed to be negligible.

Increasingly multiplex PE combined with MALDI-TOF MS has been applied to genotype mutations or SNPs in variety of genes.^{19,22–24,32,34,35} Solid phase capturable ddNTPs in SBE have been developed using biotinylated ddNTPs to generate 3'-biotinylated extension DNA products, which then are purified by streptavidin-coated magnetic beads before analysis by MS.²⁷ This solid phase capturable-SBE method has been applied for simultaneous genotyping of 17 Y-chromosome SNPs,³⁶ detection of 30 point mutations in *p53* in a single tube,³⁷ and for concurrent analysis of 40 SNPs of CYP2C9 and 50 SNPs of CYP2A13 genes.³⁸ However, the cost-effectiveness of this approach has not been evaluated. The method developed in this study employs unmodified oligonucleotide primers, allowing the use of standard magnetic beads in the purification step, and thereby substantially reducing the unit cost when compared with that of the solid phase capturable-SBE procedure.

In conclusion, a reliable multiplex system for simultaneously genotyping 21 (using Panel 1 protocol) and 16 (Panel 2 protocol) *HBB* mutations involving 2 PE reactions and subsequent MALDI-TOF MS analysis has been successfully developed. The 9-Da mass difference between wild-type and mutant allele in 3 *HBB* mutations could be unambiguously confirmed using an additional PE set (Panel 3 protocol). Purification of PE products by a magnetic bead DNA purification system resulted in high quality MALDI-TOF MS spectra. Data analysis was a

bottleneck for high-throughput genotyping using the current method, due to the high level of multiplexing. Software for automatic allele identification should be developed to calibrate each mass spectrum using extension primers in each reaction as internal references and to identify genotypes based on mass differences between primer and extension peaks. Nevertheless, this fast, accurate, and flexible multiplex genotyping system should prove useful for rapid and inexpensive screening of β -thalassaemia carriers and patients, as well as for prenatal diagnosis and molecular screening of the Thai population harboring high frequencies of mutant *HBB* alleles. This approach also could be used as a model in the development of rapid and accurate genotyping of common point mutations in single-gene defects or of SNPs as markers in multigenic disorders, and for identification of potential markers for clinical diagnosis, monitoring and prognosis.

Acknowledgments

We thank Dr. Prapin Wilairat for valuable comments and suggestions.

References

1. Patrinos GP, Giardine B, Riemer C, Miller W, Chui DH, Anagnou NP, Wajcman H, Hardison RC: Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. *Nucleic Acids Res* 2004, 32 (Database issue):D537–D41
2. Fucharoen S, Winichagoon P: Hemoglobinopathies in Southeast Asia. *Hemoglobin* 1987, 11:65–88
3. Levinson D: Thailand. Ethnic groups worldwide: A ready reference handbook. Phoenix, Oryx Press, 1998, pp. 286–289
4. Chinchang W, Viprasak V, Pung-Amritt P, Tanphaichitr VS, Yenichitsomanus PT: Molecular analysis of unknown beta-globin gene mutations using polymerase chain reaction-single strand conformation polymorphism (PCR-SSCP) technique and its application in Thai families with beta-thalassemias and beta-globin variants. *Clin Biochem* 2005, 38:987–996
5. Old JM, Petrou M, Modell B, Weatherall DJ: Feasibility of antenatal diagnosis of beta thalassaemia by DNA polymorphisms in Asian Indian and Cypriot populations. *Br J Haematol* 1984, 57:255–263
6. Maggio A, Giambona A, Cai SP, Wall J, Kan YW, Chehab FF: Rapid and simultaneous typing of hemoglobin S, hemoglobin C, and seven Mediterranean beta-thalassemia mutations by covalent reverse dot-blot analysis: application to prenatal diagnosis in Sicily. *Blood* 1993, 81:239–242
7. Sutcharitchan P, Saiki R, Fucharoen S, Winichagoon P, Erlich H, Embury SH: Reverse dot-blot detection of Thai beta-thalassaemia mutations. *Br J Haematol* 1995, 90:809–816
8. Winichagoon P, Saechan V, Sripanich R, Nopparatana C, Kanokpongsakdi S, Maggio A, Fucharoen S: Prenatal diagnosis of beta-thalassaemia by reverse dot-blot hybridization. *Prenat Diagn* 1999, 19:428–435
9. Fortina P, Dotti G, Conant R, Monokian G, Parrella T, Hitchcock W, Rappaport E, Schwartz E, Surrey S: Detection of the most common mutations causing beta-thalassemia in Mediterraneans using a multiplex amplification refractory mutation system (MARMS). *PCR Methods Appl* 1992, 2:163–166
10. Lee HH, Chang JG, Chen RT, Yang ML, Choo KB: Prenatal diagnosis of beta-thalassemic mutations in Chinese by multiple restriction fragment-single strand conformation polymorphism analysis. *Proc Natl Sci Counc Repub China B* 1994, 18:112–117
11. Ghanem N, Girodon E, Vidaud M, Martin J, Fanen P, Plassa F, Goossens M: A comprehensive scanning method for rapid detection of beta-globin gene mutations and polymorphisms. *Hum Mutat* 1992, 1:229–239
12. Losekoot M, Fodde R, Harteveld CL, van Heeren H, Giordano PC, Bernini LF: Denaturing gradient gel electrophoresis and direct sequencing of PCR amplified genomic DNA: a rapid and reliable diagnostic approach to beta thalassaemia. *Br J Haematol* 1990, 76:269–274
13. Wong C, Dowling CE, Saiki RK, Higuchi RG, Erlich HA, Kazazian HH, Jr: Characterization of beta-thalassaemia mutations using direct genomic sequencing of amplified single copy DNA. *Nature* 1987, 330:384–386
14. Chan K, Wong MS, Chan TK, Chan V: A thalassaemia array for Southeast Asia. *Br J Haematol* 2004, 124:232–239
15. Wang W, Kham SK, Yeo GH, Quah TC, Chong SS: Multiplex minisequencing screen for common Southeast Asian and Indian beta-thalassemia mutations. *Clin Chem* 2003, 49:209–218
16. Su YN, Lee CN, Hung CC, Chen CA, Cheng WF, Tsao PN, Yu CL, Hsieh FJ: Rapid detection of beta-globin gene (HBB) mutations coupling heteroduplex and primer-extension analysis by DHPLC. *Hum Mutat* 2003, 22:326–336
17. Yip SP, Pun SF, Leung KH, Lee SY: Rapid, simultaneous genotyping of five common Southeast Asian beta-thalassemia mutations by multiplex minisequencing and denaturing HPLC. *Clin Chem* 2003, 49:1656–1659
18. Ding C, Chiu RW, Lau TK, Leung TN, Chan LC, Chan AY, Charoenkwan P, Ng IS, Law HY, Ma ES, Xuk X, Wanapiraki C, Sanguanserm-srii T, Liao C, Aim MATJ, Chuin DKH, Cantora CR, Lo YMD: MS analysis of single-nucleotide differences in circulating nucleic acids: Application to noninvasive prenatal diagnosis. *Proc Natl Acad Sci USA* 2004, 101:10762–10767
19. Liao HK, Su YN, Kao HY, Hung CC, Wang HT, Chen YJ: Parallel minisequencing followed by multiplex matrix-assisted laser desorption/ionization mass spectrometry assay for beta-thalassemia mutations. *J Hum Genet* 2005, 50:139–150
20. Ross P, Hall L, Smirnov I, Haff L: High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nature Biotechnol* 1998, 16:1347–1351
21. Sun X, Ding H, Hung K, Guo B: A new MALDI-TOF based mini-sequencing assay for genotyping of SNPs. *Nucleic Acids Res* 2000, 28:E68
22. Blievernicht JK, Schaeffeler E, Klein K, Eichelbaum M, Schwab M, Zanger UM: MALDI-TOF mass spectrometry for multiplex genotyping of CYP2B6 single-nucleotide polymorphisms. *Clin Chem* 2007, 53:24–33
23. Humeny A, Bonk T, Berkholz A, Wildt L, Becker CM: Genotyping of thrombotic risk factors by MALDI-TOF mass spectrometry. *Clin Biochem* 2001, 34:531–536
24. Nakai K, Habano W, Fujita T, Nakai K, Schnackenberg J, Kawazoe K, Suwabe A, Itoh C: Highly multiplexed genotyping of coronary artery disease-associated SNPs using MALDI-TOF mass spectrometry. *Hum Mutat* 2002, 20:133–138
25. Tost J, Gut IG: Genotyping single nucleotide polymorphisms by MALDI mass spectrometry in clinical applications. *Clin Biochem* 2005, 38:335–350
26. Bleicher K, Bayer E: Various factors influencing the signal intensity of oligonucleotides in electrospray mass spectrometry. *Biol Mass Spectrom* 1994, 23:320–322
27. Kim S, Edwards JR, Deng L, Chung W, Ju J: Solid phase capturable dideoxynucleotides for multiplex genotyping using mass spectrometry. *Nucleic Acids Res* 2002, 30:e85
28. Haff LA, Smirnov IP: Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Res* 1997, 7:378–388
29. Carvalho CM, Pena SD: Optimization of a multiplex minisequencing protocol for population studies and medical genetics. *Genet Mol Res* 2005, 4:115–125
30. Roskey MT, Juhasz P, Smirnov IP, Takach EJ, Martin SA, Haff LA: DNA sequencing by delayed extraction-matrix-assisted laser desorption/ionization time of flight mass spectrometry. *Proc Natl Acad Sci USA* 1996, 93:4724–4729
31. Meyer K, Fredriksen A, Ueland PM: High-level multiplex genotyping of polymorphisms involved in folate or homocysteine metabolism by matrix-assisted laser desorption/ionization mass spectrometry. *Clin Chem* 2004, 50:391–402
32. Bray MS, Boerwinkle E, Doris PA: High-throughput multiplex SNP

- genotyping with MALDI-TOF mass spectrometry: practice, problems and promise. *Hum Mutat* 2001, 17:296–304
33. Cotter RJ: Time-of-flight mass spectrometry for the structural analysis of biological molecules. *Anal Chem* 1992, 64:1027A–1039A
34. Paracchini S, Arredi B, Chalk R, Tyler-Smith C: Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. *Nucleic Acids Res* 2002, 30:e27
35. Yang H, Wang H, Wang J, Cai Y, Zhou G, He F, Qian X: Multiplex single-nucleotide polymorphism genotyping by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal Biochem* 2003, 314:54–62
36. Mengel-Jorgensen J, Sanchez JJ, Borsting C, Kirpekar F, Morling N: MALDI-TOF mass spectrometric detection of multiplex single base extended primers. A study of 17 y-chromosome single-nucleotide polymorphisms *Anal Chem* 2004, 76:6039–6045
37. Kim S, Ulz ME, Nguyen T, Li CM, Sato T, Tycko B, Ju J: Thirty fold multiplex genotyping of the p53 gene using solid phase capturable dideoxynucleotides and mass spectrometry. *Genomics* 2004, 83:924–931
38. Misra A, Hong JY, Kim S: Multiplex genotyping of cytochrome p450 single-nucleotide polymorphisms by use of MALDI-TOF mass spectrometry. *Clin Chem* 2007, 53:933–939

Hematological abnormalities in patients with distal renal tubular acidosis and hemoglobinopathies

Sookkasem Khositseth,¹ Apiwan Sirikanaerat,² Siri Khoprasert,³ Sauwalak Opastirakul,⁴ Pornchai Kingwatanakul,⁵ Wanna Thongnoppakhun,⁶ and Pa-thai Yenchitsomanus^{7*}

Mutations of the human *SLC4A1* gene encoding erythroid and kidney isoforms of anion exchanger 1 (AE1, band 3) result in erythrocyte abnormalities or distal renal tubular acidosis (dRTA) and such mutations are observed in Southeast Asia, where hemoglobinopathies are prevalent. Genetic and hematological studies in 18 Thai patients with dRTA have shown that 12 of them (67%) carried *SLC4A1* mutations (7 G701D/G701D, 3 SAO/G701D, and 2 G701D/A858D). Of these 12 patients, three had homozygous G701D/G701D and heterozygous Hb E; one compound heterozygous SAO/G701D and heterozygous α^+ -thalassemia; and one compound heterozygous G701D/A858D and heterozygous Hb E. Of 6 patients without *SLC4A1* mutation, two each carried heterozygous or homozygous Hb E and one of the latter also had Hb H disease ($-\text{SEA}/-\alpha^{4,2}$). The blood smears of patients with homozygous G701D/G701D showed ~25% ovalocytes. Strikingly, the patients with coexistence of homozygous G701D/G701D and heterozygous Hb E had 58% ovalocytes. Similarly, the patients who had compound heterozygous SAO/G701D showed 49% ovalocytes, but the patient with coexistence of compound heterozygous SAO/G701D and heterozygous α^+ -thalassemia had 70% ovalocytes. Our previous study has shown that under metabolic acidosis, the patients with homozygous G701D/G701D or compound heterozygous SAO/G701D had reticulocytosis, indicating compensated hemolysis. A patient with compound heterozygous SAO/G701D and heterozygous α^+ -thalassemia presented with hemolytic anemia and hepatosplenomegaly which was alleviated by alkaline therapy. Taken together, the coexistence of both homozygous or compound heterozygous *SLC4A1* mutations and hemoglobinopathy has a combined effect on red cell morphology and degree of hemolytic anemia, which is aggravated by acidosis. Am. J. Hematol. 83:465–471, 2008. © 2008 Wiley-Liss, Inc.

Introduction

Human solute carrier family 4, anion exchanger, member 1 (*SLC4A1*) gene encodes both erythroid and kidney isoforms of anion ($\text{Cl}^-/\text{HCO}_3^-$) exchanger 1 (AE1) protein (see Fig. 1) [1]. The erythroid AE1 (eAE1) isoform is present in the red cell membrane while the kidney AE1 (kAE1) isoform is found at the basolateral membrane of the type A (α) intercalated cells [2,3] of the renal collecting duct which functions in anion ($\text{Cl}^-/\text{HCO}_3^-$) exchange and indirectly involves in H^+ secretion at the apical membrane of these cells. Thus, *SLC4A1* mutations show pleiotropic effects resulting in two distinct disorders—erythrocyte abnormalities and distal renal tubular acidosis (dRTA) [4].

eAE1 is the most abundant glycoprotein on the erythrocyte membrane, present as dimers or higher oligomers. In addition to anion exchange function, it serves as an anchor protein of the cytoskeleton network, binding to ankyrin, proteins 4.1 and 4.2, and a number of cytoplasmic proteins [5,6]. This binding is critical for maintenance of the biconcave disc shape of red blood cells [7]. *SLC4A1* mutations account for ~20% of hereditary spherocytosis (HS) and almost all ovalocytosis in Southeast Asia [6,8]. HS is a common hereditary hemolytic anemia with the presence of osmotically fragile spheroidal-shape erythrocytes and splenomegaly [9], whereas Southeast Asian ovalocytosis (SAO) is a morphological abnormality of erythrocytes caused by a mutational deletion of 27 base-pairs in codons 400 to 408 of exon 11 of *SLC4A1*, resulting in an in-frame deletion of 9 amino acids at the junction between the cytoplasmic and first transmembrane domains [10]. The SAO may confer a selective advantage by prevention of cerebral malaria [11].

A defect of kAE1 at the basolateral membrane of the α -intercalated cell indirectly results in a failure of H^+ secretion at the apical side of the membrane and an inability to estab-

lish a cell-to-lumen H^+ gradient [3], leading to dRTA. *SLC4A1* mutations can give rise to autosomal dominant (AD) or autosomal recessive (AR) dRTA [12,13]. The latter has been found especially in children in the Southeast Asian populations [14–17]. In Thailand, several genotypes of *SLC4A1* mutations causing AR dRTA have been reported, including the homozygous G701D/G701D mutation, and the compound heterozygous SAO/G701D, SAO/R602H, G701D/S773P, and G701D/A858D mutations [8,14,15,17,18].

Hemoglobinopathies are highly prevalent in Southeast Asia [19]. They are a group of AR disorders characterized

¹Department of Pediatrics, Faculty of Medicine, Thammasat University, Thailand; ²Department of Pediatrics, Khon Kaen Hospital, Khon Kaen, Thailand; ³Department of Pediatrics, Surajthani Hospital, Surajthani, Thailand; ⁴Department of Pediatrics, Chaingmai University, Chaingmai, Thailand; ⁵Department of Pediatrics, Faculty of Medicine, Chulalongkorn University, Chulalongkorn, Thailand; ⁶Division of Molecular Genetics, Department of Research and Development, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand; ⁷Division of Medical Molecular Biology and BIOTEC-Medical Biotechnology Unit, Department of Research and Development, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand

Contract grant sponsor: Thailand Research Fund (TRF); Contract grant number: TRG4880005.

*Correspondence to: Dr. Pa-thai Yenchitsomanus, Division of Medical Molecular Biology, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand. E-mail: grpye@mahidol.ac.th

Received for publication 16 August 2007; Revised 16 December 2007; Accepted 28 December 2007

Am. J. Hematol. 83:465–471, 2008.

Published online 10 January 2008 in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/ajh.21151

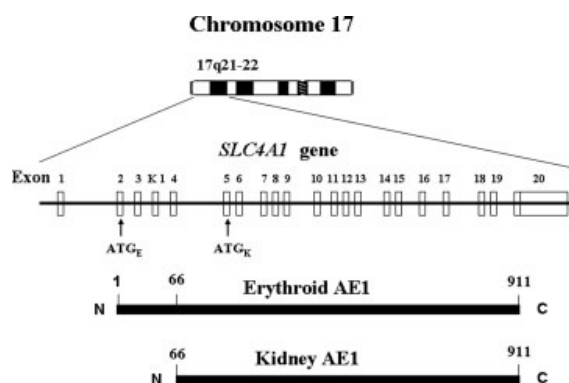


Figure 1. The human *SLC4A1* gene locates on chromosome 17 at the region q21-22. This gene encodes AE1 protein in erythrocytes and α -intercalated cells of the kidneys. Erythroid AE1 (eAE1) contains 911 amino acids while kidney AE1 (kAE1) lacks the first 65 amino acids that is present at the N-terminus of eAE1 due to an alternative transcription using the promoter in intron 3 and alternative splicing using K1 exon to join with exon 4 and others in the mRNA transcript.

by either reduced synthesis of one or more normal globin chains (thalassemias), synthesis of a structurally abnormal globin chain (hemoglobin variants) or reduced synthesis of a hemoglobin variant, e.g., Hb E [20]. Two main forms of thalassemia, α - and β -thalassemia, are caused by mutations of the α - and β -globin gene, respectively. In Thailand, the prevalence of α^0 -thalassemia (or α -thalassemia 1) carrier is ~3–4%, whereas that of α^+ -thalassemia (or α -thalassemia 2) carrier is 20–30% [21]. Approximately 5% of Thais are β -thalassemia carriers [22]. The gene frequencies of Hb E ($\beta 26 \text{ Glu} \rightarrow \text{Lys}$) among Thais are 1–32.7% [23].

Since both *SLC4A1* and globin gene mutations are present in the Thai population, it is possible that these two mutations may coexist in some individuals, which could potentially increase the severity of hematological abnormalities. Indeed, Tanphaichitr et al. [14] have reported xerocytic anemia in two Thai siblings with dRTA due to the homozygous *SLC4A1* G701D/G701D (band 3 Bangkok I) mutation with the presence of homozygous Hb E. Recently, a reversal to normal osmotic fragility test has been reported in a patient with coexistence of heterozygous SAO and β^+ -thalassemia trait [24]. We have previously described the results of *SLC4A1* mutation analysis in a group of Thai pediatric dRTA patients [8]. In the same article, we have reported that the patients with *SLC4A1* mutations presented with compensated hemolysis when they had metabolic acidosis and a patient with compound heterozygous SAO/G701D and heterozygous α^+ -thalassemia exhibited a severe hemolytic anemia which was alleviated by alkaline therapy [8]. However, hematological abnormalities in this group of patients with dRTA associated with *SLC4A1* mutations in the presence of hemoglobinopathies have not been systematically studied and documented.

Herein, we performed hematological studies and have described red cell morphological abnormalities in this group of patients. The data show that all patients with *SLC4A1* mutations had abnormal red cells, if only a degree of ovalocytosis. In those without detectable *SLC4A1* mutations, hematological abnormalities were only present if there was coexistent hemoglobinopathy such as thalassemia. If another hematological abnormality (typically thalassemia) was present in addition to an *SLC4A1* mutation, the phenotype was more marked than expected. We confirmed, in an uncontrolled study, that hemolysis in these cases improves when the acidosis is treated, irrespective of the cause of the hemolysis and irrespective of whether an *SLC4A1* mutation is present or not.

Results

Hematological findings and *SLC4A1* mutations

The hematological findings and *SLC4A1* mutations of 17 patients were previously reported [8]. One patient was added to this study (Q:II-1, Table I). The domicile of each patient is indicated in a map of Thailand (Fig. 2). Selected hematological data after treating iron deficiency are shown in Table I. Briefly, 6 patients had heterozygous Hb E (Hb AE) and 2 patients had homozygous Hb E (Hb EE) (Table I). No β -thalassemia was observed in this group of patients. Mutational deletions of the α -globin gene indicating α -thalassemia were detected in two patients (J:II-2 and Q:II-1). One patient (J:II-2) carried a single α -globin gene deletion ($-\alpha^{3.7}/\alpha\alpha$) compatible with heterozygous α^+ -thalassemia [8] and another patient (Q:II-1) carried both homozygous Hb E (Hb EE) and compound heterozygous α -globin gene deletions ($-\alpha^{\text{SEA}}/\alpha^{4.2}$), known as EFBart's disease [25].

Twelve of 18 patients (67%) were found to carry *SLC4A1* mutations: 7 were homozygous G701D/G701D, 3 compound heterozygous SAO/G701D, and 2 compound heterozygous G701D/A858D. Six patients (33%) had no *SLC4A1* mutations (Table I).

Hematological manifestations of patients with dRTA associated with *SLC4A1* mutations and coexistence of hemoglobin E or thalassemia

Of 7 patients with dRTA who had homozygous G701D/G701D mutation, 4 (A:II-1, B:II-1, C:II-1, and D:II-1) had no anemia and normal hemoglobin type (Hb AA2). Their peripheral blood smears showed ~25% (range = 21–27%) of ovalocytes and rare xerocytes (Tables I and III, and Fig. 4A). Note that the ovalocytic cells in this and other genotypes included the cells with stomatocytic appearance which was less prominent. Another three patients in this group (E:II-1, F:II-1, and G:II-1) had coexistence of heterozygous Hb E (Hb E ~28–30%) with normal hemoglobin level but ~58% (56–62%) ovalocytes, few elliptocytes, and rare xerocytes in peripheral blood smears (Tables I and III, and Fig. 4B).

Three patients (H:II-1, I:II-1, and J:II-1) with dRTA carried compound heterozygous SAO/G701D mutations. Peripheral blood smears of 2 patients (H:II-1 and I:II-1) who had Hb AA2 showed ~49% (46–52%) of ovalocytes, rare elliptocytes, and pincer cells (Tables I and III, and Fig. 4C). One patient (J:II-2) presented with hemolytic anemia (Hb 86 g/L) and hepatosplenomegaly with a serum ferritin of 107 mg/mL. His reticulocyte count was 7%, indicating hemolytic anemia, and Hb type was Hb A 96%, Hb A₂ 3% and Hb F 0.5%. Analysis of the α -globin gene demonstrated a single α -globin gene deletion ($-\alpha^{3.7}/\alpha\alpha$) or heterozygous α^+ -thalassemia. After receiving adequate alkaline therapy, he still had mild anemia (Hb 103 g/L) without hepatosplenomegaly (Table I and Fig. 3). His peripheral blood smear showed 70% of ovalocytes (Table III and Fig. 4D).

Two sisters (K:II-1 and K:II-2) carried compound heterozygous G701D/A858D mutations. Their hematological data were shown previously [8]. Peripheral blood smears of these two sisters (K:II-1 and K:II-2) contained 56% and 31% of ovalocytes, respectively, with rare xerocytes, pincer cells, echinocytes, and acanthocytes (Tables I and III, and Fig. 4E,F). K:II-1 who also carried heterozygous Hb E had a greater number of ovalocytes.

Our previous work [8] has shown that during metabolic acidosis, the patients with dRTA associated with the *SLC4A1* mutation had low hemoglobin levels and reticulocytoses, indicating that they had compensated hemolysis. Interestingly, after receiving adequate alkaline treatment,

TABLE I. *SLC4A1* Genotypes and Hematological Data of Patients with Distal Renal Tubular Acidosis

Patient	Sex/age (year)	Hb typing	<i>SLC4A1</i> genotypes	Hb (g/L)	Hct (%)	MCV (fL)	MCHC (g/dL)	RDW (%)	Retic (%)	Normal RBC (%)	Ovalocyte (%)	Target (%)	Others (%)
A:II-1	M/1.5	A/A2	G701D/G701D	133	36.8	73	35.9	15	2.1	79	21	0	0
B:II-1	M/3.5	A/A2	G701D/G701D	153	43.3	86	35.4	11.4	2.1	74	24	0	1 ^a , 1 ^b
C:II-1	M/2.5	A/A2	G701D/G701D	126	37.5	82	33.6	14	3.0	71	27	0	1 ^b , 1 ^c
D:II-2	M/1.5	A/A2	G701D/G701D	140	39.0	80	36	12	1.3	74	26	0	0
E:II-1	M/3.5	A/E	G701D/G701D	140	39.0	69	36.1	12.1	1.0	39	56	1	1 ^a , 1 ^c , 2 ^d
F:II-1	M/8.5	A/E	G701D/G701D	142	40.0	70	35	12.1	1.2	36	58	2	1 ^a , 1 ^b , 2 ^d
G:II-1	M/4	A/E	G701D/G701D	130	38.0	74	36.6	15	1.2	32	62	1	1 ^b , 4 ^d
H:II-1	M/1.5	A/A2	SAO/G701D	120	37.0	91	38.6	13.5	1.0	50	46	0	1 ^a , 1 ^b , 1 ^c , 1 ^d
I:II-1	M/1.8	A/A2	SAO/G701D	112	36.7	92	30.5	13.3	3.0	48	52	0	0
J:II-2	M/9	A/A2, - $\alpha^{3.7}/\alpha\alpha$	SAO/G701D	103	31.0	90	34	14.1	1.4	26	70	0	1 ^a , 1 ^b , 2 ^c
K:II-2	F/5	A/A2	G701D/A858D	139	38.0	82	37.1	10.9	3.0	62	31	0	2 ^a , 3 ^b , 1 ^d
K:II-1	F/5	A/E	G701D/A858D	120	35.0	80	37	10.7	4.0	41	56	0	1 ^a , 1 ^b , 1 ^c
L:II-3	F/1	A/A2	Normal	130	37.5	81	34.7	13.9	2.0	99	0	0	1 ^b
M:II-2	F/7.8	A/A2	Normal	124	37.0	78	33.5	15.8	2.5	99	0	0	1 ^b
N:II-1	F/1	A/E	Normal	123	36.4	67	33.1	14.2	1.5	91	0	8	1 ^b
O:II-2	F/8	A/E	Normal	130	40.5	74	32	14.7	2.0	89	0	8	3 ^b
P:II-1	F/9	E/E	Normal	115	31.6	57	36.4	16.5	1.5	58	0	26	1 ^b , 6 ^f
Q:II-1	F/16	E/E, - $\text{SEA}/-\alpha^{4.2}$	Normal	109	34.9	47	31.7	17.4	13.0	13	11	49	27 ^b

Hb, hemoglobin; Hct, hematocrit; MCV, mean corpuscular volume; MCHC, mean cell hemoglobin concentration; RDW, red cell distribution width; Retic, reticulocyte count.

^aPincer cell.

^bSchistocyte.

^cXerocyte.

^dElliptocyte.

^eAcanthocyte.

^fEchinocyte.

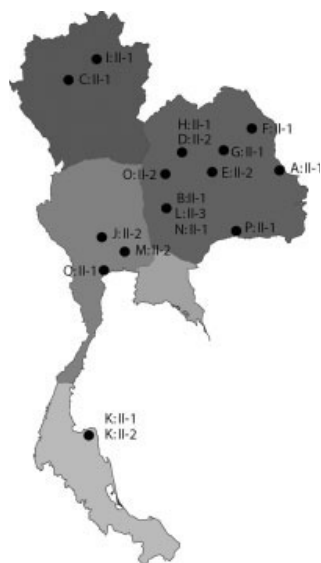


Figure 2. A map of Thailand showing the domiciles of patients indicated by the patients' codes on the map.

their hemoglobin levels were increased and reticulocytoses were abolished (Fig. 3).

Hematological manifestations of patients with dRTA but no *SLC4A1* mutation

Six patients with dRTA (L:II-3, M:II-2, N:II-1, O:II-2, P:II-1, and Q:II-1) had no *SLC4A1* mutation. Their selected hematological data are shown in Table I. The patients with Hb AA2 (L:II-3 and M:II-2) had normal blood smears (Fig. 5A) whereas those with Hb AE (N:II-1 and O:II-2) had some tar-

get cells (8%) (Fig. 5B), and those with Hb EE (P:II-1) had a greater proportion of target cells (26%) (Fig. 5C). One patient (Q:II-1) had both Hb EE and Hb H disease due to compound heterozygous α -globin gene deletions (- $\text{SEA}/-\alpha^{4.2}$). Her hemoglobin type was Hb E (82.04%), Hb F (1.90%), and Hb Bart's (1.32%), compatible with EFBart's disease, and peripheral blood smear showed numerous target cells (49%) and schistocytes (27%) (Fig. 5D). Five of these 6 patients (except for Q:II-1) had neither hemolytic anemia nor reticulocytosis during acidosis. The patient with EFBart's disease (Q:II-1) who had normal ferritin level (106 ng/mL) was slightly anemic (Hb level 109 g/L) with the presence of polychromasias and spherocytes in the blood smear during metabolic acidosis, which resolved after alkaline therapy.

Hematological manifestations of individuals with heterozygous *SLC4A1* mutations

The patients' parents who were heterozygous for *SLC4A1* mutations had no clinical phenotype of dRTA. The hematological data of some individuals are shown in Table II. Three individuals (C:I-1, C:I-2, and J:I-2) with heterozygous G701D and normal hemoglobin type (or with coexistence of heterozygous α^+ -thalassemia) did not have anemia. The peripheral blood smears of C:I-1 and C:I-2 showed normal red cells with few ovalocytes (2–4%) (Tables II and III, and Fig. 6A), whereas those of J:I-2 who also carried heterozygous α^+ -thalassemia showed a greater number of ovalocytes (17%) (Tables II and III, and Fig. 6B). Two individuals (K:I-2 and B:I-1) with heterozygous G701D and heterozygous Hb E (Hb AE) had decreased MCV and 15–16% of ovalocytes in their peripheral blood smears (Tables II and III, and Fig. 6C). An individual (E:I-1) with heterozygous G701D and homozygous Hb E (Hb EE) had a low-normal hemoglobin level and microcytosis. His peripheral blood smear showed many ovalocytes (22%) and target cells (7%) (Tables II and III, and Fig. 6D).

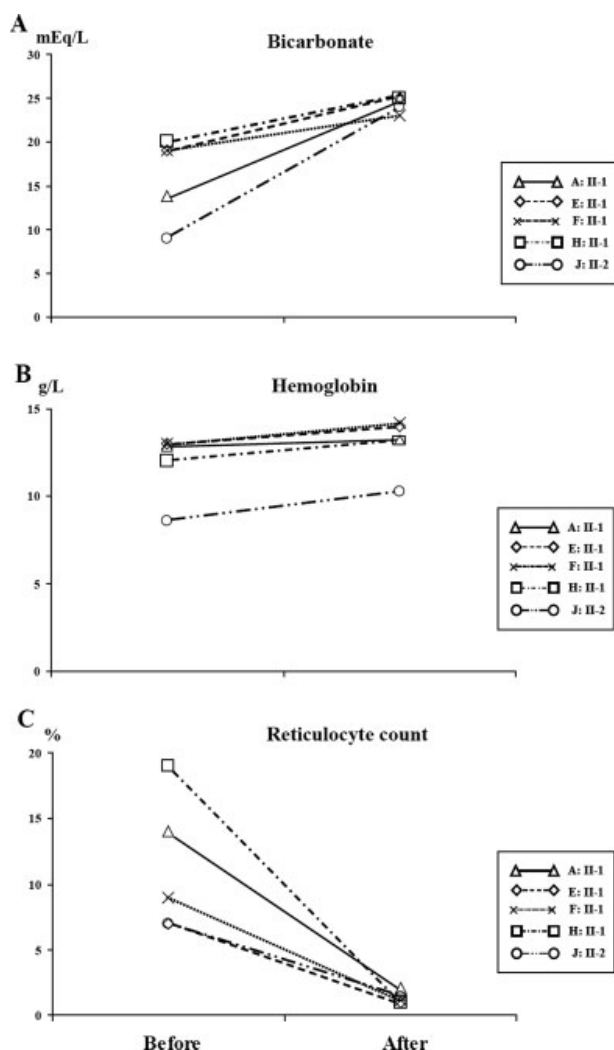


Figure 3. The effect of alkaline therapy on serum bicarbonate (A), hemoglobin (B), and reticulocyte counts (C). Their levels before and after alkaline therapy were demonstrated. The adequate alkaline therapy as indicated by normal serum bicarbonate levels (A) affected slight increases of hemoglobin levels (B) but great decreases of reticulocyte counts (C).

