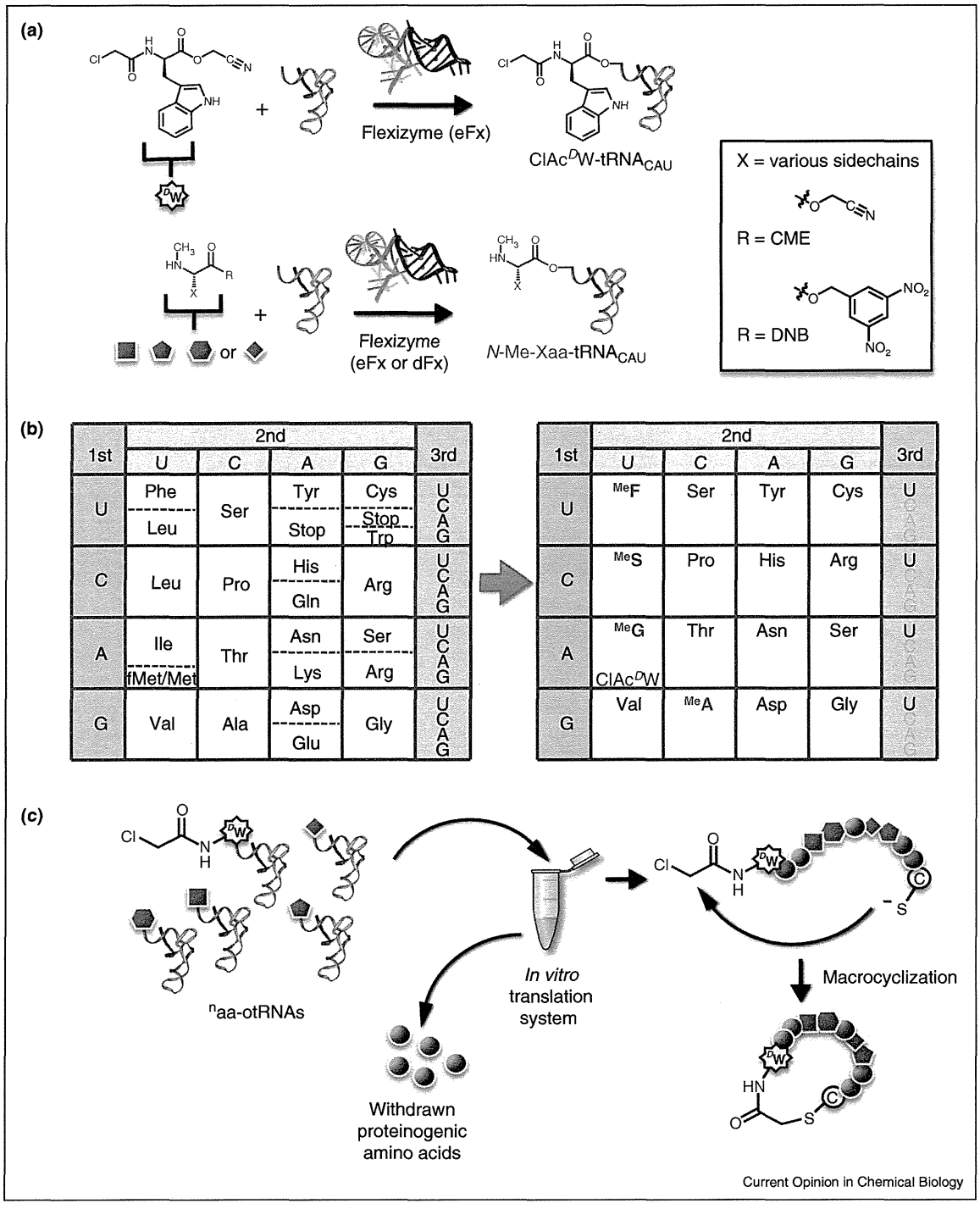
Figure 2



The FIT system [18,35••]. **(a)** Aminoacylation of tRNA using an activated *N*-acyl-D-Trp or an *N*-methyl amino acid along with the appropriate flexizyme. The flexizyme used depends on the activating group used [13,22••]. If the sidechain X is an aromatic group, the activating group shown in the rectangular box would be a cyanomethyl ester (CME), and the corresponding flexizyme would be enhanced-flexizyme (eFx). If the sidechain X is not an aromatic group, the activating group would be a dinitrobenzyl ester (DBE), and the corresponding flexizyme would be dinitro-flexizyme (dFx). **(b)** A reprogrammed genetic code that utilizes only NNU codons for elongation and the AUG codon for initiation by *N*-chloroacetyl (ClAc)-^{D}Trp. White boxes represent blank codon boxes for which the corresponding codon is not found in the random region of the mRNA sequence. **(c)** General scheme depicting genetically encoded macrocyclization of a non-standard peptide using the FIT system. Colored polygons represent non-standard amino acids (^{n}aa) depicted in **(a)**. Gray spheres represent proteinogenic amino acids. A 'C' within a black circle indicates the location of a cysteine.

the FIT system. A brief note that we previously referred to such a translation system as *withdrawn* PURE (*w*PURE); however, since we started developing more variations of the translation system differing from the original PURE system (*e.g.* the *c*PURE system that is specifically designed for macrocyclization between the peptide's N-terminus and C-terminus includes two N-terminal modifying enzymes), we renamed the *w*PURE system to the FIT system to reflect its versatility.

