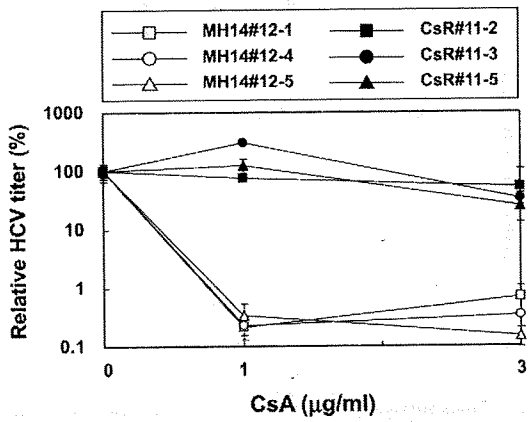


Fig. 4. Hepatitis C virus (HCV) RNA alteration contributed to cyclosporin A (CsA)-resistance. Total RNA extracted from CsA-resistant CsR#11 cells or that from wild-type MH14#12 cells as a control was transfected into Huh7 cells. Colonies established after 4-week selection with G418 were isolated, propagated individually, and tested for CsA response. Three cell clones derived from MH14#12 cells, MH14#12-1, MH14#12-4, and MH14#12-5 cells, and three cell clones from CsR#11 cells, CsR#11-2, CsR#11-3, and CsR#11-5 cells, were treated with 1 and 3 μg/mL CsA for 7 days. HCV-RNA titers were quantified by real-time RT-PCR analysis. The dots represent the means of three independent experiments.

by 1 log, in contrast to the wild-type MH14 clone, in which CsA decreased HCV-RNA by more than 2 logs under the same experimental condition (Fig. 5b). Thus, these mutations were demonstrated to confer CsA resistance; in addition to this, some cellular factors in Huh7 cells may also play minor roles in modulating the CsA sensitivities, given the result that Q86R/I252T/D320E cell clones were relatively sensitive to CsA compared with CsR#11-derived cell clones as shown in Figure 4. We next aimed to determine which of the three mutations, Q86R/I252T/D320E, was responsible for the CsA resistant phenotype, and individual mutations were engineered back into the wild-type MH14 replicon and stable replicon cells were produced as described above. Among three single amino acid mutations, the I252T mutation in NS3 resulted in a significant reduction in replication fitness (Fig. 5a), and almost failed to produce cell colonies. Cell clones harboring MH14 with both Q86R and D320E mutations, Q86R/D320E-2 and Q86R/D320E-3 cells, showed reduced sensitivity to CsA that was comparable to the levels in Q86R/I252T/D320E cells, suggesting Q86R and/or D320E mutation(s) was enough to confer the resistance. Subsequently, we treated the replicon cell clones carrying MH14 with either Q86R or D320E mutation alone, Q86R (Q86R-1 and -4 cells) and D320E (D320E-1 and -2 cells) cells, with CsA for 1 week. The titer of Q86R replicons was reduced to less than 100th by CsA treatment at a concentration of 1 μg/mL, similar to the wild type. In contrast, HCV replicon with D320E mutation in NS5A exhibited reduced sensitivity to CsA, resulting in little reduction of HCV-RNA by the treatment with 1 μg/mL CsA (Fig. 5b). Q86R mutation considerably enhanced colony

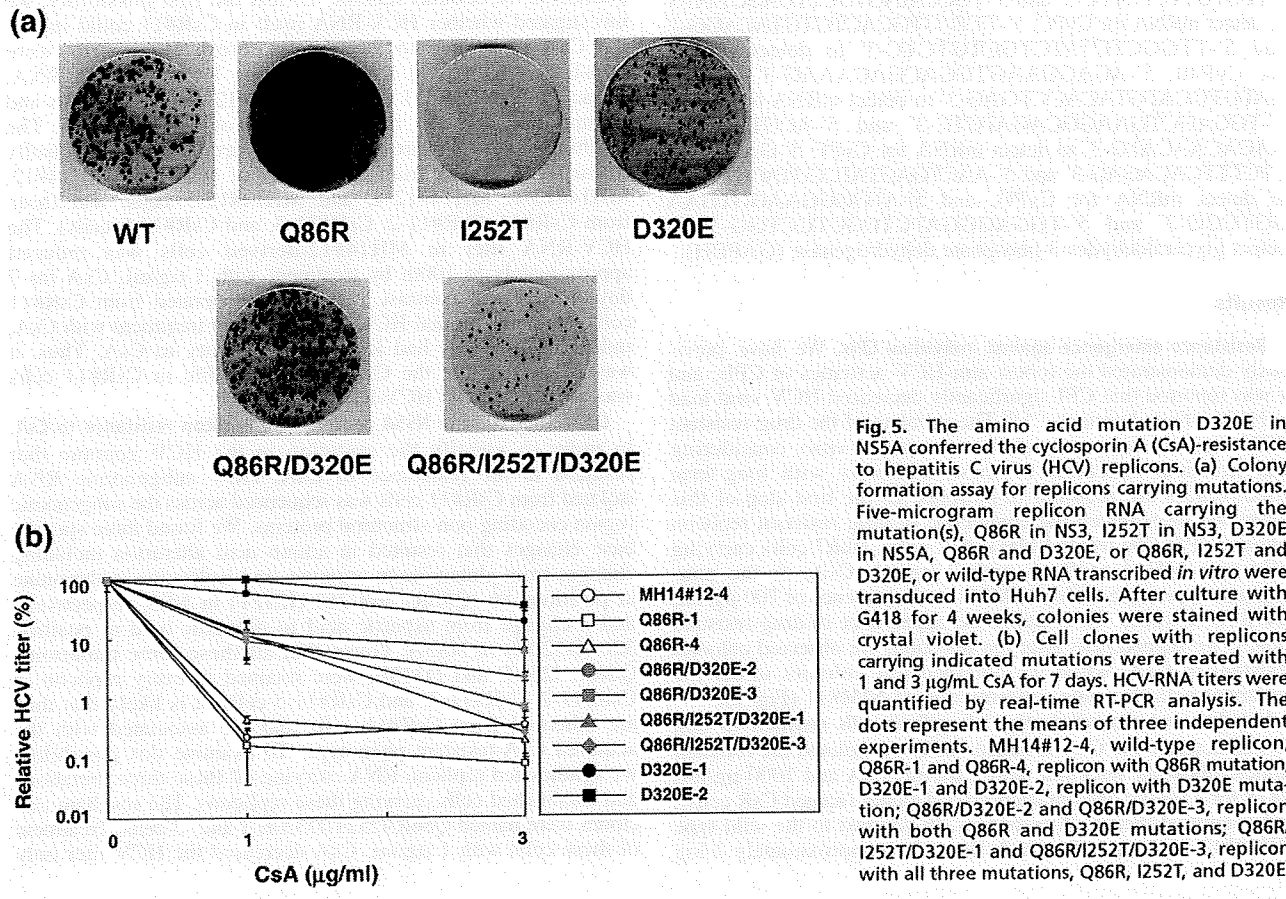


Fig. 5. The amino acid mutation D320E in NS5A conferred the cyclosporin A (CsA)-resistance to hepatitis C virus (HCV) replicons. (a) Colony formation assay for replicons carrying mutations. Five-microgram replicon RNA carrying the mutation(s), Q86R in NS3, I252T in NS3, D320E in NS5A, Q86R and D320E, or Q86R, I252T and D320E, or wild-type RNA transcribed *in vitro* were transduced into Huh7 cells. After culture with G418 for 4 weeks, colonies were stained with crystal violet. (b) Cell clones with replicons carrying indicated mutations were treated with 1 and 3 μg/mL CsA for 7 days. HCV-RNA titers were quantified by real-time RT-PCR analysis. The dots represent the means of three independent experiments. MH14#12-4, wild-type replicon; Q86R-1 and Q86R-4, replicon with Q86R mutation; D320E-1 and D320E-2, replicon with D320E mutation; Q86R/D320E-2 and Q86R/D320E-3, replicon with both Q86R and D320E mutations; Q86R/I252T/D320E-1 and Q86R/I252T/D320E-3, replicon with all three mutations, Q86R, I252T, and D320E.

doi: 10.1111/j.1349-7006.2009.01263.x

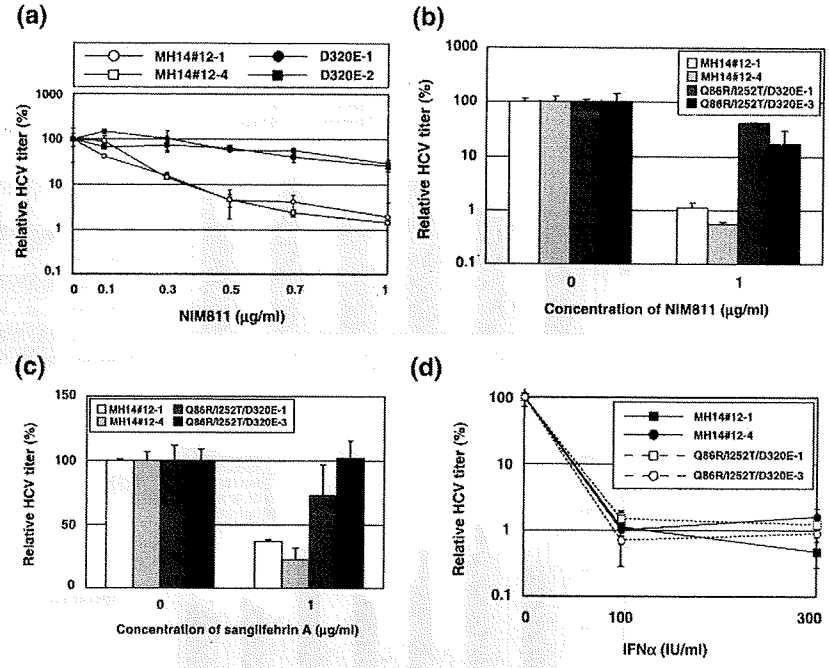


Fig. 6. Cyclosporin A (CsA)-resistant replicons demonstrated cross resistance to additional cyclophilin (CyP) inhibitors, NIM811 and sanglifehrin A (SFA), but not interferon (IFN)-α. MH14#12-1, MH14#12-4, D320E-1 and D320E-2 cells were treated with NIM811 at 0.1, 0.3, 0.5, 0.7 and 1 μg/mL (a), and MH14#12-1, MH14#12-4, Q86R/I252T/D320E-1, and Q86R/I252T/D320E-3 cells were treated with 1 μg/mL NIM811 (b), 1 μg/mL SFA (c), and 100 and 300 IU/mL IFN-α (d) for 7 days. The amounts of hepatitis C virus (HCV) RNA were quantified by real-time RT-PCR analysis. The data represent the means of three independent experiments.

formation efficiency and the D320E mutation showed little significant effect on the efficiency (Fig. 5a). Thus, the D320E mutation in NS5A was suggested as a sufficient factor to induce HCV replicon resistance to CsA, while the Q86R mutation was likely not to contribute to the resistance but to augment the efficiency of HCV replication itself.

The point mutation in NS5A conferred resistance to CPIs. Next, we examined cross-resistance between CsA and other CPIs or IFNα using the CsA-resistant replicon we produced above. Treatment with 0.1–1 μg/mL NIM811 for 7 days showed that the response to NIM811 of D320E-1 and -2 cells was less compared with that of MH14#12-1 and -4 cells, indicating that a CsA-resistant clone also acquired NIM811 resistance (Fig. 6a). A similar result was seen using Q86R/I252T/D320E cells (Fig. 6b). We then tested the anti-HCV activity of SFA, an additional CPI possessing distinct chemical backbone from those of cyclosporins.[16,17] Treatment with 1 μg/mL SFA reduced HCV replication in the wild-type cells, MH14#12-1, and -4 cells; however, it did not significantly reduce replication in Q86R/I252T/D320E cells (Fig. 6c). These results demonstrate that the CsA-resistant cells described in this study were also resistant to additional CPIs, confirming that these two compounds exerted anti-HCV effects via targeting CyP. Finally, we treated Q86R/I252T/D320E cell clones with 100 and 300 IU/mL IFNα for 7 days, and HCV-RNA titers were reduced by 2 logs in both clonal cell lines examined, Q86R/I252T/D320E-1 and Q86R/I252T/D320E-3 cells, as well as in wild-type MH14#12-derived cells, MH14#12-1, and MH14#12-4 cells (Fig. 6d). These results suggested no cross-resistance between CsA and IFNα, consistent with the previous report that the anti-HCV activity of CsA was independent of the IFNα signaling pathway.[18]

The role of CyP subtypes in HCV replication. We have previously reported that CyPB played a significant role in the efficient replication of HCV and CsA inhibited CyPB-mediated regulation of HCV replication. We have also suggested the involvement of other CyP subtypes in HCV replication.[19] To gain further insight into mechanisms underlying the anti-HCV properties of CPIs, we examined the roles of individual CyP subtypes in HCV replication in the wild-type MH14#12-1 and -4 replicon cells. In order to achieve this we knocked down CyPB with siRNAs (Fig. 7d), siCyPB-1 and -2, and found that this procedure reduced the amount of replicons to approximately half the initial level (Fig. 7c), a result consistent with the previous reports. Knockdown of CyPC, CyPE, CyPF, and CyPG (Fig. 7b) did not significantly affect the viral replication under these experimental conditions (Fig. 7a). Some groups have also suggested a role of CyPA in HCV replication.[20,21] Then, we synthesized individual siRNAs reported so far to be effective against CyPA, siCyPA-161, siCyPA-285, and siCyPA-459, and transfected them using a reagent with low cytotoxic activity to knock down endogenous CyPA (Fig. 7d). As shown in Figure 7c, the siRNAs directed against CyPA reduced HCV titers in MH14#12-1, and -4 cells. We previously observed that knockdown of CyPA little affected HCV replication in MH14 cells.[5] Here, by using a new transfection reagent with less cytotoxicity and higher knockdown efficiency, we observed the effect of CyPA knockdown on HCV replication, which suggests that CyPA-mediated regulation of HCV replication is strictly influenced by CyPA's expression level and cellular condition. Under this experimental condition, our RNAi experiments also displayed that knockdown of CyP40 (Fig. 7g), alternatively known as peptidylprolyl isomerase D (NM_005038), decreased the HCV titer (Fig. 7f) without significant cytotoxic effects, presenting CyP40 as additional cellular factor required for HCV replication.