Two individuals (H:I-1 and I:I-2) with heterozygous SAO and normal hemoglobin type had no anemia but numerous ovalocytes (83–85%) (Tables II and III, and Fig. 6E). An individual (K:I-1) with heterozygous A858D was not anemic but he had peculiar red cell morphologies including ovalocytes (20%), acanthocytes (6%), echinocytes (6%), and schistocytes (5%) (Tables II and III, and Fig. 6F).

Discussion

We present herein hematological abnormalities in patients with dRTA in the absence or presence of hemoglobinopathies—common genetic disorders in Southeast Asian populations. A majority (67%) of the patients with dRTA in this study resulted from *SLC4A1* mutations. The most common genotype was homozygous G701D/G701D, accounting for 58% of the dRTA patients associated with *SLC4A1* mutations. The remaining (42%) were compound heterozygous *SLC4A1* mutations (SAO/G701D and G701D/A858D). These homozygous and compound heterozygous *SLC4A1* mutations normally result in morphological changes of erythrocytes, predominantly ovalocytic morphology. As previously described in the studies by our and other groups

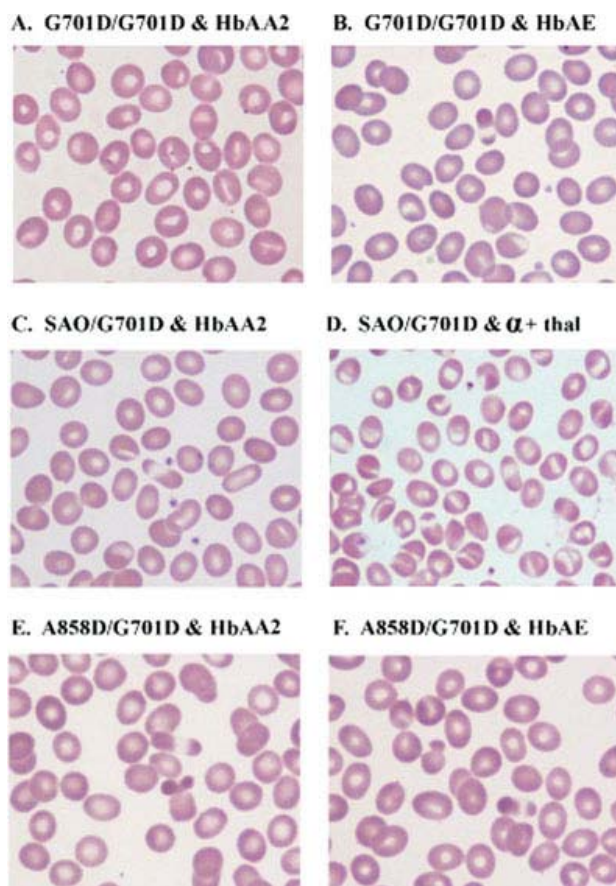


Figure 4. Peripheral blood smears of patients with dRTA-associated with *SLC4A1* mutations in the absence and presence of hemoglobinopathy: (A), homozygous G701D/G701D and Hb AA2, (B), homozygous G701D/G701D and Hb AE, (C), compound heterozygous SAO/G701D and Hb AA2, (D), compound heterozygous SAO/G701D and α^+ thalassemia, (E), compound heterozygous A858D/G701D and Hb AE, and (F), compound heterozygous A858D/G701D and Hb AA2. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

[8,26], under the condition of metabolic acidosis, these morphologically changed erythrocytes were susceptible to hemolysis, indicated by reticulocytosis or anemia. The improvement of anemia after alkali therapy has been reported [26]. However, the anemia in a majority of the patients of the present study was milder than that observed in the patients reported in the previous study. Since hemolysis and reticulocytosis in these patients could be alleviated after the correction of acidosis by alkaline therapy (Fig. 3), the pathological process leading to hematological abnormalities in these patients is likely to involve both eAE1 defect and metabolic acidosis.

By contrast, the patients with dRTA but without *SLC4A1* mutations did not have similar hematological abnormalities; under the condition of metabolic acidosis, they had neither hemolysis nor reticulocytosis. However, they might have some hematological changes attributable to the coexistence of hemoglobinopathies. We found a dRTA patient without a *SLC4A1* mutation who also had EFBart's disease caused by the coinheritance of both homozygous Hb E and Hb H disease ($--^{SEA}/\alpha^{4.2}$). To our knowledge, she is the first case with the presence of both dRTA and EFBart's diseases to be reported. During metabolic acidosis, this patient had anemia with the presence of many target cells, spherocytes, and polychromasias in her blood smear, which also improved with alkaline therapy similar to that which

has been described in our previous report [8]. This finding suggests that, to some extent, metabolic acidosis can induce a greater degree of hemolysis of the thalassemic erythrocytes. Thus, metabolic acidosis may affect erythrocytes with eAE1 defect as well as with thalassemia containing an imbalanced globin chain.

Coexistence of *SLC4A1* mutations and hemoglobinopathies usually caused more severe hematological abnormalities. Hb E, $\beta 26 \text{ Glu} \rightarrow \text{Lys}$, is the most common hemoglobinopathy widely distributed in Southeast Asians, including Thais. Normally, individuals heterozygous and homozygous for Hb E are clinically asymptomatic with variable degrees of microcytosis. In the homozygous condition, microcytosis and target cells are prominent but, anemia, if present, is mild. The coexistence of *SLC4A1* mutations with Hb E was frequent in the studied patients. The presence of both homozygous G701D/G701D and heterozygous Hb E resulted in greatly increased ovalocytosis in addition to mild hemolysis with reticulocytosis in the metabolic acidosis con-

dition. A more severe hemolysis with hepatosplenomegaly was observed in a patient with the coexistence of compound heterozygous SAO/G701D and heterozygous α^+ -thalassemia ($-\alpha^{3.7}/\alpha\alpha$), while the two separate conditions alone do not lead to this severe clinical phenotype. The combination of compound heterozygous SAO/G701D and α^+ -thalassemia may aggravate red blood cell membrane defects and decrease membrane stability in the acidic environment, leading to hemolytic anemia and hepatosplenomegaly. Hemolytic anemia in this patient improved with alkaline therapy. The clinical phenotype of hemolytic anemia and splenomegaly were also described in patients with the compound heterozygous SAO/G701D from Malaysia, Papua New Guinea, and Sarawak [16,26,27]. However, the coexistence of thalassemia, especially α^+ -thalassemia, which is a common form of thalassemia in Southeast Asia [21] and Melanesia [28,29], was not determined in these reported patients with apparently severe hematological phenotype.

Erythrocyte morphologies in the patients with compound heterozygous G701D/A858D were firstly described in this study. In the patients with compound heterozygous G701D/A858D, ovalocytic erythrocytes were prominent, but few pincer cells, schistocytes, and xerocytes or elliptocytes were observed (Table I, Fig. 4E,F). The degree of red cell morphological changes in the patients with compound heterozygous G701D/A858D was less than that recorded for the patients with compound heterozygous $\Delta V850/A858D$ or A858D/SAO [16]. Microcytes, elliptocytes and poikilocytes were found in the case of compound heterozygous $\Delta V850/A858D$ while less elliptocytic red cells were observed in the case of compound heterozygous A858D/SAO. In the individuals with heterozygous A858D mutation, echinocytes and acanthocytes were striking (Fig. 6F). Of the two individuals with heterozygous A858D previously described, one patient had acanthocytic red cells, but the other had normal red cells [16]. Thus, it may be doubtful as to whether acanthocytosis is caused by some other factor or this morphological change is corrected by some unknown cause. The milder degree of erythrocyte morphological changes with the absence of echinocytes and acanthocytes in the patients with compound heterozygous G701D/A858D as found in our study suggests that AE1 G701D may be able to compensate the effect of AE1 A858D. The patient with the coexistence of compound heterozygous G701D/A858D and heterozygous Hb E had more ovalocytic change, a lower hemoglobin level, and little more reticulocytosis than the one who had simply compound heterozy-

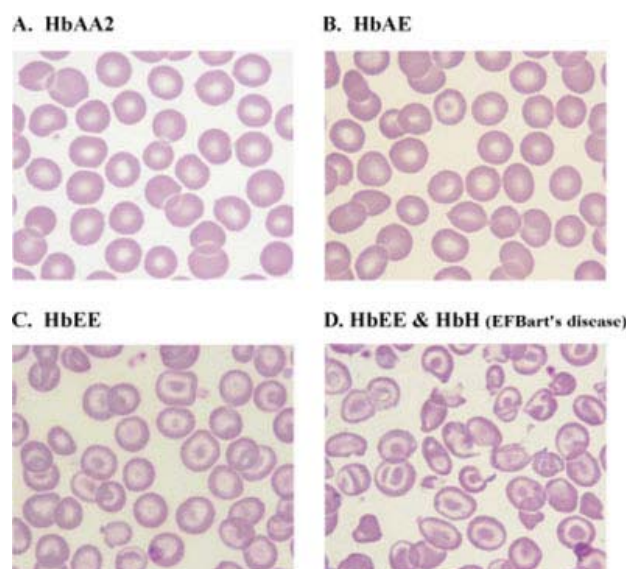


Figure 5. Peripheral blood smears of patients with dRTA without *SLC4A1* mutation in the absence and presence of hemoglobinopathy: (A), Hb AA2, (B), Hb AE, (C), Hb EE, and (D), Hb EE and Hb H (EFBart's disease). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

TABLE II. Hematological Data of Available Parents with Heterozygous *SLC4A1* Mutation

Parent	Sex/age (year)	Hb typing	<i>SLC4A1</i> genotypes	Hb (g/L)	Hct (%)	MCV (fL)	MCHC (g/dL)	RDW (%)	Retic (%)	Normal RBC (%)	Ovalocyte (%)	Target (%)	Others
C:I-1	M/34	A/A2	G701D/N	180	54	84	34	12	1.0	96	4	0	0
C:I-2	F/32	A/A2	G701D/N	140	40	87	34	12	1.0	98	2	0	0
J:I-2	F/35	A/A2, $-\alpha^{4.2}/\alpha\alpha$	G701D/N	137	41	83.9	35	12	1.5	83	17	0	0
K:I-2	F/37	A/E	G701D/N	120	35	76	36	10	1.0	81	16	3	0
B:I-1	M/32	A/E	G701D/N	140	41	75	36	11	2.0	83	15	2	0
E:I-1	M/36	EE	G701D/N	120	36	59	34	14.4	1.5	69	22	7	2 ^a
H:I-1	M/50	A/A2	SAO/N	140	41	97	35	13.4	1.0	17	83	0	0
I:I-2	F/22	A/A2	SAO/N	130	39	88	34	11.5	1.0	15	85	0	0
K:I-1	M/36	A/A2	A858D/N	140	38	88	37	10.8	2.0	63	20	0	5 ^a , 6 ^b , 6 ^c

Hb, hemoglobin; Hct, hematocrit; MCV, mean corpuscular volume; MCHC, mean cell hemoglobin concentration. RDW, red cell distribution width; Retic, reticulocyte count.

^aSchistocyte.

^bAcanthocyte.

^cEchinocyte.

TABLE III. Percentages of Ovalocytes in Peripheral Blood Smears of the Patients with dRTA and Their Available Parents

Genotypes	Ovalocyte (%)
Normal <i>SLC4A1</i> and Hb AA2	0
Normal <i>SLC4A1</i> and Hb AE	0
Normal <i>SLC4A1</i> and Hb EE	0
Normal <i>SLC4A1</i> and Hb EE Hb H	11
G701D/G701D and Hb AA2	21–27
G701D/G701D and Hb AE	56–62
SAO/G701D and Hb AA2	46–52
SAO/G701D and α^+ -thalassemia	70
A858D/G701D and Hb AA2	31
A858D/G701D and Hb AE	56
G701D/N and Hb AA2	2–4
G701D/N and Hb H	17
G701D/N and Hb AE	15–16
G701D/N and Hb EE	22
SAO/N and Hb AA2	83–85
A858D/N and Hb AA2	20

gous G701D/A858D without heterozygous Hb E, indicating an additional effect of Hb E or excessive α -globin chain or both on the red cells with this compound heterozygous condition.

The underlying mechanism of hemolysis in patients with dRTA-associated *SLC4A1* mutations in concomitance with Hb E or thalassemia is unknown. One hypothesis is that the combined damage to red cell membranes caused by structural and functional abnormalities of eAE1 in addition to an excess of α -globin (in Hb E) or β -globin (in α -thalassemia) that precipitates and binds to the defective membrane cytoskeleton may enhance the susceptibility to hemolysis. Moreover, abnormal eAE1 leads to anion imbalance, and excessive α - or β -globin can produce intracellular oxidative stress [30,31]. All these combined defects may cause hemolysis upon exposure to metabolic acidosis. Further study of red cell membrane protein defects in these combined conditions should provide a better understanding of the hemolytic mechanism.

The cause of dRTA in the 6 patients without *SLC4A1* mutation may be due to either an environmental factor or mutation of other gene. Autoimmune diseases were not observed in this group of patients who were very young, and the clinical phenotype of the disease in these patients is not compatible with the defect of carbonic anhydrase II (CAII) gene which is associated with dRTA and osteopetrosis. Mutations of *ATP6V1B1* or *ATP6V0A4* encoding B1 and a4 subunits of H^+ -ATPase can also cause AR dRTA [32,33]. Severe sensorineural hearing loss has been reported in the patients with mutations of these genes. Although clinical hearing loss was not recorded in our 6 patients without *SLC4A1* mutation, the study of *ATP6V1B1* or *ATP6V0A4* should be further carried out in these patients. It should be pointed out that these genes are not known to express in red cells. Thus, their defects would be unlikely to cause erythrocyte morphological changes.

In conclusion, under the metabolic acidosis condition, greater degrees of hematological abnormalities including ovalocytosis and hemolytic anemia were observed in the patients with the coexistence of dRTA-associated with *SLC4A1* mutations and hemoglobinopathy than in patients with the same condition but no coexisting hemoglobinopathy. The example was clearly demonstrated in the patient with presence of both compound heterozygous SAO/G701D and heterozygous α^+ -thalassemia who had hemolytic anemia and hepatosplenomegaly. The hemolysis and

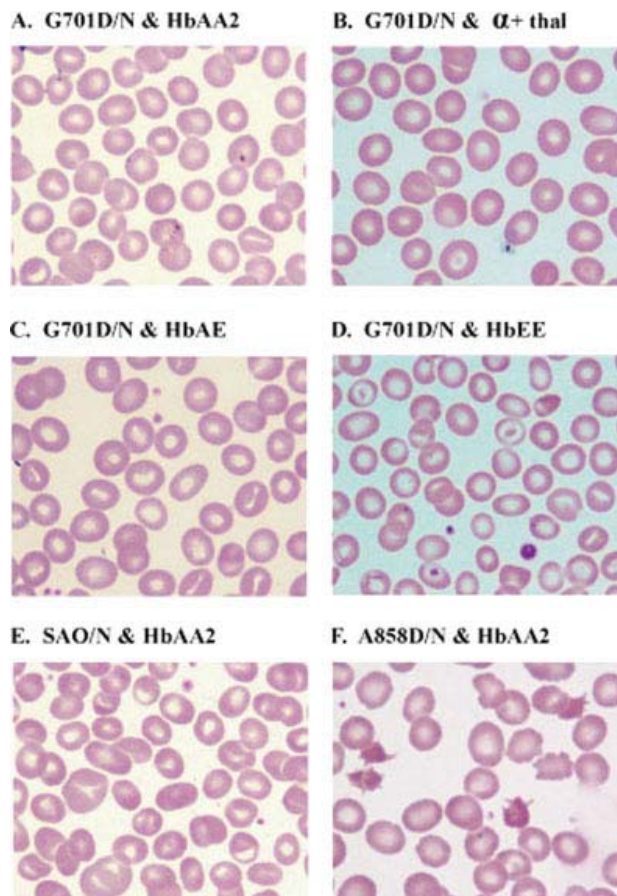


Figure 6. Peripheral blood smears of selected parents with heterozygous *SLC4A1* mutations in the absence and presence of hemoglobinopathy: (A), G701D/N and Hb AA2, (B), G701D/N and α^+ thalassemia, (C), G701D/N and Hb AE, (D), G701D/N and Hb EE, (E), SAO/N and HbAA2, and (F), A858D/N and HbAA2. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

clinical symptoms in these patients could be alleviated by an adequate alkaline therapy owing to the elimination of unfavorable condition of metabolic acidosis.

Methods

Patients with dRTA

Pediatric patients diagnosed with dRTA were recruited from six referral hospitals in different regions of Thailand (Fig. 2). The study protocol was approved by the pertinent Institutional Review Board. A total of 18 patients with dRTA, 10 males and 8 females, with mean $\pm$ SD age at diagnosis of 4.4 ± 2.9 years, were studied. The clinical manifestations and the results of *SLC4A1* analysis of 17 patients in this study have previously been reported [8]. After giving informed consent to their parents, patients' blood samples were collected for laboratory analyses. The criteria for diagnosis of dRTA were persistent hyperchloremic metabolic acidosis ($HCO_3^- < 18$ mmol/L), urinary pH > 5.5 , and absence of glucosuria. The patients with creatinine clearance lower than 60 mL/1.73 m²/min were excluded from the study. For more details of laboratory results, the reader may refer to our previous publication [8].

Hematological studies and α -globin gene analysis

Hematological studies and peripheral blood smears were performed for all patients and available parents. Reticulocyte count was conducted by methylene blue staining, and serum ferritin was measured by immunoassay. Morphological abnormalities were examined and scored from 1,000 red blood cells in one slide of each individual. The methods for

hematological studies, hemoglobin typing, and α -globin gene analysis have been reported in our recent article [8].

Analysis of *SLC4A1* mutations

Genomic DNA was prepared from peripheral blood leukocytes. The regions of all exons of *kAE1* (exons 4 to 20) as well as the kidney promoter sequence in intron 3 were amplified by PCR then subjected to initial mutation screening by denaturing high-performance liquid chromatography (DHPLC) analysis using Wave DNA Fragment Analysis System (Transgenomic[®], Crewe, UK) as previously described [8]. The fragments showing abnormal elution profiles on DHPLC were reamplified by PCR, purified using QIAquick Gel Extraction Kit (Qiagen, GmbH, Germany), and sequenced by using ABI-Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit and an automated ABI-PRISM310 sequencer (Applied Biosystems[®], Foster City, CA). Examinations of the *SLC4A1* mutations causing G701D and SAO were also conducted by using PCR-based methods as previously described [15,17].

Acknowledgments

Thammasat University Research Development Fund. PY is a TRF-Senior Research Scholar and also supported by the Pre-clinic Staff Development Fund from the Faculty of Medicine Siriraj Hospital, Mahidol University. The authors thank all the parents and patients for their participation and appreciate the technical assistances from Werasak Sasanakul, Nunghathai Sawasdee, Atchara Paemanee, Sirintra Nakjang, Thursak Sinthana, and Pongsaton Chinabondee.

References

- Schofield AE, Martin PG, Spillet D, Tanner MJ. The structure of the human red blood cell anion exchanger (EPB3, AE1, band 3) gene. *Blood* 1994;84: 2000–2012.
- Wagner S, Vogel R, Lietzke R, et al. Immunochemical characterization of a band 3-like anion exchanger in collecting duct of human kidney. *Am J Physiol* 1987;253:F213–F221.
- Alper SL, Natale J, Gluck S, et al. Subtypes of intercalated cells in rat kidney collecting duct defined by antibodies against erythroid band 3 and renal vacuolar H⁺-ATPase. *Proc Natl Acad Sci USA* 1989;86:5429–5433.
- Yenchitsomanus PT. Human anion exchanger1 mutations and distal renal tubular acidosis. *Southeast Asian J Trop Med Public Health* 2003;34:651–658.
- Low PS. Structure and function of the cytoplasmic domain of band 3: Center of erythrocyte membrane-peripheral protein interactions. *Biochim Biophys Acta* 1986;864:145–167.
- Tanner MJ. The structure and function of band 3 (AE1): Recent developments (review). *Mol Membr Biol* 1997;14:155–165.
- Tanner MJ. Band 3 anion exchanger and its involvement in erythrocyte and kidney disorders. *Curr Opin Hematol* 2002;9:133–139.
- Khositseth S, Sirikanerit A, Wongbenjarat K, et al. Distal renal tubular acidosis associated with anion exchanger 1 mutations in children in Thailand. *Am J Kidney Dis* 2007;49:841–850;e841.
- Bruce LJ, Tanner MJ. Erythroid band 3 variants and disease. *Baillieres Best Pract Res Clin Haematol* 1999;12:637–654.
- Jarolim P, Palek J, Amato D, et al. Deletion in erythrocyte band 3 gene in malaria-resistant Southeast Asian ovalocytosis. *Proc Natl Acad Sci USA* 1991; 88:11022–11026.
- Allen SJ, O'Donnell A, Alexander ND, et al. Prevention of cerebral malaria in children in Papua New Guinea by southeast Asian ovalocytosis band 3. *Am J Trop Med Hyg* 1999;60:1056–1060.
- Yenchitsomanus PT, Sawasdee N, Paemanee A, et al. Anion exchanger 1 mutations associated with distal renal tubular acidosis in the Thai population. *J Hum Genet* 2003;48:451–456.
- Yenchitsomanus P, Kittanakom S, Rungroj N, et al. Molecular mechanism of autosomal dominant and recessive distal renal tubular acidosis caused by *SLC4A1* (*AE1*) mutations. *J Mol Genet Med* 2005;1.
- Tanphaichitr VS, Sumboonnanonda A, Ideguchi H, et al. Novel AE1 mutations in recessive distal renal tubular acidosis. Loss-of-function is rescued by glycoporphin A. *J Clin Invest* 1998;102:2173–2179.
- Vasuvattakul S, Yenchitsomanus PT, Vachuanichsanong P, et al. Autosomal recessive distal renal tubular acidosis associated with Southeast Asian ovalocytosis. *Kidney Int* 1999;56:1674–1682.
- Bruce LJ, Wrong O, Toye AM, et al. Band 3 mutations, renal tubular acidosis and South-East Asian ovalocytosis in Malaysia and Papua New Guinea: Loss of up to 95% band 3 transport in red cells. *Biochem J* 2000;350 (Part 1):41–51.
- Yenchitsomanus PT, Vasuvattakul S, Kirdpon S, et al. Autosomal recessive distal renal tubular acidosis caused by G701D mutation of anion exchanger 1 gene. *Am J Kidney Dis* 2002;40:21–29.
- Sritippayawan S, Sumboonnanonda A, Vasuvattakul S, et al. Novel compound heterozygous *SLC4A1* mutations in Thai patients with autosomal recessive distal renal tubular acidosis. *Am J Kidney Dis* 2004;44:64–70.
- Weatherall DJ, Clegg JB. Inherited haemoglobin disorders: An increasing global health problem. *Bull World Health Organ* 2001;79:704–712.
- Old JM. Screening and genetic diagnosis of haemoglobin disorders. *Blood Rev* 2003;17:43–53.
- Wasi P, Pootrakul S, Pootrakul P, et al. Thalassemia in Thailand. *Ann NY Acad Sci* 1980;344:352–363.
- Fucharoen S, Winichagoon P. Hemoglobinopathies in Southeast Asia: Molecular biology and clinical medicine. *Hemoglobin* 1997;21:299–319.
- Livingstone FB. Frequencies of Hemoglobin Variants Thalassemia, the Glucose-6-phosphate dehydrogenase Deficiency, G-6-PD Variants, and Ovalocytosis in Human Population. New York, Oxford: Oxford University Press; 1985.
- Fucharoen G, Fucharoen S, Singsanan S, Sanchaisuriya K. Coexistence of Southeast Asian ovalocytosis and β -thalassemia: A molecular and hematological analysis. *Am J Hematol* 2007;82:381–385.
- Forget BG. CA: Thalassemia syndrome. In: Hoffman R, Benz EJ, Shattil SJ, et al., editors. *Hematology: Basic Principle and Practice*. Philadelphia, PA: Elsevier Churchill Livingstone, 2005. pp 557–589.
- Choo KE, Nicoli TK, Bruce LJ, et al. Recessive distal renal tubular acidosis in Sarawak caused by AE1 mutations. *Pediatr Nephrol* 2006;21:212–217.
- Wrong O, Bruce LJ, Unwin RJ, et al. Band 3 mutations, distal renal tubular acidosis, and Southeast Asian ovalocytosis. *Kidney Int* 2002;62:10–19.
- Yenchitsomanus PT, Summers KM, Bhatia KK, et al. Extremely high frequencies of α -globin gene deletion in Madang and on Kar Kar Island, Papua New Guinea. *Am J Hum Genet* 1985;37:778–784.
- Yenchitsomanus P, Summers KM, Board PG, Bhatia KK, et al. Alpha-thalassemia in Papua New Guinea. *Hum Genet* 1986;74:432–437.
- Stocks J, Offerman EL, Modell CB, Dormandy TL. The susceptibility to autooxidation of human red cell lipids in health and disease. *Br J Haematol* 1972; 23:713–724.
- Brunori MFG, Fioretti E. Formation of superoxide in the autooxidation of the isolated alpha and beta chains of human hemoglobin and its involvement in hemichrome precipitation. *Eur J Biochem* 1975;53:99–104.
- Karet FE, Finberg KE, Nelson RD, Nayir A, et al. Mutations in the gene encoding B1 subunit of H⁺-ATPase cause renal tubular acidosis with sensorineural deafness. *Nat Genet* 1999;21:84–90.
- Stover EH, Borthwick KJ, Bavalua C, et al. Novel ATP6V1B1 and ATP6V0A4 mutations in autosomal recessive distal renal tubular acidosis with new evidence for hearing loss. *J Med Genet* 2002;39:796–803.



Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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



Identification of human hnRNP C1/C2 as a dengue virus NS1-interacting protein

Sansanee Noisakran^{a,b,1}, Suchada Sengsai^{b,1}, Visith Thongboonkerd^{b,c}, Rattiyaporn Kanlaya^{b,c}, Supachok Sinchaikul^d, Shui-Tein Chen^{d,e}, Chunya Puttikhunt^{a,b}, Watchara Kasinrerak^{a,f}, Thawornchai Limjindaporn^b, Wiyada Wongwiwat^b, Prida Malasit^{a,b}, Pa-thai Yenchitsomanus^{a,b,*}

^a Medical Biotechnology Unit, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Pathumthani 12120, Thailand

^b Medical Molecular Biology Unit, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Adulyadejvikrom Building (12th Floor), Mahidol University, Bangkok 10700, Thailand

^c Medical Proteomics Unit, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

^d Institute of Biological Chemistry and Genomic Research Center, Academia Sinica, Taipei, Taiwan

^e Institute of Biochemical Sciences, College of Life Science, National Taiwan University, Taipei, Taiwan

^f Department of Clinical Immunology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai 50200, Thailand

ARTICLE INFO

Article history:

Received 22 April 2008

Available online 8 May 2008

Keywords:

hnRNP C1/C2

Dengue virus

NS1

Interacting protein

Host and viral protein interaction

Virus replication

ABSTRACT

Dengue virus nonstructural protein 1 (NS1) is a key glycoprotein involved in the production of infectious virus and the pathogenesis of dengue diseases. Very little is known how NS1 interacts with host cellular proteins and functions in dengue virus-infected cells. This study aimed at identifying NS1-interacting host cellular proteins in dengue virus-infected cells by employing co-immunoprecipitation, two-dimensional gel electrophoresis, and mass spectrometry. Using lysates of dengue virus-infected human embryonic kidney cells (HEK 293T), immunoprecipitation with an anti-NS1 monoclonal antibody revealed eight isoforms of dengue virus NS1 and a 40-kDa protein, which was subsequently identified by quadrupole time-of-flight tandem mass spectrometry (Q-TOF MS/MS) as human heterogeneous nuclear ribonucleoprotein (hnRNP) C1/C2. Further investigation by co-immunoprecipitation and co-localization confirmed the association of hnRNP C1/C2 and dengue virus NS1 proteins in dengue virus-infected cells. Their interaction may have implications in virus replication and/or cellular responses favorable to survival of the virus in host cells.

© 2008 Elsevier Inc. All rights reserved.

Introduction

Dengue virus is a mosquito-borne human pathogen which causes a serious public health concern around the world with approximately 100 million cases of dengue infection and 500,000 cases of hospitalizations per annum [1]. The fatality rate of the affected individuals is about 1–5% and occurs mostly in children [1]. However, the mechanisms involved in the pathogenesis of dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS) remain unraveled.

Dengue virus is a positive, single-stranded RNA virus in the genus *Flavivirus* of the family *Flaviviridae* and contains a 11-kb genome encoding three-structural proteins (capsid, C; premembrane, prM; and envelope, E) and seven-nonstructural proteins (NS1,

NS2A, NS2B, NS3, NS4A, NS4B, and NS5) [2]. In virus-infected cells, newly synthesized NS1 appears as a monomer in the lumen of the endoplasmic reticulum (ER) and subsequently undergoes glycosylation and dimerization as the protein is transported along the host secretory pathway to the cell surface and eventually to the extracellular milieu [3–5]. The exact roles of NS1 in each compartment are not clearly understood.

Secreted NS1 is found to activate complements in the presence or absence of specific antibodies and interact with human complement regulatory protein clusterin, potentially leading to viral and host immune complex formation and subsequent plasma leakage [6,7]. The correlation between levels of secreted NS1 and disease severity has also been observed [7,8]. Unlike the secreted form, cell surface-associated NS1 requires cross-linking of specific antibodies to induce efficient complement activation and intracellular signal transduction in response to dengue virus infection [3,7].

How the NS1 molecule functions inside virus-infected cells is still elusive. A number of previous studies propose the role of intracellular NS1 in the maturation process of dengue virus [9–11]. The NS1 molecule co-localizes with double-stranded dengue viral RNA (dsRNA), and associates with intracellular membrane structures,

* Corresponding author. Address: Medical Molecular Biology Unit, Department of Research and Development, Adulyadejvikrom Building (12th Floor), Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand. Fax: +66 2 418 4793.

E-mail address: grppe@mahidol.ac.th (P.-t. Yenchitsomanus).

¹ These authors contributed equally to this work.

which are presumed sites of virus replication, and possibly with other viral nonstructural proteins, including NS2A, NS3, NS4A, and NS5 to form viral replication complexes in virus-infected cells [9–12]. Very little is known about the interplay between dengue virus NS1, host proteins, and cellular responses during dengue virus infection. We therefore hypothesized that the intracellular NS1 may interact with host cellular proteins to facilitate its proper folding, trafficking and/or to promote favorable environment for virus production in the host cell. Biochemical and proteomic approaches were utilized in this study to identify NS1-interacting proteins and subsequently confirm the protein–protein interaction.

Materials and methods

Cell line, virus, and antibodies. A human embryonic kidney epithelial cell line, 293T, dengue virus serotype 2 strain 16681, and mouse monoclonal antibodies recognizing linear epitopes (1B2, NS1-1F, NS1-3F, and NS1-4F) or conformational epitopes (NS1-8.2 and 1A4) on dengue virus NS1 were obtained as described previously [13–15]. A mouse anti-human hnRNP C1/C2 monoclonal antibody (clone 4F4) was purchased from Santa Cruz Biotechnology, Inc., CA, USA.

Dengue virus infection and immunoprecipitation. 293T cells were infected with dengue virus at a multiplicity of infection (MOI) of 1 and harvested at indicated time points post-infection to verify the percentage of dengue virus infection by immunofluorescence staining for viral E and NS1 antigen expression and flow cytometry [15]. Mock-infected cells served as negative controls. Clear lysates were prepared from mock and dengue virus-infected cells and then subjected to immunoprecipitation using a mouse isotype-matched control IgG1 or IgG2a antibody (MOPC 21 or UPC 10, Sigma, St. Louis, MO, USA), a mouse anti-NS1 monoclonal antibody (1A4, IgG2a), or a mouse anti-human hnRNP C1/C2 (4F4, IgG1) according to a previously described method with minor modifications [15].