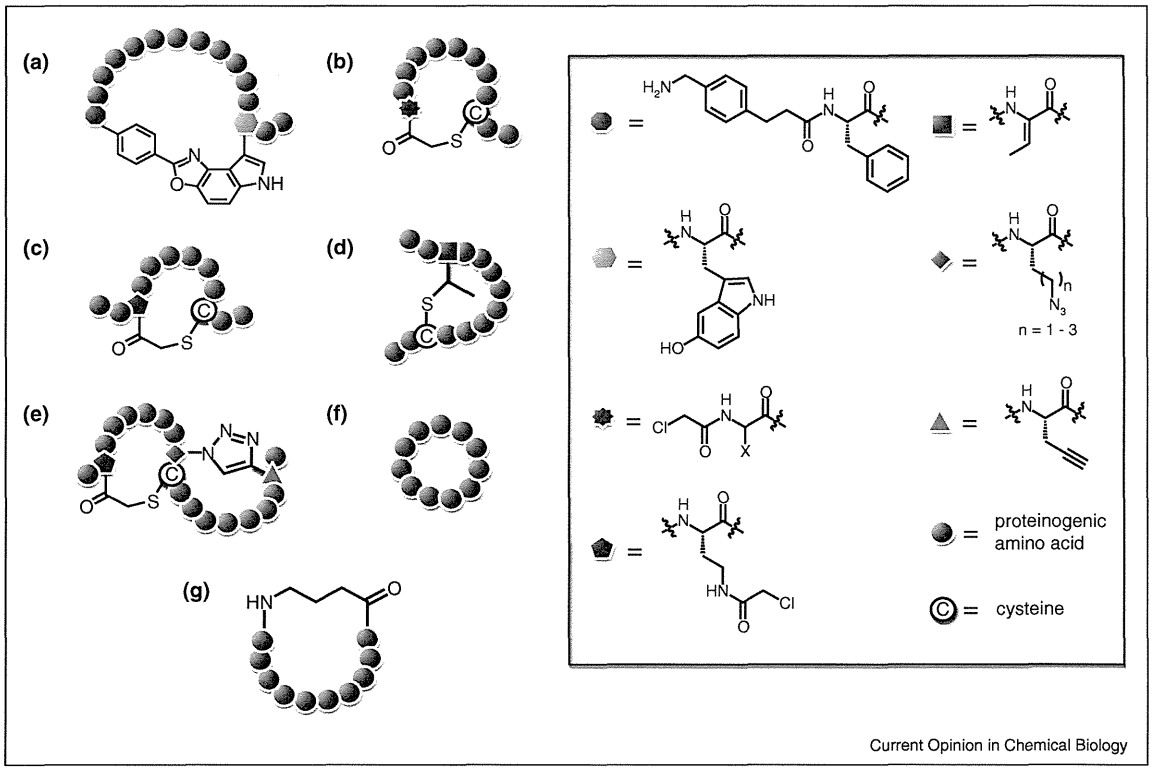
Despite the fact that otRNAs are supposedly inert with the ARSs present in the FIT system, to completely prevent the recycling of otRNAs that have been used by the translation machinery, omission of ARSs corresponding to any of the reprogrammed codons is crucial; this securely dismisses the possibility of reacylation of hydrolyzed otRNAs with contaminating proteogenic amino acids or any unexpected mischarging events. Using this technology, a variety of non-proteinogenic amino acids and α-hydroxy acids have been successfully incorporated into non-standard polypeptides and polyesters, respectively [13–20,21••,22••]. Recently, a new assortment of artificial amino acids that are chemically and biochemically valuable have been shown to be efficiently incorporated into non-standard peptides [23•,24,25••,26–28].

Reprogramming initiation and elongation

Needless to say, naturally occurring bioactive compounds have heavily influenced the inspiration for non-proteinogenic amino acid choice and design. One such non-proteinogenic amino is the D-amino acid. Unfortunately, elongation with D-amino acids has been known to occur with poor efficiency in the translation apparatus even when using the FIT system [13]. However, we have found that D-amino acid-$tRNA^{fMet}_{CAU}$ can be used for initiating translation, and while the translation efficiency is D-amino acid sidechain-dependent, *N*-acylation of the D-amino acid-$tRNA_{CAU}$ in general causes a significant increase in translation efficiency [21••] (Figure 2a). An alternative method of introducing D-amino acids is to initiate translation using an exotic peptide containing the D-amino acids [29]. It was demonstrated that a peptide bearing as many as four consecutive D-amino acids could be used to initiate translation.