CyPA was related to the CsA-resistant phenotype. We next asked which CyP subtype among CyPA, B, and 40 was related to the CsA resistance observed in our clones. To answer this question, we performed RNAi experiments in the CsA-resistant cell lines, CsR#11-2 and CsR#11-3 cells. Transfection of these cells with specific CyPB or CyP40 siRNAs resulted in the reduction of each subtype (Fig. 7d,g) and decreased the amount of HCV-RNA in CsR#11-derived cells and wild-type MH14#12-derived cells by approximately 50% (Fig. 7c,f). Thus, CyPB and CyP40 were likely to play roles in viral replication, even in the CsA-resistant cells. However, relative HCV titers were not reduced by CyPA knockdown in these CsA-resistant cells in contrast to the case with the wild-type replicon cells (Fig. 7c). A similar resistant phenotype to CyPA knockdown was observed in D320E

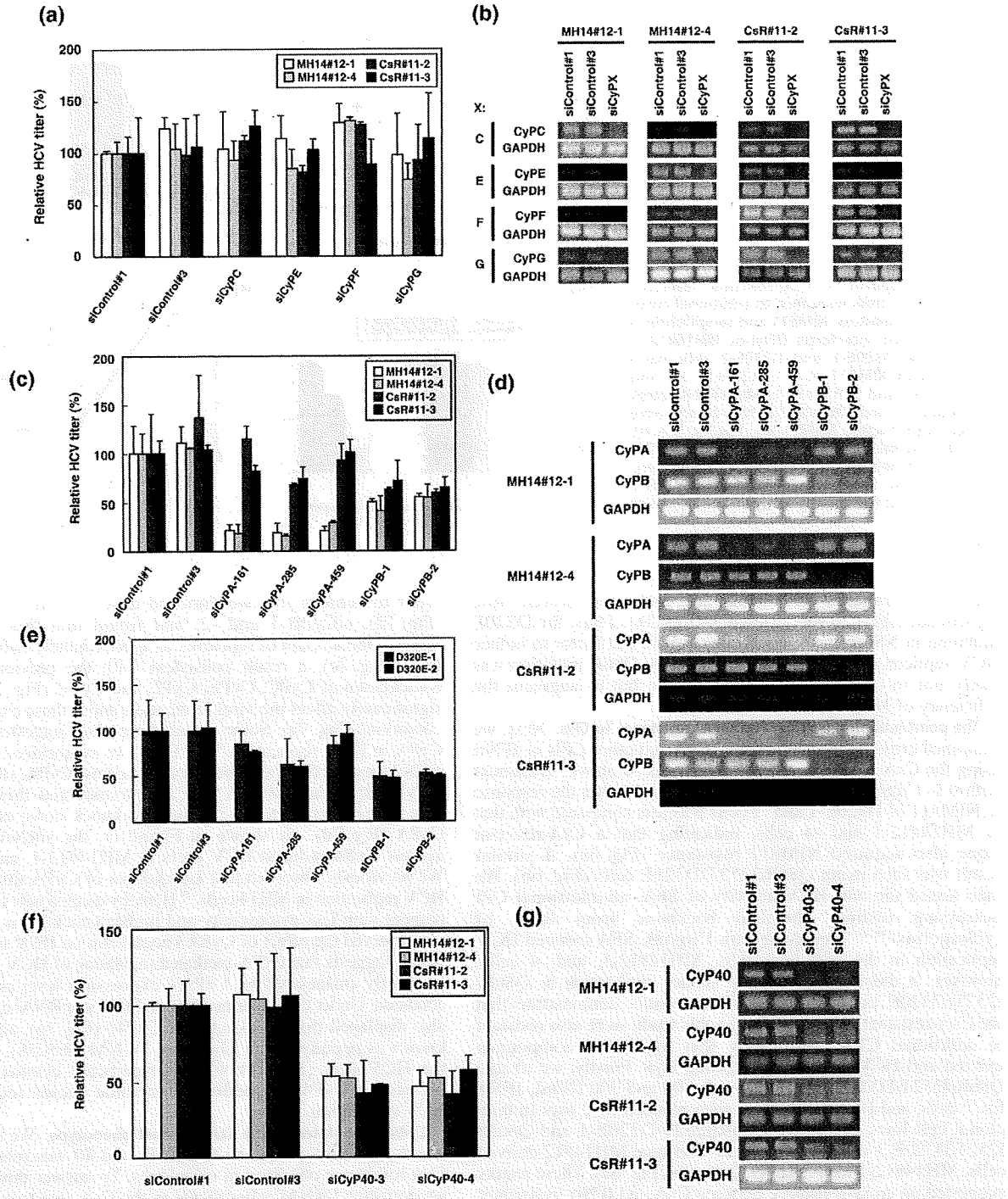


Fig. 7. Cyclophilin (CyP) subtypes related to anti-hepatitis C virus (HCV) effect of CyP inhibitor. MH14#12-derived cells, MH14#12-1 and MH14#12-4 cells, and CsR#11-2 and CsR#11-3 cells, were transfected with siRNAs specific for CyPC (siCyPC), CyPE (siCyPE), CyPF (siCyPF), and CyPG (siCyPG) (a); or those specific for CyPA (siCyPA-161, siCyPA-285, and siCyPA-459) and CyPB (siCyPB-1 and siCyPB-2) (c); or those specific for CyP40 (siCyP40-3 and siCyP40-4) (f); or randomized siRNA controls (siControl#1 and siControl#3). D320E cells were also transfected with the above siRNAs specific for either CyPA or CyPB (e). At 5 days post-transfection, the levels of HCV-RNA were quantified by real-time RT-PCR analysis. The mRNA levels of individual CyP subtypes, CyPC, CyPE, CyPF, and CyPG (CyPX corresponds to each CyP subtype indicated on the left side of the panels) (b), or CyPA and CyPB (d), or CyP40 (g) were measured using glyceraldehydes-3-phosphate dehydrogenase (GAPDH) as internal controls by RT-PCR analysis at 5 days post-transfection. The data represent the means of three independent experiments.

doi: 10.1111/j.1349-7006.2009.01263.x

cell clones (Fig. 7c), showing that CyPA was related to the CsA-resistance conferred by D320E mutation. The CsA-resistant clones obtained in this study were likely to have acquired CyPA independence for efficient HCV replication.

Discussion

Given that CPIs suppressed HCV viral load in cell culture and in patients with chronic hepatitis C,[14,15] CPIs are expected to be new anti-HCV agents. It is important to further reveal the factors related to CPI's anti-HCV activities, thinking over the practical use of CPIs with maximized efficacy and high specificity facing challenges such as side effects and the emergence of resistance to them in clinical settings. Here, we isolated and characterized a variant resistant to CPIs using a HCV subgenomic replicon system. A mutation in NS5A, D320E, was shown to confer the CPI-resistance to HCV replicon, resulting in CyPA independence for efficient viral replication. In addition, assessment of a wide range of CyP subtype knockdown experiments found CyP40 to be a new contributor to HCV replication.

Of the mutations identified, Q86R substitution in NS3 dramatically enhanced the capacity of replication. This mutation was observed as compensatory mutation[22] following the selection of replicons resistant to protease inhibitors SCH503034[23] and SCH6.[24] In addition, this mutation also appeared following the passaging of replicon cells in the absence of drug pressure.[25,26] In actuality, this mutation did not contribute to CsA resistance in the replicon cells (Fig. 5b), and thus was thought to be an adaptive mutation similar to that suggested in previous reports. I252T mutation in NS3, on the other hand, severely reduced the replicative fitness of HCV. The significance of I252T mutation under CsA treatment remains to be studied. The alteration of amino acid residue in NS5A, D320E, resulted in the conversion of HCV replicon to that of the CsA-resistant phenotype. There have been no reports of a link between NS5A and individual CyP subtype in the context of HCV replication, though mutations in NS5A were found to be keys for the acquisition of CsA resistance.[27] We have previously reported that CyPB was important for viral replication, but NS5A did not interact with CyPB in MH14 cells.[5] Indeed, in cells harboring replicons with D320E mutation, CyPB was found to contribute to viral replication but was not related to CsA resistance, as knockdown of CyPB diminished the viral titer to approximately half, similar to the case of the wild type. Therefore, other CyP molecules crucial for viral replication were suggested to be involved in the phenomenon of the CsA resistance. CyPA is another CyP subtype recently published to be critical for HCV replication in connection with viral polymerase.[20,21] Our CsA-resistant replicon cells displayed resistance to CyPA knockdown when compared to wild-type replicon, suggesting that CyPA participated in the replication process and the CsA resistance was due in part to resistance to CyPA inhibition. Therefore, it might be possible that NS5A functions coordinated with CyPA for viral replication and D320E mutation could contribute to enhancement of the relation. But NS5A was unable to bind CyPA *in vitro*.[5] NS5A might be regulated by CyPA associated with other cellular or viral factors during HCV replication. The fact that the D320 falls upon one of the two discontinuous domains needed for the interaction with NS5B to functionally modulate it[28,29] lead us to presume influence of NS5A on the reported NS5B–CyPA interaction.[21] In addition to CyPA and CyPB, which have been published to be cellular factors required for HCV replication, the results suggested that another CyP subtype, CyP40, contributed to viral replication. Acting as a molecular chaperone, it is conceivable that CyP40 directly interacts with viral proteins to boost their functions, similar to CyPA and CyPB. Heat shock protein (Hsp) 90 is a well-known chaperone forming complex with CyP40. Recently, Hsp90 was shown to be harnessed by HCV NS5A via the FK-506 binding protein 8 (FKBP8) bridge. FKBP8 is a homologous immunophilin of CyP40 that is required for viral replication.[30] This result led to the hypothesis that CyP40 serves as a linker between viral proteins and Hsp90. CyP40 is also known to associate with estrogen receptor (ESR) and we have published that ESRα escorted NS5B to replication complex (RC).[31] We also speculate CyP40 connected to ESRα may be important for the recruitment or functional reinforcement of viral and cellular factors for HCV replication in RC. Among these CyP subtypes, CyPA dependency was suggested to be one of the determinants of CsA sensitivity. Interestingly, CyPB and CyP40 play significant role in HCV replication even in CsA-resistant replicon cells. Another CPI, NIM811, is also likely to target CyPA, at least in part, to suppress HCV replication, given the cross-resistance of CsA-resistant replicon to MIN811. However, there is still also the possibility that other CyPs mediate anti-HCV effect of NIM811, which needs to be elucidated in future study.

Understanding the profile of CPI-resistance mutations in the HCV genome and the viral and cellular factors involved will aid in the progression of CPI-centered strategies preparing for the problem of drug resistance. In addition, the cells harboring CPI-resistant replicons established here may prove beneficial for further characterization of resistance mechanisms and for the screening of novel compounds with the potential of clinical application to defeat CPI-resistant variants. Also, CyP40 as a contributor to HCV replication could become another specific antiviral target. The information arising from this study is expected to contribute to the successful use of CPIs against a liver carcinogen, HCV.

Acknowledgments

We thank Novartis (Basel, Switzerland) for providing the CsA derivative, NIM811, and SFA. This work was supported by Grants-in-Aid from the Ministry of Health, Labor and Welfare of Japan. This work was also supported by Grants-in-Aid for Cancer Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan, by Grants-in-Aid for the Research for the Future Program from the Japan Society for the Promotion of Science (JSPS), and by Grants-in-Aid for the Program for Promotion of Fundamental Studies in Health Science from the Organization for Pharmaceutical Safety of Japan. K.W. is a recipient of a JSPS Postdoctoral Fellowship for Research Abroad, and K.G. is a recipient of a JSPS Research Fellowship for Young Scientists.

References

1 Sarbah SA, Younossi ZM. Hepatitis C: an update on the silent epidemic. *J Clin Gastroenterol* 2000; **30**: 125–43.
2 Manns MP, McHutchison JG, Gordon SC *et al.* Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet* 2001; **358**: 958–65.
3 Fried MW, Shiffman ML, Reddy KR *et al.* Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002; **347**: 975–82.
4 Watashi K, Hijikata M, Hosaka M, Yamaji M, Shimotohno K. Cyclosporin A suppresses replication of hepatitis C virus genome in cultured hepatocytes. *Hepatology* 2003; **38**: 1282–8.
5 Watashi K, Ishii N, Hijikata M *et al.* Cyclophilin B is a functional regulator of hepatitis C virus RNA polymerase. *Mol Cell* 2005; **19**: 111–22.
6 Manns MP, Foster GR, Rockstroh JK, Zeuzem S, Zoulim F, Houghton M. The way forward in HCV treatment – finding the right path. *Nat Rev Drug Discov* 2007; **6**: 991–1000.
7 McGovern BH, Abu Dayyeh BK, Chung RT. Avoiding therapeutic pitfalls: the rational use of specifically targeted agents against hepatitis C infection. *Hepatology* 2008; **48**: 1700–12.
8 Melnikova I. Hepatitis C therapies. *Nat Rev Drug Discov* 2008; **7**: 799–800.