Two-dimensional (2-D) gel electrophoresis and mass spectrometry. Immunoprecipitated proteins (200 µg) were subjected to the first dimensional separation by isoelectric focusing (IEF) on Immobiline DryStrip (nonlinear pH 3–10, 7-cm long; Amersham Bioscience) and the second dimensional separation in 12% SDS–polyacrylamide gel [16]. Separated protein spots were visualized by staining with Coomassie brilliant blue R-250 and the protein spot of interest was excised and in-gel digested with trypsin. Peptides were extracted from the gel pieces, purified, and analyzed by Q-TOF Ultima mass spectrometer (Micromass, Manchester, UK) according to the methods described previously [16].

Immunoblot analysis. Immunoprecipitated proteins, which had been heated or left unheated at 95 °C for 5 min in the presence or absence of 5% mercaptoethanol, were separated by electrophoresis in 10% SDS–polyacrylamide gel and transferred to a PVDF membrane (Millipore Corporation, Billerica, USA) using a SemiPhor semi-dry transphor unit (Amersham Bioscience). The membrane was processed as described previously [15], except that a mixture of mouse anti-NS1 monoclonal antibodies described above or a mouse anti-human hnRNP C1/C2 monoclonal antibody (4F4) was utilized in this study. The immunoreactive proteins were visualized by Western Lightning Chemiluminescence Reagent Plus (Perkin-Elmer Applied Biosystems, Foster City, CA, USA).

Double immunofluorescence staining. Mock and dengue virus-infected cells on glass coverslips were fixed with 3.7% formaldehyde in PBS for 7 min followed by absolute methanol for 10 min at room temperature (RT). The cells were incubated with a mouse isotype-matched control IgG1 antibody (MOPC 21) or a mouse anti-human hnRNP C1/C2 monoclonal antibody (4F4) at the dilution of 1:500 for 1 h at RT. Successive incubations of the cells for 30 min at RT

in the dark were then performed using a Cy3-conjugated rabbit anti-mouse IgG antibody (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA, USA) at the dilution of 1:4000, 10% normal mouse serum in PBS, and 20 µg/ml of a FITC-conjugated mouse anti-NS1 monoclonal antibody (1A4). The stained cells were visualized under a laser-scanning confocal microscope (LSM 510 Meta, Carl Zeiss, Jena, Germany).

Results and discussion

Determination of a suitable condition for preparation of dengue virus-infected cell lysates

Initially, HEK 293T cell line that had been infected with dengue virus at an MOI of 1 were collected daily for 3 days and assessed for

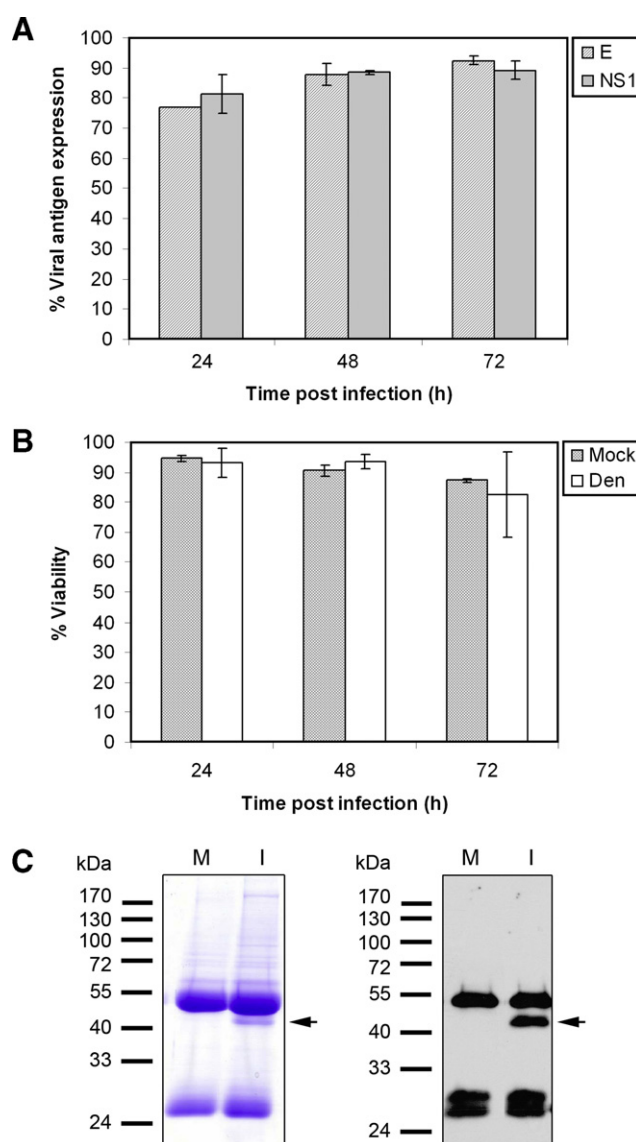


Fig. 1. Dengue virus infection of HEK 293T cells. Mock and virus-infected cells were assessed for (A) the expression of viral E and NS1 by immunofluorescence staining and flow cytometry and (B) the percentage of cell viability by trypan blue exclusion. Mock (M) and dengue virus-infected (I) cell lysates were subjected to immunoprecipitation using an anti-NS1 monoclonal antibody. Precipitated proteins were electrophoresed in a 10% SDS–polyacrylamide gel under reducing and heated conditions and visualized by Coomassie blue staining (C, left panel) or immunoblot analysis probed with an anti-NS1 specific antibody (C, right panel). Arrows indicate the dengue virus NS1 protein.

the percentage of dengue virus infection and cell viability. Mock-infected cells served as negative controls. Infection of HEK 293T cells with dengue virus for 48 h resulted in high levels of viral antigen expression (Fig. 1A) but low percentage of cell death (Fig. 1B). As a result, we employed this condition for preparing cell lysates and tested whether dengue virus NS1 could be immunoprecipitated from these samples using an anti-NS1 monoclonal antibody.

Analysis by SDS-PAGE and Coomassie blue staining revealed that, under reducing and heated conditions, a unique 46-kDa protein band was detectable in the precipitated sample from virus-infected cell lysate (Fig. 1C, left panel, lane I), but not in that from mock control (Fig. 1C, left panel, lane M). The reactivity with the anti-NS1 antibody in immunoblot analysis (Fig. 1C, right panel, lane I) strongly indicated that the specific 46-kDa protein was monomeric NS1 of dengue virus. Two additional protein bands of about 25 and 50 kDa were also observed in the precipitated samples from mock and dengue virus-infected cell lysates (Fig. 1C, left panel, lanes M and I), corresponding to light and heavy chains of IgG used for the immunoprecipitation (Fig. 1C, right panel, lanes M and I). These results therefore demonstrated that the infecting condition used in this study was adequate for establishment of dengue virus-infected samples for immunoprecipitation.

Identification of dengue virus NS1-interacting proteins by 2-D gel electrophoresis and mass spectrometry

Our results of SDS-PAGE could not show clearly any other proteins that were co-precipitated with dengue virus NS1 following immunoprecipitation with the specific antibody. This may have been due to insufficient amounts of the starting materials used for immunoprecipitation and/or the limited ability of SDS-PAGE to separate proteins with similar molecular sizes. We therefore prepared large amounts of mock and dengue virus-infected cell lysates, utilized them for immunoprecipitation using an isotype-matched control antibody or an anti-NS1 monoclonal antibody, and analyzed precipitated proteins by 2-D gel electrophoresis.

Immunoprecipitation of dengue virus-infected cell lysate with the isotype-matched control antibody resulted in the appearance of two major sets of protein spots of about 25 and 50 kDa and other minor protein spots on the 2-D SDS-polyacrylamide gel with similar patterns observed in the mock-infected cell lysate precipitated with the anti-NS1 antibody [Fig. 2A, Mock (anti-NS1) and DENV-2 (control IgG)]. Immunoblot analysis of the 25- and 50-kDa protein spots using rabbit anti-mouse immunoglobulin antibody suggested that these proteins may correspond to different isoforms

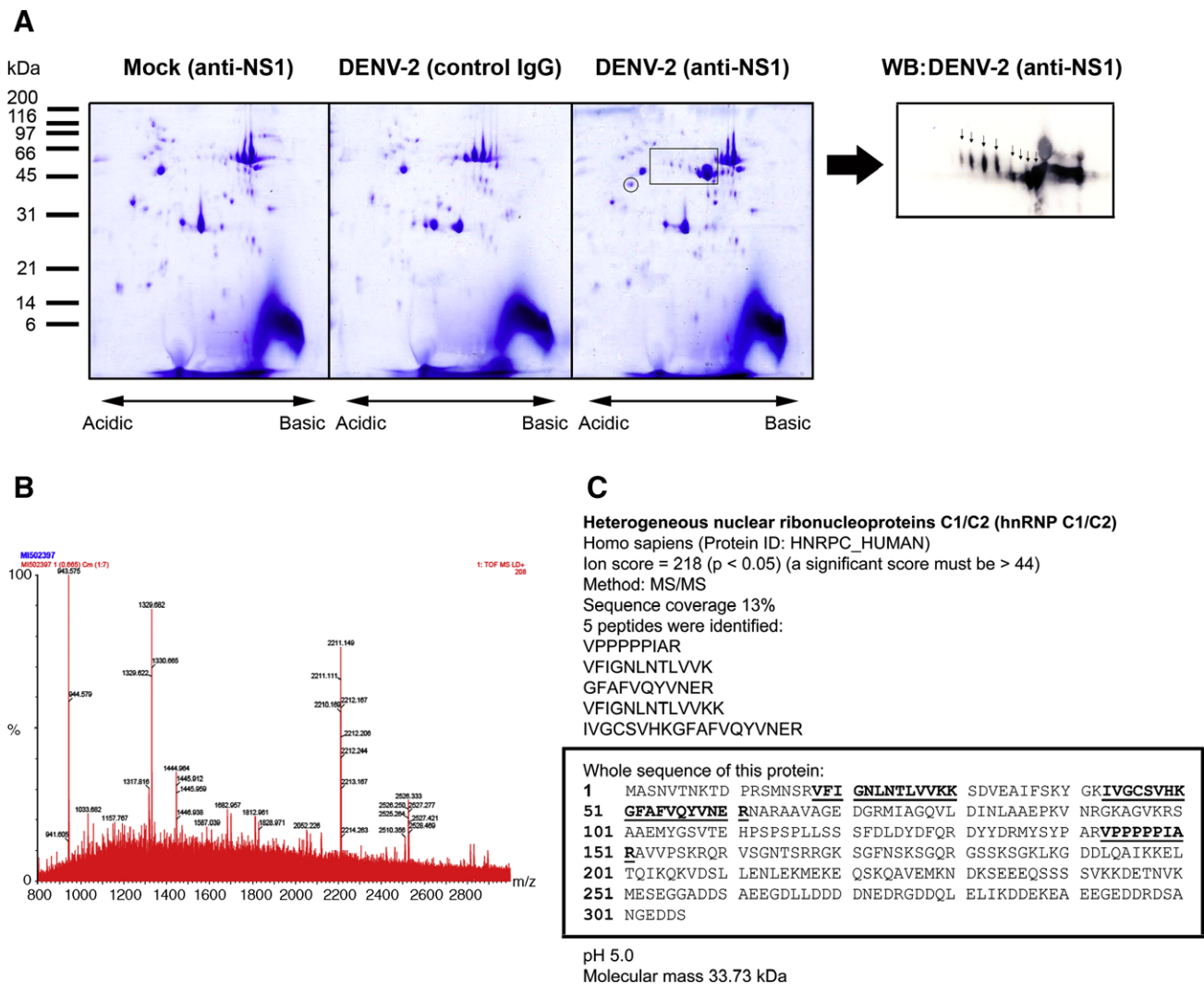


Fig. 2. Identification of hnRNP C1/C2 by 2-D gel electrophoresis and mass spectrometry. Mock and dengue virus-infected cell lysates were subjected to immunoprecipitation using either an anti-NS1 monoclonal antibody or an isotype-matched control antibody. (A) Precipitated proteins were separated by 2-D gel electrophoresis and visualized by Coomassie blue staining. A 40-kDa protein spot of interest (circled) and eight isoforms of dengue virus NS1 (indicated with a rectangle) are shown in DENV-2 (anti-NS1). The NS1 isoforms were verified by immunoblot analysis [WB: DENV-2 (anti-NS1)]. (B) The 40-kDa protein spot was processed for Q-TOF MS/MS analysis. The profile of mass/charge ratio of peptides is shown. (C) Five peptides resulting from the MS/MS analysis are shown to match the human heterogeneous nuclear ribonucleoproteins C1/C2.

of IgG light and heavy chains, respectively (data not shown). Apart from the detection of these protein spots, immunoprecipitation of dengue virus-infected cell lysate with the anti-NS1 specific antibody yielded a unique profile of eight protein spots of approximately 46 kDa as well as a single 40-kDa protein spot [Fig. 2, DENV-2 (anti-NS1)] which were not detectable in other precipitated control samples [Fig. 2, Mock (anti-NS1) and DENV-2 (control IgG)]. The former set of protein spots likely represented various isoforms of dengue virus NS1 owing to their reactivity with an anti-NS1 antibody in the immunoblot analysis [Fig. 2A—the right-most inset, WB: DENV-2 (anti-NS1)]. The unknown 40-kDa protein spot, which was co-precipitated with dengue virus NS1, was subsequently excised, in-gel digested with trypsin and processed for mass spectrometry. Spectra of the eluting peptides obtained from mass spectrometric analysis were shown in Fig. 2B. Five peptides were identified and, based on the SwissProt database, were deduced to be from human heterogeneous nuclear ribonucleoproteins (hnRNP) C1/C2 (SwissProt accession number **P07910**) (Fig. 2C). The identified peptides were common to both isoforms of hnRNP C and corresponded to sequence coverage of about 13% of the protein (Fig. 2C).

Co-immunoprecipitation of hnRNP C1/C2 and dengue virus NS1

Human hnRNP C1/C2 are members of the heterogeneous nucleolar ribonucleoprotein family which consists of 20 major hnRNP proteins (designated hnRNP A1 through U) with molecular sizes of approximately 36–120 kDa [17,18]. Alternative mRNA splicing of hnRNP C proteins with a 13-amino acid deletion occurring after glycine 106 or serine 107 generates two isoforms of proteins, hnRNP C1 (290 amino acids) and hnRNP C2 (303 amino acids) [19,20]. The hnRNP C1 and C2 proteins (41 and 43 kDa) are involved in mRNA biogenesis and contain important conserved motifs essential for RNA binding, protein–protein interaction and nuclear localization [17,18,21–23]. In order to determine whether the hnRNP C1/C2 interacts with dengue virus NS1 in virus-infected cells, immunoprecipitation of mock and dengue virus-infected cell lysates was performed using isotype-matched control antibodies or monoclonal antibodies specific against human hnRNP C1/C2 or dengue virus NS1. The presence of the two proteins in each precipitated sample was then analyzed by the immunoblot assay under nonreducing and unheated conditions. The lysates prior to immunoprecipitation were also included in this assay as controls for detection of hnRNP C1/C2 and dengue virus NS1.

As expected, hnRNP C1/C2 was observed in mock and dengue virus-infected cell lysate while dengue virus NS1 was detected only in the latter sample (Fig. 3, input). Using the anti-NS1 monoclonal antibody for immunoprecipitation, dengue virus NS1 and hnRNP C1/C2 were pulled down from virus-infected lysate, but not from mock-infected control (Fig. 3, IP: α NS1). On the contrary, immunoprecipitation with the anti-hnRNP C1/C2 monoclonal antibody yielded only the hnRNP C1/C2 in mock and dengue virus-infected samples, but no NS1 dimer was observed (Fig. 3, IP: α hnRNP). As negative controls, immunoprecipitation with either isotype-matched control antibody did not give any specific band of NS1 dimer; however, relatively low levels of hnRNP C1/C2 could be detected probably as a result of certain extents of nonspecific binding to protein G–Sepharose beads (Fig. 3, IP: IgG1 and IgG2a). Taken together, these findings suggested that the interaction between the dengue NS1 and hnRNP C1/C2 occurred in virus-infected cells. Inability to co-immunoprecipitate dengue virus NS1 using the anti-hnRNP C1/C2 antibody could be due to either masking of the hnRNP C1/C2 epitope by the interacting protein as suggested by a previous work on the Epstein-Barr virus SM protein [24], or only a small portion of the total hnRNP C1/C2 population involved in the dengue virus NS1 interaction.

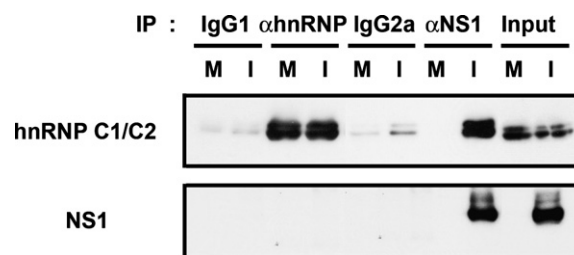


Fig. 3. Co-immunoprecipitation of hnRNP C1/C2 and dengue virus NS1. Mock (M) and dengue virus-infected (I) cell lysates were processed for immunoprecipitation (IP) using isotype-matched control antibodies (IgG1 or IgG2a), an anti-hnRNP C1/C2 monoclonal antibody (clone 4F4, IgG1), or an anti-NS1 monoclonal antibody (clone 1A4, IgG2a). Each precipitated sample was detected for the presence of hnRNP C1/C2 and dengue virus NS1 by immunoblot analysis using specific antibodies. Mock and virus-infected cell lysates prior to the immunoprecipitation (input) served as controls for the protein detection.

Co-localization of hnRNP C1/C2 and dengue virus NS1

Further investigation was carried out to determine sub-cellular distribution of hnRNP C1/C2 and dengue virus NS1 by double immunofluorescence staining and confocal microscopy. hnRNP C1/C2 was found predominantly in the nucleus and, to a lesser extent, in the cytoplasm of both mock (data not shown) and dengue virus-infected cells with an evenly distributed pattern (Fig. 4A and B, hnRNP C1/C2). Although hnRNP C1/C2 is primarily a nuclear protein in the cells, it can be translocated to the cytoplasm following the induction of apoptosis via Rho-activated kinase activation [25] or during poliovirus and human rhinovirus infection through an apoptosis-independent mechanism [26,27]. Nevertheless, an increase in the nuclear efflux of the hnRNP C1/C2 was not observed upon dengue virus infection in our study (data not shown).

Unlike the hnRNP C1/C2, dengue virus NS1 was detected as large foci mainly in the cytoplasm of dengue virus-infected cells (Fig. 4A and B, NS1), consistent with the staining pattern observed in previous studies [7,10,15]. A partial co-localization of the hnRNP C1/C2 and NS1 was detectable in some dengue virus-infected cells particularly at the perinuclear regions (Fig. 4A and B, Merge). This observation may help to explain the incapability to detect reciprocal co-immunoprecipitation with the anti-hnRNP C1/C2 antibody (Fig. 3, IP: α hnRNP). Consistent findings on the co-localization and the co-immunoprecipitation of hnRNP C1/C2 with dengue virus NS1 was observed not only in HEK 293T cell line (Fig. 3 and 4) but also in other three human cell lines (including liver HepG2, fibroblast HF and endothelial EAhy926) and primary human umbilical vein endothelial cells, HUVEC, (data not shown), confirming that this interaction is genuine, but probably transient and involved in only a part of total host cellular hnRNP C1/C2.

A previous study using a yeast two hybrid system identified human signal transducer and activator of transcription 3 β (STAT3 β) as a dengue virus NS1-interacting protein which may be implicated in the induction of IL-6 levels and subsequent plasma leakage [28]. By utilizing different approach, we demonstrated herein that hnRNP C1/C2 is additional host cellular protein that associates with dengue virus NS1 in virus-infected cells, but mechanisms of this protein–protein interaction are not known. The common structural motifs on hnRNP C1/C2 [17,22,23] may interact with dengue virus NS1 directly or indirectly through a link of other host proteins, viral proteins, and/or viral RNA resulting in the formation of viral replication complexes. This notion was supported by a previous study on the role of hnRNP C1 in the initiation of positive-strand viral RNA synthesis in poliovirus-infected cells through its interaction with the 3'-end of negative strand RNA of poliovirus and other three viral protein precursors, polypeptide 3CD, P2, and P3, crucial for poliovirus replication [29]. In addition, detection of various

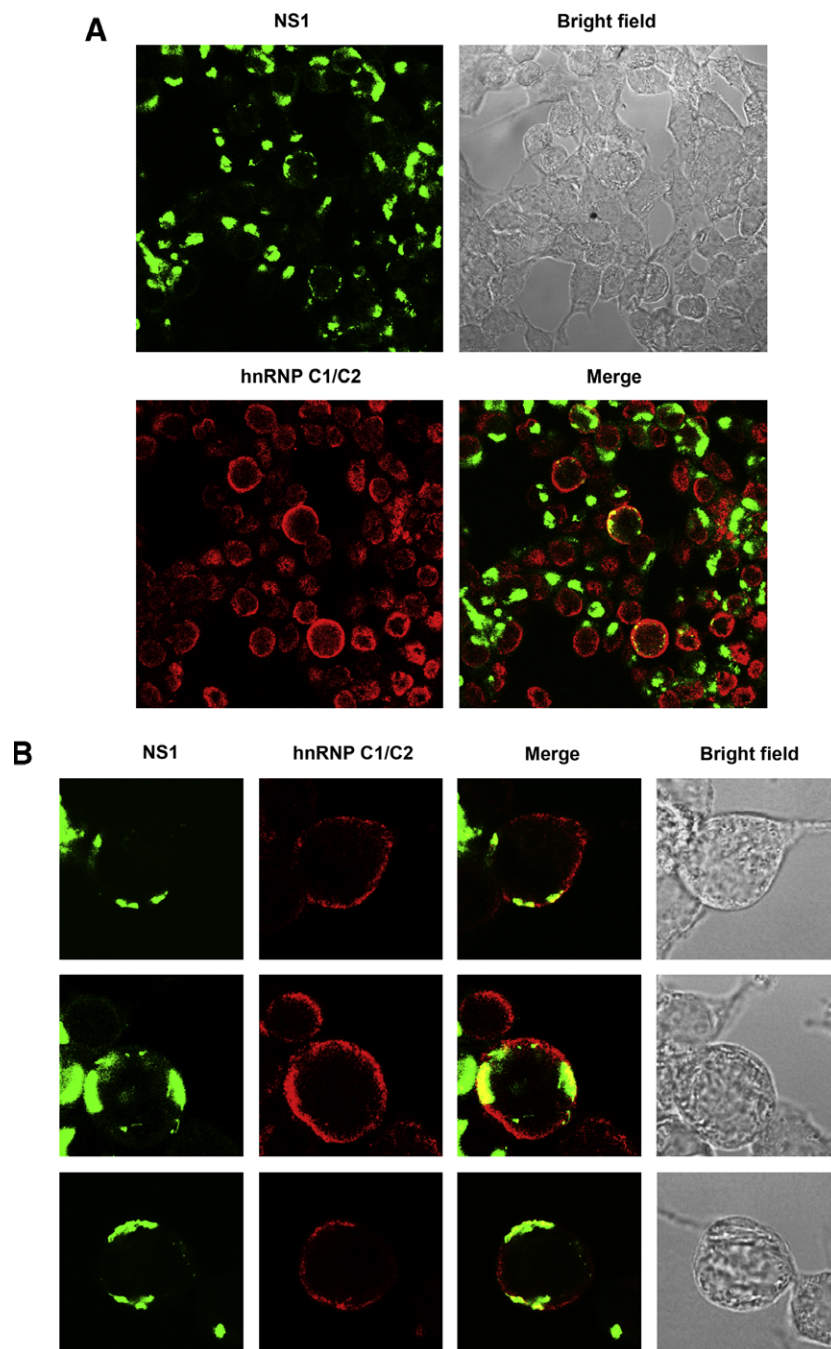


Fig. 4. Co-localization of hnRNP C1/C2 and dengue virus NS1. Dengue virus-infected cells were subjected to double immunofluorescence staining and observed for localization of hnRNP C1/C2 (red) and dengue virus NS1 (green) by using a laser-scanning confocal microscope with a 63 $\times$ objective lens (A,B). Individual images that had been captured in the same fields are merged and areas where co-localization of both the proteins occurs are shown in yellow. Three representative fields of the stained cells with a 5 $\times$ magnification are shown in (B). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

isoforms of dengue virus NS1 with distinct isoelectric points (*pI*) by 2-D gel electrophoresis in our study implied that differential post-translational modifications of NS1 might occur in virus-infected cells and give rise to generation of different isoforms of the protein, each of which potentially exerts different functions. Which particular isoform of dengue virus NS1 interacts with hnRNP C1/C2 and whether this association is involved in virus replication and host cellular responses during dengue virus infection require further investigations.

Acknowledgments

We thank Dr. Panisadee Avirutnan for technical assistance in HUVEC cultures. This work was supported by a research grant (BT-B-02-MG-B4-4801) from the National Center for Genetic Engineering and Biotechnology (to S.N.), Senior Research Scholar Grants from Thailand Research Fund (to P.M. and P.Y.), and Siriraj Graduate Thesis Scholarship from the Faculty of Medicine Siriraj Hospital, Mahidol University (to S.Se.).

References

- [1] World Health Organization, Dengue and dengue haemorrhagic fever. <<http://www.who.int/mediacentre/factsheets/fs117/en/>>, 2008 (accessed April 7).
- [2] B.D. Lindenbach, C.M. Rice, Flaviviridae: the viruses and their replication, in: D.M. Knipe, P.M. Howley, D.E. Griffin, R.A. Lamb, M.A. Martin, B. Roizman, S.E. Strauss (Eds.), *Fields Virology*, fourth ed., Lippincott Williams & Wilkins, Philadelphia, 2001, pp. 991–1041.
- [3] M.G. Jacobs, P.J. Robinson, C. Bletchly, J.M. Mackenzie, P.R. Young, Dengue virus nonstructural protein 1 is expressed in a glycosylphosphatidylinositol-linked form that is capable of signal transduction, *FASEB J.* 14 (2000) 1603–1610.
- [4] G. Winkler, V.B. Randolph, G.R. Cleaves, T.E. Ryan, V. Stollar, Evidence that the mature form of the flavivirus nonstructural protein NS1 is a dimer, *Virology* 162 (1988) 187–196.
- [5] M. Flamand, F. Megret, M. Mathieu, J. Lepault, F.A. Rey, V. Deubel, Dengue virus type 1 nonstructural glycoprotein NS1 is secreted from mammalian cells as a soluble hexamer in a glycosylation-dependent fashion, *J. Virol.* 73 (1999) 6104–6110.
- [6] T. Kurosu, P. Chaichana, M. Yamate, S. Anantapreecha, K. Ikuta, Secreted complement regulatory protein clusterin interacts with dengue virus nonstructural protein 1, *Biochem. Biophys. Res. Commun.* 362 (2007) 1051–1056.
- [7] P. Avirutnan, N. Punyadee, S. Noisakran, C. Komoltri, S. Thiemmea, K. Auethavornanan, A. Jairungsri, R. Kanlaya, N. Tangthawornchaikul, C. Puttikhunt, S. Pattanakitsakul, P. Yenchitsomanus, J. Mongkolsapaya, W. Kasinrer, N. Sittisombut, M. Husmann, M. Blettner, S. Vasanawathana, S. Bhakdi, P. Malasit, Vascular leakage in severe dengue virus infections: a potential role for the non-structural viral protein NS1 and complement, *J. Infect. Dis.* 193 (2006) 1078–1088.
- [8] D.H. Libraty, P.R. Young, D. Pickering, T.P. Endy, S. Kalayanaroop, S. Green, D.W. Vaughn, A. Nisalak, F.A. Ennis, A.L. Rothman, High circulating levels of the dengue virus nonstructural protein NS1 early in dengue illness correlate with the development of dengue hemorrhagic fever, *J. Infect. Dis.* 186 (2002) 1165–1168.
- [9] B.D. Lindenbach, C.M. Rice, Genetic interaction of flavivirus nonstructural proteins NS1 and NS4A as a determinant of replicase function, *J. Virol.* 73 (1999) 4611–4621.
- [10] J.M. Mackenzie, M.K. Jones, P.R. Young, Immunolocalization of the dengue virus nonstructural glycoprotein NS1 suggests a role in viral RNA replication, *Virology* 220 (1996) 232–240.
- [11] A.A. Khromykh, P.L. Sedlak, E.G. Westaway, cis- and trans-acting elements in flavivirus RNA replication, *J. Virol.* 74 (2000) 3253–3263.
- [12] E.G. Westaway, J.M. Mackenzie, M.T. Kenney, M.K. Jones, A.A. Khromykh, Ultrastructure of Kunjin virus-infected cells: colocalization of NS1 and NS3 with double-stranded RNA, and of NS2B with NS3, in virus-induced membrane structures, *J. Virol.* 71 (1997) 6650–6661.
- [13] C. Puttikhunt, W. Kasinrer, S. Srisa-ad, T. Duangchinda, W. Silakate, S. Moonsom, N. Sittisombut, P. Malasit, Production of anti-dengue NS1 monoclonal antibodies by DNA immunization, *J. Virol. Methods* 109 (2003) 55–61.
- [14] K. Sittisiri, Production and characterization of monoclonal antibodies to dengue 2 virus, M.Sc. Thesis, Mahidol University, 1994.
- [15] S. Noisakran, T. Dechtawewat, P. Rinkawekkan, C. Puttikhunt, A. Kanjanahaluethai, W. Kasinrer, N. Sittisombut, P. Malasit, Characterization of dengue virus NS1 stably expressed in 293T cell lines, *J. Virol. Methods* 142 (2007) 67–80.
- [16] S.N. Pattanakitsakul, K. Rungrojcharoenkit, R. Kanlaya, S. Sinchaikul, S. Noisakran, S.T. Chen, P. Malasit, V. Thongboonkerd, Proteomic analysis of host responses in HepG2 cells during dengue virus infection, *J. Proteome Res.* 6 (2007) 4592–4600.
- [17] G. Dreyfuss, M.J. Matunis, S. Pinol-Roma, C.G. Burd, hnRNP proteins and the biogenesis of mRNA, *Annu. Rev. Biochem.* 62 (1993) 289–321.
- [18] S. Pinol-Roma, hnRNP proteins and the nuclear export of mRNA, *Semin. Cell Dev. Biol.* 8 (1997) 57–63.
- [19] C.G. Burd, M.S. Swanson, M. Gorlach, G. Dreyfuss, Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: a diversity of RNA binding proteins is generated by small peptide inserts, *Proc. Natl. Acad. Sci. USA* 86 (1989) 9788–9792.
- [20] M.S. Swanson, T.Y. Nakagawa, K. LeVan, G. Dreyfuss, Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins, *Mol. Cell. Biol.* 7 (1987) 1731–1739.
- [21] J.H. Kim, B. Hahm, Y.K. Kim, M. Choi, S.K. Jang, Protein–protein interaction among hnRNPs shuttling between nucleus and cytoplasm, *J. Mol. Biol.* 298 (2000) 395–405.
- [22] B. Carpenter, C. MacKay, A. Alnabulsi, M. MacKay, C. Telfer, W.T. Melvin, G.I. Murray, The roles of heterogeneous nuclear ribonucleoproteins in tumour development and progression, *Biochim. Biophys. Acta* 1765 (2006) 85–100.
- [23] M. Gorlach, C.G. Burd, G. Dreyfuss, The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins, *J. Biol. Chem.* 269 (1994) 23074–23078.
- [24] S.C. Key, T. Yoshizaki, J.S. Pagano, The Epstein-Barr virus (EBV) SM protein enhances pre-mRNA processing of the EBV DNA polymerase transcript, *J. Virol.* 72 (1998) 8485–8492.
- [25] H.H. Lee, C.L. Chien, H.K. Liao, Y.J. Chen, Z.F. Chang, Nuclear efflux of heterogeneous nuclear ribonucleoprotein C1/C2 in apoptotic cells: a novel nuclear export dependent on Rho-associated kinase activation, *J. Cell Sci.* 117 (2004) 5579–5589.
- [26] K.E. Gustin, P. Sarnow, Effects of poliovirus infection on nucleocytoplasmic trafficking and nuclear pore complex composition, *EMBO J.* 20 (2001) 240–249.
- [27] K.E. Gustin, P. Sarnow, Inhibition of nuclear import and alteration of nuclear pore complex composition by rhinovirus, *J. Virol.* 76 (2002) 8787–8796.
- [28] J.J. Chua, R. Bhuvanathan, V.T. Chow, M.L. Ng, Recombinant non-structural 1 (NS1) protein of dengue-2 virus interacts with human STAT3beta protein, *Virus Res.* 112 (2005) 85–94.
- [29] J.E. Brunner, J.H. Nguyen, H.H. Roehl, T.V. Ho, K.M. Swiderek, B.L. Semler, Functional interaction of heterogeneous nuclear ribonucleoprotein C with poliovirus RNA synthesis initiation complexes, *J. Virol.* 79 (2005) 3254–3266.