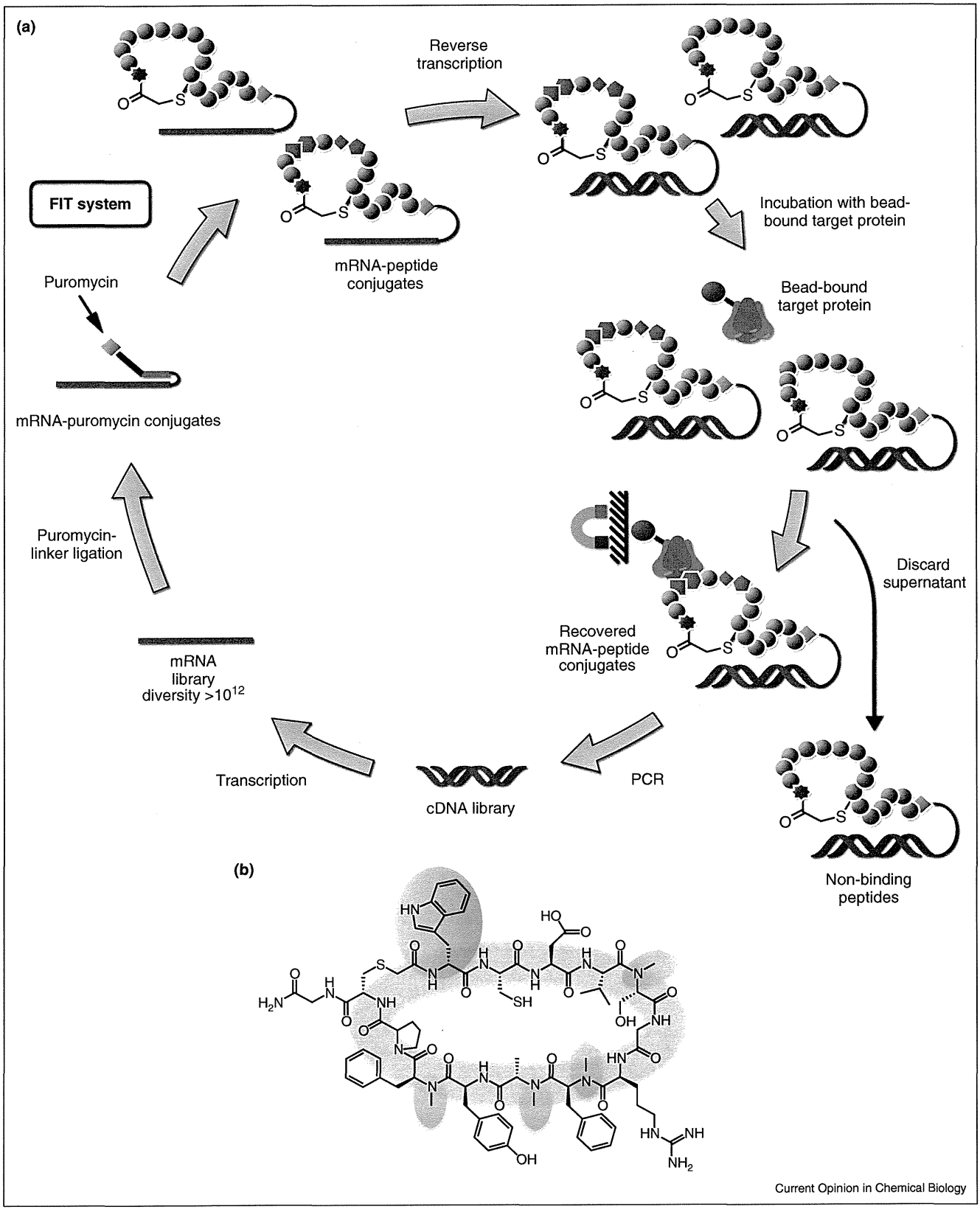
N-methylation, a deceptively small modification to amino acids, could have a tremendous impact on peptides'

Figure 3



Genetically encoded macrocyclized (exotic non-standard) peptides using the FIT system. Macrocyclization can occur via **(a)** oxidative coupling [24], **(b)** *N*-terminus-cysteine thioether bond formation [25••], inter-sidechain thioether bond formation using **(c)** S_N2 substitution [26] or **(d)** Michael addition [28], **(e)** Huisgen 1,3-dipolar cycloaddition in conjunction with thioether bond formation [27], **(f)** seamless head-to-tail polyamide backbone cyclization [37], or **(g)** backbone cyclization with an embedded γ-amino acid [38]. Colored polygons represent artificial amino acids. Gray spheres represent proteinogenic amino acids. The inset box on the right shows the chemical identities of the artificial amino acids, represented by the colored polygons, before macrocyclization. X can be any D-proteinogenic or L-proteinogenic sidechain. A 'C' within a black circle indicates the location a cysteine.

Figure 4


(a)
Reverse transcription
FIT system
Puromycin
mRNA-peptide conjugates
Incubation with bead-bound target protein
Bead-bound target protein
mRNA-puromycin conjugates
Puromycin-linker ligation
Discard supernatant
Recovered mRNA-peptide conjugates
mRNA library diversity >10^12
Transcription
cDNA library
PCR
Non-binding peptides
(b)
Current Opinion in Chemical Biology

pharmacological properties including oral availability, peptidase resistance and cell permeability [30]. It has been shown that the translation machinery is capable of utilizing *N*-methyl-amino acid-tRNAs (Figure 2a) and incorporating *N*-methyl-amino acids using genetic code reprogramming [22••,31–33,34•]. Careful reprogramming of the genetic code has resulted in a new genetic code that allows for the maximum diversity of *N*-methyl amino acids as well as proteinogenic amino acids [35••] (Figure 2b). Furthermore, these *N*-methylated peptides can be cyclized using the efficient and non-reducible thioether bond formation (*vide infra*, Figure 2c) [22••]. Expanding on the theme of incorporating *N*-substituted amino acids, the incorporation of *N*-substituted glycines (rGly) were also shown to proceed efficiently, and the resulting peptide–peptoid hybrids could, likewise, be cyclized [23•]. Nitrogen modification is not limited to the amide nitrogen (*e.g.* other recently incorporated non-standard amino acids include sidechain acylated or methylated lysines [20]).

Cyclization technologies adopted into the FIT system

In NRPSs, the thioesterase domain catalyzes linear peptide release via hydrolysis or aminolysis. Alternatively, the peptide can be released by macrocyclization. As with *N*-methylation, macrocyclization also enhances peptides' pharmacological properties. It is no surprise that many natural product peptides are macrocycles and that researchers are developing ways to introduce post-translational macrocyclization [36–43]. A unique way of genetically encoding backbone macrocyclization is by withdrawing an mRNA-encoded amino acid to promote elongation arrest and peptidyl-tRNA drop-off, which can then be utilized to promote a diketopiperazine-mediated macrocyclization [36]. This system requires the use of peptidyl deformylase (PDF) and, optionally, methionine aminopeptidase (MAP) if the initiation methionine is undesired in the final product.