9 Cordes F, Kaiser R, Selbig J. Bioinformatics approach to predicting HIV drug resistance. *Expert Rev Mol Diagn* 2006; **6**: 207–15.
10 Mo H, Lu L, Dekhtyar T *et al.* Characterization of resistant HIV variants generated by *in vitro* passage with lopinavir/ritonavir. *Antiviral Res* 2003; **59**: 173–80.
11 Molla A, Korneyeva M, Gao Q *et al.* Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nat Med* 1996; **2**: 760–6.
12 Shulman N, Winters M. A review of HIV-1 resistance to the nucleoside and nucleotide inhibitors. *Curr Drug Targets Infect Disord* 2003; **3**: 273–81.
13 Watashi K, Hijikata M, Tagawa A, Doi T, Marusawa H, Shimotohno K. Modulation of retinoid signaling by a cytoplasmic viral protein via sequestration of Sp110b, a potent transcriptional corepressor of retinoic acid receptor, from the nucleus. *Mol Cell Biol* 2003; **23**: 7498–509.
14 Flisiak R, Dumont JM, Crabbe R. Cyclophilin inhibitors in hepatitis C viral infection. *Expert Opin Invest Drugs* 2007; **16**: 1345–54.
15 Flisiak R, Horban A, Gallay P *et al.* The cyclophilin inhibitor Debio-025 shows potent anti-hepatitis C effect in patients coinfected with hepatitis C and human immunodeficiency virus. *Hepatology* 2008; **47**: 817–26.
16 Zhang LH, Liu JO. Sanglifehrin A, a novel cyclophilin-binding immunosuppressant, inhibits IL-2-dependent T cell proliferation at the G1 phase of the cell cycle. *J Immunol* 2001; **166**: 5611–8.
17 Zenke G, Strittmatter U, Fuchs S *et al.* Sanglifehrin A, a novel cyclophilin-binding compound showing immunosuppressive activity with a new mechanism of action. *J Immunol* 2001; **166**: 7165–71.
18 Goto K, Watashi K, Murata T, Hishiki T, Hijikata M, Shimotohno K. Evaluation of the anti-hepatitis C virus effects of cyclophilin inhibitors, cyclosporin A, and NIM811. *Biochem Biophys Res Commun* 2006; **343**: 879–84.
19 Ishii N, Watashi K, Hishiki T *et al.* Diverse effects of cyclosporine on hepatitis C virus strain replication. *J Virol* 2006; **80**: 4510–20.
20 Nakagawa M, Sakamoto N, Tanabe Y *et al.* Suppression of hepatitis C virus replication by cyclosporin A is mediated by blockade of cyclophilins. *Gastroenterology* 2005; **129**: 1031–41.
21 Yang F, Robotham JM, Nelson HB, Irsigler A, Kenworthy R, Tang H. Cyclophilin a is an essential cofactor for hepatitis C virus infection and the principal mediator of cyclosporine resistance *in vitro*. *J Virol* 2008; **82**: 5269–78.
22 Xavier LL, Moya A, Gonzalez-Candelas F. Mapping natural polymorphisms of hepatitis C virus NS3/4A protease and antiviral resistance to inhibitors in worldwide isolates. *Antivir Ther* 2008; **13**: 481–94.
23 Tong X, Chase R, Skelton A, Chen T, Wright-Minogue J, Malcolm BA. Identification and analysis of fitness of resistance mutations against the HCV protease inhibitor SCH 503034. *Antiviral Res* 2006; **70**: 28–38.
24 Yi M, Tong X, Skelton A *et al.* Mutations conferring resistance to SCH6, a novel hepatitis C virus NS3/4A protease inhibitor. Reduced RNA replication fitness and partial rescue by second-site mutations. *J Biol Chem* 2006; **281**: 8205–15.
25 Blight KJ, Kolykhalov AA, Rice CM. Efficient initiation of HCV RNA replication in cell culture. *Science* 2000; **290**: 1972–4.
26 Krieger N, Lohmann V, Bartenschlager R. Enhancement of hepatitis C virus RNA replication by cell culture-adaptive mutations. *J Virol* 2001; **75**: 4614–24.
27 Fernandes F, Poole DS, Hoover S *et al.* Sensitivity of hepatitis C virus to cyclosporine A depends on nonstructural proteins NS5A and NS5B. *Hepatology* 2007; **46**: 1026–33.
28 Shirota Y, Luo H, Qin W *et al.* Hepatitis C virus (HCV) NS5A binds RNA-dependent RNA polymerase (RdRP) NS5B and modulates RNA-dependent RNA polymerase activity. *J Biol Chem* 2002; **277**: 11149–55.
29 Shimakami T, Hijikata M, Luo H *et al.* Effect of interaction between hepatitis C virus NS5A and NS5B on hepatitis C virus RNA replication with the hepatitis C virus replicon. *J Virol* 2004; **78**: 2738–48.
30 Okamoto T, Nishimura Y, Ichimura T *et al.* Hepatitis C virus RNA replication is regulated by FKBP8 and Hsp90. *Embo J* 2006; **25**: 5015–25.
31 Watashi K, Inoue D, Hijikata M, Goto K, Aly HH, Shimotohno K. Anti-hepatitis C virus activity of tamoxifen reveals the functional association of estrogen receptor with viral RNA polymerase NS5B. *J Biol Chem* 2007; **282**: 32765–72.

doi: 10.1111/j.1349-7006.2009.01263.x

Bioorganic & Medicinal Chemistry Letters 18 (2008) 4638–4641

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

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

Synthesis and evaluation of 5′-modified 2′-deoxyadenosine analogues as anti-hepatitis C virus agents

Masahiro Ikejiri [a,b,*], Takayuki Ohshima [a], Akemi Fukushima [a], Kunitada Shimotohno [c], Tokumi Maruyama [a,*]

[a] Faculty of Pharmaceutical Sciences at Kagawa Campus, Tokushima Bunri University, 1314-1 Shido, Sanuki, Kagawa 769-2193, Japan
[b] Faculty of Pharmacy, Osaka Ohtani University, 3-11-1 Nishikiori-kita, Tondabayashi, Osaka 584-8540, Japan
[c] Center for Integrated Medical Research, Keio University School of Medicine, 35 Shinanomachi, Shinjuku, Tokyo 160-8582, Japan

ARTICLE INFO

Article history:
Received 9 June 2008
Revised 3 July 2008
Accepted 5 July 2008
Available online 10 July 2008

Keywords:
Antiviral agent
Hepatitis C virus
HCV
Nucleoside

ABSTRACT

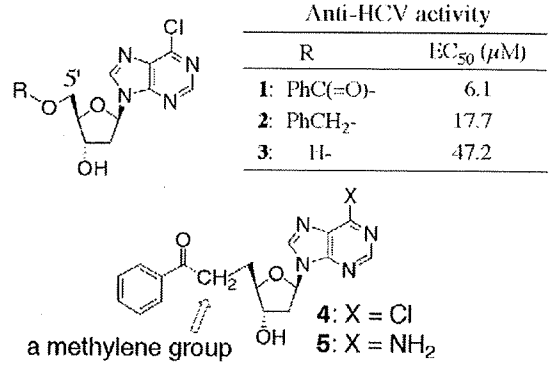
In order to study the effect of 5′-modification of 2′-deoxynucleoside on its anti-HCV activity, several analogues were synthesized and evaluated. Among the analogues, a 5′-deoxy-5′-phenacylated analogue exhibited a good anti-HCV activity with an EC_{50} of 15.1 μM. This compound is expected to operate via a type of mechanism that does not involve a generally known 5′-*O*-triphosphorylation process.

Hepatitis C virus (HCV)[1] is a major causative agent of non-A and non-B hepatitis. It is estimated to have infected >170 million individuals, that is, 3.5% of the world's population. HCV infection is a leading cause of chronic hepatitis, liver cirrhosis and hepatocellular carcinoma. Current therapy based on pegylated interferon and ribavirin is often poorly tolerated and is effective in only 50% of patients. Therefore, the development of further effective therapeutic agents against HCV is an urgent public health requirement.

In our previous study,[2] we revealed that several 5′-*O*-masked analogues of 6-chloropurine-2′-deoxyriboside, such as benzoate **1** and benzyl ether **2**, exhibit an effective anti-HCV activity in a subgenomic replicon cell line and are more potent than the corresponding unmasked analogue **3** (Fig. 1). Since it is generally accepted that most nucleoside antivirals exhibit their potency after being converted to the corresponding 5′-triphosphates,[3] the unmasked (or phosphorylated) 5′-hydroxyl group is indispensable for the antiviral activity. Accordingly, our result that the 5′-*O*-masking leads to an improvement in the anti-HCV activity appears to be inconsistent with the common understanding, interestingly.

We presume that the anti-HCV activity of certain 5′-*O*-masked analogues would arise from a new type of mechanism that does not involve the 5′-*O*-triphosphorylation process. However, there is still room for the discussion on the 5′-*O*-masking effect because certain carbon–oxygen bonds, for example, the carboxylic ester bond of compound **1** (i.e., the benzoate moiety in compound **1**), are often hydrolyzed in cultured cells; in other words, there is a possibility that compound **1** simply operates as a prodrug of **3**.[4,5] Therefore, in order to confirm the effectiveness of 5′-*O*-masking groups, particularly that of the benzoyl group of compound **1**, we planned the syntheses and anti-HCV evaluations of ketone analogues **4** and **5**, in which the 5′-oxgen atom was replaced with a methylene group to prevent the hydrolytic removal of the benzoyl group.

The synthesis of **4** began with readily available 3′-*O*-TBS-2′-deoxyadenosine (**6**)[6] (Scheme 1). First, we attempted to subject



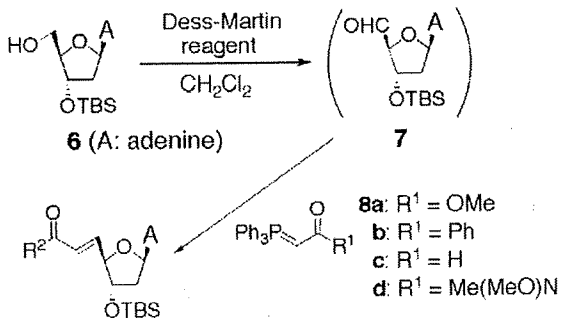
Anti-HCV activity

R	EC_{50} (μM)
1: PhC(=O)-	6.1
2: $PhCH_2$-	17.7
3: H-	47.2

Figure 1. Structures of 5′-modified analogues.

* Corresponding authors.
E-mail address: ikejim@osaka-ohtani.ac.jp (M. Ikejiri).

0960-894X/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2008.07.015



compd	R^2	yield (%)	E:Z
9	MeO–	90	E only
10	Ph–	60	13:1
11	H–	—	
12	Me(MeO)N–	—	

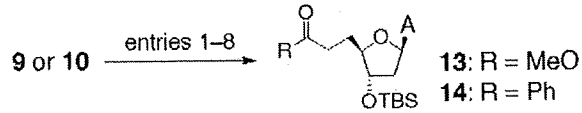
Scheme 1.

the isolated aldehyde **7** obtained via the oxidation of **6** to the following Wittig reaction; however, it was unsuccessful due to the instability of **7**. This issue was overcome by using a one-pot oxidation–Wittig reaction with Dess–Martin periodinane (DMP) and stabilized phosphorus ylide.[7] Among the four types of ylides examined (**8a–d**), two of them (**8a** and **8b**) successfully afforded the desired compounds **9** and **10** in 90% (*E*-isomer only) and 60% (*E*:*Z* = 13:1) yields, respectively, while the others (**8c** and **8d**) yielded complex mixtures. Since the Dess–Martin oxidation is not very suitable for the large-scale synthesis of **9** and **10** because of the explosive nature of DMP (and also its precursor, 2-iodoxybenzoic acid[8]), several other one-pot protocols such as Moffatt oxidation–Wittig,[9] PCC–Wittig,[10] TEMPO–BAIB–Wittig,[11] and TPAP–NMO–Wittig[12] were examined with **6** and **8a**. However, the TLC analyses of all the attempts revealed low yields and/or the formation of by-products.

With the thus-obtained products, the reduction of the C–C double bond was examined (Table 1). Compound **9** was converted to **13** under standard hydrogenation conditions (Pd/C, H_2, THF) with an excellent yield although the reaction required a long reaction time (~2 days) and comparatively large quantities of the catalyst (50 wt.%) (entry 1). In contrast, the conjugate reduction of **9** by sodium borohydride–transition metal salt (e.g., $NiCl_2$ and CuCl) systems[13] furnished **13** in a short time (1–3 h), but the yield was moderate (entries 2 and 3). The use of elemental magnesium in methanol led to a complex mixture (entry 4). In the case of **10**, chemoselective hydrogenations by Sajiki's procedures (Pd/fibroin–H_2 or Pd/C–Ph_2S–H_2)[14] were ineffective, resulting in the recovery of a large amount of the starting material (entries 5 and 6), while the 1,4-reduction with indium hydride generated in situ by using $PhSiH_3$–$In(OAc)_3$ or Bu_3SnH–$InCl_3$[15] efficiently afforded the desired product **14** in good yields (entries 7 and 8). Consequently, the conditions in the case of entries 1 and 7 were employed for routine syntheses of **13** and **14**, respectively, in view of their simple experimental procedures as well as their good yields.

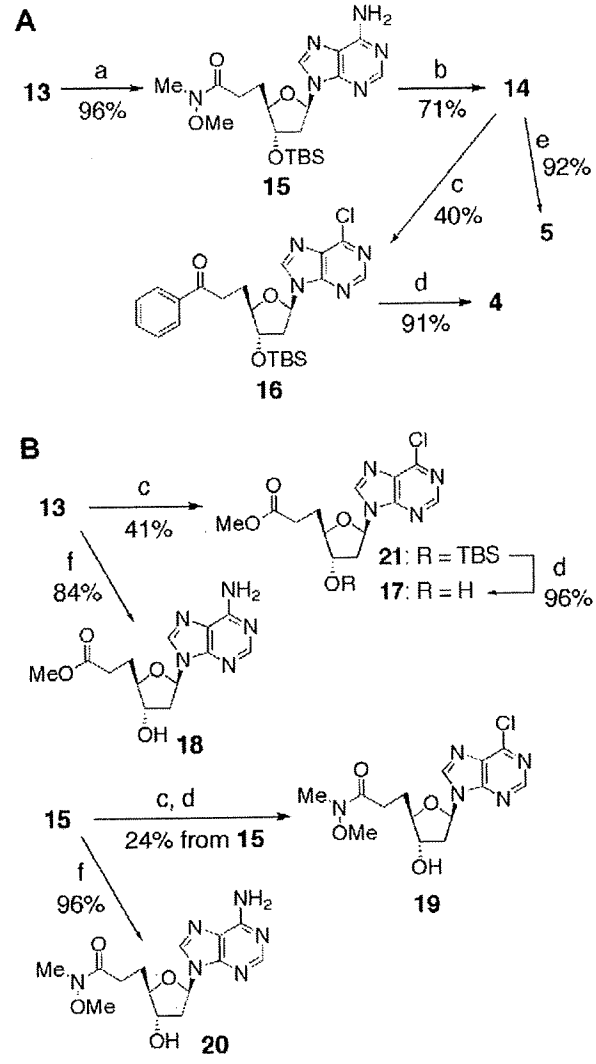
Compound **13** was readily converted to **14** with a two-step sequence, that is, via a Weinreb amide **15**, as illustrated in Scheme 2-A. Using a Grignard reagent (PhMgBr) led to a better yield (71%) than when phenyl lithium was used (58% yield). This two-step conversion will effectively serve for the synthesis of various analogues because the phenyl moiety of **14** can be easily replaced with other groups by changing the type of Grignard reagent. The amino group of **14** was substituted by a chloro group to afford **16** (40% yield),

Table 1
Chemoselective reduction of C–C double bond



Entry	Substrate	Conditions	Product	Yield (%)
1	**9**	Pd/C, H_2, THF, rt, 2 d	**13**	94
2	**9**	$NiCl_2$, $NaBH_4$, MeOH, 0 °C, 1 h	**13**	50
3	**9**	CuCl, $NaBH_4$, MeOH, 0 °C, 2.5 h	**13**	76
4	**9**	Mg, MeOH, reflux, 2 h	Complex mix.	
5	**10**	Pd/fibroin, H_2, MeOH, rt, 2 d	**14**	Trace[a]
6	**10**	Pd/C, Ph_2S, H_2, MeOH, rt, 2 d	**14**	31[b]
7	**10**	$PhSiH_3$, $In(OAc)_3$, EtOH, rt, over night	**14**	84
8	**10**	Bu_3SnH, $InCl_3$, *i*-PrOH −78 °C to rt, 2 h	**14**	93

[a] 93% of **10** was recovered.
[b] 56% of **10** was recovered.