Sensitization to Fas-mediated apoptosis by dengue virus capsid protein

Thawornchai Limjindaporn^{a,b,c,*,1}, Janjuree Netsawang^{b,c,1}, Sansanee Noisakran^{c,e},
Somchai Thiemmecha^{b,c}, Wiyada Wongwiwat^{b,c}, Sangkab Sudsaward^{a,c},
Panisadee Avirutnan^{b,c,e}, Chunya Puttikhunt^{c,e}, Watchara Kasinrer^{d,e},
Rungtawan Sriburi^f, Nopporn Sittisombut^{e,f}, Pa-thai Yenchitsomanus^{b,c,e}, Prida Malasit^{b,c,e}

^a Department of Anatomy, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand

^b Department of Immunology, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand

^c Medical Molecular Biology Unit, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand

^d Department of Clinical Immunology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, Thailand

^e Medical Biotechnology Unit, National Center for Genetic Engineering and Biotechnology,
National Science and Technology Development Agency, Bangkok, Thailand

^f Department of Microbiology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

Received 26 July 2007

Available online 13 August 2007

Abstract

Dengue fever (DF) and dengue hemorrhagic fever (DHF) are important public health problems in tropical regions. Abnormal hemostasis and plasma leakage are the main patho-physiological changes in DHF. However, hepatomegaly, hepatocellular necrosis and fulminant hepatic failure are occasionally observed in patients with DHF. Dengue virus-infected liver cells undergo apoptosis but the underlying molecular mechanism remains unclear. Using a yeast two-hybrid screen, we found that dengue virus capsid protein (DENV C) physically interacts with the human death domain-associated protein Daxx, a Fas-associated protein. The interaction between DENV C and Daxx in dengue virus-infected liver cells was also demonstrated by co-immunoprecipitation and double immunofluorescence staining. The two proteins were predominantly co-localized in the cellular nuclei. Fas-mediated apoptotic activity in liver cells constitutively expressing DENV C was induced by anti-Fas antibody, indicating that the interaction of DENV C and Daxx involves in apoptosis of dengue virus-infected liver cells.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Dengue virus capsid protein; Human Daxx; Protein interaction; Apoptosis

Dengue virus is a mosquito-borne member of the family *Flaviviridae*. Infection with dengue virus produces variable clinical illness ranging from non-specific viral infection to dengue hemorrhagic fever (DHF) and dengue shock syndrome (DHF/DSS) [1]. DHF patients generally present with continuous fever lasting 2–7 days, with hemorrhagic tendencies, plasma leakage, thrombocytopenia and hemo-

concentration. Liver injury with an elevation of amino-transferases and reactive hepatitis are common in adult patients with dengue virus infection [2,3]. Councilman bodies in liver biopsies of DHF/DSS cases correspond to hepatocytes in apoptosis and this may be a key element in the pathophysiology of hepatic failure associated with DHF/DSS [4].

Apoptosis of liver HepG2 cells during dengue virus infection has been observed [4–8]. Because the capsid protein of other members of the family *Flaviviridae* have multiple functions including either induction or inhibition of apoptosis [9–12], we suspected that dengue virus capsid protein (DENV C) might be involved in apoptosis. DENV C is a

* Corresponding author. Address: Department of Anatomy, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand. Fax: +66 2 418 4793.

E-mail address: limjindaporn@yahoo.com (T. Limjindaporn).

¹ The first two authors contributed equally to this work.

small, highly positively charged, 12-kDa protein that is required for the maturation of viral particles and assembly of the nucleocapsid core [13]. A proportion of DENV C localizes to the nucleus [14–16] and interacts with heterogeneous nuclear ribonucleoprotein K (hnRNP K) [17]. By using a yeast two-hybrid screen, we showed in this study that DENV C interacts with human death domain-associated protein Daxx, a modulator of Fas-mediated apoptosis [18]. The interaction between DENV C and Daxx was examined by immunoprecipitation, and co-localization studies in dengue virus-infected HepG2 cells. Given the pro-apoptotic role of Daxx, the effect of DENV C expression on viability of HepG2 cells was examined. HepG2 cells expressing DENV C became sensitive to Fas-mediated apoptosis, suggesting the interaction with Daxx may be physiologically significant and relevant to pathogenesis of dengue virus infection.

Materials and methods

Yeast two-hybrid screening. Two-hybrid screening was performed as described by Finley and Brent [19]. DENV C was amplified from pBluescript II KS (S1SP6-4497), a plasmid which contains the 5' portion of dengue virus serotype 2 strain 16681 [20], with a pair of primers, 5'-GCG CTG AAT TC ATG AAT GAC CAA CGG AAA-3' and 5'-GAT ACG GGA TCC CTA TCT GCG TCT CCT AT-3'. The PCR was carried out in a GeneAmp PCR System 9700 (Applied Biosystems), starting with an initial denaturation step of 94 °C for 5 min and followed by 30 cycles of denaturation at 94 °C for 45 s, annealing at 60 °C for 45 s, extension at 72 °C for 2 min, and one cycle of final extension at 72 °C for 10 min. Subsequently, the amplified DNA was cloned as an in-frame fusion with the LexA DNA binding domain in the yeast expression vector pEG202, which contains a *HIS3* selectable marker. The resulting bait plasmid, pEG-C, was transformed into *S. cerevisiae* strain RFY 206 (MATa *his3Δ200 leu2-3 lys2Δ201 ura3-52 trp1Δ::hisG*) containing a Lexop-*lacZ* reporter plasmid, pSH18-34, under *URA3* selection. A galactose inducible HeLa cell cDNA prey library in plasmid pJG4-5 containing a *TRP1* selectable marker was transformed into strain RFY 231 (MATa *his3 leu2::3Lexop-LEU2 ura3 trp1 LYS2*) [21]. The bait strain was mated with the library strains and plated on galactose drop-out medium lacking histidine, tryptophan, uracil, and leucine (gal/raf -u, -h, -w, -l). The production of a DENV C binding protein by a prey plasmid was expected to activate the 3Lexop-*LEU2* reporter. Putative positive clones were patched to four indicator plates: (glu/-u, -h, -w, -l), (gal/raf -u, -h, -w, -l), (glu/X-Gal -u, -h, -w), and (gal/raf/X-Gal -u, -h, -w). Prey plasmids were rescued from clones exhibiting a galactose-inducible *Leu*⁺ *lacZ*⁺ phenotype by transforming into a *Trp*⁻ *Escherichia coli* strain KC8 [19]. To verify the interaction, recovered prey plasmids were introduced into yeast strain RFY 231 along with the *lacZ* plasmid and bait plasmid and again tested on the indicator plates.

Co-immunoprecipitation. HepG2 cells were grown in DMEM (Gibco-BRL) supplemented with 10% heat inactivated fetal bovine serum (Hyclone), 1 mM sodium pyruvate (Sigma), 1 mM non-essential amino acids (Gibco-BRL), and 1.2% penicillin G-streptomycin at 37 °C in 5% CO₂. Up to 3 × 10⁶ HepG2 cells were infected with dengue virus strain 16681 at a MOI of 1 for 32 h. The cell pellets were lysed with RIPA buffer. Five microgram of purified mouse anti-DENV C monoclonal antibody produced in the laboratory or 5 μg of rabbit anti-Daxx polyclonal antibody (Santa Cruz Biotechnology) or 5 μg of isotype-matched control antibodies were added to lysates. The mixture was incubated at 4 °C overnight and then protein G Sepharose beads (Amersham-Pharmacia Biosciences) were added. The bound proteins were eluted, subjected to SDS-PAGE and performed immunoblot analysis. The membranes were incubated with mouse anti-DENV C antibody or rabbit anti-Daxx antibody and followed by probing either with horseradish peroxidase (HRP)-conjugated rabbit anti-mouse Ig antibody or HRP-conjugated

swine anti-rabbit Ig antibody, respectively. The protein bands were detected by using ECL reagents (Amersham-Pharmacia Biosciences).

Co-localization. HepG2 cells were grown on cover slips as described above. Subconfluent monolayer of these cells were infected with dengue virus strain 16681 at a MOI 1 for 32 h, washed, fixed with 4% formaldehyde in PBS for 20 min, and permeabilized with 0.2% Triton X-100 in PBS for 10 min at room temperature. After washing three times with 0.1% Triton X-100 in PBS, cover slips were incubated both with mouse anti-DENV C antibody and rabbit anti-Daxx antibody for an hour. After washing, the cover slips were incubated both with Alexa 488-conjugated rabbit anti-mouse Ig antibody (Molecular Probes) and Cy3-conjugated donkey anti-rabbit Ig antibody (Jackson ImmunoResearch laboratories) as secondary antibodies at room temperature for an hour. Fluorescent images were captured with a confocal microscope (model LSM 510, Carl Zeiss).

Generation of HepG2 cells constitutively expressing DENV C. The DENV C gene was amplified from pBluescript II KS (S1SP6-4497) [20] using primers CapBamHI: 5'-GTAGGATCCATGAATGACCAACGG AAAAAG-3' and CapXhoI: 5'-GCACTCGAG CTATCTGCGTCTCTCT AT-3'. The PCR was carried out as described above for construction of the two-hybrid bait plasmid. The PCR product was cloned into the plasmid pcDNA3.1/His C (Invitrogen). The integrity of positive clones was verified by DNA sequencing and an evaluation of DENV C expression. HepG2 cells were transfected with either plasmid pcDNA3.1 his c/ DENV C or empty vector pcDNA3.1/His C using Lipofectamine (Invitrogen). Twenty-four hours post transfection, stable cell colonies were selected in DMEM containing 1 mg/ml G418 (Calbiochem) for 1 month. The isolated G418-resistant colonies were maintained in DMEM containing 0.5 mg/ml G418 and examined for the expression of DENV C. To compare the level of DENV C expression in dengue virus-infected HepG2 cells and G418-resistant clones, Western blot analysis was performed using totally 75 μg of total protein and probed with β-actin (Santa Cruz Biotechnology) and DENV C antibodies.

Apoptosis assays. Up to 1 × 10⁶ of stable HepG2 transformants were treated with 0.5 μg/ml anti-Fas mAb (Sigma) and 1 μg/ml cycloheximide for 24 h in culture medium containing 2% FBS. Two clones of HepG2 cells constitutively expressing DENV C; HepG2C1 and HepG2C6, and a clone of HepG2 cells containing the empty vector; HepG2His, were examined for the presence of apoptotic cells by harvesting both detached and adherent cells, staining for cell surface phosphatidyl serine, assaying for caspase-3 activation, and performing the DNA fragmentation assay.

For annexin V/propidium iodide staining, samples were suspended in annexin V buffer and incubated on ice with FITC-conjugated annexin V (BD Biosciences) for 15 min. Immediately prior to analysis, propidium iodide was added and samples were analyzed by flow cytometry (Becton–Dickinson).

In the assessment of caspase-3 activation, the cells were lysed with RIPA buffer and subjected to SDS-PAGE analysis as previously described. The blots were incubated with goat anti-caspase-3 antibody (Santa Cruz Biotechnology) followed by a HRP-conjugated rabbit anti-goat Ig antibody. Protein bands were detected with ECL reagents.

The DNA fragmentation assay was performed using the SDS-high salt extraction method [22]. Cell pellets (~10⁶ cells) were suspended in 80 μl of PBS and gently mixed with 300 μl of buffer containing 10 mM Tris-HCl (pH 7.6), 10 mM EDTA, and 0.6% SDS. The lysates were mixed with 100 μl of 5 M NaCl and incubated overnight at 4 °C. Supernatants were treated sequentially with RNase A (1 mg/ml) and proteinase K (0.2 mg/ml) for 30 min at 37 °C. Precipitated DNA was resuspended in TE buffer, electrophoresed in 1.5% agarose prepared in Tris-borate-EDTA buffer, and stained with ethidium bromide.

Results and discussion

DENV C interacts with Daxx in a yeast two hybrid system

To identify human proteins that interact with DENV C, 10⁷ clones from a HeLa cDNA library were screened using DENV C as bait. Sixty putative positive clones were

obtained. Restriction endonuclease analysis indicated that 23 unique genes were represented (data not shown). Sequence analysis of the inserts showed that one of the cDNA inserts encoded an amino acid stretch (529–740) of Daxx. We demonstrated the specificity of the two-hybrid interaction between DENV C and Daxx by showing galactose-dependent activation of the reporters in cells containing the Daxx prey along with the DENV C bait but not a domain III of dengue virus envelope protein DENV E bait (Fig. 1A). After sequencing the prey plasmid containing Daxx obtained from our yeast two-hybrid screening, the region of Daxx that interacts with DENV C was the

211 carboxyl-terminal residues, a region which binds to multiple cellular proteins including Fas and Pml [18,23–25] (Fig. 1B). Interestingly, the nucleocapsid protein of hantavirus, which, like dengue virus, causes a hemorrhagic fever, has also shown to interact with the 240 carboxyl-terminal residues of Daxx [26].

Daxx is a 740 amino acid protein that contains two amino-terminal amphipathic helices (PAH1, PAH2), a coiled-coiled domain (CC), an acidic domain (D/E), and a carboxyl-terminal serine/proline/threonine rich domain (S/P/T). It is localized both in the cytoplasm and in the nucleus, although the majority is present in the nucleus.

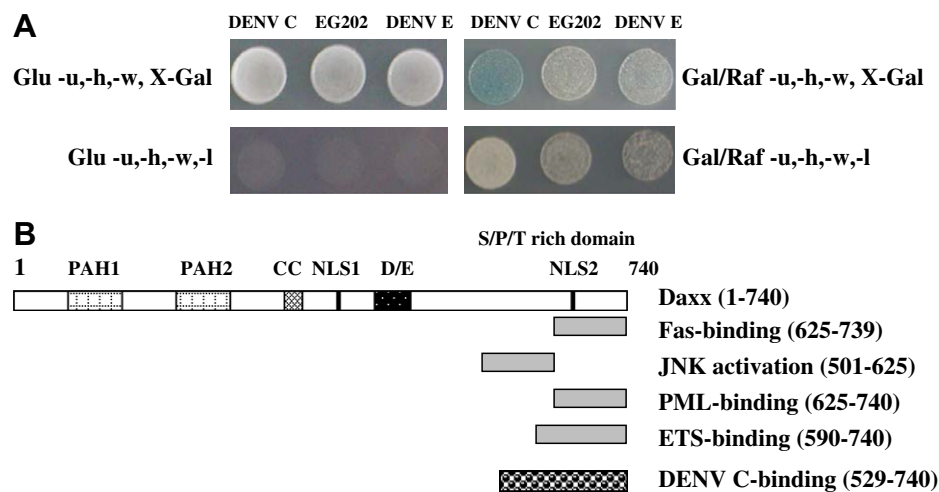


Fig. 1. DENV C interacts with Daxx in a yeast two-hybrid system. (A) Yeast strain RFY231 was co-transformed with a bait plasmid, a Daxx prey plasmid and *lacZ* reporter plasmid. The bait plasmids were pEG-C, expressing the LexA-DENV C fusion protein, the empty bait plasmid pEG202, or an unrelated bait plasmid pEG-E. A specific interaction was indicated by galactose-dependent β -galactosidase expression, as evidenced by blue colonies on the galactose containing X-Gal plate, and by galactose-dependent growth on the leucine deficient plate. (B) The region of Daxx that interacts with DENV C was the 211 carboxyl-terminal residues, a region which binds to Fas and Pml. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

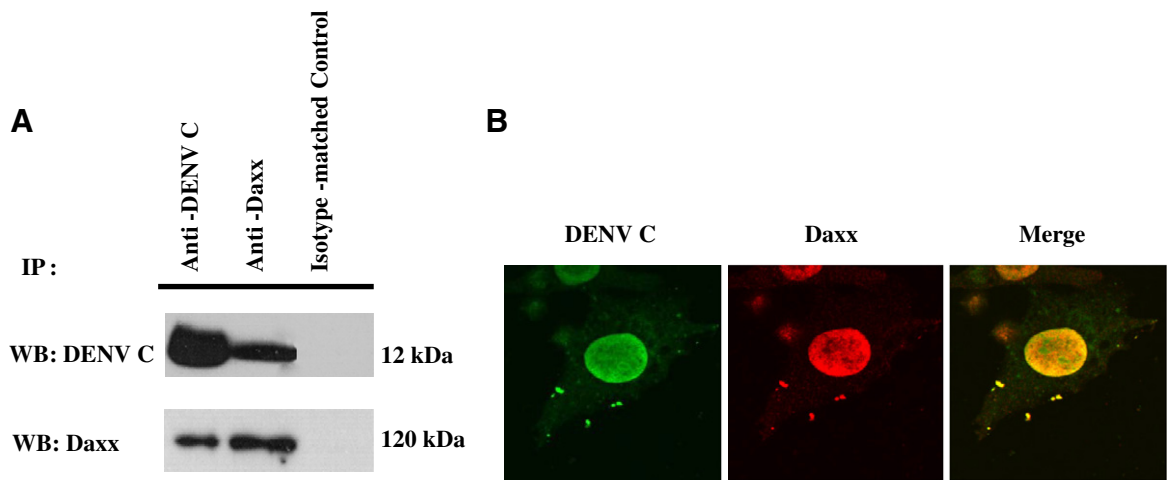


Fig. 2. Interaction between DENV C and Daxx in dengue virus-infected HepG2 cells. (A) The cell lysates from dengue virus-infected HepG2 cells were immunoprecipitated either with purified mouse anti-DENV C antibody or rabbit anti-Daxx antibody or isotype-matched control antibodies. The complexes were detected either with mouse anti-DENV C antibody or rabbit anti-Daxx antibody. (B) Dengue virus-infected HepG2 cells were fixed and immunostained with mouse anti-DENV C antibody (green in A) and rabbit anti-Daxx antibody (red in B). The merged image (yellow in C) demonstrated co-localization between DENV C and Daxx. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In the cytoplasm, Daxx interacts with Fas at the plasma membrane [18]. Stimulation of Fas causes translocation of Daxx from the nucleus to the cytoplasm, where it interacts with ASK1 and promotes JNK activation [27]. In the

nucleus, Daxx shuttles between two different sub-nuclear structures, nucleoplasm and the promyelocytic leukemia nuclear bodies (PML-NB) [23,24]. In the nucleoplasm, Daxx acts as a transcriptional repressor [28,29]. Daxx asso-

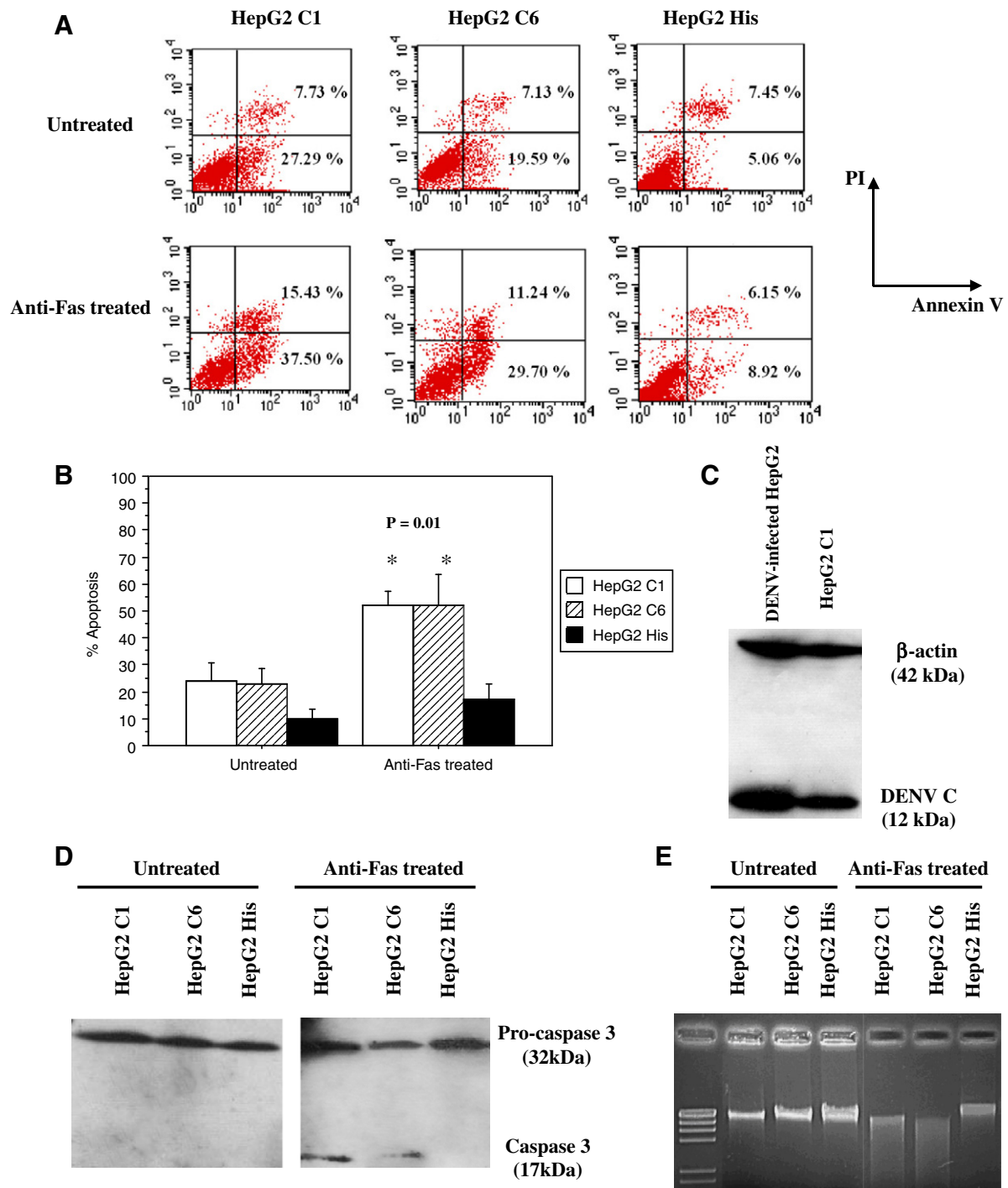


Fig. 3. Sensitization of Fas-mediated apoptosis by DENV C. (A) The untreated and anti-Fas antibody treated of HepG2C1, HepG2C6, and HepG2His were incubated with FITC-conjugated annexin V. Immediately prior to analysis, propidium iodide was added and samples were analyzed by flow cytometry. (B) The percentage of cells undergoing apoptosis as measured by the percent of annexin V positive from three independent experiments was shown. (C) The expression levels of DENV C in dengue virus-infected HepG2 and HepG2C1 were compared. (D) In the assessment of caspase-3 activation, untreated and anti-Fas antibody treated HepG2C1, HepG2C6, and HepG2His were incubated with goat anti-caspase-3 antibody. Three independent experiments were conducted and gave the similar results. (E) The DNA fragmentation assay using the pellets from untreated and anti-Fas antibody treated of HepG2C1, HepG2C6, and HepG2His. Similar results were obtained from three independent experiments.

ciates with proteins that are critical for transcriptional repression, such as histone deacetylase II [30]. The ability of Daxx to repress transcription is inhibited by its localization to the PML-NB [28].

DENV C associates with Daxx in dengue virus-infected HepG2 cells

Clinical studies have implicated liver in the pathogenesis of dengue virus infection [2,3]. In addition, a transformed liver cell line, HepG2, is permissive to dengue virus and infection of HepG2 cells with dengue virus results in the induction of apoptosis [4–8]. In order to confirm DENV C-Daxx interaction in liver cells, the lysates from dengue virus-infected HepG2 cells were tested for coimmunoprecipitation of the two proteins. As shown in Fig. 2A, anti-DENV C antibody pulled down Daxx protein and, conversely, anti-Daxx antibody co-precipitated DENV C in immunoprecipitation. Immunoprecipitation with isotype-matched control antibodies did not bring down Daxx or DENV C. These results suggest that DENV C and Daxx interact during dengue virus infection of HepG2 cells.

As further evidence of interaction, dengue virus-infected cells were examined for co-localization of the proteins by an immunofluorescence confocal microscopy. Daxx was primarily localized in the nucleus of dengue virus-infected HepG2 cells (Fig. 2B). A similar nuclear distribution of DENV C was detected in these cells. Co-localization of DENV C and Daxx was clearly evident when the DENV C and Daxx images were superimposed (Fig. 2B). These data support the physical interaction of DENV C and Daxx and further suggest that association occurs in the nucleus of dengue virus-infected HepG2 cells.

DENV C expression sensitizes HepG2 cells to Fas-mediated apoptosis

Dengue virus infection induces apoptosis in HepG2 cells [4–8] and Daxx has been implicated in a myriad of interactions controlling apoptosis and cell death [31,32]. It is possible that DENV C, through an interaction with Daxx, may influence apoptosis. To test this possibility, HepG2 cells constitutively expressing DENV C were generated and examined for apoptosis. Two DENV C expressing cell lines were constructed; HepG2C1 and HepG2C6. The DENV C expression in HepG2C1 is comparable to dengue virus-infected HepG2 cells (Fig. 3C). While both expressed DENV C, expression was about twofold higher in HepG2C1 than HepG2C6 and no DENV C expression was detected in the negative control strain HepG2His.

As an indicator of apoptosis, cells were stained with FITC-conjugated annexin V and propidium iodide and analyzed by flow cytometry [33]. As seen in Fig. 3A and B, cell lines HepG2C1 and HepG2C6 had a 2- to 3-fold increase in the intensity of annexin V staining relative to the control cells, HepG2His, but comparable numbers of

propidium iodide-stained cells. This result suggests that DENV C-expressing cells are in an early stage apoptosis before any evidence of caspase 3 activation or DNA laddering is apparent (Fig. 3D and E).

Since Daxx participates in Fas-mediated apoptosis, the effect of Fas activation in DENV C-expressing cells was examined. Treatment with anti-Fas antibody had no effect on HepG2His control cells as measured by annexin V staining, caspase activation or DNA laddering (Fig. 3). This is in agreement with prior work demonstrating that HepG2 cells do not respond to Fas activation [7]. However, both HepG2C1 and HepG2C6 showed a significant increase in staining with annexin V and propidium iodine upon treatment with anti-Fas antibody (Fig. 3A and B). Apoptosis in response to this treatment was also evident in the development of activation of caspase 3 and DNA ladder in these cells (Fig. 3D and E).

The foregoing indicates that DENV C affects the HepG2 cells and sensitizes them to Fas-mediated apoptosis. Presumably this is due to the interaction of DENV C with Daxx, which may disrupt one or more of the many interactions of Daxx with other proteins controlling cell death [31,32]. DENV C may not act alone in inducing apoptosis. The apoptotic process in HepG2 cells can be induced by the dengue virus ectodomain of the small membrane protein ApoptoM [5]. Apoptosis pathways other than the Fas-mediated pathway are also involved in these cells, including activation of NF- κ B and TNF-related apoptosis-inducing ligand [4,7]. However, the apoptotic process appears to be independent of p53 [8].

In summary, this work is the first to demonstrate an interaction between DENV C and human Daxx and the pro-apoptotic function of DENV C in liver cells. Future studies will be directed toward elucidating the detailed molecular mechanisms by which DENV C contributes to Fas-mediated apoptosis.

Acknowledgments

We appreciate Dr. Russell L Finley Jr., Wayne State University, and Dr. William A Fonzi, Georgetown University, for reagents, reading the manuscript and discussion. This work is financially supported by Thailand Research Fund, Siriraj Grant for Research and Medical Education, and Siriraj Chalearmprakit fund to T.L., by Royal Golden Jubilee Ph.D. Program to J.N. and W.W. and by the Senior Research Scholar Program of Thailand Research Fund to P.Y. and P.M.

References

- [1] D.J. Gubler, Dengue and dengue hemorrhagic fever, Clin. Microbiol. Rev. 11 (1998) 480–496.
- [2] L.J. Souza, J.G. Alves, R.M. Nogueira, C. Gicovate Neto, D.A. Bastos, E.W. Siqueira, J.T. Souto Filho, A. Cezario Tde, C.E. Soares, C. Carneiro Rda, Aminotransferase changes and acute hepatitis in patients with dengue fever: analysis of 1,585 cases, Braz. J. Infect. Dis. 8 (2004) 156–163.

- [3] A. Khongphatthanayothin, P. Lertsapcharoen, P. Supachokchaiwatana, P. Satupan, K. Thongchaiprasit, Y. Poovorawan, C. Thisyakorn, Hepatosplanchnic circulatory dysfunction in acute hepatic infection: the case of dengue hemorrhagic fever, *Shock* 24 (2005) 407–411.
- [4] P. Marianneau, A. Cardona, L. Edelman, V. Deubel, P. Despres, Dengue virus replication in human hepatoma cells activates NF- κ B which in turn induces apoptotic cell death, *J. Virol.* 71 (1997) 3244–3249.
- [5] A. Catteau, O. Kalinina, M.C. Wagner, V. Deubel, M.P. Courageot, P. Despres, Dengue virus M protein contains a proapoptotic sequence referred to as ApoptoM, *J. Gen. Virol.* 84 (2003) 2781–2793.
- [6] P. Marianneau, A.M. Steffan, C. Royer, M.T. Drouet, A. Kirn, V. Deubel, Differing infection patterns of dengue and yellow fever viruses in a human hepatoma cell line, *J. Infect. Dis.* 178 (1998) 1270–1278.
- [7] T. Matsuda, A. Almasan, M. Tomita, K. Tamaki, M. Saito, M. Tadano, H. Yagita, T. Ohta, N. Mori, Dengue virus-induced apoptosis in hepatic cells is partly mediated by Apo2 ligand/tumour necrosis factor-related apoptosis-inducing ligand, *J. Gen. Virol.* 86 (2005) 1055–1065.
- [8] T. Thongtan, S. Panyim, D.R. Smith, Apoptosis in dengue virus infected liver cell lines HepG2 and Hep3B, *J. Med. Virol.* 72 (2004) 436–444.
- [9] A. Ruggieri, T. Harada, Y. Matsuura, T. Miyamura, Sensitization to Fas-mediated apoptosis by hepatitis C virus core protein, *Virology* 229 (1997) 68–76.
- [10] H. Marusawa, M. Hijikata, T. Chiba, K. Shimotohno, Hepatitis C virus core protein inhibits Fas- and tumor necrosis factor α -mediated apoptosis via NF- κ B activation, *J. Virol.* 73 (1999) 4713–4720.
- [11] C.S. Hahn, Y.G. Cho, B.S. Kang, I.M. Lester, Y.S. Hahn, The HCV core protein acts as a positive regulator of fas-mediated apoptosis in a human lymphoblastoid T cell line, *Virology* 276 (2000) 127–137.
- [12] J.S. Yang, M.P. Ramanathan, K. Muthumani, A.Y. Choo, S.H. Jin, Q.C. Yu, D.S. Hwang, D.K. Choo, M.D. Lee, K. Dang, W. Tang, J.J. Kim, D.B. Weiner, Induction of inflammation by West Nile virus capsid through the caspase-9 apoptotic pathway, *Emerg. Infect. Dis.* 8 (2002) 1379–1384.
- [13] R.J. Kuhn, W. Zhang, M.G. Rossmann, S.V. Pletnev, J. Corver, E. Lenches, C.T. Jones, S. Mukhopadhyay, P.R. Chipman, E.G. Strauss, T.S. Baker, J.H. Strauss, Structure of dengue virus: implications for flavivirus organization, maturation, and fusion, *Cell* 108 (2002) 717–725.
- [14] M. Tadano, Y. Makino, T. Fukunaga, Y. Okuno, K. Fukai, Detection of dengue 4 virus core protein in the nucleus. I. A monoclonal antibody to dengue 4 virus reacts with the antigen in the nucleus and cytoplasm, *J. Gen. Virol.* 70 (Pt 6) (1989) 1409–1415.
- [15] R. Bulich, J.G. Aaskov, Nuclear localization of dengue 2 virus core protein detected with monoclonal antibodies, *J. Gen. Virol.* 73 (Pt 11) (1992) 2999–3003.
- [16] S.H. Wang, W.J. Syu, K.J. Huang, H.Y. Lei, C.W. Yao, C.C. King, S.T. Hu, Intracellular localization and determination of a nuclear localization signal of the core protein of dengue virus, *J. Gen. Virol.* 83 (2002) 3093–3102.
- [17] C.J. Chang, H.W. Luh, S.H. Wang, H.J. Lin, S.C. Lee, S.T. Hu, The heterogeneous nuclear ribonucleoprotein K (hnRNP K) interacts with dengue virus core protein, *DNA Cell Biol.* 20 (2001) 569–577.
- [18] X. Yang, R. Khosravi-Far, H.Y. Chang, D. Baltimore, Daxx, a novel Fas-binding protein that activates JNK and apoptosis, *Cell* 89 (1997) 1067–1076.
- [19] R. Brent, R.L. Finley Jr., Understanding gene and allele function with two-hybrid methods, *Annu. Rev. Genet.* 31 (1997) 663–704.
- [20] R. Sriburi, P. Keelapang, T. Duangchinda, S. Pruksakorn, N. Maneekarn, P. Malasit, N. Sittisombut, Construction of infectious dengue 2 virus cDNA clones using high copy number plasmid, *J. Virol. Methods* 92 (2001) 71–82.
- [21] J. Gyuris, E. Golemis, H. Chertkov, R. Brent, Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2, *Cell* 75 (1993) 791–803.
- [22] B. Hirt, Selective extraction of polyoma DNA from infected mouse cell cultures, *J. Mol. Biol.* 26 (1967) 365–369.
- [23] S. Zhong, P. Salomoni, S. Ronchetti, A. Guo, D. Ruggero, P.P. Pandolfi, Promyelocytic leukemia protein (PML) and Daxx participate in a novel nuclear pathway for apoptosis, *J. Exp. Med.* 191 (2000) 631–640.
- [24] A.M. Ishov, A.G. Sotnikov, D. Negorev, O.V. Vladimirova, N. Neff, T. Kamitani, E.T. Yeh, J.F. Strauss 3rd, G.G. Maul, PML is critical for ND10 formation and recruits the PML-interacting protein daxx to this nuclear structure when modified by SUMO-1, *J. Cell Biol.* 147 (1999) 221–234.
- [25] R. Perlman, W.P. Schiemann, M.W. Brooks, H.F. Lodish, R.A. Weinberg, TGF- β -induced apoptosis is mediated by the adapter protein Daxx that facilitates JNK activation, *Nat. Cell Biol.* 3 (2001) 708–714.
- [26] X.D. Li, T.P. Makela, D. Guo, R. Soliymani, V. Koistinen, O. Vapalahti, A. Vaheiri, H. Lankinen, Hantavirus nucleocapsid protein interacts with the Fas-mediated apoptosis enhancer Daxx, *J. Gen. Virol.* 83 (2002) 759–766.
- [27] H.Y. Chang, H. Nishitoh, X. Yang, H. Ichijo, D. Baltimore, Activation of apoptosis signal-regulating kinase 1 (ASK1) by the adapter protein Daxx, *Science* 281 (1998) 1860–1863.
- [28] H. Li, C. Leo, J. Zhu, X. Wu, J. O'Neil, E.J. Park, J.D. Chen, Sequestration and inhibition of Daxx-mediated transcriptional repression by PML, *Mol. Cell Biol.* 20 (2000) 1784–1796.
- [29] R. Li, H. Pei, D.K. Watson, T.S. Papas, EAP1/Daxx interacts with ETS1 and represses transcriptional activation of ETS1 target genes, *Oncogene* 19 (2000) 745–753.
- [30] A.D. Hollenbach, C.J. McPherson, E.J. Mientjes, R. Iyengar, G. Grosveld, Daxx and histone deacetylase II associate with chromatin through an interaction with core histones and the chromatin-associated protein Dek, *J. Cell Sci.* 115 (2002) 3319–3330.
- [31] P. Salomoni, A.F. Khelifi, Daxx: death or survival protein? *Trends Cell Biol.* 16 (2006) 97–104.
- [32] J.S. Michaelson, The Daxx enigma, *Apoptosis* 5 (2000) 217–220.
- [33] M. van Engeland, L.J. Nieland, F.C. Ramaekers, B. Schutte, C.P. Reutelingsperger, Annexin V-affinity assay: a review on an apoptosis detection system based on phosphatidylserine exposure, *Cytometry* 31 (1998) 1–9.