The flexibility of the FIT system was demonstrated by the incorporation of an artificial amino acid at the initiation position, an internal elongation position(s), or both internal and elongation positions for the purpose of genetically encoding macrocyclization. Genetically encoded macrocyclization can take place via chemically assisted oxidative coupling or spontaneous thioether formation [24–27] (Figure 3a–c and e). Thermal isomerization of a vinylglycine residue to a dehydrobutrine residue also affords access to thioether formation through an inter-sidechain Michael addition by a nearby cysteine [28] (Figure 3d). Bicyclic structures are also accessible and can be precisely formed using thioether formation coupled with the orthogonal Huisgen 1,3-dipolar cycloaddition [27] (Figure 3e). Seamless head-to-tail polyamide backbone macrocyclization can be genetically encoded [37] (Figure 3f) with the option of inserting a novel γ-amino acid in the middle of standard α-amino acids [38] (Figure 3g). Among this abundance of FIT system-based macrocyclization methods, the *N*-terminus-to-sidechain thioether formation [25••] was the preferred method in a recently performed *in vitro* selection of non-standard peptides (*vide infra*).

The random non-standard peptide integrated discovery (RaPID) system

Several years ago, it was anticipated that the introduction of non-standard amino acids into a translation system would inevitably lead to the discovery of bioactive non-standard peptides with enhanced pharmacological properties using *in vitro* selection [44]. With the ability to translate natural product-like peptidomimetics with high efficiency and fidelity, we have the means to identify bioactive compounds using *in vitro* selection by an appropriate display method, like mRNA display [45,46]. Toward this goal, the FIT system was integrated with the mRNA display format, and referred to as the random non-standard peptide integrated discovery system [35••] (Figure 4a). To date, more than a dozen of bioactive non-standard cyclic peptides have been discovered using the RaPID system in our laboratory and some have recently been reported in literature [35••,47•]. Among them, Yamagishi *et al.* [35••] recently reported an *in vitro* selection for *N*-methylated cyclic peptide inhibitors of a formerly non-druggable enzyme, E6AP. E6AP is known to ubiquitinate p53, a protein known to promote apoptosis, and other potentially tumor-related proteins.

The reprogrammed genetic code designed by Yamagishi *et al.* was thoroughly optimized for the RaPID system. For example, to create an mRNA library free from random stop codons, only NNU codons (where N is any of the four nucleobases and U is uracil) were used in the random region of the mRNA. The threat of misincorporation of proteinogenic amino acids forced the generation of multiple variations and optimization of reprogrammed genetic codes. In addition to the incorporation rate of *N*-alkyl amino acids being lower than the incorporation rate of proteinogenic acids [48], the choice of codons and the corresponding tRNAs [49,50] can affect the efficiency and sometimes the fidelity of *N*-methyl amino acid incorporation. While the incorporation of multiple non-standard amino acid adds yet another layer of complexity [51], library construction using a reprogrammed genetic

(Figure 4 Legend) The RaPID system [35••]. **(a)** A schematic representation of single cycle of a generalized RaPID system. **(b)** CM_{11}-11, an NRP-like inhibitor of E6AP discovered using the RaPID system. Non-standard *N*-methyl groups, D-amino acid and macrocyclic peptide backbone are shown within orange ovals, a blue oval and a green doughnut, respectively.

code containing low non-standard amino acid content could result in the selection of peptides possessing high affinity, but lacking the desired non-standard amino acid [43]. The final reprogrammed genetic code designed by Yamagishi *et al.* [35••] shown in Figure 2b resulted in the highest fidelity and still retained a high *N*-methyl amino acid content. Since UGG, the codon for tryptophan, was also eliminated; peptide library synthesis was initiated using a D-tryptophan to ensure that tryptophan is represented. As previously mentioned, the reprogrammed genetic code was optimized to accommodate elongation using any of 4 different *N*-methyl amino acids and 11 different proteogenic amino acids. Four out of 16 codon boxes represent an *N*-methyl amino acid, which gives a 25% chance of incorporating an *N*-methyl amino acid at any position in the random region. The peptides isolated at the end of the selection possessed an average of four *N*-methyl residues out of eleven residues in the random region. The cyclic peptide with the highest affinity, CM_{11}-1 (K_d = 0.60 nM), bears the three notable non-canonical features found in nonribosomal peptide cyclosporin A – a D-amino acid, *N*-methylation, and a macrocyclic backbone (Figure 4b). Most importantly, *in vitro* assays show that CM_{11}-1 inhibits E6AP poly-ubiquitination of target proteins, p53 and Prx1.

Conclusions and perspectives

Standard ribosomal synthesis is limited to twenty canonical amino acids. The FIT system has addressed and diminished these limitations allowing us to produce custom natural product-like peptidomimetics. The RaPID system has firmly placed itself amidst the growing repertoire of novel techniques for the *in vitro* selection of non-standard peptides. The promiscuity of flexizyme allows us to quickly conjugate tRNAs with amino acids from an incredibly diverse repertoire of non-standard amino acids. Initiation using a non-standard amino acid has proven to be extremely robust and has allowed us to produce stable D-amino acid-containing macrocyclic peptides that have been found to be bioactive. The above example of macrocyclic *N*-methyl-peptides against E6AP has opened a new door for the discovery of non-druggable enzyme inhibitors. However, the RaPID system has no limitation for targets of choice. The RaPID system could be used not only for the discovery of enzyme inhibitors, but also disruptors of protein–protein interaction based on various different natural product-like peptidomimetic scaffolds. More examples of such *in vitro* selections are forthcoming in the near future.