Scheme 2. Reagents: (a) Me(MeO)NH·HCl, *n*-BuLi, THF; (b) PhMgBr, THF; (c) *t*-BuONO, Et_4NCl, CCl_4–CH_2Cl_2; (d) TBAF, AcOH, THF; (e) $Et_3N{\cdot}3HF$, THF; (f) TAS-F, MeCN.

which was subsequently treated with a mixture of tetrabutylammonium fluoride (TBAF) and acetic acid, giving the desired product **4** in 91% yield. A moderate yield of **16** was mainly obtained due to the competitive elimination of its nucleobase moiety. The other desired compound **5** was prepared in 92% yield by exposing **14** to triethylamine trihydrofluoride.

Since we are interested in the structure–activity relationship (SAR) of not only the benzoyl moiety but also the methyl ester and Weinreb amide moieties contained in the synthetic intermediates, we conducted syntheses of the corresponding analogues **17–20**, as shown in Scheme 2-B. 6-Chloropurine analogues **17** and **19** were prepared from **13** and **15**, respectively, under conditions almost identical to those used in the synthesis of **4** (i.e., *t*-BuONO–Et_4NCl and TBAF–AcOH). The conversion to **18** and **20** was effectively accomplished by the treatment of **13** and **15** with tris(dimethylamino)sulfonium difluorotrimethylsilicate (TAS-F),[16] while that with TBAF led to a mixture of the desired product and certain tertabutylammonium salts that were difficult to separate.

The synthesized nucleoside analogues mentioned above were assayed for their ability to inhibit HCV RNA replication in a subgenomic replicon Huh7 cell line (LucNeo#2),[17] and the result is presented in Table 2 and Figure 2. These cells contain an HCV subgenomic replicon RNA encoding a luciferase reporter gene as a marker. The antiviral potency of the analogues against the HCV replicon is expressed as EC_{50}, which was quantified by a luciferase assay after a two-day incubation period with the corresponding compound. In addition, the associated cytotoxicity (expressed as CC_{50} in Table 2) was evaluated in a tetrazolium (XTT)-based assay according to the manufacturer's protocol.

As shown in Table 2, the ketone analogue **4** exhibited an antiviral activity against the HCV replicon with an EC_{50} of 15.1 μM (entry 1), which is nearly comparable to that of benzoate analogue **1** (entry 7). The cytotoxicity of **4** was somewhat high (CC_{50}: 76.3 μM), but was not high enough to exert an influence on the EC_{50} value because the cytotoxicity at 15 μM was considerably low (ca. 0–2%) (Fig. 2A). Thus, the decrease in the luciferase activity with **4** results from its anti-HCV activity, not its cytotoxicity. Interestingly, compounds **17** and **19** also exhibited anti-HCV activities (entries 3 and 5, respectively). In contrast, the 6-amino analogues **5**, **18**, and **20** did not exhibit any significant anti-HCV activity (entries 2, 4, and 6).[18]

To confirm the anti-HCV potency of compound **4**, subgenomic replicon RNA levels were quantified by real-time RT-PCR analysis (Fig. 2B). Exposing the replicon cells to 12.5 and 25 μM of **4** reduced the replicon RNA amount up to approximately 60% and 25%, respectively. This result is almost consistent with that of the luciferase assay with **4**.

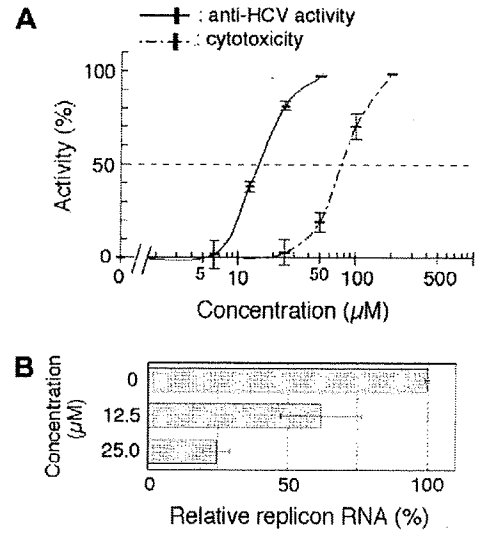
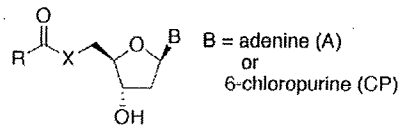


Figure 2. Anti-HCV activity and cytotoxicity of **4**: (A) result of luciferase assay and XTT assay; (B) result of real-time RT-PCR.

Taking these data into account, it appears that the phenacyl group ($BzCH_2$–) equipped at the C5′ position as well as the benzoyloxy group (BzO–) is effective functional group for anti-HCV activity; this should be noteworthy because the 5′-phenacyl group is expected to operate without being converted to the corresponding 5′-hydroxyl group (or 5′-triphosphate group). This result strongly supports our hypothesis that the 5′-*O*-masking group can contribute to the anti-HCV activity not only as a unit for the prodrug system but also as a part of the substrate. Although the detailed mechanism is unclear and the biological activity is still insufficient, the antiviral potency of such 5′-modified analogues is of great interest because they are likely to operate via a pathway that does not involve the 5′-*O*-phosphorylation process. We hope that the present study will contribute to developing a new class of HCV therapeutic agents.

Table 2
Inhibitory potency (EC_{50}) and cytotoxicity (CC_{50}) of the synthesized analogues in HCV replicon assay



Entry	Compound	R	X	B	EC_{50}[a] (μM)	CC_{50}[a] (μM)
1	**4**	Ph	CH_2	CP	15.1 ± 0.4	76.3 ± 5.2
2	**5**	Ph	CH_2	A	>200	–
3	**17**	MeO	CH_2	CP	32.9 ± 1.6	>200
4	**18**	MeO	CH_2	A	>200	–
5	**19**	Me(MeO)N	CH_2	CP	40.4 ± 1.4	>200
6	**20**	Me(MeO)N	CH_2	A	>200	–
7	**1**	Ph	O	CP	6.1[b]	111[b]

[a] EC_{50}: 50% effective concentration; CC_{50}: 50% cytotoxic concentration.
[b] Extracts obtained from our previous study (Ref. 2).

Acknowledgments

This research was partly supported by a Grant-in-Aid for Young Scientists (B) (20790106) from the Ministry of Education, Culture, Sports, Science and Technology of Japan and by a Grant-in-Aid from Mitsubishi Chemical Corporation Fund.

Supplementary data

Supplementary data associated with this article (experimental details and spectroscopic data of new compounds **4–6**, **9**, **10**, **13–21**) can be found, in the online version, at doi:10.1016/j.bmcl.2008.07.015.

References

1. Recent reviews: (a) Gordon, C. P.; Keller, P. A. *J. Med. Chem.* **2005**, *48*, 1; (b) De Francesco, R.; Migliaccio, G. *Nature* **2005**, *436*, 953; (c) De Clercq, E. *Nat. Rev. Drug Discov.* **2007**, *6*, 1001.
2. Ikejiri, M.; Ohshima, T.; Kato, K.; Toyama, M.; Murata, T.; Shimotohno, K.; Maruyama, T. *Bioorg. Med. Chem.* **2007**, *15*, 6882.
3. Arimilli, M. N.; Dougherty, J. P.; Cundy, K. C.; Bischofberger, N.. In *Advances in Antiviral Drug Design*; De Clercq, E., Ed.; Jai Press Inc.: Stamford Connecticut, 1999; Vol. 3, pp 69–91. and also see Refs. 1a and c and references therein.

4. Several 5′-*O*-acyl nucleoside analogues have been reported as prodrugs of the corresponding deacylated analogues Parang, K.; Wiebe, L. I.; Knaus, E. E. *Curr. Med. Chem.* **2000**, *7*, 995.
5. Not only compound **1** but also **2** might operate as a prodrug of **3** since *O*-dealkylated metabolism is caused in some cases. Silverman, R. B. *The Organic Chemistry of Drug Design and Drug Action, Second Edition*; Elsevier, 2004, Chapter 7.
6. Somu, R. V.; Wilson, D. J.; Bennett, E. M.; Boshoff, H. I.; Celia, L.; Beck, B. J.; Barry, C. E., III; Aldrich, C. C. *J. Med. Chem.* **2006**, *49*, 7623.
7. Barrett, A. G. M.; Hamprecht, D.; Ohkubo, M. *J. Org. Chem.* **1997**, *62*, 9376.
8. A one-pot oxidation–Wittig reaction with 2-iodoxybenzoic acid is also reported Crich, D.; Mo, X.-S. *Synlett* **1999**, 67.
9. Rapp, M.; Haubrich, T. A.; Perrault, J.; Mackey, Z. B.; McKerrow, J. H.; Chiang, P. K.; Wnuk, S. F. *J. Med. Chem.* **2006**, *49*, 2096.
10. Bressette, A. R.; Glover, L. C., IV *Synlett* **2004**, 738.
11. Vatèle, J.-M. *Tetrahedron Lett.* **2006**, *47*, 715.
12. MacCoss, R. N.; Balskus, E. P.; Ley, S. V. *Tetrahedron Lett.* **2003**, *44*, 7779.
13. (a) Narisada, M.; Horibe, I.; Watanabe, F.; Takeda, K. *J. Org. Chem.* **1989**, *54*, 5308; (b) Satoh, T.; Nanba, K.; Suzuki, S. *Chem. Pharm. Bull.* **1971**, *19*, 817.
14. (a) Ikawa, T.; Sajiki, H.; Hirota, K. *Tetrahedron* **2005**, *61*, 2217; (b) Mori, A.; Mizusaki, T.; Miyakawa, Y.; Ohashi, E.; Haga, T.; Maegawa, T.; Monguchi, Y.; Sajiki, H. *Tetrahedron* **2006**, *62*, 11925.
15. (a) Miura, K.; Yamada, Y.; Tomita, M.; Hosomi, A. *Synlett* **2004**, 1985; (b) Inoue, K.; Ishida, T.; Shibata, I.; Baba, A. *Adv. Synth. Catal.* **2002**, *344*, 283.
16. Kang, S. B.; De Clercq, E.; Lakshman, M. K. *J. Org. Chem.* **2007**, *72*, 5724.
17. (a) Watashi, K.; Hijikata, M.; Hosaka, M.; Yamaji, M.; Shimotohno, K. *Hepatology* **2003**, *38*, 1282; (b) Murata, T.; Hijikata, M.; Shimotohno, K. *Virology* **2005**, *340*, 105; (c) Goto, K.; Watashi, K.; Murata, T.; Hishiki, T.; Hijikata, M.; Shimotohno, K. *Biochem. Biophys. Res. Commun.* **2006**, *343*, 879. And also see Ref. 2.
18. A similar SAR trend was observed in our previous study. Ikejiri, M.; Saijo, M.; Morikawa, S.; Fukushi, S.; Mizutani, T.; Kurane, I.; Maruyama, T. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2470.

FEBS Letters 582 (2008) 351–358

Expressed sequence tags from cynomolgus monkey (*Macaca fascicularis*) liver: A systematic identification of drug-metabolizing enzymes

Yasuhiro Uno[a,*], Yutaka Suzuki[b,*], Hiroyuki Wakaguri[b], Yoshiko Sakamoto[a], Hitomi Sano[a], Naoki Osada[c], Katsuyuki Hashimoto[c], Sumio Sugano[b], Ituro Inoue[a,d]

[a] *Division of Genetic Diagnosis, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan*
[b] *Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo, 4-6-1 Shirokanedai, Minatoku, Tokyo 108-8639, Japan*
[c] *Department of Biomedical Resources, National Institute of Biomedical Innovation, Ibaraki, Osaka, Japan*
[d] *Division of Molecular Life Science, School of Medicine, Tokai University, Shimokasuya 134, Isehara, Kanagawa 259-1193, Japan*

Received 1 October 2007; revised 14 December 2007; accepted 18 December 2007

Available online 31 December 2007

Edited by Takashi Gojobori

Abstract **The liver, a major organ for drug metabolism, is physiologically similar between monkeys and humans. However, the paucity of identified genes has hampered a deep understanding of drug metabolism in monkeys. To provide such a genetic resource, 28655 expressed sequence tags (ESTs) were generated from a cynomolgus monkey liver full-length enriched cDNA library, which contained 23 unique ESTs homologous to human drug-metabolizing enzymes. Our comparative genomics approach identified nine lineage-specific candidate ESTs, including three drug-metabolizing enzymes, which could be important for understanding the physiological differences between monkeys and humans.**

Keywords: Cynomolgus monkey; Drug metabolism; Drug-metabolizing enzyme; Expressed sequence tags; Lineage-specific gene; Liver

1. Introduction

Cynomolgus monkeys have been used as an animal model for the investigation of human physiology and disease because of their close genetic and physiological similarities to humans. Application of this animal model includes predicting metabolic fate of newly developed drugs due to pharmacokinetics similar to humans. However, we now know that differences in metabolic properties are occasionally seen for some drugs between monkeys and humans [1–7] possibly due to differences in genetic components essential for drug metabolism between the two lineages, such as lineage-specific genes and alternatively spliced transcripts. However, limited numbers of lineage-specific genes identified in monkeys have hampered complete knowledge of lineage differences in drug metabolism.

An expressed sequence tag (EST)-sequencing approach has been a rapid and efficient way to identify novel cDNAs that provide a basis to investigate genetic components essential to various physiological functions. In non-human primates, efforts have been made for the comprehensive identification of ESTs in chimpanzees [8], rhesus monkeys [9,10], and cynomolgus monkeys [11–13]. However, liver tissue has not been extensively sequenced for ESTs, thus only limited genetic information is available on liver physiological function such as drug metabolism. With the completion of a draft of the rhesus monkey genome sequence [14], EST analysis of macaques should be more feasible and accurate.

To provide a monkey genetic resource, 28655 ESTs from cynomolgus monkey liver were generated. These macaque ESTs analyzed against the rhesus genome identified 1064 unique ESTs, most of which (77.0%) matched the human RefSeq database. cDNAs highly homologous to human drug-metabolizing genes were identified, including those of cytochrome P450 (CYP), UDP-glucuronosyltransferase (UGT), glutathione *S*-transferase (GST), sulfotransferase (SULT), and flavin-containing monooxygenase (FMO). Moreover, our method to select lineage-specific ESTs successfully identified novel transcripts related to drug metabolism. This genetic information should help in discerning various physiological characteristics, including drug metabolism in monkeys.