Metabolic Enzymes, Antioxidants, and Cytoskeletal Proteins Are Significantly Altered in Vastus Lateralis Muscle of K-Depleted Cadaveric Subjects

Ratree Tavichakorntrakool,[†] Pote Sriboonlue,[†] Vitoon Prasongwattana,[†] Anucha Puapairoj,[‡] Pa-thai Yenchitsomanus,^{§,||} Supachok Sinchaikul,[⊥] Shui-Tein Chen,^{⊥,¶} Chaisiri Wongkham,[†] and Visith Thongboonkerd^{*,∇}

Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand, Department of Pathology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand, Medical Molecular Biology Unit, Office for Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, Medical Biotechnology Unit, National Center for Genetic Engineering and Biotechnology (BIOTEC), Bangkok, Thailand, Institute of Biological Chemistry and Genomic Research Center, Academia Sinica, Taipei, Taiwan, Institute of Biochemical Sciences, College of Life Science, National Taiwan University, Taipei, Taiwan, and Medical Proteomics Unit, Office for Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

Received November 3, 2008

Abstract: Molecular mechanisms underlying myopathy caused by prolonged potassium (K) depletion remain poorly understood. In the present study, we examined proteome profile of vastus lateralis muscle obtained from cadaveric subjects who had K depletion (KD) (muscle K < 80 $\mu\text{mol/g}$ wet weight) compared to those who had no KD (NKD) (muscle K $\geq$ 80 $\mu\text{mol/g}$ wet weight) ($n = 6$ per group). Muscle proteins were extracted, resolved by 2-DE, and visualized with CBB-R250 stain. Spot matching and intensity analysis revealed significant changes in levels of 11 (6 increased and 5 decreased) protein spots in the KD group. Q-TOF MS and MS/MS analyses identified these altered proteins as metabolic enzymes (aldehyde dehydrogenase 1A1, uridine diphosphoglucose pyrophosphorylase, enolase 1, cytosolic malate dehydrogenase, and carbonic anhydrase III), antioxidants (peroxiredoxin-3 isoform b), cytoskeletal proteins (slow-twitch skeletal troponin I and myosin light chain 2), and others. These altered proteins are involved in many cellular functions, including bioenergetics, acid–base regulation, oxidative stress response, and muscle contractility. Vali-

dation was done by Western blot analysis, which confirmed the increased level of peroxiredoxin-3 and decreased level of troponin-I in the KD muscle. Linear regression analysis also revealed a significant negative correlation between peroxiredoxin-3 level and muscle K content ($r = -0.887$; $p < 0.001$), as well as a significant positive correlation between troponin-I level and muscle K content ($r = 0.618$; $p < 0.05$). Our results implicate the important roles these altered proteins play in the development of KD-associated myopathy.

Keywords: Antioxidants • Cytoskeletal proteins • Metabolic enzymes • Pathophysiology • Potassium depletion • Proteome • Proteomics • Skeletal muscle

Introduction

Maintenance of homeostasis in normal cells requires a normal range of potassium (K) level in blood circulation. Skeletal muscle is the main organ responsible for K storage in human body. Therefore, the most acceptable method for determination of total body K status is the measurement of muscle K content and K depletion (or deficiency) is widely defined when muscle K content is $<80 \mu\text{mol/g}$ wet weight.¹ As muscle is the major pool for K storage, prolonged depletion of K in this important storage pool would result to systemic effects of K deficiency, and in turn, can cause muscular disorder, namely “hypokalemic myopathy”. Although this metabolic disorder has been recognized for decades, its molecular mechanisms underlying this type of myopathy remain poorly understood.

Previously, several studies had examined alterations in skeletal muscles during K depletion. These studies, however, focused most of their attention on physiological changes, particularly roles of Na, K-ATPase and the balance of cellular cations.^{1–4} Recently, our group has evaluated changes in muscle proteome profile of BALB/c mice fed with K-depleted

* To whom correspondence should be addressed. Visith Thongboonkerd, MD, FRCPT, Medical Proteomics Unit, Office for Research and Development, 12th Floor Adulyadej Vikrom Building, 2 Prannok Road, Siriraj Hospital, Bangkoknoi, Bangkok 10700, Thailand. Phone/Fax: +66-2-4184793. E-mail: thongboonkerd@dr.com (or) vthongbo@yahoo.com.

[†] Department of Biochemistry, Faculty of Medicine, Khon Kaen University.

[‡] Department of Pathology, Faculty of Medicine, Khon Kaen University.

[§] Medical Molecular Biology Unit, Office for Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University.

^{||} Medical Biotechnology Unit, National Center for Genetic Engineering and Biotechnology (BIOTEC).

[⊥] Institute of Biological Chemistry and Genomic Research Center, Academia Sinica.

[¶] Institute of Biochemical Sciences, College of Life Science, National Taiwan University.

[∇] Medical Proteomics Unit, Office for Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University.

Altered Proteome in K⁺-Depleted Muscle

(KD) diet for 8 weeks compared to those fed with normal chow.⁵ Comparative analysis revealed increased levels of 16 proteins in the KD muscles, most of which were muscle enzymes playing significant roles in many metabolic pathways. However, there was an argument that those data might be less relevance as the artificial induction of chronic K depletion in a murine model might not represent pathophysiological processes of K deficiency in humans.

We, therefore, have extended our study to characterize K deficiency in humans. Our recent study analyzing muscle K contents in cadaveric subjects, who resided in the northeastern part of Thailand and died from vehicle accident, revealed that approximately 23% of the recruited subjects (7 out of 30 cases) had K depletion ($K < 80 \mu\text{mol/g}$ wet weight),⁶ consistent to the previous surveys, which indicated that the northeastern part of Thailand is a high prevalent area of K deficiency.^{7,8} This population was therefore our focus in the present study aiming for analyzing changes in proteome profile of skeletal muscle in human subjects when muscle K was chronically depleted. Specimens were collected within 6 h of death to obtain the most reliable data.⁹ Proteins extracted from left vastus lateralis muscles were analyzed by 2-DE-based proteomics approach, comparing between KD and non-K-depleted (NKD) subjects.

Materials and Methods

Cadaveric Subjects and Sample Collection. All experiments were reviewed and approved by Khon Kaen University Ethics Committee. Totally, 30 vehicular accident cadaveric males who were born and resided in the northeastern region of Thailand were recruited. Their relatives were interviewed for past health history and those who had one or more of the following conditions were excluded from the study; (i) underlying medical illness prior to the death; (ii) evidence of injury to left vastus lateralis muscle; (iii) no information on past health history; (iv) no permission from the cadaveric relatives.

Autopsies of the muscle tissues were conducted at the Department of Forensic Medicine, Srinakarin Hospital, Khon Kaen University. To minimize the effects of postmortem autolysis of tissues, all autopsies were done within 6 h of the death, according to our previous study.⁹ Approximately 1 cm³ of the left vastus lateralis muscle was taken from each subject, weighed, and then subjected to measurement of K content and proteome analysis.

Measurement of Muscle K Content. Approximately 35 mg of the muscle sample was used for the measurement of K content, which was performed in triplicate for each subject. Briefly, muscle sample was homogenized at room temperature in 2.1 mL of 5% trichloroacetic acid (TCA) using a glass homogenizer, and then centrifuged at 1100g for 10 min. The clear supernatant was subsequently analyzed for K content with a flame photometer using lithium as an internal standard.

Sample Preparation and Two-Dimensional Gel Electrophoresis (2-DE). Muscle tissues were briskly frozen in liquid nitrogen and ground to powder using prechilled mortar and pestle. The ground samples were then resuspended in a lysis buffer containing 7 M urea, 2 M thiourea, 4% (w/v) 3-[(3-cholamidopropyl)dimethyl-ammonio]-1-propanesulfonate (CHAPS), 2% (v/v) ampholytes (pH 3–10) and 40 mM dithiothreitol (DTT), and the mixtures were incubated at 4 °C for 30 min. After centrifugation at 12 000g for 10 min, the supernatants were saved and the protein concentrations were determined by spectrophotometry using the Bradford method.¹⁰

Immobilized pH gradient (IPG) strips (linear pH 3–10, 7 cm long) (GE Healthcare; Uppsala, Sweden) were rehydrated overnight with 200 μg of total protein derived from individual subjects as mentioned above and premixed with a rehydration buffer containing 7 M urea, 2 M thiourea, 2% (w/v) CHAPS, 0.5% (v/v) ampholytes (pH 3–10), 18 mM dithiothreitol (DTT), and 0.002% bromophenol blue (to make the final volume of 150 μL per strip). The first-dimensional separation was performed in Ettan IPGphor II Isoelectric Focusing Unit (GE Healthcare) at 20 °C, using stepwise mode to reach 6500 Vh. After completion of the isoelectric focusing (IEF), the separated proteins were equilibrated with a buffer containing 6 M urea, 65 mM DTT, 29.3% glycerol, 75 mM Tris-HCl (pH 8.8), 2% sodium dodecyl sulfate (SDS) and 0.002% bromophenol blue for 15 min, and then with another buffer containing 6 M urea, 135 mM iodoacetamide, 29.3% glycerol, 75 mM Tris-HCl, 2% SDS and 0.002% bromophenol blue for further 15 min. The equilibrated IPG strips were then transferred onto 12% acrylamide slab gels (8 × 9.5 cm) and the second-dimensional separation was performed in Hoefer miniVE system (GE Healthcare) with the current of 20 mA/gel for approximately 1.5 h. The resolved protein spots were then visualized using Coomassie Brilliant Blue R-250 stain.

Matching and Analysis of Protein Spots. Image Master 2D Platinum software (GE Healthcare) was used for matching and analysis of protein spots on 2-D gels. A reference gel was created by combining all of the spots appeared in individual gels into one image. The reference gel was then used for matching of corresponding protein spots among different gels. Background subtraction was performed and the intensity volume of each spot was normalized with total intensity volume (summation of the intensity volumes obtained from all spots within the same 2-D gel).

In-Gel Trypsin Digestion. Protein spots whose intensity levels significantly differed between the KD and NKD groups were excised from the 2-D gels and the gel pieces were washed twice with 200 μL of 50% acetonitrile (ACN)/25 mM NH₄HCO₃ buffer (pH 8.0) for 15 min. The gel pieces were then washed once with 200 μL of 100% ACN and dried using a Speed Vac concentrator (Savant, Holbrook, NY). Dried gel pieces were swollen with 10 μL of 1% (w/v) trypsin (Promega, Madison WI) in 25 mM NH₄HCO₃. The gel pieces were then crushed with siliconized blue stick and incubated at 37 °C for at least 16 h. Peptides were subsequently extracted twice with 50 μL of 50% ACN/5% trifluoroacetic acid (TFA); the extracted solutions were then combined and dried with the Speed Vac concentrator. The peptide pellets were then resuspended in 10 μL of 0.1% TFA and the resuspended solutions were purified using ZipTip_{C18} (Millipore, Bedford, MA). Ten microliters of sample was drawn up and down in the ZipTip for 10 times and then washed with 10 μL of 0.1% formic acid by drawing up and expelling the washing solution 3 times. The peptides were finally eluted with 5 μL of 75% ACN/0.1% formic acid.

Protein Identification by Q-TOF MS and MS/MS Analyses. The proteolytic samples were premixed 1:1 with the matrix solution (5 mg/mL α -cyano-4-hydroxycinnamic acid in 50% ACN, 0.1% (v/v) TFA and 2% (w/v) ammonium citrate) and spotted onto the 96-well MALDI sample plate. The samples were analyzed by Q-TOF Ultima MALDI instrument (Micro-mass, Manchester, U.K.), which was fully automated with predefined probe motion pattern and the peak intensity threshold for switching over from MS survey scanning to MS/MS, and from one MS/MS to another. Within each sample well,

technical notes

parent ions that met the predefined criteria (any peak within the m/z of 800–3000 range with intensity above 10 count $\pm$ include/exclude list) were selected for CID MS/MS using argon as the collision gas and a mass dependent ± 5 V rolling collision energy until the end of the probe pattern was reached. The MS and MS/MS data were extracted and outputted as the searchable .txt and .pkl files, respectively, for independent searches using the MASCOT search engine (<http://www.matrixscience.com>), assuming that all peptides were monoisotopic. Fixed modification was carbamidomethylation at cysteine residues, whereas variable modification was oxidation at methionine residues. Only one missed trypsin cleavage was allowed, and peptide mass tolerances of 100 and 50 ppm were allowed for peptide mass fingerprinting (PMF) and MS/MS ions search, respectively, against the NCBI mammalian protein database. Threshold for significant hits was probability-based MOWSE (molecular weight search) scores >69 or ions scores >37 for PMF and MS/MS analyses, respectively.

Western Blot Analysis. Equally, 20 μ g of total proteins extracted from vastus lateralis muscle (as described above) was resolved in 12% SDS-PAGE and transferred onto polyvinylidene fluoride membranes. The membranes were incubated at 4 °C overnight with primary antibodies to peroxiredoxin-3 (Santa Cruz Biotechnology, Inc.; Santa Cruz, CA), troponin-I (Santa Cruz) or to β -actin (Abcam, Inc.; Cambridge, U.K.); all with a dilution of 1:4000 in 5% skimmed milk/PBS. After washing, the membrane was incubated further with respective secondary antibodies conjugated with horseradish peroxidase with a dilution of 1:8000. Reactive protein bands were then visualized with ECL Plus Western blot detection system (GE Healthcare). Band intensity was measured by ImageQuant TL software (GE Healthcare).

Statistical Analysis. All data are reported as Mean $\pm$ SD. Statistical analysis was performed with SPSS software (version 11.0) (SPSS Corporation, Chicago, IL). To test for differences between 2 sample groups, an unpaired t -test was used. To test for correlations between two parameters, a linear regression analysis was performed. P -values <0.05 were considered statistically significant.

Results

Totally, 12 cadaveric subjects were included for proteome analysis. On the basis of K content in left vastus lateralis muscle, the subjects were divided into two groups including 6 non-K-depleted (NKD) (muscle K content ≥ 80 μ mol/g wet weight) and 6 K-depleted (KD) (muscle K content <80 μ mol/g wet weight) subjects. Their actual values were 90.77 ± 3.08 versus 72.09 ± 2.18 , respectively, $p < 0.0001$ (Figure 1). All were males with ages of 39.17 ± 4.73 and 38.67 ± 4.55 years for NKD and KD groups, respectively ($p = 0.940$). The intervals or durations from death to specimen collection were 3.88 ± 0.47 and 3.68 ± 0.38 h ($p = 0.747$) and ambient temperatures during sample collection were 28.50 ± 0.89 and 29.00 ± 0.73 °C ($p = 0.674$) for NKD and KD groups, respectively.

Muscle proteome was analyzed by a gel-based method to define differentially expressed proteins in vastus lateralis muscle of KD subjects compared to NKD samples. Extracted proteins were resolved in 2-DE and visualized with Coomassie Brilliant Blue R-250 stain. With the use of 2-D analysis software, approximately 270 protein spots were visualized in each 2-D gel. Quantitative intensity analysis and statistics revealed significantly differential expression of 11 protein spots (labeled with numbers in Figure 2), including 6 up-regulated and 5

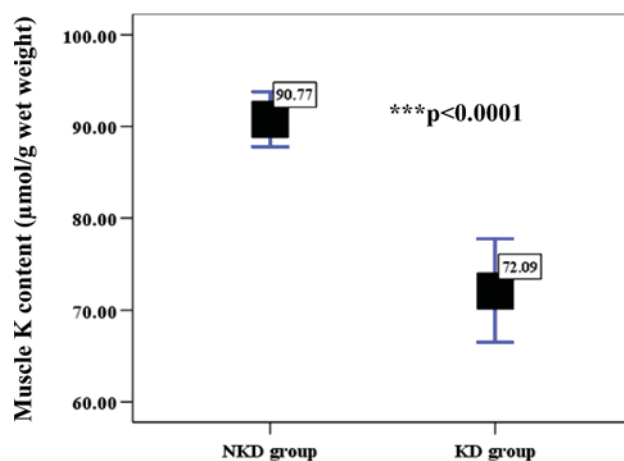


Figure 1. K contents in left vastus lateralis muscle in male cadaveric subjects included into this study. On the basis of cutoff level established in previous studies,^{1,6} the subjects were grouped as KD (K-depleted; muscle K < 80 μ mol/g wet weight) and NKD (non-K-depleted; muscle K ≥ 80 μ mol/g wet weight) groups.

down-regulated spots. The degrees of alterations ranged from 1.21- to 7.94-fold increases and 0.11- to 0.54-fold decreases (Table 1). These altered proteins were then subjected to mass spectrometric protein identification. With the use of Q-TOF MS and MS/MS analyses, 10 of 11 altered proteins were successfully identified. Their identities, identifiers, PMF scores, MS/MS ions scores, percentages of sequence coverage (%Cov), numbers of matched peptides, and intensity data are summarized in Table 1.

The identified proteins were then classified into four functional groups based on their major roles in biological processes, including metabolic enzymes, antioxidants, cytoskeletal proteins and miscellaneous proteins, whose roles remain unclear. Many of these proteins have several synonyms. We also performed database search using the information appeared in the well-established protein databases including NCBI (<http://www.ncbi.nlm.nih.gov/>), Swiss-Prot/TrEMBL (<http://www.expasy.org/>) and HumanCyc (<http://humancyc.org/>) to obtain additional details for their functions. Their synonyms and details of their main functions, as well as metabolic pathways involved, are summarized in Table 2.

To confirm our proteomic data, Western blot analysis of randomly selected altered proteins was performed. Figure 3 illustrates that relative band intensity levels of peroxiredoxin-3 and troponin-I in Western blots normalized with β -actin band were significantly increased and decreased, respectively, in the KD muscle samples compared to the NKD group. We also performed a linear regression analysis as to examine whether intensity levels of peroxiredoxin-3 and troponin-I correlated well with muscle K contents. Figure 4A demonstrates that there was a significant negative (reverse) correlation between the spot intensity levels of peroxiredoxin-3 and K contents in vastus lateralis muscle ($r = -0.887$; $p < 0.001$), whereas Figure 4B illustrates a significant positive correlation between the spot intensity levels of troponin-I and K contents ($r = 0.618$; $p < 0.05$).

Discussion

The present study employed a gel-based proteomics approach and identified 11 proteins whose abundance levels were

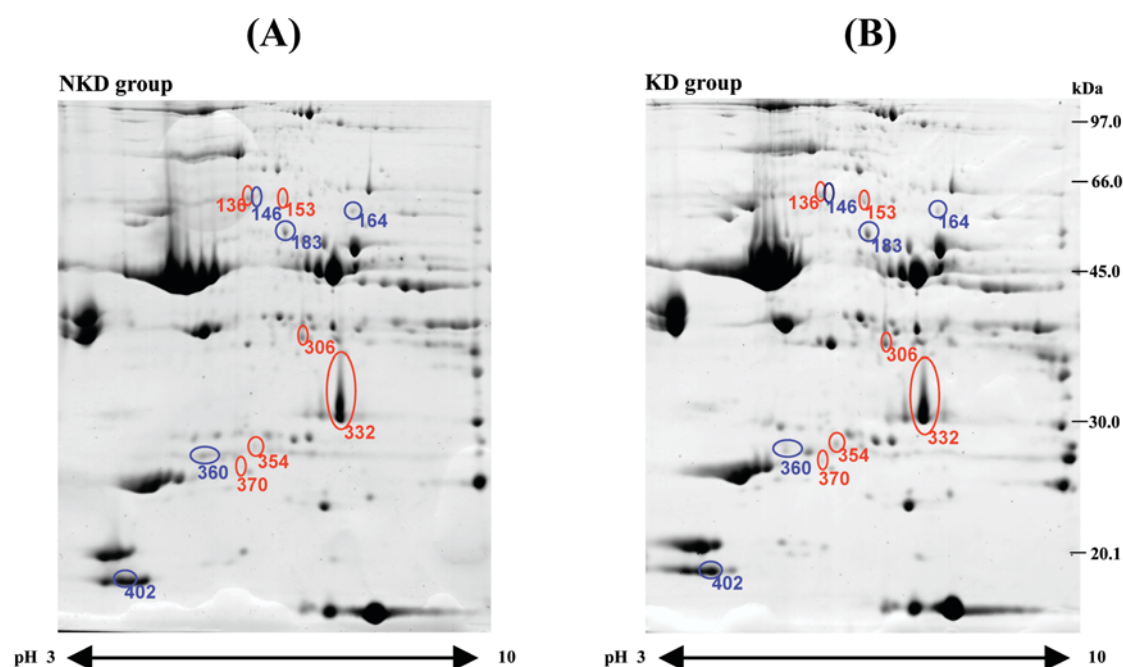


Figure 2. Representative 2-D proteome maps showing 11 differentially expressed proteins between non-K-depleted (NKD) and K-depleted (KD) groups (totally 6 gels per group). Numbers of protein spots labeled in these maps correspond to those reported in Table 1. Red-circled spots denote the increased proteins, whereas blue-circled spots denote the decreased proteins.

Table 1. Summary of Differentially Expressed Proteins Between non-K-Depleted (NKD) and K-Depleted (KD) Groups^a

spot no.	protein	NCBI ID	scores (MS, MS/MS)	%cov (MS, MS/MS)	no. of matched peptides (MS, MS/MS)	pI	MW (kDa)	intensity levels (KD/NKD)	p-value
Metabolic Enzymes									
153	Aldehyde dehydrogenase 1A1	gil21361176	95, NA	25, NA	13, NA	6.30	55.45	1.94	0.035
164	Uridine diphosphoglucose pyrophosphorylase	gil881394	77, NA	28, NA	11, NA	8.16	55.84	0.28	0.049
183	Enolase 1	gil4503571	155, 135	50, 8	18, 2	7.01	47.48	0.21	0.029
306	Cytosolic malate dehydrogenase	gil5174539	102, 86	44, 8	11, 2	6.91	36.63	1.21	0.049
332	Carbonic anhydrase III, muscle specific	gil13436164	163, 177	54, 28	14, 5	6.86	29.84	1.28	0.018
Antioxidants									
370	Peroxiredoxin-3 isoform b	gil32483377	78, 47	37, 5	9, 1	7.04	26.11	7.94	0.008
Cytoskeletal Proteins									
360	Slow-twitch skeletal troponin I	gil339965	NA, 50	NA, 12	NA, 2	9.55	21.82	0.14	0.023
402	Myosin light chain 2	gil28372499	110, NA	72, NA	13, NA	4.91	19.12	0.54	0.034
Miscellaneous									
136	Similar to tripartite motif-containing 4, isoform CRA_b	gil119572531	186, 24	62, 3	20, 1	5.99	37.57	2.28	0.019
146	Similar to tripartite motif-containing 4, isoform CRA_b	gil119572531	162, 14	56, 4	17, 1	5.99	37.57	0.11	0.008
354	Unidentified	NA	NA, NA	NA, NA	NA, NA	NA	NA	2.52	0.046

^a NA = Not applicable.

significantly altered in the KD muscle. To the best of our knowledge, this is the first time that proteomics has been applied to characterize the altered muscle proteome during K depletion in humans. Some of these altered proteins have been demonstrated to be involved in many cellular functions, particularly energy production, acid–base regulation, oxidative stress response, and muscle contractility. Functional significances for some of these proteins are discussed as follows.

Aldehyde dehydrogenase is an important metabolic enzyme in an oxidative pathway of alcohol metabolism and plays role in catalyzing the oxidation of aldehyde to acid using NAD as a coenzyme.¹¹ Cytosolic malate dehydrogenase catalyzes the reversible oxidation of malate to oxaloacetate, utilizing the NAD/NADH cofactor system in citric acid cycle.¹² It also plays

a crucial role in malate-aspartate shuttle in cellular metabolism and is important for transporting NADH equivalents across the mitochondrial membrane, controlling citric acid cycle pool, and providing contractile function.¹² Our present study identified increased levels of both aldehyde dehydrogenase and cytosolic malate dehydrogenase, suggesting their roles in the replenishment of energy source in the low-K condition.