Acknowledgements

This work was supported by a Grants-in-Aid from the Japan Society for Promotion of Science (JSPS), the Specially Promoted Research (21000005) and the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization (NEDO) to HS, and GCOE Chemistry Innovation through Cooperation of Science and Engineering and The JSPS Postdoctoral Fellowship For Foreign Researchers (P11344) to CJH.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Nolan EM, Walsh CT: **How nature morphs peptides scaffolds into antibiotics.** *Chembiochem* 2009, **10**:34-53.
2. Fischbach MA, Lai JR, Roche ED, Walsh CT, Liu DR: **Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes.** *Proc Natl Acad Sci USA* 2007, **104**:11951-11956.
3. Evans BS, Chen Y, Metcalf WW, Zhao H, Kelleher N: **Directed evolution of nonribosomal peptide synthetase AdmK generates new andrimid derivatives in vivo.** *Chem Biol* 2011, **18**:601-607.
4. Chatterjee C, Patton GC, Cooper L, Paul M, van der Donk WA: **Engineering dehydro amino acids and thioesters into peptides using lacticin 481 synthetase.** *Chem Biol* 2006, **13**:1109-1117.
5. Levengood MR, Knerr PJ, Oman TJ, van der Donk WA: **In vitro mutagenesis of lantibiotic analogues containing nonproteinogenic amino acids.** *J Am Chem Soc* 2009, **2**:12024-12025.
6. Shi Y, Yang X, Garg N, van der Donk WA: **Production of lantipeptides in *Escherichia coli*.** *J Am Chem Soc* 2011, **113**:2338-2341.
7. Tianero MD, Donia MS, Young TS, Schultz PG, Schmidt EW: **Ribosomal route to small-molecule diversity.** *J Am Chem Soc* 2012, **134**:418-425.
8. Forster AC, Tan Z, Nalam MNL, Lin H, Qu H, Cornish VW, Blacklow SC: **Programming peptidomimetic syntheses by translating genetic codes designed *de novo*.** *Proc Natl Acad Sci USA* 2003, **100**:6353-6357.
9. Frankel A, Roberts RW: **In vitro selection for sense codon suppression.** *RNA* 2003, **9**:780-786.
10. Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T: **Cell-free translation reconstituted with purified components.** *Nat Biotechnol* 2001, **19**:751-755.
11. Josephson K, Hartman MCT, Szostak JW: **Ribosomal synthesis of unnatural peptides.** *J Am Chem Soc* 2005, **127**:11727-11735.
12. Hartman MCT, Josephson K, Lin C, Szostak JW: **An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides.** *PLoS ONE* 2007, **2**:e972.
13. Murakami H, Ohta A, Ashigai H, Suga H: **A highly flexible tRNA acylation method for non-natural peptide synthesis.** *Nat Methods* 2006, **3**:357-359.
14. Ohuchi M, Murakami H, Suga H: **The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus.** *Curr Opin Chem Biol* 2007, **11**:537-542.
15. Ohta A, Yamagishi Y, Suga H: **Synthesis of biopolymers using genetic code reprogramming.** *Curr Opin Chem Biol* 2008, **12**:159-167.
16. Kang TJ, Suga H: **Ribosomal synthesis of nonstandard peptides.** *Biochem Cell Biol* 2008, **86**:92-99.
17. Morimoto J, Hayashi Y, Iwasaki K, Suga H: **Flexizymes: their evolutionary history and the origin of catalytic function.** *Acc Chem Res* 2011, **44**:1359-1368.
18. Goto Y, Katoh T, Suga H: **Flexizymes for genetic code reprogramming.** *Nat Protoc* 2011, **6**:779-790.
19. Ohta A, Murakami H, Suga H: **Polymerization of α-hydroxy acids by ribosomes.** *Chembiochem* 2008, **9**:2773-2778.
20. Kang TJ, Yuzawa S, Suga H: **Expression of histone H3 tails with combinatorial lysine modifications under the reprogrammed genetic code for the investigation on epigenetic markers.** *Chem Biol* 2008, **15**:1166-1174.

21. Goto Y, Murakami H, Suga H: **Initiating translation with D-amino**
•• **acids.** *RNA* 2008, **14**:1390-1398.
The authors used genetic code reprogramming to demonstrate that initiation of peptide synthesis can performed using D-amino acids. Pre-acylation of the D-amino acids was found to be crucial for the efficient translation of the non-standard peptides.

22. Kawakami T, Murakami H, Suga H: **Messenger RNA-**
•• **programmed incorporation of multiple N-methyl-amino acids into linear and cyclic peptides.** *Chem Biol* 2008, **15**:32-42.
The FIT system was used to incorporate as many as ten consecutive *N*-methyl amino acids. By reprogramming initiation with *N*-chloroacetyl (ClAc)-Phe, cyclic peptides containing as many as four different *N*-methyl-amino acids were cleanly produced.

23. Kawakami T, Murakami H, Suga H: **Ribosomal synthesis of**
• **polypeptoids and peptoid–peptide hybrids.** *J Am Chem Soc* 2008, **130**:16861-16863.
The FIT system was used to synthesize polypeptoids and cyclic peptoid–peptide hybrids. Incorporation of four consecutive rGly monomers appeared slightly less efficient than incorporation of four consecutive *N*-methyl amino acids. Cyclic peptoid–peptide hybrids containing as many as three rGly monomers were cleanly produced.

24. Yamagishi Y, Ashigai H, Goto Y, Murakami H, Suga H: **Ribosomal synthesis of cyclic peptides with a fluorogenic oxidative coupling reaction.** *Chembiochem* 2009, **10**:1469-1472.

25. Goto Y, Ohta A, Sako Y, Yamagishi Y, Murakami H, Suga H:
•• **Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides.** *ACS Chem Biol* 2008, **3**:120-129.
Various amino acids and *N*-acyl-amino acids were used to initiate peptide synthesis in methionine-depleted translation systems. Antitumor cyclic peptide G7-18NATE was successfully produced upon translation initiated with a *N*-chloroacetyl (ClAc)-tryptophan followed by spontaneous macrocyclization. A small cyclic peptide library encoded by 160 distinct mRNAs was also successfully produced.

26. Sako Y, Goto Y, Murakami H, Suga H: **Ribosomal synthesis of peptidase-resistant peptides closed by a nonreducible inter-side-chain bond.** *ACS Chem Biol* 2008, **3**:241-249.