2. Materials and methods

2.1. cDNA library construction and EST sequencing

Liver samples were collected from three adult cynomolgus monkeys (two males and one female) and used to generate a full-length enriched cDNA library using the pME18S-FL3 vector by the oligo-capping method as previously described [15]. Purified DNA was sequenced using the ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit, Version 2.0 (Applied Biosystems, Foster City, CA), followed by electrophoresis with ABI-3700 DNA Analyzer (Applied Biosystems) according to the manufacturer's instructions. Primers (5′-GGATGTTGCCTTTACTTCTA-3′ and 5′-TTTTTTTTTTTT-TTTTTV-3′) were used for single-pass sequencing of 5′ and 3′-ends for each cDNA, respectively.

*Corresponding authors.
E-mail addresses: unoxx001@pharm.hokudai.ac.jp (Y. Uno), ysuzuki@k.u-tokyo.ac.jp (Y. Suzuki).

Abbreviations: CYP, cytochrome P450; EST, expressed sequence tag; FMO, flavin-containing monooxygenase; GST, glutathione *S*-transferase; ORF, open reading frame; SULT, sulfotransferase; UGT, UDP-glucuronosyltransferase

0014-5793/$32.00
doi:10.1016/j.febslet.2007.12.031

2.2. Sequence data analysis

Vector sequence was trimmed and sequence quality was inspected using Phred (University of Washington). Only EST sequences longer than 200 bases were used. Generated EST sequences were first computationally mapped to the *Macaca mulatta* genomic sequence (rheMac2, UCSC Genome Browser). Computational mapping was carried out as previously described by consequential use of sequence alignment programs, BLAT and SIM4 [16]. Only ESTs over the entire sequence length that mapped perfectly at unique positions on the macaque genome were regarded as "mapped". Further information for each cDNA is presented in our database, DBTSS (http://dbtss.hgc.jp), and a user manual has been published [16].

The macaque genomic sequences to which our cynomolgus ESTs mapped were examined for any corresponding human genomic and RefSeq sequence. If any, the corresponding macaque EST was correlated with the human RefSeq gene. Based on information from the correlated human RefSeq gene, GO (Gene Ontology) classification was carried out for macaque ESTs using GO slim (http://www.geneontology.org/) for "Biological Process", Molecular Function", and "Cellular Component".

2.3. Identification of putative macaque-specific transcripts

To identify macaque ESTs that do not match to human genes, the ESTs were analyzed by either a genome- or cDNA-based approach. In the genome-based approach, we selected the EST sequence located outside human-macaque alignable regions according to the genome–genome alignment in the UCSC Genome Browser. In the cDNA-based approach, ESTs were first searched with the human RefSeq database using BLASTN (cut-off = 1.0e−100). Those ESTs with no hits were clustered with each other (cut-off = 0.0; >98% identity) and clusters containing more than 10 ESTs were selected. Those clustered cDNAs were searched against the human RefSeq database again (1.0e−50), and the generated sequence alignments were further manually inspected. For the macaque-specific transcript candidates, complete sequences were determined by primer-walking.

3. Results and discussion

3.1. Sequencing and clustering of macaque liver ESTs

A full-length cDNA library was constructed from cynomolgus monkey (*Macaca fascicularis*) livers using the oligo-capping method [15]. One-pass sequencing at 5′ and 3′-ends of the liver cDNA clones and sequence processing generated a total of 28 655 high quality ESTs (deposited in GenBank under Accession Nos. BB873801–BB902455). Only 3′ ESTs (27 959 entries) were further analyzed. Of these ESTs, 14 727 (53%) were successfully mapped to 1064 different regions in the rhesus macaque genome. Of the 1064 regions, 819 (77%) reside in genomic regions highly homologous to human RefSeq genes as revealed by a genome–genome comparison, and were annotated with human RefSeq genes (Table 1). Clustering of 27 959 ESTs was carried out by calculating the number of ESTs that mapped to the same region, which should represent the cluster size for the corresponding gene. This analysis for the 1064 mapped regions in the genome indicated that these 1064 unique ESTs consisted of 525 contigs (49.3%) and 539 singletons (50.7%). The number of members in each cluster ranged up to 4354, with a 26.9 average. The gene expression profile based on our EST data reflected liver functional characteristics because the most abundantly expressed genes were hepatocyte-specific markers, such as albumin, fibrinogen gamma and beta polypeptides, haptoglobin, and alcohol dehydrogenase, all of which comprised more than half of the identified ESTs (Table 1). Such high redundancy of hepatic ESTs from the non-normalized cDNA libraries has been also seen for human libraries [17–19].

3.2. Functional classification of ESTs

Provisional functional classification was carried out using GO slim terms based on the human RefSeq genes that correlated with our macaque ESTs (Fig. 1). Out of 819 unique ESTs that matched a human RefSeq entry, 786 (96.0%) were assigned to at least one main category; Biological Process, Molecular Function, and Cellular Component, to which 520, 458, and 373 sequences (48.9%, 43.0%, and 35.1%) were classified, respectively. Sequences from 133 ESTs (16.2%) were annotated into all three categories. The largest EST groups include metabolism, transcription, protein biosynthesis, electron transport, transport, signal transduction, and lipid metabolism (Fig. 1A) for Biological Process, and protein binding, transferase activity, and nucleotide binding for Molecular Function (Fig. 1B).

3.3. ESTs relevant to drug metabolism

Our major objective was identification of cDNAs important for drug metabolism, namely those encoding drug-metabolizing enzymes, which belong to CYP, UGT, GST, SULT, and FMO families. The 446 ESTs in 23 clusters were highly homologous to genes for such drug-metabolizing enzymes in humans (Table 2). For the CYP family, only ESTs for the CYP1 to CYP4 subfamilies are indicated in the list because of their importance in drug metabolism. The CYP family contained 231 entries (51.8%), the largest group among the ESTs for drug-metabolizing enzymes. CYP, a phase I drug-metabolizing enzyme, is involved in hydroxylation of a large number of

Table 1
Genes abundantly expressed in liver (>100 reads)

Contig number	ESTs	Human RefSeq ID	Annotation
15043	4354	NM_000477	Albumin
15223	3763	NM_000509	Fibrinogen gamma chain
15577	1097	NM_005141	Fibrinogen beta chain
10429	277	NM_005143	Haptoglobin
19271	253	NM_000668	Alcohol dehydrogenase IB (class I), beta polypeptide
17733	245	NM_001085	Serpin peptidase inhibitor, clade A, member 3
5070	227	NM_000035	Aldolase B, fructose-bisphosphate
6405	158	NM_016413	Carboxypeptidase B2 (plasma, carboxypeptidase U)
5438	154	NM_000638	Vitronectin
11141	130	NM_000354	Serpin peptidase inhibitor, clade A, member 7
12017	125	NM_001622	Alpha-2-HS-glycoprotein
15222	109	NM_000508	Fibrinogen alpha chain
17083	106	NM_001756	Serpin peptidase inhibitor, clade A, member 6

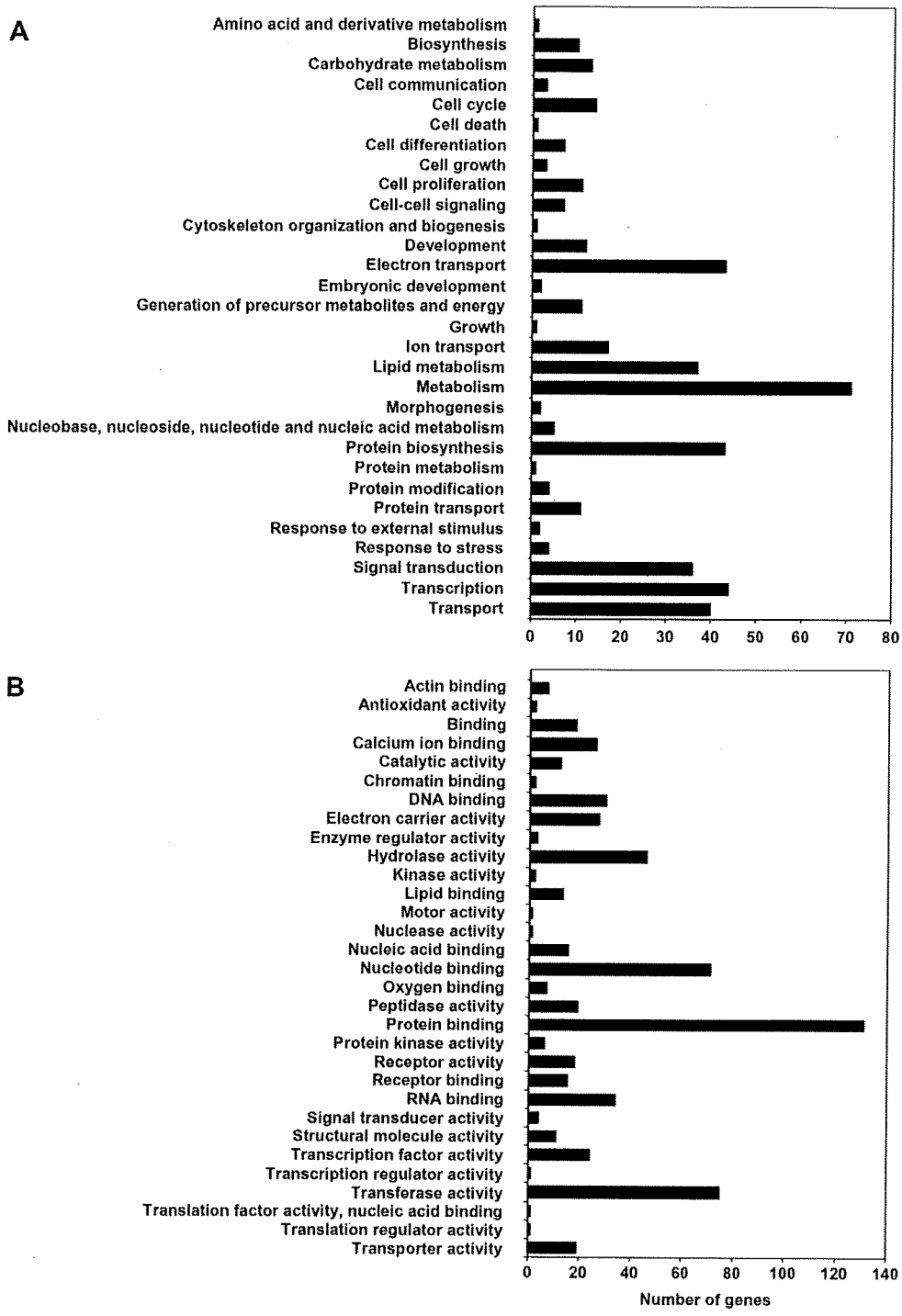


Fig. 1. Functional classification of cynomolgus liver ESTs. All non-redundant ESTs were assigned to each functional category as described in Section 2. Biological process (A) and Molecular function (B) are shown.

drugs [20]. Among the CYP ESTs identified, 124 (53.7%) belonged to the CYP2C subfamily that is important for metabolism of ~20% of all prescribed drugs such as tolbutamide, phenytoin, and warfarin [21]. Fifty-two ESTs belonged to the CYP3A subfamily. In humans, genes in the CYP3A subfamily (especially *CYP3A4*) are essential for drug metabolism, and are involved in the metabolism of more than half the currently prescribed drugs. Moreover, human CYP3A4 and CYP3A5 occupy more than half of the total CYP protein content in liver [20], contributing substantially overall drug

Table 2
Cynomolgus ESTs highly homologous to human drug-metabolizing enzyme families, CYP, UGT, GST, SULT, and FMO

Family	Contig number	Number of ESTs	Accession number	Matched human cDNA
CYP	18729	89	NM_0007700	CYP2C8
	12912	38	NM_017460	CYP3A4
	463	30	NM_000106	CYP2D6
	19109	21	NM_000769	CYP2C19
	19111	14	NM_000771	CYP2C9
	12910	14	NM_000777	CYP3A5
	19262	13	NM_000773	CYP2E1
	7671	8	NM_023944	CYP4F12
	8451	3	NM_000775	CYP2J2
	8410	1	NM_000778	CYP4A11
UGT	15423	75	NM_001074	UGT2B7
	2531	41	NM_019093	UGT1A3
	15424	30	NM_050394	UGT2B28
GST	19270	9	NM_145740	GSTA1
	14165	8	NM_000846	GSTA2
	9619	7	NM_000851	GSTM5
	1616	7	NM_145792	MGST1
	19167	3	NM_004832	GSTO1
	17697	2	NM_145870	GSTZ1
SULT	10631	10	NM_001055	SULT1A1
	19266	2	NM_001054	SULT1A2
	3203	1	NM_006588	SULT1C2
FMO	10051	20	NM_001002294	FMO3

metabolism in humans. Thirty ESTs were highly similar to human CYP2D6. In the human genome, three *CYP2D* genes are present including one functional *CYP2D* gene (*CYP2D6*) and two pseudogenes (*CYP2D7* and *CYP2D8*). CYP2D6 accounts for 5% of the total hepatic CYP content and is responsible for the metabolism of 25% of all drugs oxidized by CYPs [20]. In cynomolgus monkeys, CYP2D17, which is highly homologous to human CYP2D6, has been isolated [22]. Meanwhile, marmoset is known to have two functional CYP2Ds with different metabolic properties, CYP2D19 and CYP2D30 [23]. Further in-depth analysis of our EST clones could reveal whether *CYP2D17* is the only *CYP2D* gene expressed in cynomolgus monkey liver. Characterization of these CYP EST clones is currently in progress, such as full-length sequencing, tissue expression patterns, and metabolic assays, the outcome of which has been partly published [24,25].

Clusters for other drug-metabolizing enzymes of UGT, GST, SULT, and FMO families contained 146, 36, 13, and 20 ESTs, respectively (Table 2). UGT, a phase II drug-metabolizing enzyme, catalyzes the conjugation of various drugs to assist drug excretion and is composed of UGT1A, UGT2A, and UGT2B subfamilies in humans. The 146 ESTs for the UGT family were grouped into three clusters. Forty-one EST sequences were highly homologous to human UGT1A3. The *UGT1A* gene locus contains 13 distinct first exons (promoters) followed by exons 2–5 that are shared among all 13 transcripts, giving rise to nine different proteins (four pseudogenes) in humans [26]. Considering that these ESTs were 3′ cDNAs, the 41 EST clones possibly encode multiple *UGT1A* genes. Because only four *UGT1A* genes have been identified for macaques, further sequence analysis of these UGT1A ESTs could lead to the identification of novel *UGT1A* genes in this lineage. Seventy-five and 30 ESTs matched human UGT2B7 and UGT2B28, respectively. Initial full-length sequencing of the UGT2B EST clones revealed that the clones contained the UGT2B33 cDNA (GenBank Accession No. AB371703) newly identified in cynomolgus monkeys as well as the previously identified cDNAs for cynomolgus UGT2B9, UGT2B18, UGT2B19, UGT2B20, UGT2B23, and UGT2B30 (GenBank Accession Nos. U91582, AF016310, AF112112, AF072223, AF112113, and AF401657, respectively). In contrast to *UGT1A*, *UGT2B* genes have been frequently duplicated in many mammalian lineages [26]; therefore, some of these *UGT2Bs* are possibly lineage-specific genes as discussed below.