Uridine diphosphoglucose pyrophosphorylase plays role in mammalian carbohydrate interconversion. It catalyzes the transfer of glucose moiety from glucose-1-phosphate to UTP, resulting to the formation of UDP-glucose.¹³ Muscle glycogen provides a quick source of energy for either aerobic or anaerobic metabolisms, which can be exhausted in less than an hour during vigorous activity/exercise. Enolase 1 is an important

Table 2. Brief Descriptions of Significantly Altered Proteins in KD Muscle

spot no.	protein/synonyms	function/catalytic activity
Metabolic Enzymes		
153	Aldehyde dehydrogenase 1A1^c EC 1.2.1.3 Aldehyde dehydrogenase, cytosolic ALDH CLASS 1 ALHDII ALDH-E1 Aldehyde dehydrogenase 1 family, member A1 Retinal dehydrogenase 1 Acetaldehyde dehydrogenase 1 Aldehyde dehydrogenase 1, soluble Aldehyde dehydrogenase, liver cytosolic	Aldehyde dehydrogenase is the second enzyme of the major oxidative pathway of alcohol metabolism. $\text{Aldehyde} + \text{NAD}^+ + \text{H}_2\text{O} \rightarrow \text{acid} + \text{NADH}$ $\text{L-lactaldehyde} + \text{NAD}^+ + \text{H}_2\text{O} \rightarrow \text{L-lactate} + \text{NADH}$
164	Uridine diphosphoglucose pyrophosphorylase^a EC 2.7.7.9. UTP-glucose-1-phosphate uridylyltransferase. UDP-glucose pyrophosphorylase UDPGP UGPase	Its role in glycogen synthesis, plays an important function to synthesis of UDP-glucose from glucose-1-phosphate. In muscle and liver tissue, UDP-glucose is a direct precursor for glycogen synthesis. $\alpha\text{-D-glucose-1-phosphate} + \text{UTP} \rightarrow \text{UDP-glucose} + \text{PPi}$
183	Enolase 1^{a,c} EC 4.2.1.11 Alpha-enolase 2-phospho-D-glycerate hydro-lyase Non-neural enolase NNE Phosphopyruvate hydratase C-myc promoter-binding protein MBP-1 MPB-1	Its role in glycolytic pathway via ATP production. $2\text{-phospho-D-glycerate} \rightarrow \text{phosphoenolpyruvate} + \text{H}_2\text{O}$ Cofactor: Mg. Required Mg for catalysis and stabilizing the dimeric structure.
306	Cytosolic malate dehydrogenase^c EC 1.1.1.37 Malate dehydrogenase 1, NAD (soluble) Cytosolic malate dehydrogenase Soluble malate dehydrogenase	It catalyzes the reversible oxidation of malate to oxaloacetate, utilizing the NAD/NADH cofactor system in the citric acid cycle. It may participate in the malate-aspartate shuttle. $\text{Malate} + \text{NAD}^+ \leftrightarrow \text{oxaloacetate} + \text{NADH}$
332	Carbonic anhydrase III,^c EC 4.2.1.1 Carbonic anhydrase 3 CA-III	Its role in acid–base regulation, plays a parts in reversible hydration of carbon dioxide. $\text{HCO}_3^- + \text{H}^+ \leftrightarrow \text{CO}_2 + \text{H}_2\text{O}$ Cofactor: Zn
Antioxidants		
370	Peroxiredoxin-3 isoform b^c Peroxiredoxin 3 or PRX III Antioxidant protein 1 AOP-1 MER5 protein homologue, HBC189 Thioredoxin-dependent peroxide reductase precursor	Peroxiredoxin 3 is an antioxidant protein. It is a putative alkylhydroperoxide reductase that functions as an antioxidant.
Cytoskeletal Proteins		
360	Slow-twitch skeletal troponin I,^c Troponin I, slow skeletal muscle Troponin I, slow-twitch isoform Troponin I, skeletal, slow	Troponin I is the inhibitory subunit of troponin, the thin filament regulatory complex which confers calcium-sensitivity to skeletal muscle actomyosin ATPase activity.
402	Myosin light chain 2,^c Myosin regulatory light chain 2	It may modulate the interaction between myosin and actin.
Miscellaneous		
136	Similar to tripartite motif-containing ^d , isoform CRA_b	Its function has not been identified.
146	Similar to tripartite motif-containing ^d , isoform CRA_b	Its function has not been identified.
354	Unidentified	NA ^e

^a Swiss-Prot database. ^b TrEMBL database. ^c HumanCyc database. ^d NCBI database. ^e NA = Not applicable.

enzyme in glycolytic pathway that catalyzes the conversion of 2-phospho-D-glycerate to phosphoenolpyruvate. This enzyme requires Mg as a cofactor for stabilizing its dimeric structure.^{14–16} Both uridine diphosphoglucose pyrophosphorylase and enolase 1 are involved in carbohydrate metabolism; their down-regulation in the KD subjects may be related to the energy depletion. The decrease in ATP production and coordinated declines of K and Mg levels may reflect the decrease of Na, K-ATPase activity and muscle contraction.¹⁷ Likewise, the ATP and Mg depletion may reduce the activity of Ca-ATPase and may be associated with the reduction of sarcoplasmic reticulum Ca reuptake and muscle function.^{18–20}

Carbonic anhydrase III plays role in acid–base regulation by reversible catalyzing the reaction of H^+ and HCO_3^- to CO_2 and H_2O .^{21–23} Chronic K depletion is associated with intracel-

lular acidosis,^{24,25} and thus, its increase may be for controlling the acid–base homeostasis in the KD muscle. On the other hand, the increase of this enzyme in muscle may be a result of muscle damage and may be used as a biomarker, as for the case of its serum level, which can be used for the evaluation of myocardial infraction,²⁶ muscular dystrophy,²⁷ other neuromuscular disorders,²⁷ and after intensive exercise.²⁸ Moreover, the increased carbonic anhydrase III may have another diagnostic value for neurological disorders, in which creatine kinase level is normal.²⁷

Peroxiredoxin-3 is a mitochondrial putative alkylhydroperoxide reductase that acts as an antioxidant. It plays an important role in eliminating peroxides generated during metabolism and may participate in the signaling cascades of growth factors and tumor necrosis factor-alpha by regulating

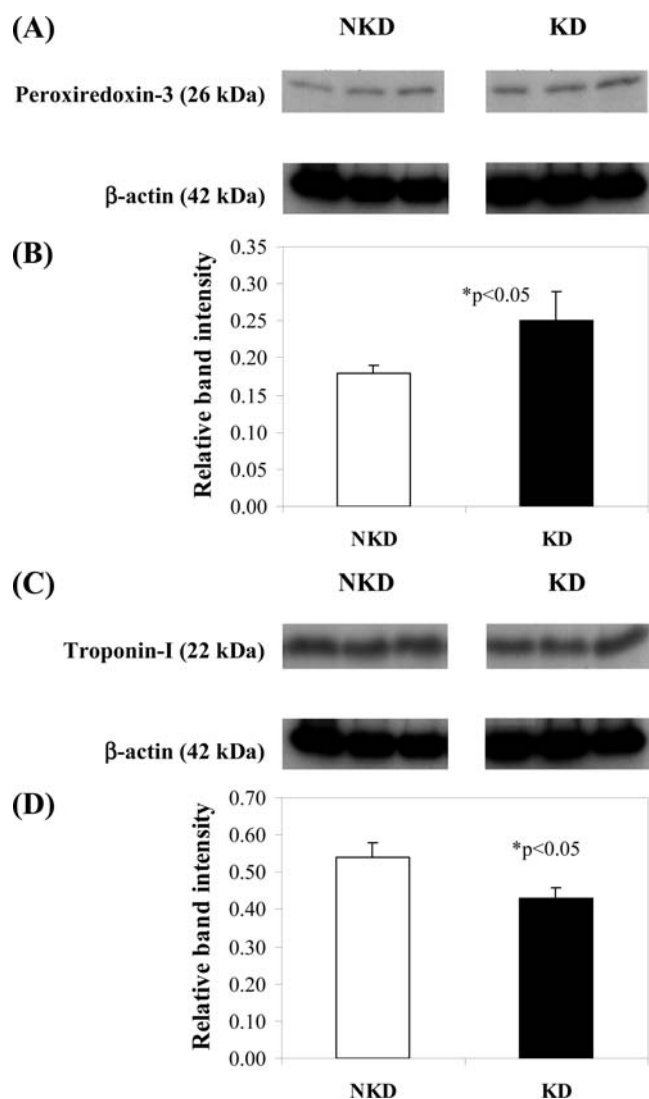
Altered Proteome in K^+ -Depleted Muscle

Figure 3. Confirmation of the proteomic data by Western blot analysis. Panels A and C show immunoreactive bands of peroxiredoxin-3 and troponin-I, respectively, with β -actin as the loading control. Panels B and D show the band intensity levels of peroxiredoxin-3 and troponin-I, respectively, normalized with the band intensity of β -actin. Comparative analysis revealed significant increase of the relative band intensity of peroxiredoxin-3 and significant decrease of the relative band intensity of troponin-I in KD muscle samples compared to NKD samples, consistent with the proteomic data shown in Figure 2 and Table 1.

the intracellular concentration of H_2O_2 .²⁹ Likewise, both reduced K and increased intracellular Na contents⁶ may favor the cellular accumulation of Ca via the reduced Na/Ca exchange mechanism and lead to the mitochondrial swelling and cellular damage.^{30–32} This mitochondrial Ca overload may lead to the increased production of reactive oxygen species, causing peroxidation of membrane lipids and possibly decrease in ATP synthesis. The overload of mitochondrial Ca also increases the risk of hyper-permeabilization of transition pore, which ultimately leads to cell apoptosis or necrosis. Local increase in cytosolic Ca may lead to phospholipase A_2 activation, resulting in the degradation of both cellular and organelle membranes.³² Therefore, the increase in peroxiredoxin-3 level in KD muscle may be a biological response to control oxidative stress during K depletion.

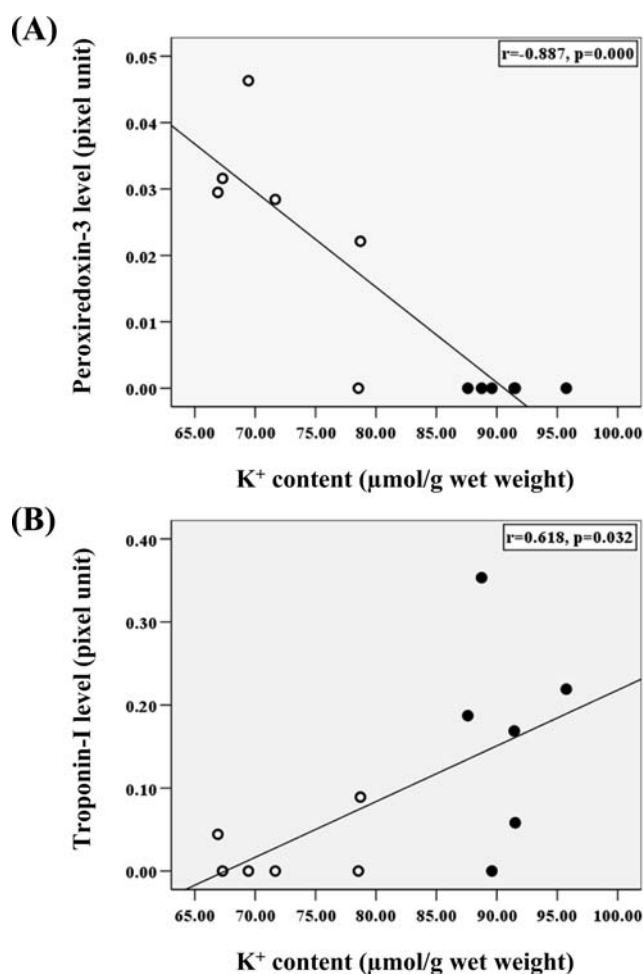


Figure 4. Linear regression analysis revealed a significant negative (reverse) correlation between levels of muscle peroxiredoxin-3 and muscle K contents (A) and a significant positive correlation between levels of muscle troponin-I and muscle K contents (B). (○) = values in KD vastus lateralis muscle; (●) = values in NKD vastus lateralis muscle.

Troponin I is an inhibitory subunit of troponin, the thin filament regulatory complex, which confers Ca-sensitivity to actomyosin ATPase activity.³³ Troponin I has two isoforms: slow- and fast-twitch skeletal troponin I. They could provide insight into differential types of injury of specific muscle fiber types and may be present in blood samples of patients with various skeletal muscle disorders.^{34,35} The fast-twitch skeletal muscle also has myosin, which is a multisubunit protein consisting of two heavy chains and a series of light chains. Myosin light chains can be divided into phosphorylated light chain (MLC2) (also called the regulatory light chain) and nonphosphorylated light chains (MLC1 and MLC3) (also called the alkali light chains). Among these, MLC2 is expressed specifically in skeletal muscle.³⁶ Phosphorylation of MLC2 by the enzyme MLC kinase, in the presence of Ca and calmodulin, increases actin-activated myosin ATPase activity and thereby regulates the contractile activity.^{37–40} The lower MLC phosphorylation found in elderly people could impair their performance by decreasing the rate and extent of force development during contraction.⁴¹ Our previous study⁶ reported that K depleted muscles also had the increase of intracellular Ca. Sarcoplasmic reticulum is capable of storing quite large quantities of Ca. However, over buffering capacity may lead to a local

activation of calpain, leading to degradation of cytoskeletal, myofibrillar and membrane proteins.³² In our present study, the decreases of slow-twitch skeletal troponin I and MLC2 may aggravate or be make it more prone to the impairment of muscle contractility in these KD subjects.

Comparing the data obtained from our present study on human subjects to those reported in our previous study on a murine model of prolonged K depletion,⁵ it was not unexpected that there were no common changes identified from these two studies. These differences were most likely due to the following. First, the animal model was not able to simulate all the clinical features and metabolic profiles of K depletion in humans. Second, the degree and duration of K depletion were somewhat different. Finally, adaptive responses to metabolic changes between these two species significantly differed.

In summary, we identified several proteins with significantly altered levels during K depletion in vastus lateralis muscle in humans. These altered proteins play important roles in many cellular functions, including bioenergetics, acid–base regulation, oxidative stress response, and muscle contractility. Our findings may be helpful for explaining the pathogenic mechanisms and pathophysiology of myopathy associated with chronic K depletion commonly encountered in the Northeastern Thais.

Abbreviations: ACN, acetonitrile; CHAPS, 3-[(3-cholamidopropyl) dimethyl-ammonio]-1-propanesulfonate; DTT, dithiothreitol; IEF, isoelectric focusing; IPG, immobilized pH gradient; KD, K-depleted; MALDI, matrix-assisted laser desorption/ionization; MLC, myosin light chain; MOWSE, molecular weight search; MS, mass spectrometry; MS/MS, tandem mass spectrometry; NKD, non-K-depleted; PMF, peptide mass fingerprinting; Q-TOF, quadrupole time-of-flight; SDS, sodium dodecyl sulfate; TCA, trichloroacetic acid; TFA, trifluoroacetic acid.

Acknowledgment. We wish to express our deep appreciation to all cadaveric subjects for the invaluable specimens and their relatives for permission of specimen collection. We also thank Drs. Banchoop Sripa, Sukkid Yasothornsrikul, Patcharee Boonsiri, and Sontaya Simasatiansophon for their advice, and are grateful to the Core Facilities for Proteomics and Structural Biology Research, Institute of Biological Chemistry, Academia Sinica, Taiwan. This project was supported by a research grant from Faculty of Medicine, Khon Kaen University, Thailand (to V. Prasongwattana) and by The Thailand Research Fund, Commission on Higher Education, Mahidol University, the National Research Council of Thailand, Siriraj Grant for Research and Development, and the National Center for Genetic Engineering and Biotechnology (to V. Thongboonkerd).

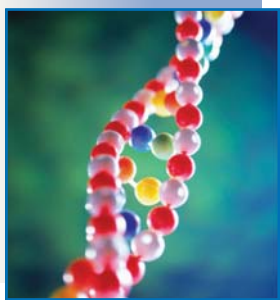
References

- (1) Dorup, I.; Skajaa, K.; Clausen, T.; Kjeldsen, K. Reduced concentrations of potassium, magnesium, and sodium-potassium pumps in human skeletal muscle during treatment with diuretics. *Br. Med. J.* **1988**, *296*, 455–458.
- (2) Dorup, I.; Clausen, T. Correlation between magnesium and potassium contents in muscle: role of Na(+)-K+ pump. *Am. J. Physiol.* **1993**, *264*, C457–C463.
- (3) Prasongwattana, V.; Tavichakorntrakool, R.; Sriboonlue, P.; Wongkham, C.; Bovornpadungkitti, S.; Premgamone, A.; Reungjui, S. Correlation between Na, K-ATPase activity and potassium and magnesium contents in skeletal muscle of renal stone patients. *Southeast Asian J. Trop. Med. Public Health* **2001**, *32*, 648–653.
- (4) Norgaard, A.; Kjeldsen, K.; Clausen, T. Potassium depletion decreases the number of 3H-ouabain binding sites and the active Na-K transport in skeletal muscle. *Nature (London)* **1981**, *293*, 739–741.

- (5) Thongboonkerd, V.; Kanlaya, R.; Sinchaikul, S.; Parichatanond, P.; Chen, S. T.; Malasit, P. Proteomic identification of altered proteins in skeletal muscle during chronic potassium depletion: Implications for hypokalemic myopathy. *J. Proteome Res.* **2006**, *5*, 3326–3335.
- (6) Tavichakorntrakool, R.; Prasongwattana, V.; Sriboonlue, P.; Pua-pairoj, A.; Wongkham, C.; Wiangsimma, T.; Khunkitti, W.; Triam-jangarun, S.; Tanratanaujit, M.; Chamsuwan, A.; Khunkitti, W.; Yenchitsomanus, P. T.; Thongboonkerd, V. K+, Na+, Mg2+, Ca2+, and water contents in human skeletal muscle: correlations among these monovalent and divalent cations and their alterations in K+-depleted subjects. *Transl. Res.* **2007**, *150*, 357–366.
- (7) Lelamali, K.; Khunkitti, W.; Yenrudi, S.; Panichaphongse, V.; Huiprasert, L.; Sitprija, V.; Tungsanga, K. Potassium depletion in a healthy north-eastern Thai population: no association with tubulo-interstitial injury. *Nephrology* **2003**, *8*, 28–32.
- (8) Sriboonlue, P.; Prasongwattana, V.; Tungsanga, K.; Tosukhowong, P.; Phantumvanit, P.; Bejrputra, O.; Sitprija, V. Blood and urinary aggregator and inhibitor composition in controls and renal-stone patients from northeastern Thailand. *Nephron* **1991**, *59*, 591–596.
- (9) Tavichakorntrakool, R.; Prasongwattana, V.; Sriboonlue, P.; Pua-pairoj, A.; Pongsul, J.; Khuntikeo, N.; Triam-jangarun, S.; Han-panich, W.; Yenchitsomanus, P. T.; Wongkham, C.; Thongboonkerd, V. Serial analyses of postmortem changes in human skeletal muscle: A case study of alterations in proteome profile, histology, electrolyte contents, water composition, and enzyme activity. *Proteomics Clin. Appl.* **2008**, *2*, 1255–1264.
- (10) Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248–254.
- (11) Abriola, D. P.; Fields, R.; Stein, S.; MacKerell, A. D., Jr.; Pietruszko, R. Active site of human liver aldehyde dehydrogenase. *Biochemistry* **1987**, *26*, 5679–5684.
- (12) Lo, A. S.; Liew, C. T.; Ngai, S. M.; Tsui, S. K.; Fung, K. P.; Lee, C. Y.; Waye, M. M. Developmental regulation and cellular distribution of human cytosolic malate dehydrogenase (MDH1). *J. Cell Biochem.* **2005**, *94*, 763–773.
- (13) Jue, T.; Rothman, D. L.; Shulman, G. I.; Tavittian, B. A.; DeFronzo, R. A.; Shulman, R. G. Direct observation of glycogen synthesis in human muscle with 13C NMR. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 4489–4491.
- (14) Kornblatt, M. J. Mechanism of rabbit muscle enolase: identification of the rate-limiting steps and the site of Li+ inhibition. *Arch. Biochem. Biophys.* **1996**, *330*, 12–18.
- (15) Maughan, D. W.; Henkin, J. A.; Vigoreaux, J. O. Concentrations of glycolytic enzymes and other cytosolic proteins in the diffusible fraction of a vertebrate muscle proteome. *Mol. Cell. Proteomics* **2005**, *4*, 1541–1549.
- (16) Piast, M.; Kustrzeba-Wojcicka, I.; Matusiewicz, M.; Banas, T. Molecular evolution of enolase. *Acta Biochim. Pol.* **2005**, *52*, 507–513.
- (17) Clausen, T. Na+-K+ pump regulation and skeletal muscle contractility. *Physiol. Rev.* **2003**, *83*, 1269–1324.
- (18) Byrd, S. K.; McCutcheon, L. J.; Hodgson, D. R.; Gollnick, P. D. Altered sarcoplasmic reticulum function after high-intensity exercise. *J. Appl. Physiol.* **1989**, *67*, 2072–2077.
- (19) Byrd, S. K.; Bode, A. K.; Klug, G. A. Effects of exercise of varying duration on sarcoplasmic reticulum function. *J. Appl. Physiol.* **1989**, *66*, 1383–1389.
- (20) Byrd, S. K. Alterations in the sarcoplasmic reticulum: a possible link to exercise-induced muscle damage. *Med. Sci. Sports Exerc.* **1992**, *24*, 531–536.
- (21) Moynihan, J. B. Carbonic anhydrase activity in mammalian skeletal and cardiac muscle. *Biochem. J.* **1977**, *168*, 567–569.
- (22) Berg, J. T.; Ramanathan, S.; Gabrielli, M. G.; Swenson, E. R. Carbonic anhydrase in mammalian vascular smooth muscle. *J. Histochem. Cytochem.* **2004**, *52*, 1101–1106.
- (23) Tripp, B. C.; Smith, K.; Ferry, J. G. Carbonic anhydrase: new insights for an ancient enzyme. *J. Biol. Chem.* **2001**, *276*, 48615–48618.
- (24) Wetzel, P.; Kleinke, T.; Papadopoulos, S.; Gros, G. Inhibition of muscle carbonic anhydrase slows the Ca(2+) transient in rat skeletal muscle fibers. *Am. J. Physiol.: Cell Physiol.* **2002**, *283*, C1242–C1253.
- (25) Tricarico, D.; Mele, A.; Conte, C. D. Carbonic anhydrase inhibitors ameliorate the symptoms of hypokalaemic periodic paralysis in rats by opening the muscular Ca2+-activated-K+ channels. *Neuromuscul. Disord.* **2006**, *16*, 39–45.
- (26) Vaananen, H. K.; Syrjala, H.; Rahkila, P.; Vuori, J.; Melamies, L. M.; Myllyla, V.; Takala, T. E. Serum carbonic anhydrase III and myoglobin concentrations in acute myocardial infarction. *Clin. Chem.* **1990**, *36*, 635–638.

- (27) Ohta, M.; Itagaki, Y.; Itoh, N.; Hayashi, K.; Nishitani, H.; Ohta, K. Carbonic anhydrase III in serum in muscular dystrophy and other neurological disorders: relationship with creatine kinase. *Clin. Chem.* **1991**, *37*, 36–39.
- (28) Takala, T. E.; Rahkila, P.; Hakala, E.; Vuori, J.; Puranen, J.; Vaananen, K. Serum carbonic anhydrase III, an enzyme of type I muscle fibres, and the intensity of physical exercise. *Pflugers Arch.* **1989**, *413*, 447–450.
- (29) Kang, S. W.; Chae, H. Z.; Seo, M. S.; Kim, K.; Baines, I. C.; Rhee, S. G. Mammalian peroxiredoxin isoforms can reduce hydrogen peroxide generated in response to growth factors and tumor necrosis factor- α . *J. Biol. Chem.* **1998**, *273*, 6297–6302.
- (30) Knochel, J. P.; Schlein, E. M. On the mechanism of rhabdomyolysis in potassium depletion. *J. Clin. Invest.* **1972**, *51*, 1750–1758.
- (31) Knochel, J. P. Neuromuscular manifestations of electrolyte disorders. *Am. J. Med.* **1982**, *72*, 521–535.
- (32) Gissel, H. The role of Ca²⁺ in muscle cell damage. *Ann. N.Y. Acad. Sci.* **2005**, *1066*, 166–180.
- (33) Corin, S. J.; Juhasz, O.; Zhu, L.; Conley, P.; Kedes, L.; Wade, R. Structure and expression of the human slow twitch skeletal muscle troponin I gene. *J. Biol. Chem.* **1994**, *269*, 10651–10659.
- (34) Schantz, P. G. Plasticity of human skeletal muscle with special reference to effects of physical training on enzyme levels of the NADH shuttles and phenotypic expression of slow and fast myofibrillar proteins. *Acta Physiol. Scand. Suppl.* **1986**, *558*, 1–62.
- (35) Simpson, J. A.; Labugger, R.; Collier, C.; Brison, R. J.; Iscoe, S.; Van Eyk, J. E. Fast and slow skeletal troponin I in serum from patients with various skeletal muscle disorders: a pilot study. *Clin. Chem.* **2005**, *51*, 966–972.
- (36) Faerman, A.; Shani, M. The expression of the regulatory myosin light chain 2 gene during mouse embryogenesis. *Development* **1993**, *118*, 919–929.
- (37) Metzger, J. M.; Moss, R. L. Myosin light chain 2 modulates calcium-sensitive cross-bridge transitions in vertebrate skeletal muscle. *Biophys. J.* **1992**, *63*, 460–468.
- (38) Matsumura, F. Regulation of myosin II during cytokinesis in higher eukaryotes. *Trends Cell Biol.* **2005**, *15*, 371–377.
- (39) Hernandez, O. M.; Jones, M.; Guzman, G.; Szczesna-Cordary, D. Myosin essential light chain in health and disease. *Am. J. Physiol.: Heart Circ. Physiol.* **2007**, *292*, H1643–H1654.
- (40) Fujita, H.; Sasaki, D.; Fukuda, K.; Ishiwata, S. Myosin light chain 2 modulates MgADP-induced contraction in rabbit skeletal and bovine cardiac skinned muscle. *J. Physiol.* **2002**, *542*, 221–229.
- (41) Gelfi, C.; Vigano, A.; Ripamonti, M.; Pontoglio, A.; Begum, S.; Pellegrino, M. A.; Grassi, B.; Bottinelli, R.; Wait, R.; Cerretelli, P. The human muscle proteome in aging. *J. Proteome Res.* **2006**, *5*, 1344–1353.

PR800941G



© PHOTODISC

Variable-Length Haplotype Construction for Gene–Gene Interaction Studies

A Nonparametric Classification Approach

BY ANUNCHAI ASSAWAMAKIN, NACHOL CHAIYARATANA, CHANIN LIMWONGSE, SARAVUDH SINSOMROS, PA-THAI YENCHITSOMANUS, AND PRAKARNKIAT YOUNGKONG

Genetic epidemiology is a research field that aims to identify genetic polymorphisms that are involved in disease susceptibility. In this article, a variable-length haplotype construction for gene–gene interaction (VarHAP) technique is proposed. The technique will involve nonparametric classification where haplotypes inferred from multiple single nucleotide polymorphism (SNP) data are the classifier inputs. Usual candidate polymorphisms include restriction fragment-length polymorphisms (RFLPs), variable number of tandem repeats (VNTRs), and SNPs. In recent years, SNPs are the most common choices because of its simplicity and cost reduction in identification protocols. SNPs in diploid organisms are excellent biallelic genetic markers for various studies including genetic association, gene–gene interaction, and gene–environment interaction. The availability of multiple SNPs on the same gene can also lead to haplotype analysis where genotypes of interest can be phased into pairs of haplotypes. In this article, a VarHAP technique is proposed. The technique involves nonparametric classification where haplotypes inferred from multiple SNP data are the classifier inputs for case–control studies.

Traditional techniques for identification of relationship between a single SNP and disease susceptibility status involve various univariate statistical tests including χ^2 and odds ratio tests [1], [2]. However, many complementary computational techniques have been developed in the past decade to handle problems that involve multiple SNPs. Heidema et al. [3] have categorized these multilocus techniques, which are capable of identifying a candidate SNP set from possible SNPs, into parametric and nonparametric methods. Examples of parametric method cover logistic regression techniques [4] and neural networks [5]. On the other hand, examples of nonparametric method include a set-association approach [6], combinatorial techniques [7]–[9], and recursive partitioning techniques [10], [11]. In some of the aforementioned parametric [4], [5] and nonparametric [9]–[11] methods, pattern recognition and classification approaches have been successfully implemented as their core engines.

In addition to single and multiple SNP analysis, haplotype analysis has also gained attention from genetic epidemiologists. Haplotypes provide a record of evolutionary history

more accurately than individual SNPs. Furthermore, haplotypes can capture the patterns of linkage disequilibrium (LD), a phenomenon where SNPs that are located in close proximity tend to travel together, in genomes more accurately. Therefore, haplotypes may enable susceptible gene identification in complex diseases more effectively than individual SNPs [12]. In lieu of this evidence, haplotype analysis should also be considered in addition to direct genotype analysis. Many computational techniques use haplotypes, which are inferred from multiple SNPs, as problem inputs. For instance, Sham et al. [13] proposes a logistic regression technique that produces a mapping model between haplotypes and disease status, whereas Becker et al. [14] combine haplotype explanation probabilities of given genotypes from multiple gene or unlinked region data into a scalar statistic for a permutation test. Nonetheless, haplotypes have rarely been used as inputs for nonparametric classifiers for genetic association and interaction studies.

Haplotype classification is thus the main interest of this article. The chosen architecture for nonparametric classifier is the multifactor dimensionality reduction (MDR) technique [9]. Similar to the original MDR technique, the proposed technique would be able to identify appropriate candidate SNPs from possible candidate SNPs and can be used in case–control genetic interaction studies. However, this technique would also be able to handle the situation where disease susceptibility is detectable in different haplotype backgrounds.

MDR, Haplotype Inference, and Haplotype Explanation Probability

MDR

MDR is a classifier-based technique that is capable of identifying the best genetic marker combination among possible markers for the separation between case and control samples. Similar to the other classification systems, a k -fold cross-validation technique provides a means to determine the classification accuracy of the candidate marker model. Basically, the combined case and control samples are randomly divided into k folds, where $k - 1$ folds of the samples are used to construct a decision table for the classifier whereas the remaining fold of samples is used to identify the prediction capability of the constructed decision table. The decision table construction

Digital Object Identifier 10.1109/EMEB.2009.932902

Examples of nonparametric method include a set-association approach, combinatorial techniques, and recursive partitioning techniques.

and testing procedure is repeated k times. Hence, the samples in each fold will always be used both to construct and test the decision table. The number of cells in a decision table is given by G^{n_c} , where n_c is the number of candidate markers selected from possible markers, and G is the number of possible genotypes according to the marker. For an SNP, which is a biallelic marker, G is equal to three. During the decision table construction, each cell in the table is filled with case and control samples that have their genotype corresponding to the cell label. The ratio between the number of case and control samples will provide the decision for each cell whether the corresponding genotype is a disease-predisposing or protective genotype. An example of decision table construction is illustrated in Figure 1. The prediction accuracy of the decision table is subsequently evaluated by counting the number of case and control samples in the testing fold that their disease status can be correctly identified using the constructed decision rules. The process of decision table construction and evaluation must be

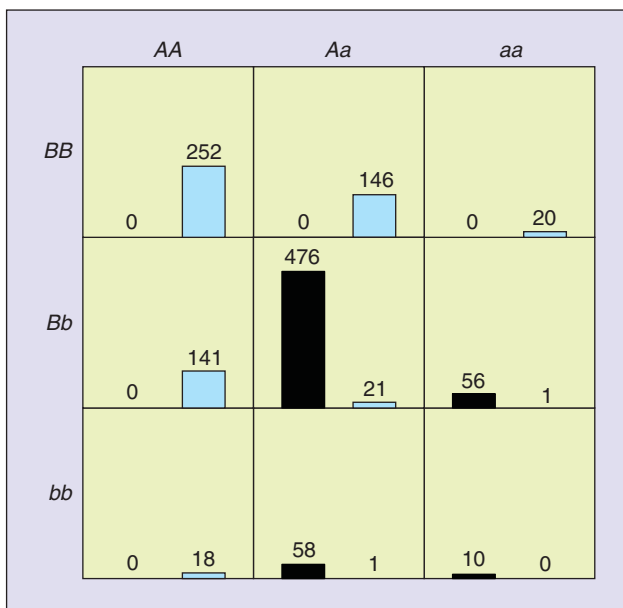


Fig. 1. An MDR decision table that is constructed using 1,200 case-control samples. The genotype of each sample is determined from two SNPs. The table consists of nine cells, where each cell represents a unique genotype. The left (black) bar in each cell represents the number of case samples whereas the right (pale blue) bar represents the number of control samples. The cells with genotypes $AaBb$, $aaBb$, $Aabb$, and $aabb$ are labeled as predisposing genotypes whereas the cells with genotypes $AABB$, $AaBB$, $aaBB$, $AABb$, and $AAbb$ are labeled as protective genotypes.

cycled through all or some of possible $2^{n_m} - 1$ combinations, where n_m is the total number of available markers in the study. The best genetic marker combination is determined from three criteria: prediction accuracy, cross-validation consistency, and a sign test p value. Each time that a testing fold is used for prediction accuracy determination, the accuracy of the interested marker combination model can be compared with that of other models that also contain the same number of markers. The model that consistently ranks first in comparison with other choices with the same amount of markers would have a high cross-validation consistency. The nonparametric sign test p value is calculated from the number of testing folds with accuracy $\geq 50\%$. This single-tailed p value is given by

$$p = \sum_{i=n_a}^{n_f} \binom{n_f}{i} \left(\frac{1}{2}\right)^{n_f}, \quad (1)$$

where n_f is the total number of cross-validation folds, and n_a is the number of cross-validation folds with testing accuracy $\geq 50\%$ [9]. Among the three criteria, the prediction accuracy is the main criterion for decision making, whereas the other criteria are used only as auxiliary measures. The cross-validation consistency generally confirms that the high-rank model can be consistently identified regardless of how the samples are divided for cross-validation. On the other hand, a sign test p value indicates the number of testing folds with acceptable prediction accuracy and hence describes the usability of the model in the classification task. In a situation where two or more models with different number of markers are equally good in terms of prediction accuracy, cross-validation consistency, and sign test p value, the most parsimonious model (the combination with the least number of markers) will be the best model.

Haplotype Inference

With the availability of multiple SNPs from the same gene, haplotypes can be inferred from given genotypes. Let 0 and 1 denote the major (common) and minor (rare) alleles at an SNP location in a haplotype. A genotype can then be represented by a string, which consists of characters 0, 1, and 2. In the genotype string, 0 denotes a homozygous wild-type site, 1 denotes a heterozygous site, and 2 denotes a homozygous variant or homozygous mutant site. A genotype with all homozygous sites or a single heterozygous site can always be phased into one pair of haplotypes. On the other hand, a genotype with multiple heterozygous sites can be phased into multiple haplotype pairs. For example, genotype 0102 leads to haplotypes 0001 and 0101, whereas genotype 0112 leads to two possible haplotype pairs: 0001/0111 and 0011/0101. Many algorithms exist for haplotype inference [15]–[17]. In this article, an expectation-maximization algorithm [15] is chosen because of its simplicity and implementation efficacy. Regardless of the inference technique employed, the usual result

from an inference algorithm covers haplotype frequencies and possible haplotype phases of each genotype.

Haplotype Explanation Probability

In a genomic region with multiple heterozygous sites, multiple pairs of haplotypes can be inferred from a given genotype. The probability of a genotype to be phased into one specific pair of haplotypes would depend on the frequencies of haplotypes constituting the pairs [14]. This probability is given by

$$w_{ij} = \frac{f_i f_j}{\sum_{(h_k, h_l) \in H} f_k f_l}, \quad (2)$$

where w_{ij} is the probability for haplotype pair ij , f_i denotes the frequency of the i th haplotype, h_k is the k th haplotype, and H represents the set of haplotype explanations that are compatible with the genotype of interest. For example, genotype 0110 can be phased into two haplotype pairs: 0000/0110 (h_1/h_4) and 0010/0100 (h_2/h_3). If the frequencies for haplotypes 0000, 0010, 0100, and 0110, are 0.5, 0.2, 0.2, and 0.1, respectively, the probabilities for the pairs 0000/0110 and 0010/0100 are 0.556 and 0.444. Obviously, the probability of a genotype with all homozygous sites or a single heterozygous site to be phased into a pair of haplotypes would be equal to one. In genetic interaction studies, where the number of genes or unlinked regions is greater than one, the haplotype explanation probabilities from all regions can be combined together. An overall contribution by one sample to haplotype configuration $(h_j^1, h_j^2, \dots, h_j^{n_u})$ in a study with n_u genes/unlinked regions is given by

$$c_{(h_j^1, h_j^2, \dots, h_j^{n_u})} = 2 \prod_{i=1}^{n_u} w_{jk}^i \frac{(1 + \delta_{jk}^i)}{2}, \quad (3)$$

where $c_{(h_j^1, h_j^2, \dots, h_j^{n_u})}$ is the contribution value and δ is defined as $\delta_{jk} = 1$ for $j = k$ and $\delta_{jk} = 0$ for $j \neq k$. In the previous example where haplotypes from only one region are considered, $c_{h_1} = 0.556$, $c_{h_2} = 0.444$, $c_{h_3} = 0.444$, and $c_{h_4} = 0.556$. Note that the sum of contribution values is equal to two; this reflects the fact that each genotype is made up of two haplotypes. Becker et al. [14] used this contribution value in the construction of a contingency table, where a χ^2 test statistic was subsequently calculated. Using Monte Carlo simulations, an estimated p value is then obtained. Similar to the model exploration strategy in MDR, the process of contingency table construction and p -value calculation can also be cycled through all or some of the possible interaction models. The model with appropriate candidate SNPs taken from possible SNPs is the one with minimum p value and is said to be the best model for the explanation of interactions. This statistics-based procedure can be found as an integral part of the FAMHAP program (software for single-marker analysis and joint analysis of unphased genotype data from tightly linked markers) [18].