27. Sako Y, Morimoto J, Murakami H, Suga H: **Ribosomal synthesis of bicyclic peptides via two orthogonal inter-side-chain reactions.** *J Am Chem Soc* 2008, **130**:7232-7234.

28. Goto Y, Iwasaki K, Torikai K, Murakami H, Suga H: **Ribosomal synthesis of dehydrobutyrine- and methyllanthionine-containing peptides.** *Chem Commun* 2009, **23**:3419-3421.

29. Goto Y, Suga H: **Translation initiation with initiator tRNA charged with exotic peptides.** *J Am Chem Soc* 2009, **131**:5040-5041.

30. Chatterjee J, Gilon C, Hoffman A, Kessler H: **N-methylation of peptides: a new perspective in medicinal chemistry.** *Acc Chem Res* 2008, **41**:1331-1342.

31. Tan Z, Forster AC, Blacklow SC, Cornish VW: **Amino acid backbone specificity of the *Escherichia coli* translation machinery.** *J Am Chem Soc* 2004, **126**:12752-12753.

32. Frankel A, Millward SW, Roberts RW: **Encodamers: unnatural peptide oligomers encoded in RNA.** *Chem Biol* 2003, **10**:1043-1050.

33. Merryman C, Green R: **Tranformation of aminoacyl tRNAs for the in vitro selection of "drug-like" molecules.** *Chem Biol* 2004, **11**:575-582.

34. Subtelny AO, Hartman MCT, Szostak JW: **Ribosomal synthesis**
• **of *N*-methyl peptide.** *J Am Chem Soc* 2008, **130**:6131-6136.
Linear peptides containing multiple incorporations of *N*-methyl Leu, *N*-methyl Thr, and *N*-methyl Val were ribosomally synthesized. In particular, pairwise incorporations produced excellent yields with relatively minor truncation products. Notably, in the FIT system, *N*-methyl Leu, *N*-methyl Thr, and *N*-methyl Val were poor substrates for flexizyme, and significant amounts of peptide could not be produced.

35. Yamagishi Y, Shoji I, Miyagawa S, Kawakami T, Katoh T, Goto Y,
•• Suga H: **Natural product-like macrocyclic *N*-methyl-peptide inhibitors against a ubiquitin ligase uncovered from a ribosome-expressed de novo library.** *Chem Biol* 2011, **18**:1562-1570.
In this article, the authors perform an *in vitro* selection for *N*-methylated cyclic peptide binders of E6AP. The cyclic peptide with the highest affinity, CM_{11}-1, contained four *N*-methyl groups. CM_{11}-1 was shown to not only bind to E6AP, but also inhibit enzyme activity.

36. Kang TJ, Hayashi Y, Suga H: **Synthesis of the backbone cyclic peptide sunflower trypsin inhibitor-1 promoted by the induced peptidyl-tRNA drop-off.** *Angew Chem Int Ed* 2011, **50**:2159-2161.

37. Kawakami T, Ohta A, Ohuchi M, Ashigai H, Murakami H, Suga H: **Diverse backbone-cyclized peptides via codon reprogramming.** *Nat Chem Biol* 2009, **5**:888-890.

38. Ohshiro Y, Nakajima E, Goto Y, Fuse S, Takahashi T, Doi T, Suga H: **Ribosomal synthesis of backbone-macro peptide containing γ-amino acids.** *Chembiochem* 2011, **12**:1183-1187.

39. Seebeck FP, Szostak JW: **Ribosomal synthesis of dehydroalanine-containing peptides.** *J Am Chem Soc* 2006, **128**:7150-7151.

40. Seebeck FP, Ricardo A, Szostak JW: **Artificial lantipeptides from *in vitro* translations.** *Chem Commun* 2011, **47**:6141-6143.

41. Millward SW, Takahashi TT, Roberts RW: **A general route for post-translational cyclization of mRNA display libraries.** *J Am Chem Soc* 2005, **127**:14142-14143.

42. Young T, Young DD, Ahmad I, Louis JM, Benkovic SJ, Schultz PG: **Evolution of cyclic peptide protease inhibitors.** *Proc Natl Acad Sci USA* 2011, **94**:12297-12302.

43. Millward SW, Fiacco S, Austin RJ, Roberts R: **Design of cyclic peptides that bind protein surfaces with antibody-like affinity.** *ACS Chem Biol* 2007, **2**:625-634.

44. Frankel A, Li S, Starck SR, Roberts R: **Unnatural RNA display libraries.** *Curr Opin Struct Biol* 2003, **13**:506-512.

45. Nemoto N, Miyamoto-sato E, Husumi Y, Yanagawa H: **In vitro virus: bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro.** *FEBS Lett* 1997, **414**:405-408.

46. Roberts R, Szostak J: **RNA-peptide fusions for the *in vitro* selection of peptides and proteins.** *Proc Natl Acad Sci USA* 1997, **94**:12297-12302.

47. Hayashi Y, Morimoto J, Suga H: **In vitro selection of Anti-Akt2**
• **thioether-macrocyclic peptides leading to isoform-selective inhibitors.** *Chem Biol* 2012 doi: 10.1021/cb200388k.
Using the RaPID system, the authors isolated and characterized cyclic peptides that bind to Akt2. When tested for binding to Akt1 or Akt3, the cyclic peptide Pakti-1 was found to be isoform selective for Akt2.

48. Pavlov MY, Watts RE, Tan Z, Cornish VW, Ehrenberg M, Forster AC: **Slow peptide bond formation by proline and other *N*-alkylamino acids in translation.** *Proc Nat Acad Sci USA* 2009, **106**:50-54.