Other EST sequences were highly homologous to six human genes, two genes, and one gene in the GST, SULT, and FMO families, respectively (Table 2). GST is another phase II enzyme, catalyzing the conjugation of electrophilic substrates to glutathione and is composed of at least 16 genes for cytosolic, mitochondrial, and microsomal GSTs in humans [27]. SULT is also a gene family comprising at least 10 human genes, catalyzing sulfate conjugation of a wide variety of drugs [28]. FMO is a family of flavoproteins, catalyzing oxygenation of various drugs containing sulfur, nucleophilic nitrogen, and phosphorus heteroatoms [29]. Full-length sequencing and functional characterization of these EST clones are currently under investigation, by which novel genes could be identified because limited numbers of genes have been identified for these enzymes in monkeys. These results suggest that our EST-sequencing approach successfully identified a number of cDNA clones for various drug-metabolizing enzymes in monkeys.

3.4. Identification of lineage-specific genes

In order to better utilize monkeys as an animal model, it is essential to understand similarities or differences in genes expressed between monkeys and humans. The EST data should provide essential information on lineage-specific genes and transcripts. To identify macaque-specific transcripts, 27959 3′ ESTs were analyzed by either a genome- or cDNA-based approach. In the genome-based approach, we found 77 EST clusters, for which at least a part of the sequences were located outside human-macaque alignable regions. In the cDNA-based approach, we identified 12 clusters containing >10 ESTs that were unmatched to any human RefSeq genes according to BLASTN (cut-off = 1.0e−100). Clones available for the 10 remaining candidate clusters after subsequent manual inspection, along with clones for the 29 clusters randomly selected from 77 candidates in the genome-based approach, were subjected to full-length sequencing (excluding 1 overlapping clone). Sequence analysis of these 38 clones confirmed that nine clones contained lineage-specific candidate genes. Of these, six clones matched to human RefSeq sequences (Table 3): two clones lack a portion of human genome sequence and the other four matched to more than one member of a gene family. Thus, these four clones were potentially lineage-specific genes and were further characterized as described below.

One candidate clone (Qlv-U097A-G10) encoded CYP2C76 with a relatively low homology (~80%) to members of the human CYP2C subfamily, CYP2C8, CYP2C9, CYP2C18, and CYP2C19 (Table 3). The extent of homology was much lower than those for other ESTs (~95%). Our characterization

Table 3
Potential lineage-specific ESTs in cynomolgus monkey

Clone ID	GenBank Accession number	Nucleotide (bp)	ORF[a] (Number of amino acids)	Cynomolgus sequence	The most highly homologous human RefSeq cDNAs	Genome- or cDNA-based approach	Aligned location
Novel member of gene family							
Qlv-U042A-F11	AB362497	1637	454	None	CFH, CFHR3/4	Genome/cDNA	Intergenic
Qlv-U097A-G10	AB362507	1986	489	CYP2C76	CYP2C8/9/18/19	cDNA	Intergenic
Qlv-U346A-B11[b]	AB371605	1758	472	CYP2A23	CYP2A6/7/13	cDNA	Intergenic
Qlv-U405A-G11	AB362508	2225	528	UGT2B19	UGT2B4	cDNA	Intergenic
Partially unmatched to human genome							
Qlv-U244A-C6[b]	AB362499	1612	305	None	TSPAN12	Genome	Intergenic
Qlv-U258A-D7[b]	AB362500	1984	89	None	SS18L1	Genome	Intergenic
Unmatched to human genome							
Qlv-U050A-D10	AB362503	1700	34	None	None	Genome	Intron
Qlv-U295A-A3	AB362504	2278	118	None	None	Genome	Intergenic
Qlv-U389A-C1	AB362506	2043	90	None	None	Genome	Intergenic

[a]The longest ORF was selected.
[b]Transcript variants with different exon-intron structure from human homologs.

of CYP2C76 (GenBank Accession No. DQ074807) at the RNA, protein, and genomic level revealed that this CYP2C did not have any human ortholog because the corresponding genes were not found in the human genome [24]. Moreover, this CYP2C76 was at least partly responsible for lineage differences in drug metabolism [30]. These results confirmed that our comparative genomic approach succeeded in identifying macaque-specific transcripts that are absent in humans.

One clone (Qlv-U405A-G11) identified as a lineage-specific candidate contained the cDNA for UGT2B19 previously reported [31]. Cynomolgus UGT2B19 as well as UGT2B30 cDNAs were both highly homologous (92%) to human UGT2B4 cDNA [32]. A phylogenetic comparison (Fig. 2) indicated that the 1-to-1 orthologous relationship to the human UGT2Bs could not be determined for these cynomolgus UGT2Bs, raising the possibility that *UGT2B19* might be a lineage-specific gene. *UGT2B19* is expressed in cynomolgus monkey liver and prostate and has enzymatic activity to xenobiotics (1-naphthol) and steroids (testosterone) [31]. The UGT2B subfamily consisted of a number of member genes including a lineage-specific candidate [26], suggesting that UGT2B19 and other functional UGT2B enzymes in cynomolgus monkeys contribute not only to overall drug metabolism in monkey liver but also possibly to differences in drug metabolism.

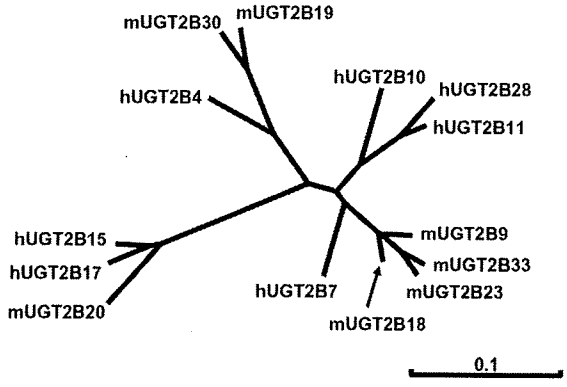


Fig. 2. A phylogenetic comparison of UGTs between macaque and human. The phylogenetic tree was based on amino acid sequence using the Clustal W program. Deduced amino acid sequences were used for cynomolgus monkeys (m) and human (h).

Another lineage-specific candidate clone (Qlv-U346A-B11) was cynomolgus CYP2A23 variant (tentatively named CYP2A23v), containing exons 1–8 with a partial intron 8 sequence and thus, lacking the entire exon 9 as compared to a complete CYP2A23 transcript. CYP2A23 and another cynomolgus CYP2A, CYP2A24, were both highly homologous (~95%) to the three human CYP2As, specifically CYP2A6, CYP2A7, and CYP2A13, indicating the difficulty in determining the orthologous relationship of CYP2A23 and CYP2A24 to human CYP2As [25]. This novel CYP2A23 transcript variant encodes a protein of 472 amino acids and lacks a part of a heme-binding region essential for CYP proteins (Fig. 3). The protein generated from this transcript, therefore, might not function as a drug-metabolizing enzyme. A similar transcript variant was also identified for CYP2C76 and UGT2B19 (data not shown). It remains to be determined whether the presence of these transcripts lacking a functional domain is limited to the animals that provided liver samples for the cDNA library construction and what roles these transcript variants play in drug metabolism.

Other than those for drug metabolism, one lineage-specific candidate (Qlv-U042A-F11) had high sequence homology to complement factor H (CFH) family genes in humans (Table 3). CFH (also called Factor H), an important complement regulator, forms a gene family along with CFH-related proteins (CFHL1-5) in humans [33]. This macaque transcript contained an open reading frame (ORF) of 454 amino acids. CFH and other genes important for immune response and T cell-mediated immunity such as *immunoglobulin-like* genes and MHC-related genes have been identified in macaques as the genes that went under positive selection [13,14,34], and thus, our finding of lineage-specific CFH-like sequence in macaques is not surprising. Further analysis of this CFH-like sequence indicated that the first 19 amino acids and the remaining amino acids were highly similar to CFH-related proteins (CFHR3 and CFHR4) and CFH in humans, respectively (data not shown), raising the possibility that this novel transcript might be a hybrid of CFH and CFH-related genes. In humans, a hybrid transcript of CFH and CFHR1 has been identified and implicated in atypical haemolytic uraemic syndrome [35].

```
mCYP2A23v   1:MLASGLLLVALLACLTVMVLMSVWQQRNSKGKLPPGPTPLPFIGNYLQLNTEQMYNSLMKISERYGPVFTIHLGPRRVVVLCGYDAVKEALVDQAEEFSG 100
mCYP2A23    1:MLASGLLLVALLACLTVMVLMSVWQQRNSKGKLPPGPTPLPFIGNYLQLNTEQMYNSLMKISERYGPVFTIHLGPRRVVVLCGYDAVKEALVDQAEEFSG 100
mCYP2A24    1:MLASGLLLVALLACLTVMVLMSVWQQRNSKGKLPPGPTPLPFIGNYLQLNTEQMCNSIMKISERYGPVFTIHLGPRRVVVLCGYDAVKEALVDQAEEFSG 100
hCYP2A6     1:MLASGMLLVALLVCLTVMVLMSVWQQRKSKGKLPPGPTPLPFIGNYLQLNTEQMYNSLMKISERYGPVFTIHLGPRRVVVLCGHDAVREALVDQAEEFSG 100
hCYP2A7     1:MLASGLLLVALLACLTVMVLMSVWQQRKSRGKLPPGPTPLPFIGNYLQLNTEHICDSIMKFSECYGPVFTIHLGPRRVVVLCGHDAVREALVDQAEEFSG 100
hCYP2A13    1:MLASGLLLVTLLACLTVMVLMSVWRQRKSRGKLPPGPTPLPFIGNYLQLNTEQMYNSLMKISERYGPVFTIHLGPRRVVVLCGHDAVKEALVDQAEEFSG 100
              *****.***.**.***********.** *.**********************....*.**.**.****************** ***.***********

mCYP2A23v 101:RGEQATFDWLFKGYGVVFSNGERAKQLRRFSIATLRDFGVGKRGIEERIQEEAGFLIEALRDTQGANIDPTFFLSRTVSNVISSIVFGDRFDYEDKEFLS 200
mCYP2A23  101:RGEQATFDWLFKGYGVVFSNGERAKQLRRFSIATLRDFGVGKRGIEERIQEEAGFLIEALRDTQGANIDPTFFLSRTVSNVISSIVFGDRFDYEDKEFLS 200
mCYP2A24  101:RGEQATFDWVFKGYGVVFSNGERAKQLRRFSIATLRDFGVGKRGIEERIQEEAGFLIEALRDTHGANIDPTFFLSRTVSNVISSIVFGDRFDYKDKEFLS 200
hCYP2A6   101:RGEQATFDWVFKGYGVVFSNGERAKQLRRFSIATLRDFGVGKRGIEERIQEEAGFLIDALRGTGGANIDPTFFLSRTVSNVISSIVFGDRFDYKDKEFLS 200
hCYP2A7   101:RGEQATFDWVFKGYGVAFSNGERAKQLLRFAIATLRDFGVGKRGIEERIQEESGFLIEAIRSTHGANIDPTFFLSRTVSNVISSIVFGDRFDYEDKEFLS 200
hCYP2A13  101:RGEQATFDWLFKGYGVAFSNGERAKQLRRFSIATLRGFGVGKRGIEERIQEEAGFLIDALRGTHGANIDPTFFLSRTVSNVISSIVFGDRFDYEDKEFLS 200
              ********* ******.**********.**.******.***************.****.*.* * ******************************.******

mCYP2A23v 201:LLRMMLGSFQFTATSAGQLYEMFSSVMKHLPGPQQQAFKELQGLEDFIAKKVEHNRRTLDPNSPRDFIDSFLIRMQEEEKNPNTEFHLKNLVLTSLNLFF 300
mCYP2A23  201:LLRMMLGSFQFTATSAGQLYEMFSSVMKHLPGPQQQAFKELQGLEDFIAKKVEHNRRTLDPNSPRDFIDSFLIRMQEEEKNPNTEFHLKNLVLTSLNLFF 300
mCYP2A24  201:LLGMMLAIFQFTSTSTGQLYEMFSSVMKHLPGPQQQAFKELQGLEDFIAKKVEHNQRTLDPNSPRDFIDSFLIRMQEEEKNPNTEFYLKNLMMTTLNLFI 300
hCYP2A6   201:LLRMMLGIFQFTSTSTGQLYEMFSSVMKHLPGPQQQAFQLLQGLEDFIAKKVEHNQRTLDPNSPRDFIDSFLIRMQEEEKNPNTEFYLKNLVMTTLNLFI 300
hCYP2A7   201:LLSMMLGIFQFTSTSTGQLYEMFSSVMKHLPGPQQQAFKLLQGLEDFIAKKVEHNQRTLDPNSPQDFIDSFLIHMQEEEKNPNTEFYLKNLMMSTLNLFI 300
hCYP2A13  201:LLRMMLGSFQFTATSTGQLYEMFSSVMKHLPGPQQQAFKELQGLEDFIAKKVEHNQRTLDPNSPRDFIDSFLIRMQEEEKNPNTEFYLKNLVMTTLNLFF 300
              **.***. **** **.**********************..***************.********.********.************.****....****

mCYP2A23v 301:GGTETVSTTLRYGFLLLMKHPEVEAKVHEEIDRVIGKNRQPKFEDWAKMPYTEAVIHEIQRFGDMLPFGVAHRVIKDTKFRDFFLPKGTEVFPMLGSVLK 400
mCYP2A23  301:GGTETVSTTLRYGFLLLMKHPEVEAKVHEEIDRVIGKNRQPKFEDWAKMPYTEAVIHEIQRFGDMLPFGVAHRVIKDTKFRDFFLPKGTEVFPMLGSVLK 400
mCYP2A24  301:AGTETVSTTLRYGFLLLMKYPEVEAKVHEEIDRVIGKNRQPKFEDRVKMPYMEAVIHEIQRFGDVIPMSLARRVNKDTKFRDFFLPKGTEVFPMLGSVLR 400
hCYP2A6   301:GGTETVSTTLRYGFLLLMKHPEVEAKVHEEIDRVIGKNRQPKFEDRAKMPYMEAVIHEIQRFGDVIPMSLARRVKKDTKFRDFFLPKGTEVFPMLGSVLR 400
hCYP2A7   301:AGTETVSTTLRYGFLLLMKHPEVEAKVHEEIDRVIGKNRQPKFEDRTKMPYMEAVIHEIQRFGDVIPMSLARRVKKDTKFRDFFLPKGTEVFPMLGSVLR 400
hCYP2A13  301:AGTETVSTTLRYGFLLLMKHPEVEAKVHEEIDRVIGKNRQPKFEDRAKMPYTEAVIHEIQRFGDMLPMGLAHRVNKDTKFRDFFLPKGTEVFPMLGSVLR 400
              ******************.**************************..**** ************  *. .* ** ************************.

mCYP2A23v 401:DPKFFSNPQDFNPQHFLDEKGQFKKSDAFVPFSIGKRPLFAARPLLTPAGASLTQLPSVLCSLVFLQLGSSC                              472
mCYP2A23  401:DPKFFSNPQDFNPQHFLDEKGQFKKSDAFVPFSIGKRNCFGEGLARMELFLFLTTIMQNFRFKSPQSPKDIDVSPKHMGFATIPPNYTMSFLPR      494
mCYP2A24  401:DPRFFSNPQDFNPQHFLDEKGQFKKSDAFVPFSIGKRNCFGEGLARMELFLFFTTIMQNFRFKSPQSPKDIDVSPKHAGFATIPRNYTMSFLPR      494
hCYP2A6   401:DPSFFSNPQDFNPQHFLNEKGQFKKSDAFVPFSIGKRNCFGEGLARMELFLFFTTVMQNFRLKSSQSPKDIDVSPKHVGFATIPRNYTMSFLPR      494
hCYP2A7   401:DPSFFSNPQDFNPQHFLDDKGQFKKSDAFVPFSIGKRNCFGEGLARMELFLFFTTVMQNFRLKSSQSPKDIDVSPKHVVFATIPRNYTMSFLPR      494
hCYP2A13  401:DPRFFSNPRDFNPQHFLDKKGQFKKSDAFVPFSIGKRYCFGEGLARMELFLFFTTIMQNFRFKSPQSPKDIDVSPKHVGFATIPRNYTMSFLPR      494
              ** *****.********..*******************..*.............*. ..... .. .......
```