VarHAP

VarHAP is proposed for case-control interaction studies. Similar to MDR, this technique is also a classifier-based technique. However, instead of using a genotype data analysis as a means to identify the best SNP combination, the decision table for classification is constructed from the haplotype contribution value as described earlier. As a result, haplotypes with different lengths must be inferred during the search for the best model. The number of decision cells during the consideration on

haplotypes constructed from a specific set of SNPs is governed by the total number of possible haplotype configurations as illustrated in Figure 2. In brief, VarHAP would maintain the ability to find the best SNP combination while also be able to identify possible disease-predisposing and protective haplotype configurations.

Since VarHAP is essentially a classification system, the principal criterion for choosing the optimal SNP combination model is still the prediction accuracy. However, using haplotype contribution value as a means for decision rule construction, an additional model selection criterion that exploits the nature of haplotype can be formulated. This criterion can be referred to as haplotype propagation capability. Basically, if a haplotype constructed from a specific set of SNPs is related to disease susceptibility status, haplotypes constructed from a SNP set, which is a

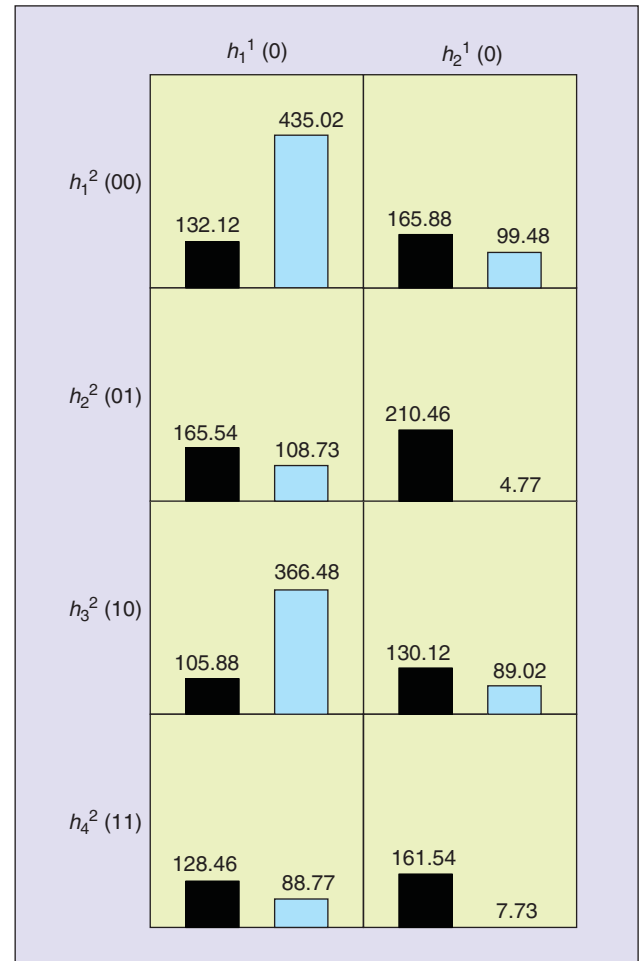


Fig. 2. A VarHAP decision table that is constructed from 1,200 case-control samples. Haplotypes in the first gene are obtained from one SNP whereas haplotypes in the second gene are inferred from two SNPs. The table consists of eight cells, where each cell represents a unique haplotype configuration. The left (black) bar in each cell represents the accumulative contribution from case samples whereas the right (pale blue) bar represents the accumulative contribution from control samples. The cells with haplotype configurations (h_1^1, h_1^2) , (h_1^1, h_2^2) , (h_2^1, h_2^2) , (h_2^1, h_3^2) , (h_1^1, h_4^2) , and (h_2^1, h_4^2) are labeled as predisposing haplotype configurations whereas the cells with haplotype configurations (h_1^1, h_1^2) and (h_1^1, h_3^2) are labeled as protective haplotype configurations.

Haplotypes may enable susceptible gene identification in complex diseases more effectively than individual SNPs.

superset of the previously specified SNPs, should also predict the same relationship. This implies that predisposing and protective haplotypes in a low-order model must be able to propagate into haplotypes in high-order models. For example, consider a single-gene problem with four possible SNPs: X1, X2, X3, and X4. If haplotypes in the model with SNPs (X2, X4) are related to disease susceptibility, haplotypes in the models with SNPs (X1, X2, X4), (X2, X3, X4), and (X1, X2, X3, X4) should

produce the same result. The haplotype propagation capability, which is a dichotomous criterion, can be determined from the evidence that the sign test p value and the prediction accuracy can be maintained throughout the process of increasing the model order. Again, in a situation where two or more models with different number of SNPs are equally good in terms of both prediction accuracy and haplotype propagation capability, the most parsimonious model will be the best model.

Table 1. Description of two-locus disease models.

Model	d_{22}	d_{21}	d_{20}	d_{12}	d_{11}	d_{10}	d_{02}	d_{01}	d_{00}	p_1	p_2	ϕ
Ep-1	ϕ	ϕ	0	ϕ	ϕ	0	0	0	0	0.210	0.210	0.707
Ep-2	ϕ	ϕ	0	0	0	0	0	0	0	0.600	0.199	0.778
Ep-3	ϕ	0	0	0	0	0	0	0	0	0.577	0.577	0.900
Ep-4	ϕ	ϕ	0	ϕ	0	0	ϕ	0	0	0.372	0.243	0.911
Ep-5	ϕ	ϕ	0	ϕ	0	0	0	0	0	0.349	0.349	0.799
Ep-6	0	ϕ	ϕ	ϕ	0	0	ϕ	0	0	0.190	0.190	1.000
Het-1	ψ	ψ	ϕ	ψ	ψ	ϕ	ϕ	ϕ	0	0.053	0.053	0.495
Het-2	ψ	ψ	ϕ	ϕ	ϕ	0	ϕ	ϕ	0	0.279	0.040	0.660
Het-3	ψ	ϕ	ϕ	ϕ	0	0	ϕ	0	0	0.194	0.194	1.000
S-1	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	0	0.052	0.052	0.522
S-2	1	1	1	ϕ	ϕ	0	ϕ	ϕ	0	0.228	0.045	0.574
S-3	1	1	ϕ	1	ϕ	0	ϕ	0	0	0.194	0.194	0.512

d_{ij} is the penetrance of a genotype carrying i disease alleles at locus 1 and j disease alleles at locus 2. p_1 is the frequency of the disease allele at locus 1 and p_2 is the frequency of the disease allele at locus 2. $\psi = 2\phi - \phi^2$.

Table 2. MDR, VarHAP, and FAMHAP results from the weak LD case study.

Two-Locus Model	MDR Prediction Accuracy (%)	VarHAP Prediction Accuracy (%)	Correct Model Identification Technique	Alternative Model Identification Technique
Ep-1	98.00	73.92	MDR(2), VarHAP(2), FAMHAP(2)	FAMHAP(2)
Ep-2	98.58	78.39	MDR(2), VarHAP(4), FAMHAP(2)	FAMHAP(2)
Ep-3	99.50	87.50	MDR(2), VarHAP(2), FAMHAP(2)	FAMHAP(2)
Ep-4	99.25	78.96	MDR(2), VarHAP(2), FAMHAP(2)	FAMHAP(2)
Ep-5	98.42	75.19	MDR(2), VarHAP(3), FAMHAP(2)	FAMHAP(2)
Ep-6	100.00	85.10	MDR(2), VarHAP(2), FAMHAP(2)	FAMHAP(2)
Het-1	93.75	73.29	MDR(2), VarHAP(2), FAMHAP(2)	FAMHAP(2)
Het-2	97.33	78.40	MDR(2), VarHAP(2), FAMHAP(2)	FAMHAP(2)
Het-3	100.00	84.40	MDR(2), VarHAP(2), FAMHAP(2)	FAMHAP(2)
S-1	94.00	72.98	MDR(2), VarHAP(2), FAMHAP(2)	FAMHAP(2)
S-2	97.58	79.81	MDR(2), VarHAP(2), FAMHAP(2)	FAMHAP(2)
S-3	96.75	79.15	MDR(2), VarHAP(2), FAMHAP(2)	FAMHAP(2)

A 10-fold cross-validation technique is used in MDR and VarHAP. The prediction accuracy is obtained for the identified principal interaction model. Estimated p values in FAMHAP results are equal to zero whereas sign test p values in MDR and VarHAP results are less than 0.001 in all two-locus problems. The number in each bracket denotes the order of the identified model (the number of SNPs in the model).

**A genotype with all homozygous sites
or a single heterozygous site can always
be phased into one pair of haplotypes.**

Data Sets

The performance of the proposed VarHAP technique is evaluated through benchmark trials. Twelve simulated data sets, which represent various gene–gene interaction phenomena, including epistasis and heterogeneity, are considered [14], [19]. Each data set contains 600 case samples and 600 control samples. Each sample consists of ten total SNPs from two genes, where five SNPs exist in each gene. All SNPs in control samples are in Hardy–Weinberg equilibrium [20]. Only one SNP from each gene is interacted with one another. The two-locus interaction models are illustrated in Table 1. The epistatic models Ep-1–Ep-6 and the heterogeneity models Het-1–Het-3 have been discussed by Neuman and Rice [21], who also provides examples of diseases for which these models may be applicable. The heterogeneity models S-1 and S-2 and the epistatic model S-3 have been investigated by Schork et al. [22]. From Table 1, if the frequency of the disease allele at a locus is greater than 0.5, the major allele is the disease allele. Otherwise, the minor allele is the disease allele. These interaction models describe disease susceptibility status in terms of penetrance. Penetrance of a genotype with a specific number of disease alleles is the probability that a subject with this genotype has the disease. The test data sets are simulated by a genomeSIM package [23] with the default setting. As a result, it is also possible to vary the LD pattern among SNPs in the same gene. This leads to two main case studies that need to be explored: strong LD and weak LD cases. In the strong LD case, the susceptibility-causative SNP in each gene and its two adjacent SNPs are in LD, where the D' value [24] is in the range of 0.80–0.95. In contrast, the D' value for each pairwise LD measurement between susceptibility-

causative SNP and its adjacent SNPs is in the range of 0.50–0.60 in the weak LD case. In the strong LD case, an interaction detection technique should be able to identify both the actual two-locus model that directly leads to disease susceptibility and other alternative models that consist of SNPs in strong LD patterns. The ability to detect these other models is important. This is because it is not always straightforward to identify SNPs that are responsible for disease susceptibility in real case–control interaction studies. In contrast, an interaction detection technique should narrow the search to the original two-locus model in weak LD case since it is the only usable model.

Results and Discussions

VarHAP is benchmarked against MDR and FAMHAP. Since the test data contains ten SNPs, all three techniques have to explore $2^{10} - 1 = 1,023$ possible SNP combination models. An initial investigation reveals that using minimum p value as the sole model selection criterion, FAMHAP reports a large number of models with the estimated p value equaling to zero. As a result, haplotype propagation capability is also implemented as an additional model selection criterion. Further, the parsimony criterion is also used when there is a tie between multiple models with different number of SNPs. The results from all three techniques in weak and strong LD case studies are summarized in Tables 2 and 3, respectively.

The prediction accuracy of MDR is higher than that of VarHAP in both case studies. This is because VarHAP uses contribution values that are obtained from inferred haplotypes instead of inferred diplotypes—pairs of haplotypes that together describe correct phases of given genotypes—to create decision

Table 3. MDR, VarHAP, and FAMHAP results from the strong LD case study.

Two-Locus Model	MDR Prediction Accuracy (%)	VarHAP Prediction Accuracy (%)	Correct Model Identification Technique	Alternative Model Identification Technique
Ep-1	98.00	73.92	MDR(2), VarHAP(2), FAMHAP(2)	VarHAP(2), FAMHAP(2)
Ep-2	98.58	77.02	MDR(2), VarHAP(4), FAMHAP(2)	VarHAP(4), FAMHAP(2)
Ep-3	99.50	87.50	MDR(2), VarHAP(2), FAMHAP(2)	FAMHAP(2)
Ep-4	99.25	78.96	MDR(2), VarHAP(2), FAMHAP(2)	VarHAP(2), FAMHAP(2)
Ep-5	98.42	75.87	MDR(2), VarHAP(3), FAMHAP(2)	VarHAP(3), FAMHAP(2)
Ep-6	100.00	85.10	MDR(2), VarHAP(2), FAMHAP(2)	FAMHAP(2)
Het-1	93.75	75.41	MDR(2), VarHAP(3), FAMHAP(2)	VarHAP(3), FAMHAP(2)
Het-2	97.33	78.40	MDR(2), VarHAP(2), FAMHAP(2)	VarHAP(2), FAMHAP(2)
Het-3	100.00	84.40	MDR(2), VarHAP(2), FAMHAP(2)	FAMHAP(2)
S-1	94.00	72.98	MDR(2), VarHAP(2), FAMHAP(2)	VarHAP(2), FAMHAP(2)
S-2	97.58	79.81	MDR(2), VarHAP(2), FAMHAP(2)	VarHAP(2), FAMHAP(2)
S-3	96.75	79.15	MDR(2), VarHAP(2), FAMHAP(2)	VarHAP(2), FAMHAP(2)

The explanation for how the results are obtained and displayed is the same as that given in Table 2.

MDR is a classifier-based technique that is capable of identifying the best genetic marker combination among possible markers for the separation between case and control samples.

rules. Consider a situation where disease susceptibility can be determined from a single SNP where the predisposing genotype is the homozygous variant. In other words, the disease susceptibility can be described by a recessive genetic model. MDR can easily classify the heterozygous and homozygous wild-type genotypes as protective genotypes. However, VarHAP would only correctly classify both homozygous genotypes since each genotype is made up from two copies of the same haplotype: two major alleles for the homozygous wild type and two minor alleles for the homozygous variant. VarHAP would partially misclassify samples with heterozygous genotype. This is because VarHAP identifies the major allele as the protective allele and the minor allele as the predisposing allele. To increase the prediction accuracy of VarHAP, it may be necessary to construct decision tables from diplotype information instead of haplotype contribution values. Nonetheless, this will also rapidly increase the dimensions of decision tables in VarHAP.

In the weak LD case study, both MDR and VarHAP are able to identify correct sets of SNPs that lead to disease susceptibility. On the other hand, FAMHAP reports both actual and alternative interaction models. This is undesirable since it would not be possible to further explain disease susceptibility from multiple candidate models in the absence of strong LD among SNPs. In other words, FAMHAP is quite sensitive in this situation. Further analysis reveals that MDR is marginally better than VarHAP in two epistasis problems: Ep-2 and Ep-5. MDR correctly identifies models that contain two SNPs, whereas the models located by VarHAP contain a few extra SNPs. Nonetheless, these two models identified by VarHAP are still useful to susceptibility explanation.

All three techniques are able to locate correct interaction models in the strong LD case study. However, only FAMHAP and VarHAP are capable of identifying alternative models. Since MDR suggests one candidate model for each fixed-number SNP set, it would not be possible for MDR to produce any alternative models. Recall that these alternative models are equally important since SNPs in the principal two-locus interaction model and SNPs from an alternative model are in strong LD. This implies that disease susceptibility can be explained using either the original interaction model or the alternative models. This disadvantage in MDR can be overcome if the cross-validation consistency criterion can be replaced by other decision criteria. In this case study, FAMHAP is marginally better than VarHAP in terms of identification of alternative models in three epistasis and heterogeneity problems: Ep-3, Ep-6, and Het-3. This means that FAMHAP is at its best when SNPs are in strong LD. Nonetheless, the overall results from both case studies suggest that VarHAP is the best technique. This is concluded from the fact that VarHAP does not report ambiguous results in weak LD case study and is also capable of producing alternative models in strong LD case study. This is crucial because it is impossible to know beforehand

whether susceptibility-causative SNPs are in weak or strong LD with other SNPs in real case-control interaction studies. In other words, a technique that performs satisfactorily in both weak and strong LD cases would have an advantage over a technique that functions well in only one scenario.

Conclusions

In this article, a nonparametric pattern recognition/classification technique for case-control, gene-gene interaction studies is presented. Instead of using direct genotype inputs in classification, inferred haplotypes, which are obtained through an expectation-maximization algorithm [15], are used as inputs. Each case-control sample contributes values derived from inferred haplotypes to decision tables that are constructed and tested for all possible gene-gene interaction models. This technique primarily uses prediction accuracy obtained from k -fold cross-validation as a means for identifying candidate SNPs, which are responsible for disease susceptibility. This technique also employs haplotype propagation capability as an additional criterion. If the selection procedure ends in a tie between two or more models with different number of SNPs, the most parsimonious model is reported as the interaction model. Since haplotypes with different lengths must be constructed during model identification, the proposed technique can be referred to as a VarHAP technique. VarHAP has been benchmarked against two interaction model detection programs namely MDR [9] and FAMHAP [14], [18] in 12 two-locus epistasis and heterogeneity problems [14], [19]. The results reveal that FAMHAP reports multiple ambiguous models in the presence of weak LD among input SNPs whereas MDR is not suitable for alternative interaction model identification when input SNPs are in strong LD. In contrast, VarHAP emerges as the most suitable technique in both situations involving weak and strong LD. Suggestions for further improvement of MDR and VarHAP are also included.

VarHAP, which is implemented in Java, and the simulated data sets used in the article are available upon request. In addition to the use of the genomeSIM package [23], the data sets can also be generated by a SNaP package [25]. Readers might also be interested in applying the techniques discussed in this article to examples of case-control data sets, which are available from the Wellcome Trust Case Control Consortium [26].

Acknowledgments

Anunchai Assawamakin was supported by the Thailand Research Fund (TRF) through the Royal Golden Jubilee Ph.D. program (Grant No. PHD/4.I.MU.45/C.1), the National Center for Genetic Engineering and Biotechnology (BIOTEC), and the National Science and Technology Development Agency (NSTDA). Nachol Chaiyaratana was supported by the Thailand Research Fund through the Research Career Development Grant (Grant No. RSA5180006). Chanin Limwongse was supported

by the Mahidol Research Grant. Prakarnkiat Youngkong was supported by the Commission on Higher Education (CHE). The authors acknowledge Scott M. Dudek at the Vanderbilt University for providing an access to the genomeSIM package.



Anunchai Assawamakin received his B.Sc. degree in pharmacy from Mahidol University and is a Ph.D. student at the Division of Molecular Genetics, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University. His research interests include human genetics and bioinformatics.



Nachol Chaiyaratana received his B.Eng. and Ph.D. degrees in control engineering from the University of Sheffield. He is an associate professor of electrical engineering at King Mongkut's University of Technology North Bangkok and an adjunct professor of genetic epidemiology at Mahidol University. His research interests include evolutionary computation, machine learning, and bioinformatics.



Chanin Limwongse received his M.D. degree from Mahidol University. He is the head of the division of molecular genetics at the Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University. His research interests include complex diseases and human genetics.



Saravudh Sinsomros received his B.Eng. degree in electrical engineering from Thammasat University and is an M.Eng. student at the Department of Electrical Engineering, Faculty of Engineering, King Mongkut's University of Technology North Bangkok. His research interests include machine learning and bioinformatics.



Pa-Thai Yenchitsomanus received his B.Sc. degree in medical technology, M.S. degree in biochemistry, and Ph.D. degree in human genetics from Chiang Mai University, Mahidol University, and Australian National University, respectively. He is a professor of medicine and the head of division of medical molecular biology at the Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University. His research interests include complex diseases and molecular genetics.



Prakarnkiat Youngkong received his B.S. and M.S. degrees in electrical engineering from Rensselaer Polytechnic Institute and University of California Los Angeles, respectively. He is a Ph.D. student at the Department of Electrical Engineering, Faculty of Engineering, King Mongkut's University of Technology North Bangkok. His current research interests include machine learning and bioinformatics.

Address for Correspondence: Nachol Chaiyaratana, Department of Electrical Engineering, Faculty of Engineering, King Mongkut's University of Technology North Bangkok, 1518 Piboolsongkram Road, Bangsue, Bangkok 10800, Thailand. E-mail: nchl@kmutnb.ac.th or n.chaiyaratana@gmail.com.

References

- [1] C. M. Lewis, "Genetic association studies: Design, analysis and interpretation," *Brief. Bioinform.*, vol. 3, pp. 146–153, June 2002.
- [2] G. Montana, "Statistical methods in genetics," *Brief. Bioinform.*, vol. 7, pp. 297–308, Sept. 2006.
- [3] A. G. Heidema, J. M. A. Boer, N. Nagelkerke, E. C. M. Mariman, D. L. van der A, and E. J. M. Feskens, "The challenge for genetic epidemiologists: How to analyze large numbers of SNPs in relation to complex diseases," *BMC Genet.*, vol. 7, p. 23, Apr. 2006.
- [4] N. Nagelkerke, J. Smits, S. Le Cessie, and H. van Houwelingen, "Testing goodness-of-fit of the logistic regression model in case-control studies using sample reweighting," *Stat. Med.*, vol. 24, pp. 121–130, Jan. 2005.
- [5] M. D. Ritchie, B. C. White, J. S. Parker, L. W. Hahn, and J. H. Moore, "Optimization of neural network architecture using genetic programming improves detection and modeling of gene-gene interactions in studies of human diseases," *BMC Bioinform.*, vol. 4, p. 28, July 2003.
- [6] J. Hoh, A. Wille, and J. Ott, "Trimming, weighting, and grouping SNPs in human case-control association studies," *Genome Res.*, vol. 11, pp. 2115–2119, Dec. 2001.
- [7] M. R. Nelson, S. L. R. Kardia, R. E. Ferrell, and C. F. Sing, "A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation," *Genome Res.*, vol. 11, pp. 458–470, Mar. 2001.
- [8] R. Culverhouse, T. Klein, and W. Shannon, "Detecting epistatic interactions contributing to quantitative traits," *Genet. Epidemiol.*, vol. 27, pp. 141–152, Sept. 2004.
- [9] L. W. Hahn, M. D. Ritchie, and J. H. Moore, "Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions," *Bioinformatics*, vol. 19, pp. 376–382, Feb. 2003.
- [10] K. L. Lunetta, L. B. Hayward, J. Segal, and P. van Eerdewegh, "Screening large-scale association study data: Exploiting interactions using random forests," *BMC Genet.*, vol. 5, p. 32, Dec. 2004.
- [11] A. Bureau, J. Dupuis, K. Falls, K. L. Lunetta, B. Hayward, T. P. Keith, and P. van Eerdewegh, "Identifying SNPs predictive of phenotype using random forests," *Genet. Epidemiol.*, vol. 28, pp. 171–182, Feb. 2005.
- [12] E. K. Silverman, "Haplotype thinking in lung disease," *Proc. Amer. Thorac. Soc.*, vol. 4, pp. 4–8, Jan. 2007.
- [13] P. C. Sham, F. V. Rijsdijk, J. Knight, A. Makoff, B. North, and D. Curtis, "Haplotype association analysis of discrete and continuous traits using mixture of regression models," *Behav. Genet.*, vol. 34, pp. 207–214, Mar. 2004.
- [14] T. Becker, J. Schumacher, S. Cichon, M. P. Baur, and M. Knapp, "Haplotype interaction analysis of unlinked regions," *Genet. Epidemiol.*, vol. 29, pp. 313–322, Dec. 2005.
- [15] L. Excoffier and M. Slatkin, "Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population," *Mol. Biol. Evol.*, vol. 12, pp. 921–927, Sept. 1995.
- [16] M. Stephens, N. J. Smith, and P. Donnelly, "A new statistical method for haplotype reconstruction from population data," *Amer. J. Hum. Genet.*, vol. 68, pp. 978–989, Apr. 2001.
- [17] T. Niu, Z. S. Qin, X. Xu, and J. S. Liu, "Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms," *Amer. J. Hum. Genet.*, vol. 70, pp. 157–169, Jan. 2002.
- [18] T. Becker and M. Knapp, "A powerful strategy to account for multiple testing in the context of haplotype analysis," *Amer. J. Hum. Genet.*, vol. 75, pp. 561–570, Oct. 2004.
- [19] M. Knapp, S. A. Seuchter, and M. P. Baur, "Two-locus disease models with two marker loci: The power of affected-sib-pair tests," *Amer. J. Hum. Genet.*, vol. 55, pp. 1030–1041, Nov. 1994.
- [20] G. H. Hardy, "Mendelian proportions in a mixed population," *Science*, vol. 28, pp. 49–50, July 1908.
- [21] R. J. Neuman and J. P. Rice, "Two-locus models of disease," *Genet. Epidemiol.*, vol. 9, no. 5, pp. 347–365, 1992.
- [22] N. J. Schork, M. Boehnke, J. D. Terwilliger, and J. Ott, "Two-trait-locus linkage analysis: A powerful strategy for mapping complex genetic traits," *Amer. J. Hum. Genet.*, vol. 53, pp. 1127–1136, Nov. 1993.
- [23] S. M. Dudek, A. A. Motsinger, D. R. Velez, S. M. Williams, and M. D. Ritchie, "Data simulation software for whole-genome association and other studies in human genetics," in *Pacific Symposium on Biocomputing 2006*, R. B. Altman, A. K. Dunker, L. Hunter, T. Murray, and T. E. Klein, Eds. Singapore: World Scientific, 2006, pp. 499–510.
- [24] R. C. Lewontin, "On measures of gametic disequilibrium," *Genetics*, vol. 120, pp. 849–852, Nov. 1988.
- [25] M. Nothnagel, "Simulation of LD block-structured SNP haplotype data and its use for the analysis of case-control data by supervised learning methods," *Amer. J. Hum. Genet.*, vol. 71, (Suppl), Article No. A2363, Oct. 2002.
- [26] The Wellcome Trust Case Control Consortium, "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls," *Nature*, vol. 447, pp. 661–678, June 2007.

DATABASE IN BRIEF

Thailand Mutation and Variation Database (ThaiMUT)

Uttapong Ruangrit^{1,7,†}, Metawee Srikumool^{2,†}, Anunchai Assawamakin^{1,3,†}, Chumpol Ngamphiw¹, Suparat Chuechote¹, Vilasinee Thaiprasarnsup², Gallissara Agavatpanitch¹, Ekawat Pasomsab⁴, Pa-thai Yenchitsomanus³, Surakameth Mahasirimongkol⁵, Wasun Chantratita⁴, Prasit Palittapongarnpim^{1,6}, Bunyarit Uyyanonvara⁷, Chanin Limwongse^{3,*}, and Sissades Tongsim^{1,*}

¹*Biostatistics and Informatics Laboratory, Genome Institute, National Center for Genetic Engineering and Biotechnology (BIOTEC), Pathumtani, Thailand;* ²*Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand;* ³*Division of Molecular Genetics, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand;* ⁴*Department of Pathology, Faculty of Medicine, Ramathibodhi Hospital, Mahidol University, Bangkok, Thailand;* ⁵*Department of Medical Sciences, Ministry of Public Health, Center for International Cooperation, Foreign Affairs Section, Nontaburi, Thailand;* ⁶*Department of Microbiology, Faculty of Science Mahidol University, Bangkok, Thailand;* ⁷*Department of Information and Computer Technology, Sirindhorn International Institute of Technology (SIIT), Pathumtani, Thailand.*

[†] Equally contributing first authors

*Correspondence to: Sissades Tongsim, Biostatistics and Informatics Laboratory, Genome Institute, National Center for Genetic Engineering and Biotechnology, 113 Thailand Science Park, Pathumtani 12120, THAILAND
Tel.: +66-2-564-6700 ext. 5551, Fax: +66-2-564-6584; E-mail: sissades@biotec.or.th, siclw@mahidol.ac.th

Communicated by Alastair F. Brown

With the completion of the human genome project, novel sequencing and genotyping technologies had been utilized to detect mutations. Such mutations have continually been produced at exponential rate by researchers in various communities. Based on the population's mutation spectra, occurrences of Mendelian diseases are different across ethnic groups. A proportion of Mendelian diseases can be observed in some countries at higher rates than others. Recognizing the importance of mutation effects in Thailand, we established a National and Ethnic Mutation Database (NEMDB) for Thai people. This database, named Thailand Mutation and Variation database (ThaiMUT), offers a web-based access to genetic mutation and variation information in Thai population. This NEMDB initiative is an important informatics tool for both research and clinical purposes to retrieve and deposit human variation data. The mutation data cataloged in ThaiMUT database were derived from journal articles available in PubMed and local publications. In addition to collected mutation data, ThaiMUT also records genetic polymorphisms located in drug related genes. ThaiMUT could then provide useful information for clinical mutation screening services for Mendelian diseases and pharmacogenomic researches. ThaiMUT can be publicly accessed from <http://gi.biotec.or.th/thaimut>. © 2008 Wiley-Liss, Inc.

Received 18 July 2007; accepted revised manuscript 18 February 2008.

KEY WORDS: Thailand's NEMBD; database; SNP; genetic variation

INTRODUCTION

The existence of human mutation databases, such as Human Gene Mutation Database (HGMD; Cooper, et al., 1998; Krawczak, et al., 2000) and Locus-Specific Database (LSDB; Claustres, et al., 2002), has influenced the occurrence of regional initiatives in the discovering and cataloging of genetic mutations and variations. The importance of collecting mutations and variations, which affect genetic diseases in various ethnic communities, was also raised during a HUGO meeting as the Mutation Database Initiative (MDI; Cotton, 2000). Complex interplay among population history, mating patterns and natural selection resulted in an accumulation of population specific mutation which underlies many common genetic diseases in a particular population. Discovery of novel mutations in an ethnic-specific population and storing them are consequential to disease detection, mutation screening and genetic counseling. Different ethnics exhibit varying mutation spectrums. The varying spectrums could be the cause of the different occurrences of Mendelian diseases across populations. As an example, hemoglobin genes reveal various mutation spectrums in different populations and even within the same population (Old, et al., 2001; Sirichotiyakul, et al., 2003). It has been known that β -Thalassemia is highly prevalent in Southeast Asia and Mediterranean while Sickle Cell Anemia is prevalent in African descendents (Clark and Thein, 2004). Armed with the mutation spectrum information in a given population, researchers can devise an appropriate scheme for mutation detection in that particular population. Such mutation information will also be beneficial for carrier screening in Mendelian diseases. Consequently, practitioners and researchers will benefit from a well-maintained mutation database.

Ethnic specific mutation databases are necessary in the studies of comprehensive demographic history and patterns of migration flow (Patrinos, 2006). Comparison of available spectrums of genetic mutations across populations could facilitate genotype-phenotype correlation studies. Apart from the collection of ethnic specific mutations, public genotypic databases had revealed more than 10 million single nucleotide polymorphisms (SNPs) in human genome in different populations (Smigielski, et al., 2000). However, validating the existence of these SNPs in a certain population and capturing their frequencies are required to determine the population significance of these SNPs (Thorisson, et al., 2005). Upon completion of these tasks, researchers would be able to further investigate predisposing genetic factors influencing diseases in their specific population.

This system is regularly maintained and updated to promote the utilization of this database in the Thai human genetic society. It allows direct personal submission, which must go through a manual assessment by the Thai human genetic research community before publishing the results on ThaiMUT. This establishment was prepared to ensure that the data and web application services available on this website will be beneficial not only to the local research community but also research communities outside Thailand.