49. Subtelny AO, Hartman MCT, Szostak JW: **Optimal codon choice can improve the efficiency and fidelity of *N*-methyl amino acid incorporation into peptides by in vitro translation.** *Angew Chem Int Ed* 2011, **50**:3164-3167.

50. Zhang B, Tan Z, Gartenmann Dickson L, Nalam MNL, Cornish VW, Forster AC: **Specificity of translation for *N*-alkyl amino acids.** *J Am Chem Soc* 2007, **129**:11316-11317.

51. Forster AC: **Low modularity of aminoacyl-tRNA substrates in polymerization by the ribosome.** *Nucleic Acids Res* 2009, **37**:3747-3755.

ACS chemical biology

Articles

pubs.acs.org/acschemicalbiology

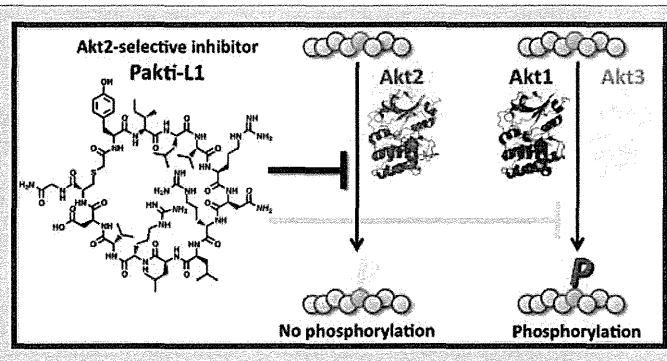
In Vitro Selection of Anti-Akt2 Thioether-Macrocyclic Peptides Leading to Isoform-Selective Inhibitors

Yuuki Hayashi,[†] Jumpei Morimoto,[‡] and Hiroaki Suga*[,†]

[†]Department of Chemistry, Graduate School of Science and [‡]Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan

Supporting Information

ABSTRACT: The Akt kinase family, consisting of three isoforms in humans, is a well-validated class of drug target. Through various screening campaigns in academics and pharmaceutical industries, several promising inhibitors have been developed to date. However, due to the mechanistic and structural similarities of Akt kinases, it is yet a challenging task to discover selective inhibitors against a specific Akt isoform. We here report Akt-selective and also Akt2 isoform-selective inhibitors based on a thioether-macrocyclic peptide scaffold. Several anti-Akt2 peptides have been selected from a library by means of an *in vitro* display system, referred to as the RaPID (Random nonstandard Peptide Integrated Discovery) system. Remarkably, the majority of these "binding-active" anti-Akt2 peptides turned out to be "inhibitory active", exhibiting IC_{50} values of approximately 100 nM. Moreover, these peptides are not only selective to the Akt kinase family but also isoform-selective to Akt2. Particularly, one referred to as Pakti-L1 is able to discriminate Akt2 250- and 40-fold over Akt1 and Akt3, respectively. This proof-of-concept case study suggests that the RaPID system has a tremendous potential for the discovery of unique inhibitors with high family- and isoform-selectivity.



In human cells, the Akt family, belonging to the serine/threonine (S/T) kinase family, plays critical roles in regulating various signal transduction pathways.[1] Because misregulation of Akt causes alterations of apoptosis, cell proliferation, and metabolisms depending upon the Akt isoforms, it is of great interest for academics, as well as pharmaceutical industries, to develop inhibitors against this kinase family. Three isoforms, Akt1, Akt2, and Akt3, are known in the human Akt family. Akt1 activates the translocation of nuclear factors, such as NF-κB (nuclear factor of kappa light polypeptide gene enhancer in B-cells),[2] which suppress apoptosis in a transcription-independent manner; thereby it promotes cell survival. In fact, elevated expression and activation of Akt1 have been observed in human gastric, breast, prostate, and ovarian cancers, implying that it is a major factor in the cancer development, *i.e.*, Akt1 is an oncogene.[3,4] Similar events involving Akt2 have been also observed in various carcinomas,[5−7] consistent with a view that Akt2 plays a similar role as Akt1 in developing malignant phenotype of these cancers. However, data generated by experiments involving Akt1 or Akt2-deficient transgenic mice have suggested that Akt2 likely plays a specific role in the insulin receptor signal transduction, implying that Akt2 is an important drug target for controlling diabetes mellitus.[8] The function of Akt3 is least understood among the Akt family, but it is expressed predominantly in the brain.[9] Presumably, it contributes to the development of the brain through the activation of growth factor mediated signal transductions.[10]

Because of the therapeutic significance of Akt in cancers and other diseases, a number of screening campaigns of molecules have been conducted to yield potent Akt inhibitors (representative Akt inhibitors along with their potency and selectivity against a specific Akt isoform over other isoforms and kinase family are summarized in Supplementary Table 1). There are four classes of inhibitors: (I) those that compete for binding to the ATP-binding site, (II) those that bind to the pleckstrin homology domain (PH domain), (III) those that bind to an allosteric site(s), and (IV) those derived from substrate peptide sequences and others.

The class I inhibitors, represented by A674563 and GSK690693, have remarkable potencies against all Akt isoforms with low nanomolar IC_{50}'s.[11−15] However, since all kinase families have structurally similar ATP-binding sites, they generally show broad spectra of inhibitory activities against many S/T kinases with comparable potencies, thus interacting with various cellular kinases in a nonselective manner. On the other hand, since the PH domain of the Akt family is unique over other kinases, the class II inhibitors are inherently Akt-selective, but because all Akt members have nearly identical PH domains, they are so far unable to distinguish a specific isoform from other isoforms.[16−21] Representative class III inhibitors are

Received: September 23, 2011
Accepted: January 11, 2012
Published: January 24, 2012