Fig. 3. Alignment of the amino acid sequences deduced from cynomolgus monkey (m) and human (h) CYP2A cDNAs. The sequences were aligned using the Clustal W program. Asterisks and dots under the sequences indicate identical amino acids and conservatively unchanged amino acids, respectively. A black line under the amino acid sequences indicates the putative heme-binding region. The CYP2A23 variant (mCYP2A23v) newly identified lacks half of the putative heme-binding region.

These results suggest that our approach of lineage-specific gene identification successfully identified potential lineage-specific genes or transcripts, possibly relevant to the immune system. Further investigation of other ESTs should help make better use of the macaque for immunological studies.

Three candidate genes were unmatched to any human RefSeq sequence, and thus could be apparent lineage-specific genes (Table 3). The two candidate genes (Qlv-U295A-A3 and Qlv-U389A-C1) reside in intergenic regions of the macaque genome, which might be novel genes in monkeys. This was supported by the fact that these two sequences did not match any human ESTs by BLAST (data not shown). The two transcripts contained relatively small ORFs (<100 amino acids). Transcripts with small ORFs have been identified in mice and humans, some of which could be actually translated in vitro [36,37]. Alternatively, these mRNAs might be functioning as non-coding RNAs. A large proportion of transcripts are non-coding RNAs, including those having essential functions in transcriptional and translational control [38,39].

4. Conclusion

The data presented here provide an overview of genes expressed in cynomolgus liver to investigate liver physiology for macaques. ESTs for genes encoding a variety of drug-metabolizing enzymes hold great promise in deepening our understanding of drug metabolism in monkeys, which in turn helps to elucidate lineage differences between monkeys and humans. Indeed, our characterization of CYP2C ESTs has identified lineage-specific CYP2C76, which is responsible for lineage differences in drug metabolism [24,30]. Furthermore, the ESTs generated in this study can be a resource for the production of microarrays. Given that our cDNA library was generated with RNAs from only three animals, the EST sequencing using the library originated from the RNA samples of more animals would be useful for the identification of the allelic variants expressed in liver.

Many drug-metabolizing enzyme genes are confined to gene families, many of which have been subjected to gene duplication or gene loss during evolution, resulting in family size differences [40]. This indicates that lineage-specific genes could be identified for gene families even between evolutionarily close lineages such as monkeys and humans. Moreover, physiological differences should partly result from differences at the transcriptional level, for example, by alternative splicing and non-coding RNAs [41]. Further investigation of our EST data will lead to the identification of lineage-specific transcripts generated by alternative splicing and lineage-specific gene gain/loss, as the efforts for identifying such transcripts have succeeded partly in macaques [9,13]. The identified lineage-specific transcripts and genes will help lead to a better understanding

of the physiological differences between monkeys and humans, leading to more efficient utilization of monkeys as an animal model.

References

[1] Stevens, J.C., Shipley, L.A., Cashman, J.R., Vandenbranden, M. and Wrighton, S.A. (1993) Comparison of human and rhesus monkey *in vitro* phase I and phase II hepatic drug metabolism activities. Drug Metab. Dispos. 21, 753–760.

[2] Sharer, J.E., Shipley, L.A., Vandenbranden, M.R., Binkley, S.N. and Wrighton, S.A. (1995) Comparisons of phase I and phase II *in vitro* hepatic enzyme activities of human, dog, rhesus monkey, and cynomolgus monkey. Drug Metab. Dispos. 23, 1231–1241.

[3] Guengerich, F.P. (1997) Comparisons of catalytic selectivity of cytochrome *P*450 subfamily enzymes from different species. Chem. Biol. Interact. 106, 161–182.

[4] Shimada, T., Mimura, M., Inoue, K., Nakamura, S., Oda, H., Ohmori, S. and Yamazaki, H. (1997) Cytochrome P450-dependent drug oxidation activities in liver microsomes of various animal species including rats, guinea pigs, dogs, monkeys, and humans. Arch. Toxicol. 71, 401–408.

[5] Weaver, R.J., Dickins, M. and Burke, M.D. (1999) A comparison of basal and induced hepatic microsomal cytochrome P450 monooxygenase activities in the cynomolgus monkey (*Macaca fascicularis*) and man. Xenobiotica 29, 467–482.

[6] Bogaards, J.J., Bertrand, M., Jackson, P., Oudshoorn, M.J., Weaver, R.J., van Bladeren, P.J. and Walther, B. (2000) Determining the best animal model for human cytochrome P450 activities: a comparison of mouse, rat, rabbit, dog, micropig, monkey and man. Xenobiotica 30, 1131–1152.

[7] Narimatsu, S., Kobayashi, N., Masubuchi, Y., Horie, T., Kakegawa, T., Kobayashi, H., Hardwick, J.P., Gonzalez, F.J., Shimada, N., Ohmori, S., Kitada, M., Asaoka, K., Kataoka, H., Yamamoto, S. and Satoh, T. (2000) Species difference in enantioselectivity for the oxidation of propranolol by cytochrome P450 2D enzymes. Chem. Biol. Interact. 127, 73–90.

[8] Sakate, R., Osada, N., Hida, M., Sugano, S., Hayasaka, I., Shimohira, N., Yanagi, S., Suto, Y., Hashimoto, K. and Hirai, M. (2003) Analysis of 5′-end sequences of chimpanzee cDNAs. Genome Res. 13, 1022–1026.

[9] Magness, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., Agy, M.B., Proll, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G. and Iadonato, S.P. (2005) Analysis of the *Macaca mulatta* transcriptome and the sequence divergence between *Macaca* and human. Genome Biol. 6, R60.

[10] Li, Y. and Su, B. (2006) No accelerated evolution of 3′UTR region in human for brain-expressed genes. Gene 383C, 38–42.

[11] Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirata, M., Suto, Y., Hirai, M., Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K. (2001) Assignment of 118 novel cDNAs of cynomolgus monkey brain to human chromosomes. Gene 275, 31–37.

[12] Osada, N., Hida, M., Kusuda, J., Tanuma, R., Hirata, M., Suto, Y., Hirai, M., Terao, K., Sugano, S. and Hashimoto, K. (2002) Cynomolgus monkey testicular cDNAs for discovery of novel human genes in the human genome sequence. BMC Genomics 3, 36.

[13] Chen, W.H., Wang, X.X., Lin, W., He, X.W., Wu, Z.Q., Lin, Y., Hu, S.N. and Wang, X.N. (2006) Analysis of 10,000 ESTs from lymphocytes of the cynomolgus monkey to improve our understanding of its immune system. BMC Genomics 7, 82.

[14] The Rhesus Macaque Genome Sequencing and Analysis Consortium (2007) Evolutionary and biomedical insights from the rhesus macaque genome. Science 316, 222–234.

[15] Suzuki, Y. and Sugano, S. (2003) Construction of a full-length enriched and a 5′-end enriched cDNA library using the oligo-capping method. Method Mol. Biol. 221, 73–91.

[16] Yamashita, R., Suzuki, Y., Wakaguri, H., Tsuritani, K., Nakai, K. and Sugano, S. (2006) DBTSS: database of human transcription start sites, progress report 2006. Nucl. Acid Res. 34, D86–89.

[17] Xu, X.R., Huang, J., Xu, Z.G., Qian, B.Z., Zhu, Z.D., Yan, Q., Cai, T., Zhang, X., Xiao, H.S., Qu, J., Liu, F., Huang, Q.H., Cheng, M., Li, N.G., Du, J.J., Hu, W., Shen, K.T., Lu, G., Fu, G., Zhong, M., Xu, S.H., Gu, W.Y., Huang, W., Zhao, X.T., Hu, G.X., Gu, J.R., Chen, Z. and Han, Z.G. (2001) Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. Proc. Natl. Acad. Sci. USA 98, 15089–15094.

[18] Yu, Y., Zhang, C., Zhou, G., Wu, S., Qu, X., Wei, H., Xing, G., Dong, C., Zhai, Y., Wan, J., Ouyang, S., Li, L., Zhang, S., Zhou, K., Zhang, Y., Wu, C. and He, F. (2001) Gene expression profiling in human fetal liver and identification of tissue- and developmental-stage-specific genes through compiled expression profiles and efficient cloning of full-length cDNAs. Genome Res. 11, 1392–1403.

[19] Otsuka, M., Arai, M., Mori, M., Kato, M., Kato, N., Yokosuka, O., Ochiai, T., Takiguchi, M., Omata, M. and Seki, N. (2003) Comparing gene expression profiles in human liver, gastric, and pancreatic tissues using full-length-enriched cDNA libraries. Hepatol. Res. 27, 76–82.

[20] Guengerich, F.P. (2005) Human cytochrome P450 enzymes in: (Ortiz de Montellano, P., Ed.), third ed, Cytochrome P450: Structure, Mechanism, and Biochemistry, pp. 377–530, Kluwer Academic/Plenum Publishers, New York.

[21] Goldstein, J.A. (2001) Clinical relevance of genetic polymorphisms in the human CYP2C subfamily. Brit. J. Clin. Pharmacol. 52, 349–355.

[22] Mankowski, D.C., Laddison, K.J., Christopherson, P.A., Ekins, S., Tweedie, D.J. and Lawton, M.P. (1999) Molecular cloning, expression, and characterization of CYP2D17 from cynomolgus monkey liver. Arch. Biochem. Biophys. 372, 189–196.

[23] Hichiya, H., Kuramoto, S., Yamamoto, S., Shinoda, S., Hanioka, N., Narimatsu, S., Asaoka, K., Miyata, A., Iwata, S., Nomoto, M., Satoh, T., Kiryu, K., Ueda, N., Naito, S., Tucker, G.T. and Ellis, S.W. (2004) Cloning and functional expression of a novel marmoset cytochrome P450 2D enzyme, CYP2D30: comparison with the known marmoset CYP2D19. Biochem. Pharmacol. 68, 165–175.

[24] Uno, Y., Fujino, H., Kito, G., Kamataki, T. and Nagata, R. (2006) CYP2C76, a novel CYP in cynomolgus monkey, is a major CYP2C in liver, metabolizing tolbutamide and testosterone. Mol. Pharmacol. 70, 477–486.

[25] Uno, Y., Hosaka, S., Matsuno, K., Nakamura, C., Kito, G., Kamataki, T. and Nagata, R. (2007) Characterization of cynomolgus monkey cytochrome P450 (CYP) cDNAs: Is *CYP2C76* the only monkey-specific *CYP* gene responsible for species differences in drug metabolism? Arch. Biochem. Biophys. 466, 98–105.

[26] Mackenzie, P.I., Bock, K.W., Burchell, B., Guillemette, C., Ikushiro, S., Iyanagi, T., Miners, J.O., Owens, I.S. and Nebert, D.W. (2005) Nomenclature update for the mammalian UDP glycosyltransferase (*UGT*) gene superfamily. Pharmacogenet. Genomics 15, 677–685.

[27] Hayes, J.D., Flanagan, J.U. and Jowsey, I.R. (2005) Glutathione transferases. Annu. Rev. Pharmacol. Toxicol. 45, 51–88.

[28] Blanchard, R.L., Freimuth, R.R., Buck, J., Weinshilboum, R.M. and Coughtrie, M.W. (2004) A proposed nomenclature system for the cytosolic sulfotransferase (SULT) superfamily. Pharmacogenetics 14, 199–211.

[29] Cashman, J.R. and Zhang, J. (2006) Human flavin-containing monooxygenases. Annu. Rev. Pharmacol. Toxicol. 46, 65–100.

[30] Uno, Y., Kumano, T., Kito, G., Nagata, R., Kamataki, T. and Fujino, H. (2007) CYP2C76-mediated species difference in drug metabolism: A comparison of pitavastatin metabolism between monkeys and humans. Xenobiotica 37, 30–43.

[31] Belanger, G., Barbier, O., Hum, D.W. and Belanger, A. (1999) Molecular cloning, expression and characterization of a monkey steroid UDP-glucuronosyltransferase, UGT2B19, that conjugates testosterone. Eur. J. Biochem. 260, 701–708.