ETHNIC BACKGROUND

Thailand covers an area of 514,000 km² in the center of the Southeast Asian peninsula. It is bordered by Myanmar (Burma), Lao People's Democratic Republic (Laos), Cambodia and Malaysia, and has 2,420 km of coastline on the Gulf of Thailand and the Andaman Sea. Thailand stretches 1,650 kilometers from north to south, and 780 kilometers from east to west at its widest part (<http://www.un.or.th/thailand/geography.html>). The estimated population is 64 million, of which approximately 9.3 million live in Bangkok, the capital city, and its vicinity. The official language of the country is called "Thai" for which 94% of population use as their first language. Four major dialects of the Thai language are the dialects used in the central, northern, southern and northeastern regions. Northeastern dialect is closely related to the Lao language. In the four southern most provinces of Pattani, Satun, Yala and Naratiwat situated near the Malaysian border, majority of the population there is Muslim speaking "Pattani" Malay. In the mountainous area of the northern region, there are approximately 525,000 highland people or hill tribes who speak distinct languages. Ten to fifteen percent of populations have Chinese origin due to steady flows of immigration from China to Thailand during 1850 toward the end of the World War II. Thus, the Chinese population in Thailand was established as commerce and artisan communities throughout the country (<http://www.un.or.th/thailand/population.html>).

In terms of ethnic-specific Mendelian disease backgrounds, Thalassemia is the most common genetic disease in Thailand. Mutation carriers were estimated at 30–40% of the Thai population. The high prevalence of Thalassemia carriers results in more than 12,000 new cases of severely affected births annually in Thailand (Lagampan, et al., 2004). For other genetic related diseases, cancer is also a major health problem and has been the most common

cause of death since 1999 (Ministry of Public Health, 2004). In Thai men, liver cancer is the most common disease followed by lung cancer. On the other hand, cervix and breast cancer are the two top cancer types prevalent in Thai women (Sriplung, et al., 2003).

SYSTEM DESIGN AND IMPLEMENTATION

The design of ThaiMUT is based on a three-tier architecture model (client, application server and database). Figure 1 depicts the overall architectural design of ThaiMUT. On the client layer, we use PHP-based scripts as a CGI on the Apache web server to render the graphical web-interface. Mutations, SNPs and other human genome reference information are tabulated in a MySQL relational database. To simplify the SQL complexity, we created a web interface using CGI scripts in the application server to compose the queries on behalf of the users. Users will only need to enter a keyword such as gene name, disease name then the web server will transform the requests to query language (SQL) to get the pertinent data from MySQL.

The content and structure of ThaiMUT follows guidelines given by <http://www.hgvs.org> and by Scriver's recommendation (Scriver, et al., 1999). ThaiMUT was constructed in such a way that mutations/SNPs can be incorporated into and queried from the database. We began by incorporating large number of Thai mutation reports excerpted from literatures in PubMed to our MySQL database. Some of the mutations not in PubMed were collected from local publications and personal communication with researchers. In addition to mutations, validated SNPs in Thai population were also cataloged. These SNPs were obtained from (Mahasirimongkol, et al., 2006) which compared amongst Thais and Northeast Asian populations (Chinese and Japanese) their allele frequencies and linkage disequilibrium (LD) patterns from 188 drug related genes. This data set was obtained by genotyping SNPs 280 individuals from 4 major geographical regions in Thailand. Allele frequencies from these drug-related SNPs were systematically made available for the first time through the ThaiMUT database. Flanking sequences of these SNPs were compared against the latest build of reference sequence (RefSeq) obtained from NCBI database (Pruitt, et al., 2007). They can be visualized along with SNPs from other populations in various public domain databases namely dbSNP (Smigielski, et al., 2000), JSNP (Hirakawa, et al., 2002) and HapMap (Thorisson, et al., 2005).

Related genomics and genetic information are supplied along with the mutations and SNPs, for example: locus IDs, gene names, OMIM associated Mendelian diseases, nomenclatures, allele frequencies and their publication references with link-outs to PubMed. In order to assist scientists to compare SNPs across different populations, ThaiMUT integrated the latest public domain SNP databases. Therefore, SNP maps from each database can be presented all at once along with Thai SNPs. For graphical viewing, a W3 standard Scalable Vector Graphics (SVG; <http://www.adobe.com/svg>) was adopted; users can visualize the location and characteristic of selected SNPs from a comparative view of different populations. An apache web-server was set up to interact with the underlying database and to provide web-based graphical output to users. Users can query for both mutations and SNPs via regular web search box or intuitively select a locus from ideogram view to get the required information.

To regularly maintain data accuracy and to alert a curator to update the database, ThaiMUT features a direct submission of unpublished data from researchers. However, that data will be marked as unpublished and will be unmarked later when publication is officially made. The submission form complies with the guideline given by the Human Genome Variation Society (HGVS; <http://www.hgvs.org>). Most novel mutations and SNPs are expected to be submitted by members of the Thai Genetic Society who share common interests in Mendelian diseases and genetic epidemiological studies.

DATABASE ACCESS

ThaiMUT database is publicly available from <http://gi.biotech.or.th/thaimut>. Most popular web browsers, e.g., Internet Explorer (IE6 on Windows XP and IE7 on Vista) and Firefox, should be able to view mutation/variation contents in ThaiMUT. To access ThaiMUT, the JavaScript feature must be enabled. By default, web browsers enable JavaScript but disable other security-prone features such as ActiveX and pop-ups blocking. Since ThaiMUT relies only on JavaScript, most users should be able to access the web site. To graphically visualize SNPs on genes, Internet Explorer users are required to install SVG (Scalable Vector Graphic) viewer plug-in made available by Adobe software (<http://www.adobe.com/svg>).

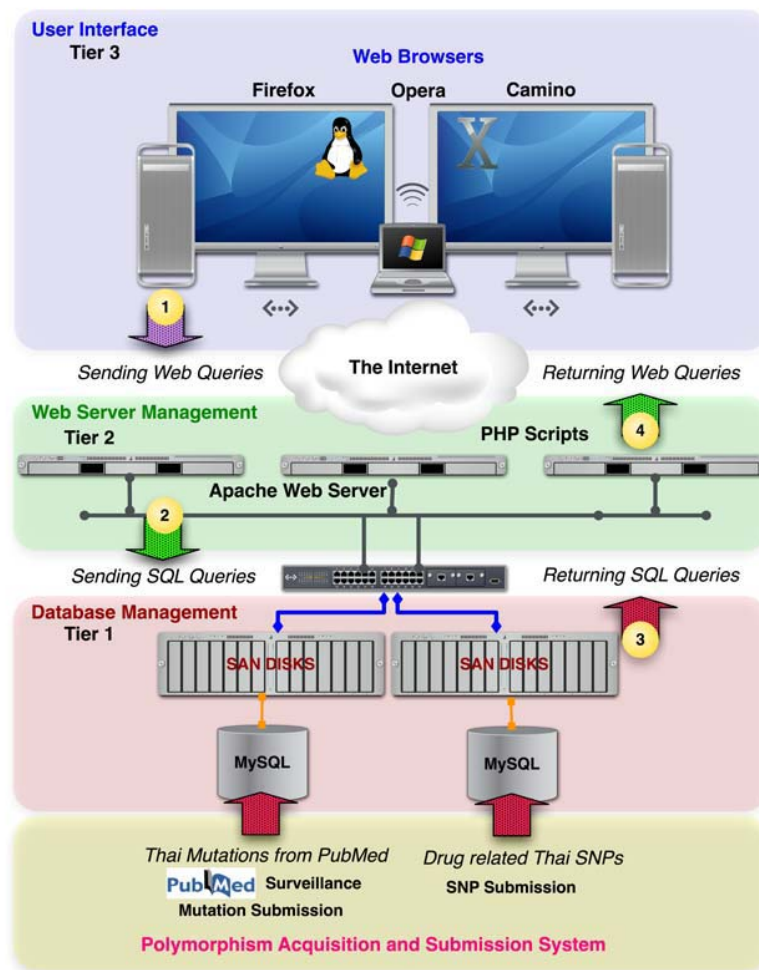


Figure 1. Implementation of ThaiMUT database system: 1) Users search for mutation/variation via web interface implemented using PHP scripts 2) Apache web server receives search requests and converts them to SQL commands issued to MySQL database 3) MySQL returns queried results back to Apache server 4) PHP scripts process raw outputs and transform them to html or SVG results that web browsers can render.

QUERYING THE DATABASE

For convenience and efficiency, both mutation and variation information in ThaiMUT must be queried through intuitive web interface similar to other genomic database interfaces. We have made the interface using Ext JS2.0 (extjs.com) which is a JavaScript library used in construction of many web applications. Four main feature search schemes (described below) are presented as tabs in the control frame of ThaiMUT. The database summary such as number, types of mutations is displayed in ThaiMUT welcome page. Figure 2 illustrates the ThaiMUT interface.

1. **Basic Search:** This is the most recommended feature search to explore mutations and variations in ThaiMUT. The search box can accept any string of gene name or locus ID, disease name, chromosome

number, OMIM number, nomenclature, title and author of reference article. The basic search is separately provided for mutation as well as variation.

2. **Advanced Search:** Similar to basic search, this feature accepts strings of all types. It also offers the search by mutation types and a range of years that an article was published. These queries can also be combined, e.g., user can search by gene name *and/or* author of the article at the same time.
3. **Alphabetical Search:** This feature arranges gene symbols or names in alphabetical order. Clicking on each alphabet, i.e., the initial of the gene name, will narrow the search space. All available genes with mutations found will be displayed.
4. **Chromosomal Search:** This feature offers a list of 22 autosomes plus the X and Y chromosomes. Users can directly jump to a chromosome of interest and visualize where mutations/variations are on the chromosome. Users can click on one of the colored boxes indicating where mutations/variations occurred on the chromosome.



Figure 2. ThaiMUT web interface to the database: 1) Two tabs for basic search; one for mutation and the other for variation 2) Mutation advanced search 3) Search mutation by gene name ordered in alphabetical 4) Mutation search by chromosome, 5) Submission system and 6) Online help and contact.

DATA SUBMISSION

ThaiMUT encourages the human genetic research community both in Thailand and other countries to submit their discovered polymorphisms to the database. Users can click on the submission tab (see Fig. 2) to access the submission system. To submit either mutation(s) or variation(s), users are required to register their emails and contact information first. ThaiMUT will issue a password as in an alert box which can be copied and used to login to use the forms. The forms are provided separately for mutation and variation. In the design of ThaiMUT database, other types of mutations or variations such as STRs, microsatellites, minisatellites can be submitted and stored to the database. The submission form strictly follows proposed mutation entry and quality control form

(<http://www.hgvs.org/entry.html>), which is a recommendation by the MDI/HGVS. For security reasons, all submitted data would not be disclosed for public viewing unless they have been published or verified by a group of committees (e.g., appointed by the Thai Human Genetic Society), who are responsible for quality control of submission entries.

RESULTS AND DISCUSSION

Using the described design, the ThaiMUT database is now available for public access. There are now 119 causative genes identified in Thais registered in the database. A total of 589 mutations are listed, which can be categorized as 507 substitutions, 59 deletion, 20 insertions and 3 indels. Clearly, this is a much smaller collection than those in some other national and international mutation databases (George, et al., 2007). However, the ThaiMUT database is another example of a collaborative endeavor to gather and make available to the public the crucial resource regarding endemic genotyping information. This effort will eventually contain most of the identified causative mutation as well as SNP variants found in populations which are descendants of Thai ancestry. Currently, a number of causative mutations described for Thais can be readily searched for and used by this database. Moreover, attempts were made to expand the number of SNPs in the registry to maximize the opportunity for the future research in pharmacogenomics and disease association. Clinicians searching for information about a given genetic disease can easily browse through the ThaiMUT web interface to find gene information, mutation data, as well as information on local laboratories performing a particular genetic test. A comprehensive list of genes whose mutations are available in ThaiMUT is also displayed. This can indirectly serve as a national diagnostic laboratory service directory.

At present, data on 1312 SNPs from 188 drug-related gene loci have been deposited into the ThaiMUT database. Both common and rare alleles were included in this dataset. The study utilized tagSNPs, a minimum number of SNPs representing other SNPs in high linkage disequilibrium region, and common SNPs in these drug related gene loci to determine genetic distance between Thais and East Asian populations. Such tagSNPs were derived from the HapMap East Asian data set whose SNPs collocate with those of Thais'. Majority of these gene loci reveals similar haplotype spectrum and frequencies between Thai and East Asian populations suggesting that the tagSNPs could be informative for an indirect association study in these drug-related genes loci. However, for the minority of these loci that have different haplotype spectrum and frequencies, a small set of samples from a studied population should be used for SNPs discovery and tagSNPs selection before genotyping candidate polymorphisms for indirect association study. Underlying similarity information within these drug-related genes was deposited in the database for further utilization. Heterozygosity from multiple data sets in a certain population will be helpful in reducing the genotyping cost. In other words, we can avoid genotyping of uninformative (non-polymorphic) loci in the population. SNP discovery of a given genomic region is also important. Knowing the extent of population sequence variation diversity will be tremendously useful for the design of genetic association study (Tocharoentanaphol, et al., 2008).

With increasing amounts of genotype data in the future, the ThaiMUT database is certain to hold a central place for Thai genetic researchers for disease association as well as pharmacogenomics studies. Information contained in the mutation database can benefit both researchers and clinicians in the following ways. First, the data serves as a local reference for potential functionally significant variants that may be rare and have not been reported in other populations. This view justifies further functional studies of those particular variants rather than disregarding them as rare non-pathogenic polymorphisms. Second, researchers who are interested in a particular disorder can have an overview of spectrum of mutation found in Thailand as compared to those found elsewhere. This may give rise to a genetic mutation screening if a specific mutation is commonly found within our population. Third, using this database as an information portal, investigators having patients with similar genetic disorders or interested in similar genes can contact one another, share their resources and eventually undertake collaborative research projects. Moreover, this would potentially lead to a better cost-effective use of scarce resources for both physicians in Thailand and those neighboring countries interested in rare genetic diseases.

ACKNOWLEDGMENTS

The authors would like to thank all researchers who published mutations and variations that have been catalogued in ThaiMUT. We are also indebted to doctors from various hospitals in Thailand for sample collections as well as to their patients for their invaluable contributions. We would like to thank in advance the Thai Genetics Society who has kindly agreed to provide us with their mutation data in the future. We also thank the National Center for Genetic Engineering and Biotechnology (BIOTEC) for hosting the database and for supporting this research through the ThaiSNP Database Project (Grant No. BT-B-02-NT-BC-4728) and the Consolidation of BIOTEC Genome Database Project (Grant No. BT-B-02-IT-BC-4932). We also acknowledge the Thailand Research Fund (TRF) through the Royal Golden Jubilee Ph.D. Program (Grant No. PHD/4.I.MU.45/C.1) for supporting A.A. Finally we acknowledge the Thailand Center for Excellence in Life Sciences for funding support to S.M., E.P., and W.C.

REFERENCES

- Clark BE, Thein SL. 2004. Molecular diagnosis of haemoglobin disorders. *Clin Lab Haematol* 26(3):159-76.
- Claustres M, Horaitis O, Vanevski M, Cotton RG. 2002. Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. *Genome Res* 12(5):680-8.
- Cooper DN, Ball EV, Krawczak M. 1998. The human gene mutation database. *Nucleic Acids Res* 26(1):285-7.
- Cotton RG. 2000. Progress of the HUGO mutation database initiative: a brief introduction to the human mutation MDI special issue. *Hum Mutat* 15(1):4-6.
- George RA, Smith TD, Callaghan S, Hardman L, Pierides C, Horaitis O, Wouters MA, Cotton RG. 2007. General mutation databases: analysis and review. *J Med Genet* 45(2):65-70.
- Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, Nakamura Y. 2002. JSNP: a database of common gene variations in the Japanese population. *Nucleic Acids Res* 30(1):158-62.
- Krawczak M, Ball EV, Fenton I, Stenson PD, Abeyasinghe S, Thomas N, Cooper DN. 2000. Human gene mutation database a biomedical information and research resource. *Hum Mutat* 15(1):45-51.
- Lagampan S, Lapvongwattana P, Tharapan C, Nonthikorn J. 2004. Health Belief Model Teaching Program For Thalassemia Education in High School Students. *Chula Med Journal*.
- Mahasirimongkol S, Chantratita W, Promso S, Pasomsab E, Jinawath N, Jongjaroenprasert W, Lulitanond V, Krittayapoositpot P, Tongsima S, Sawanpanyalert P and others. 2006. Similarity of the allele frequency and linkage disequilibrium pattern of single nucleotide polymorphisms in drug-related gene loci between Thai and northern East Asian populations: implications for tagging SNP selection in Thais. *J Hum Genet* 51(10):896-904.
- Ministry of Public Health T. 2004. Public Health Statistics A.D. 2003.
- Old JM, Khan SN, Verma I, Fucharoen S, Kleanthous M, Ioannou P, Kotea N, Fisher C, Riazuddin S, Saxena R and others. 2001. A multi-center study in order to further define the molecular basis of beta-thalassemia in Thailand, Pakistan, Sri Lanka, Mauritius, Syria, and India, and to develop a simple molecular diagnostic strategy by amplification refractory mutation system-polymerase chain reaction. *Hemoglobin* 25(4):397-407.
- Patrinos GP. 2006. National and ethnic mutation databases: recording populations' genography. *Hum Mutat* 27(9):879-87.
- Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35(Database issue):D61-5.
- Sriver CR, Nowacki PM, Lehvaslaiho H. 1999. Guidelines and recommendations for content, structure, and deployment of mutation databases. *Hum Mutat* 13(5):344-50.

- Sirichotiyakul S, Saetung R, Sanguanserm Sri T. 2003. Analysis of beta-thalassemia mutations in northern Thailand using an automated fluorescence DNA sequencing technique. *Hemoglobin* 27(2):89-95.
- Smigielski EM, Sirotkin K, Ward M, Sherry ST. 2000. dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res* 28(1):352-5.
- Sriplung H, Sontipong S, Martin N. 2003. *Cancer in Thailand 3 (1995-1997)*(Bangkok Medical Publisher).
- Thorisson GA, Smith AV, Krishnan L, Stein LD. 2005. The International HapMap Project Web site. *Genome Res* 15(11):1592-3.
- Tocharoentanaphol C, Promso S, Zelenika D, Lowhnoo T, Tongsima S, Sura T, Chantratita W, Matsuda F, Mooney S, Sakuntabhai A. 2008. Evaluation of resequencing on number of tag SNPs of 13 atherosclerosis-related genes in Thai population. *J Hum Genet* 53(1):74-86.



Prothrombin haplotype associated with kidney stone disease

Journal:	<i>Journal of the American Society of Nephrology</i>
Manuscript ID:	JASN-2010-01-0111
Manuscript Type:	Original Article - Basic Research
Date Submitted by the Author:	26-Jan-2010
Complete List of Authors:	Rungroj, Nanyawan; Faculty of Medicine Siriraj Hospital, Department of Research and Development Sritippayawan, Suchai; Faculty of Medicine Siriraj Hospital, Medicine Thongnoppakhun, Wanna; Faculty of Medicine Siriraj Hospital, Medical Molecular Biology Paemanee, Atchara; National Center for Genetic Engineering and Biotechnology (BIOTEC), Medical Biotechnology SAWASDEE, Nunghathai; Mahidol University, Medical Molecular Biology Nettuwakul, Choochai; Faculty of Medicine Siriraj Hospital, Medical Molecular Biology Sudtachat, Nirinya; National Center for Genetic Engineering and Biotechnology (BIOTEC), Medical Biotechnology Ungsupravate, Duangporn; Faculty of Medicine Siriraj Hospital, Medical Molecular Biology Praihirunkit, Pairao; Mahidol University, Institute of Molecular Biosciences Chuawattana, Duangporn; Faculty of Medicine Siriraj Hospital, Medicine Akkarapatumwong, Varaporn; Mahidol University, Institute of Molecular Biosciences Sombat, Borvornpadungkitti; Khon Kaen Hospital, Surgery Susaengrat, Wattanachai; Khon Kaen Hospital, Medicine Vasuvattakul, Somkiat; Faculty of Medicine Siriraj Hospital, Medicine YENCHITSOMANUS, Pa-thai; Faculty of Medicine Siriraj Hospital, Medical Molecular Biology
Keywords:	kidney stones, genetic renal disease, molecular genetics

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



For Peer Review

Prothrombin haplotype associated with kidney stone disease

Nanyawan Rungroj,^{1,2} Suchai Sritippayawan,³ Wanna Thongnoppakhun,² Atchara Paemanee,^{1,4} Nunghathai Sawasdee,¹ Choochai Nettuwakul,¹ Nirinya Sudtachat,^{1,4} Duangporn Ungsupravate,¹ Pairao Praihirunkit,⁵ Duangporn Chuawattana,³ Varaporn Akkarapatumwong,⁵ Sombat Borvornpadungkitti,⁶ Wattanachai Susaengrat,⁶ Somkiat Vasuvattakul,³ Prida Malasit,^{1,4} Pa-thai Yenchitsomanus^{1,4}

¹Division of Medical Molecular Biology, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

²Division of Molecular Genetics, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

³Division of Nephrology, Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

⁴Medical Biotechnology Unit, National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Bangkok, Thailand

⁵Institute of Molecular Biosciences, Mahidol University, Salaya, Nakorn Pathom, Thailand

⁶Khon Kaen Hospital, Khon Kaen, Thailand

Running title: *F2* and kidney stone disease

Word count: 172 words (abstract) and 2899 words (text)

Corresponding author:

Pa-thai Yenchitsomanus, Ph.D.

Division of Medical Molecular Biology,
Department of Research and Development,
Faculty of Medicine Siriraj Hospital,
Mahidol University,

Bangkok 10700,

THAILAND

Tel: (+66) 2418 4793 Fax: (+66) 2418 4793

E-mail: grpye@mahidol.ac.th, ptyench@gmail.com

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

Genetic and environmental factors may involve in pathogenesis of kidney stone disease. To determine genetic variations associated with the disease, we performed a case-control association study in 112 subjects each of patient and control groups by genotyping 67 single nucleotide polymorphisms (SNPs) within 8 genes including *TFF1*, *S100A8*, *S100A9*, *S100A12*, *AMBP*, *SPP1*, *UMOD*, and *F2*, encoding trefoil factor 1, calgranulin (A, B, and C), bikunin, osteopontin, Tamm-Horsfall protein, and prothrombin, respectively. Significant differences between the case and control groups in allele and genotype frequencies of 8 SNPs in *F2* were found while those of the remaining 7 genes were not. Interestingly, frequencies of two *F2* haplotypes were significantly different between the case and control groups, one haplotype (TGCCGCCGCG) associated with increased kidney stone risk ($P = 0.0013$, OR 1.612, 95% CI 1.203-2.160) and the other (CGTTCCGCTA) with reduced disease risk ($P = 0.0007$, OR 0.464, 95% CI 0.296- 0.727); these significant differences were maintained after Bonferroni's correction. These findings indicate that variations of *F2* influence susceptibility or protection to kidney stone disease.

INTRODUCTION

Kidney stone is a common urological disease characterized by the presence of hardened mineral deposits in the kidney. This disease is a worldwide health problem occurring in all geographical and ethnic groups^{1,2} with increased prevalence in some Western populations^{3,4} and its recurrences with relapse rates are 50% in 5-10 years and 75% in 20 years.^{5,6} The causes of kidney stone are heterogeneous, ranging from monogenic defect to complex interaction between genetic and environmental factors.⁷ Previous studies have shown that genetic variations of several human genes involved in kidney stone associated hypercalciuria. Some of these genes are calcitonin receptor (CTR)⁸, vitamin D receptor (VDR)⁹, urokinase¹⁰, interleukin (IL-1 β and IL-Ra)^{11,12}, E-cadherin¹³, androgen-oestrogen receptor¹⁴, vascular endothelial receptor growth factor (VEGF)¹⁵, and calcium-sensing receptor (CaSR)¹⁶. However, contributions of these genes as the cause of kidney stone are small and the major cause of the disease is still unknown.

Approximately 80% of kidney stones contain calcium oxalate (CaOx) and/or phosphate (CaP), whereas other forms, such as struvite, uric acid and cystine stones, are less common.⁷ Calcium stone formation is a multi-step process consisting of crystal nucleation, growth, aggregation, and retention, though the physiochemical mechanisms of these processes are still in dispute. Defects in the amount and/or function of urinary stone-inhibitor proteins have been proposed to be involved in kidney stone formation.¹⁷⁻¹⁹ Important urinary stone-inhibitor proteins that have been described include trefoil factor 1, calgranulin (A, B, and C), bikunin, osteopontin, Tamm-Horsfall protein, and urinary prothrombin fragment 1. These proteins might be found as matrix proteins in human kidney stone and were shown to influence the formation of kidney stone containing calcium oxalate and/or calcium phosphate. They can inhibit kidney stone formation at different stages by coating the surface of growing calcium crystals or by forming complex with calcium and oxalate.^{20,21} Trefoil factor 1^{19,22}— a novel potent inhibitor, and calgranulin²³— a member of S100 family of calcium binding proteins, inhibits CaOx crystal growth and aggregation. Bikunin, a light chain of inter-alpha-trypsin inhibitor present in human

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

urine, was found to retard crystal nucleation and inhibit crystal aggregation.²⁴ Osteopontin, a major component in the urinary stone matrix, inhibits nucleation, growth, and aggregation of CaOx crystals and also reduces the binding of the crystals to the renal epithelial cells *in vitro*.^{25,26} Tamm-Horsfall protein, one of the most abundant proteins in urine, is capable of either an inhibitor or a promoter of crystal formation depending on its molecular properties.²⁷ Urinary prothrombin fragment 1, a fragment of prothrombin that is incorporated into human stones and crystals induced in urine²⁸, is a potent inhibitor of CaOx crystal growth, aggregation, and nucleation *in vitro*.²⁹⁻³¹

Kidney stone is an important public health problem in the northeastern (NE) population of Thailand.³²⁻³⁴ The etiology and pathogenesis of kidney stone in this population is still obscure. However, it is most likely to be unique in terms of its etiology and pathogenesis from what was reported in other ethnic groups because it is not associated with the conditions of increased urinary solutes, which are known to be stone promoters, such as hypercalciuria, hyperoxaluria, and hyperuricosuria.^{35,36} We have recently reported evidence suggesting a genetic contribution to kidney stone in the NE Thai population as it was found to have familial aggregation with a high relative risk ($\lambda_R = 3.18$) among members of the affected families.³⁷ The study of genetic polymorphisms, such as single nucleotide polymorphism (SNP), has come into view as a tool for identifying genes associated with many diseases. To further investigate genetic factor contributing to kidney stone formation in the NE Thai population, we conducted a case-control association study by genotyping 67 SNPs distributed within and flanking 8 candidate genes including *TFF1*, *S100A8*, *S100A9*, *S100A12*, *AMBP*, *SPPI1*, *UMOD*, and *F2*, encoding trefoil factor 1, calgranulins (A, B, and C), bikunin, osteopontin, Tamm-Horsfall protein, and urinary prothrombin fragment 1, respectively. In this paper, we firstly report association between *F2* haplotype and kidney stone disease.

RESULTS

Patients and normal control subjects

A group of 112 patients with kidney stone (mean age 48.26 ± 11.69 ; age range of 22-80 years) and a group of 112 normal control subjects with no history of kidney stone (mean age 48.01 ± 12.87 ; age range of 22-80 years) were selected for the study. Clinical characteristics of the patients are presented in Table 1.

Genotyping

A total of 67 SNPs in 8 candidate genes were selected for genotyping in this study (Supplementary Table S1). Majority of them, 65 SNPs, were successfully typed by primer extension (PE) reaction and DHPLC analysis, but two SNPs in *F2*, SNP5 (rs2070852) and SNP9 (rs2282687), were failed to be typed by the DHPLC analysis. SNP typing in *TFF1*, *S100A9-12-8*, *AMBP*, *SPPI*, *UMOD*, and *F2* genes by PE and DHPLC analyses were conducted as 5, 6, 3, 3, 2, and 2 sets of multiplex PE reaction, respectively, as shown in Supplementary Table S2.

Heteroduplex analysis by DHPLC method was then set and optimized to genotype two SNPs in *F2*, SNP5 (rs2070852) and SNP9 (rs2282687). PCR fragment containing SNP5 was amplified using a re-designed primer pair (5'-GAACTACCGAGGGCATGTGA-3' and 5'-GCGGCAGAAATTCTCCTGTA-3') giving the product size of 229 bp while PCR fragment containing SNP9 was amplified using primer pair as shown in Supplementary Table S3 giving the product size of 271 bp. The PCR products were denatured by heating, followed by slowly re-annealing and then applied for analysis in the first and/or second screening by DHPLC.

Altogether 67 SNPs including 10, 15, 9, 14, 9, and 10 SNPs in *TFF1*, *S100A9-12-8*, *AMBP*, *SPPI*, *UMOD*, and *F2* genes were genotyped in 224 samples, 112 cases and 112 controls. Examples of SNP genotyping for each multiplex PE set and heteroduplex analyses are shown in Supplementary

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figures S1-S7. For homozygous genotype of SNPs analyzed by PE and DHPLC, two peaks in the chromatogram were observed; the first peak represented unextended primer, followed by a peak of extension product. In contrast, for heterozygous genotype of SNP, three peaks in the chromatogram was observed; the first peak was unextended primer, followed by two peaks of different extension products. SNP calling from each pattern of DHPLC elution profile obtained from multiplex PE reaction and heteroduplex analyses was verified by DNA sequencing.

18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Individual SNP analysis

Genotyping of 67 SNPs of 8 candidate genes analyzed by PE or heteroduplex and DHPLC method in 112 controls and 112 cases was obtained. Data of SNP genotyping from both groups were analyzed for their association with the kidney stone disease. SNP information, allele frequencies, and association analysis results for SNPs of each gene are listed in Supplementary Table S1. Ten monomorphic SNPs were observed including SNP4 (rs1063933) of *S100A9*, SNP6 (rs3006475), SNP8 (rs2233864), and SNP9 (rs3014883) of *S100A12*, SNP15 (rs12040625) of *S100A8*, SNP4 (novel 2), SNP5 (novel 3), and SNP12 (rs4660) of *SPP1*, SNP3 (rs9940449) and SNP10 (rs8062123) of *UMOD*, which showed only one genotype in all cases and/or controls studied. These SNPs were excluded for HWE and further analyses. Tests for deviation from HWE and for association were carried out using web-based De Finetti program. All remaining, 57 SNPs, were in HWE ($P > 0.05$) according to Pearson's goodness-of-fit chi-squared test ($df = 1$). The results of allele frequencies revealed significant differences ($P < 0.05$) between control and case groups for 3 SNPs (SNP7, SNP10, and SNP11) of the *S100A9-12-8* genes cluster and for 7 SNPs (SNP1, SNP3, SNP4, SNP5, SNP7, SNP8, and SNP10) of the *F2* gene.

54
55
56
57
58
59
60

Association analysis of SNPs in *F2* gene

Since individual SNP analysis showing significant differences of allele frequencies between case and control groups of SNPs in the *F2* gene, we further investigated to show how important of the

association of these SNPs to the disease. Individual SNPs of the *F2* gene were also analyzed by using web-based SNPStats program. The results revealed that genotype frequencies of 8 SNPs (SNP1, SNP3, SNP4, SNP5, SNP7, SNP8, SNP9, and SNP10) differed significantly between normal controls and cases ($P < 0.05$) as shown in Table 2. The model of inheritance considered by less value of Akaike's Information Criterion (AIC) was recessive for all of these SNPs. The control group had significantly higher proportions of homozygous genotypes of minor alleles for all eight SNPs than the case group ($P = 0.0036-0.0160$), indicating that homozygous genotypes of minor alleles were protective to kidney stone. The odds ratios (ORs) were between 0.15-0.40.

Haplotype association analysis of SNPs in *F2* gene

More samples from both case and control groups were selected for SNP genotyping to confirm the result of the association analysis of the *F2* gene in addition to the data that were obtained from 224 subjects. A group of 52 patients with kidney stone (mean age 51.9 ± 11.4 years) and a group of 104 controls (mean age 53.3 ± 13.7 years) were analyzed for the SNPs in *F2* gene. The Haploview software package was used to estimate the pair-wise linkage disequilibrium (LD) and haplotype blocks structure and identify haplotype-tagging SNPs (htSNPs) using the confidence intervals (CI) algorithm. One LD block spanning 19 kb containing SNPs 1-10 with four htSNPs (SNP1, SNP2, SNP3 and SNP10) was defined and five haplotypes were obtained from the data of 164 cases and 216 controls (Figure 1 and Table 3). The most common haplotype in studied samples was TGCCGCCGCG (frequency = 0.559). This haplotype was found to be significantly frequent in the case group (frequency = 0.625) than in the control group (frequency = 0.509) with P -value of 0.0013 (OR 1.612, 95% CI 1.203-2.160), indicating that this haplotype was susceptibility or increased risk of the disease. In contrast, haplotype CGTTCCGCTA was more represented in the control group (frequency = 0.178) than in the case group (frequency = 0.091) with P -value of 0.0007 (OR 0.464, 95% CI 0.296-0.727), indicating that this haplotype was protective or reduced risk of the disease. The P -values after