[32] Girard, C., Barbier, O., Turgeon, D. and Belanger, A. (2002) Isolation and characterization of the monkey *UGT2B30* gene that encodes a uridine diphosphate-glucuronosyltransferase enzyme active on mineralocorticoid, glucocorticoid, androgen and oestrogen hormones. Biochem. J. 365, 213–222.

[33] Male, D.A., Ormsby, R.J., Ranganathan, S., Giannakis, E. and Gordon, D.L. (2000) Complement factor H: sequence analysis of

221 kb of human genomic DNA containing the entire *fH*, *fHR-1* and *fHR-3* genes. Mol. Immunol. 37, 41–52.
[34] Geraghty, D.E., Daza, R., Williams, L.M., Vu, Q. and Ishitani, A. (2002) Genetics of the immune response: identifying immune variation within the MHC and throughout the genome. Immunol. Rev. 190, 69–85.
[35] Venables, J.P., Strain, L., Routledge, D., Bourn, D., Powell, H.M., Warwicker, P., Diaz-Torres, M.L., Sampson, A., Mead, P., Webb, M., Pirson, Y., Jackson, M.S., Hughes, A., Wood, K.M., Goodship, J.A. and Goodship, T.H. (2006) Atypical haemolytic uraemic syndrome associated with a hybrid complement gene. PLoS Med. 3, e431.
[36] Oyama, M., Itagaki, C., Hata, H., Suzuki, Y., Izumi, T., Natsume, T., Isobe, T. and Sugano, S. (2004) Analysis of small human proteins reveals the translation of upstream open reading frames of mRNAs. Genome Res. 14, 2048–2052.
[37] Frith, M.C., Forrest, A.R., Nourbakhsh, E., Pang, K.C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Bailey, T.L. and Grimmond, S.M. (2006) The abundance of short proteins in the mammalian proteome. PLoS Genet. 2, e52.
[38] Gustincich, S., Sandelin, A., Plessy, C., Katayama, S., Simone, R., Lazarevic, D., Hayashizaki, Y. and Carninci, P. (2006) The complexity of the mammalian transcriptome. J. Physiol. 575, 321–332.
[39] Prasanth, K.V. and Spector, D.L. (2007) Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. Genes Dev. 21, 11–42.
[40] Demuth, J.P., Bie, T.D., Stajich, J.E., Cristianini, N. and Hahn, M.W. (2006) The evolution of mammalian gene families. PLoS ONE 1, e85.
[41] Mattick, J.S. and Makunin, I.V. (2006) Non-coding RNA. Hum. Mol. Genet. 15, R17–R29.

BMC Genomics

BioMed Central

Research article

Open Access

Large-scale analysis of *Macaca fascicularis* transcripts and inference of genetic divergence between *M. fascicularis* and *M. mulatta*

Naoki Osada*1, Katsuyuki Hashimoto1, Yosuke Kameoka1, Makoto Hirata1, Reiko Tanuma1, Yasuhiro Uno2, Itsuro Inoue3, Munetomo Hida4, Yutaka Suzuki5, Sumio Sugano5, Keiji Terao6, Jun Kusuda1 and Ichiro Takahashi1

Address: 1Department of Biomedical Resources, National Institute of Biomedical Innovation, Ibaraki, Japan, 2Pharmacokinetics and Bioanalysis Center, Shin Nippon Biomedical Laboratories, Ltd., Kainain, Japan, 3Division of Genetic Diagnosis, Institute of Medical Science, University of Tokyo, Tokyo, Japan, 4International Research and Educational Institute for Integrated Medical Sciences, Tokyo Women's Medical University, Tokyo, Japan, 5Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan and 6Tsukuba Primate Center for Medical Science, National Institute of Biomedical Innovation, Tsukuba, Japan

Email: Naoki Osada* - nosada@nibio.go.jp; Katsuyuki Hashimoto - khashi@nih.go.jp; Yosuke Kameoka - ykameoka@nibio.go.jp; Makoto Hirata - mhirata@nibio.go.jp; Reiko Tanuma - tanumark@nibio.go.jp; Yasuhiro Uno - unoxx001@pharm.hokudai.ac.jp; Itsuro Inoue - ituro@ims.u-tokyo.ac.jp; Munetomo Hida - hida@imcir.twmu.ac.jp; Yutaka Suzuki - ysuzuki@hgc.jp; Sumio Sugano - ssugano@ims.u-tokyo.ac.jp; Keiji Terao - terao@nibio.go.jp; Jun Kusuda - jkusuda@nibio.go.jp; Ichiro Takahashi - ichiro-t@nibio.go.jp

* Corresponding author

Published: 24 February 2008

BMC Genomics 2008, **9**:90 doi:10.1186/1471-2164-9-90

Received: 27 September 2007

Accepted: 24 February 2008

This article is available from: http://www.biomedcentral.com/1471-2164/9/90

Abstract

Background: Cynomolgus macaques (*Macaca fascicularis*) are widely used as experimental animals in biomedical research and are closely related to other laboratory macaques, such as rhesus macaques (*M. mulatta*). We isolated 85,721 clones and determined 9407 full-insert sequences from cynomolgus monkey brain, testis, and liver. These sequences were annotated based on homology to human genes and stored in a database, QFbase http://genebank.nibio.go.jp/qfbase/.

Results: We found that 1024 transcripts did not represent any public human cDNA sequence and examined their expression using *M. fascicularis* oligonucleotide microarrays. Significant expression was detected for 544 (51%) of the unidentified transcripts. Moreover, we identified 226 genes containing exon alterations in the untranslated regions of the macaque transcripts, despite the highly conserved structure of the coding regions. Considering the polymorphism in the common ancestor of cynomolgus and rhesus macaques and the rate of PCR errors, the divergence time between the two species was estimated to be around 0.9 million years ago.

Conclusion: Transcript data from Old World monkeys provide a means not only to determine the evolutionary difference between human and non-human primates but also to unveil hidden transcripts in the human genome. Increasing the genomic resources and information of macaque monkeys will greatly contribute to the development of evolutionary biology and biomedical sciences.

Background

Genomic resources and information about primates are valuable for evolutionary and biomedical studies to determine how and why phenotypes specific to humans, as well as human diseases, have been formed. Moreover, they are important for extrapolating the results of laboratory experiments to medical research because the physiology of primates is more similar to that of humans as compared with other common experimental animals such as rodents. The cynomolgus macaque (*Macaca fascicularis*), also known as the long-tailed or crab-eating macaque, is an Old World monkey living in Southeast Asia. It is bred in laboratories worldwide and is one of the most popular primates used for laboratory animal studies, such as those on infectious diseases, immunology, pharmacology, tissue engineering, gene therapy, senescence, and learning [1]. Cynomolgus macaques, rhesus macaques (*M. mulatta*), and Japanese macaques (*M. fuscata*) are widely used for experimental studies and are closely related to each other [2-4]. The US government funded genome sequencing of the rhesus macaque because it is the most common laboratory animal bred in the US, and in 2007, the draft sequence of the rhesus macaque was published [5].

Since cynomolgus and rhesus monkeys are very closely related at the genetic level, we aim to determine the extent to which the rhesus macaque genome sequence can be used as a reference for biomedical studies involving cynomolgus macaques. At the chromosomal level, a previous study suggested that a pericentric chromosome inversion occurred in the cynomolgus lineage after splitting from rhesus macaques [6]. At the nucleotide sequence level, the genetic divergence between cynomolgus and rhesus monkeys has been measured using mitochondrial DNA sequences [2,3] or a limited number of loci on the chromosomes [4,7]. Thus, the divergence of a sufficient number of loci between cynomolgus and rhesus macaques would assist in determining the degree of genetic divergence between them. In addition, recent studies have shown that there is a considerable amount of genetic diversity within the species themselves [5-10], which also hampers the measurement of the genetic divergence. Because the divergence between the two macaques is very recent (much later than the divergence between humans and chimpanzees), we must consider the segregation of polymorphisms in the common ancestral population to estimate the correct species divergence time [11,12]. By analyzing the number of loci in the two species, we can determine the history of divergence between them, including the ancestral population size, divergence time between species, and possible gene flow [13,14].

We have constructed full-length-enriched cDNA libraries from cynomolgus monkey brain, testis, and liver using the oligo-capping method. Many comparative genomics projects have focused on sequencing of the genome or expressed sequenced tags (ESTs), and full-length cDNA sequences are uniquely informative resources for accurately predicting the full structure of transcripts in the genome [15]. Furthermore, because cynomolgus and rhesus macaques are very closely related, transcriptome data from cynomolgus macaques is useful for annotating the genome sequence of other macaques whose transcriptome data is less than 1% of that from humans and whose full-length cDNA data is scarce.

Along with the cynomolgus macaque cDNA sequencing project, we have published a part of our results, such as novel gene findings [16-19], search for fast-evolving genes [19], molecular evolution of 5'-untranslated regions (UTRs) [20], and evolution of brain-expressed genes [21]. In this study, we summarize the final sequencing project and present novel findings with an expanded dataset. In total, 85,721 ESTs and 9407 full-length sequences were determined, annotated, and stored in an in-house database and the public databases (DDBJ/EMBL/GenBank). Our study focused on the divergence between the cynomolgus and rhesus macaque genes. We did not intensively analyze the divergence between humans and cynomolgus monkeys, because a study on rhesus genome has investigated this thoroughly [5]; it also identified and discussed positively selected genes or extensively duplicated genomic regions during the evolution of *Catarrhine* primates.

Results

Summary of cDNA sequences

We constructed several oligo-capped cDNA libraries from cynomolgus monkey testis, liver, and seven anatomical parts of the brain (cerebellar cortex, parietal lobe, occipital lobe, frontal lobe, temporal lobe, medulla oblongata, and brain stem). The oligo-capping method selectively amplifies full-length cDNAs with a cap structure and poly(A) tail [22]. We sequenced the 5'- or 3'-end of 85,721 clones, yielding 63,395 and 22,326 sequences of 5'- and 3'-ESTs, respectively, after filtering the vector and low quality sequences. These EST sequences grouped into 16,466 clusters with 11,016 singletons (BLAST e-value: 1e-30). We classified them based on homology to the 26,575 non-redundant human RefSeq sequences (see methods). Of the 85,721 EST sequences, 68,257 (80%) were homologous to the human RefSeq gene set and were clustered into 9065 types of genes, indicating that our EST sequences would cover about 34% of the known human transcripts (Table 1). In particular, when we limited the human reference genes to the validated protein-coding genes (*i.e.*, RefSeq accession beginning with NM), 47% of the human reference genes were represented in the macaque cDNAs.

Table 1: Summary of cDNA clones

Library	# of isolated clones	# of full-sequenced clones
Brain: Parietal Lobe (QnpA)	8063 (5890)	649 (336)
Brain: Frontal Lobe (QflA)	13,215 (9286)	2493 (1768)
Brain: Temporal Lobe (QtrA)	6797 (6039)	1078 (862)
Brain: Occipital Lobe (QorA)	5458 (4518)	634 (606)
Brain Stem (QbsA, B)	2776 (1993)	359 (301)
Brain: Medulla Oblongata (QmoA)	4485 (3645)	1146 (912)
Brain: Cerebellar Cortex (QccE)	11,734 (9028)	731 (563)
Testis (QtsA)	10,867 (8510)	2316 (2175)
Liver (Qlv)	22,326 (20,833)	0 (0)
Total	85,721 (69,742)	9407 (7523)
Averaged Length		1882 bp

*Numbers of genes with the RefSeq homologs are shown in parentheses.

In parallel to EST sequencing, we determined about 9500 full-insert sequences of the cDNA clones. About 2500 clones whose 5'-EST sequences were not homologous to the public cDNA sequences and 7000 clones whose 5'-EST sequences were homologous to the human RefSeq sequences were chosen [16-21]. Out of the 9407 full-insert sequences, 7407 sequences were homologous to 5384 types of human genes (Table 1). The averaged length of the full-insert sequences was 1864 bp, excluding the length of the poly(A) tail. The macaque sequences were annotated for gene function and homologous locus in the human genome using information from the Entrez Gene [23] and Gene Ontology (GO) databases [24].

Database construction

All cDNA sequences and annotations were deposited in the public databases and stored in a simple in-house database, QFbase [25]. On the QFbase website, users can search the macaque clones by keywords and BLAST searches. For each human gene, the distribution of the macaque homologs is represented graphically and users can easily retrieve information of the objective macaque cDNA clones. The entries are further linked to the gene annotation in the outside databases, GenBank [26], Ensembl [27], OMIM [28], and H-InvDB [29]. The cDNA sequences were mapped on the human and rhesus genome sequences using the UCSC genome browser [30]. Moreover, 4665 human-macaque orthologous alignments are provided in the QFbase. For each alignment, the non-synonymous substitution rate (*Ka*) and the synonymous substitution rate (*Ks*) between the human and macaque cDNA sequences were estimated. Non-synonymous substitutions are nucleotide changes that replace amino acids between species whereas synonymous substitutions cause no amino acid changes. The relative pace of protein evolution was thus determined using *Ka*/*Ks*, assuming that the *Ks* value reflects the neutral mutation rate [31]. Using the database, users can sort the alignments according to the *Ka* and *Ks* values. For example, users can determine the *Ka* and *Ks* values of a particular gene or view the list of the 100 most rapidly evolved genes between humans and cynomolgus monkeys. The cDNA clones are distributed through the Human Science Research Resource Bank in Japan (Tokyo, Japan). Further information is available at the QFbase website.

Analysis of unidentified transcripts

Of the 9407 full-sequenced cDNAs, about 2000 were not homologous to the human reference gene sequences (RefSeq, built on Sep 14, 2006; BLAST: E = 1e-60). These Non-RefSeq transcripts clustered into 1245 non-redundant transcripts, which were further classified as shown in Figure 1. The list of the Non-RefSeq transcripts is provided in Additional file 1. We filtered 11 junk sequences and 210 known transcripts. The 210 transcripts matched with the unannotated human cDNA sequences in the database and were called as orphan transcripts. These may help in further annotation of the human genome.

After removing the junk sequences and orphan transcripts, the remaining 1024 transcripts were referred as the unidentified transcripts although 40% (406/1024) of the transcripts showed homology to human ESTs (BLAST: E = 1e-60), because no full cDNA sequence of humans has been registered in the public databases. One of the advantages of full-length cDNAs is that we can determine the splicing pattern and reading direction of the transcripts in the genome. We categorized the unidentified transcripts as anti-transcript, intronic spliced transcript, intronic single-exon transcript, intergenic spliced transcript, or intergenic single-exon transcript. Among the intergenic transcripts, 82 were located within 5 kb of the genic regions with the same direction as the genes. Of these, 6 were mapped on the upstream regions and 76 were mapped on the downstream regions of the known genes. The result showed they may be hidden extensions of the