ACS Publications © 2012 American Chemical Society

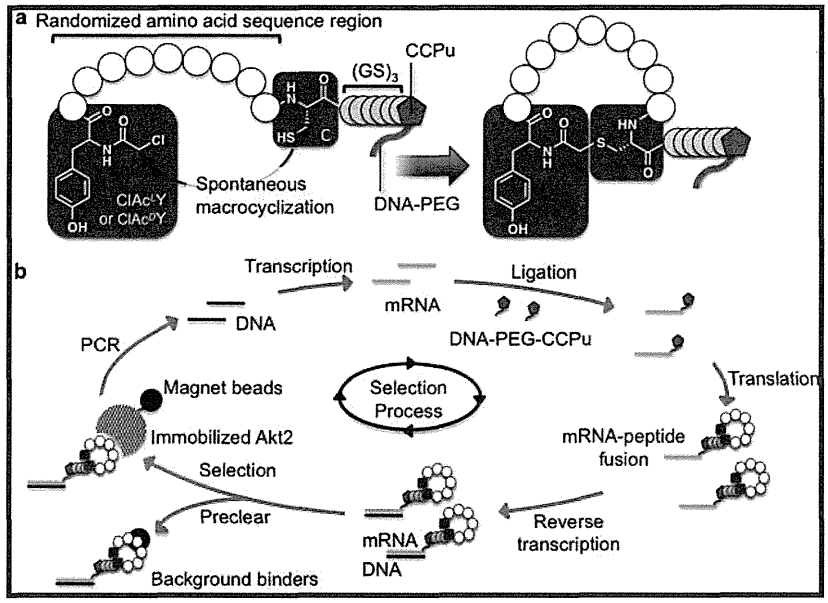


Figure 1. Schematic presentation of the macrocyclic peptide library and RaPID selection. (a) A macrocyclic peptide library closed by a thioether bond. The randomized amino acid sequence region (white circles) consists of 4–12 amino acid residues. Spontaneous cyclization takes place between ClAc^{L}Y or ClAc^{D}Y initiator and sulfhydryl group of a cysteine (C) residue, both shown in dark gray boxes. Although the C residue is designated right after the random region in each peptide, a C residue or residues can also be appeared from the random region to form a smaller ring sized head and a tail containing the unreacted cysteine residue(s). The $(GS)_3$-linker peptide is shown in gray circles. 5′-CCPu-3′ and DNA-PEG regions in DNA-PEG-CCPu are shown in a dark gray pentagon and a dark gray line, respectively. (b) Selection using RaPID system. DNA and RNA libraries are shown in black and gray lines, respectively. The full-length Akt2 protein and Ni^{2+}-NTA magnetic beads are shown in a black-dotted circle and black circles, respectively.

MK2206, Akti-1, Akti-2, and Akti-1/2 (Akt inhibitor VIII), all of which are categorized as allosteric inhibitors.[22−25] Among them, Akti-1/2 has been extensively studied by the X-ray structure complexed with Akt1.[26] This inhibitor binds to the interface between "N-lobe" and "C-lobe" in the kinase domain and further interacts with the PH domain, resulting in stabilization of the ternary interactions. This ternary complex, referred to as a "PH-in" form, prevents phosphorylation of T308 and S473, inhibiting activation of the kinase domain. The IC_{50} value of Akti-1/2 is 58 nM against Akt1 compared with 4- and 40-fold elevated IC_{50} values against Akt2 and Akt3, respectively. On the other hand, Akti-2, which is structurally similar to Akti-1/2, exhibits a greater isoform-selectivity (70-fold against Akt1 over Akt2), but its potency dropped to 325 nM IC_{50}. The class IV inhibitors competitively bind to the protein substrate-binding site of Akt.[27−30] Although the isoform selectivity of this class of inhibitors is unknown in literature, they display marginal inhibitory potencies even *in vitro*.

Even though simultaneous inhibition of all members of the Akt family by ATP-competitive inhibitors gave a maximal efficacy for caspase-3/7 activation,[31] these inhibitors may have a risk of side effects caused by undesired inhibitions of non-Akt kinases. An approach to devise inhibitors binding to the PH domain has yielded Akt-selective inhibitors, but it has been yet difficult to install an isoform-selectivity into the inhibitors. Allosteric inhibitors have given the Akt- and isoform-selective properties, but it remains difficult to achieve high selectivity and potency simultaneously against a specific isoform. Therefore, it is still a difficult challenge to devise Akt- and isoform-selective inhibitors.

We here report a potent Akt2 isoform-selective inhibitor with a 100 nM IC_{50} discovered from a library of thioether-macrocyclic peptides with a complexity of 10^{12}, by means of an *in vitro* display system, referred to as RaPID (Random nonstandard Peptide Integrated Discovery) system (Yamagishi *et al. Chem. Biol.*, in press).

RESULTS AND DISCUSSION

Thioether-Macrocyclic Peptide Libraries. In 2008, we reported a new strategy to ribosomally express thioether-macrocyclic peptides involving an initiation codon reprogramming. To facilitate such a codon reprogramming for this study, we have devised a FIT (Flexible In vitro Translation) system where a custom-made *in vitro* translation system lacking methionine (M) is supplied with *N*-(2-chloroacetyl)-aminoacyl-tRNA$^{fMet}_{CAU}$ (ClAc-aa-tRNA$^{fMet}_{CAU}$) prepared by a flexizyme (flexible tRNA acylation ribozyme),[32−35] resulting in a reassignment of AUG initiator codon from M to the ClAc-aa. In this setup, the peptide translation of mRNA is initiated by the ClAc-aa, and a cysteine (C) residue installed at a downstream position in the sequence reacts with the 2-chloroacetyl group, giving macrocyclic peptide closed by a thioether bond (Figure 1a).[36,37] The cyclization takes place spontaneously without an additional reagent, yielding a clean desired product almost regardless of the length and sequence composition. In fact, we were able to prepare a macrocyclic peptide library with various lengths from random sequences of mRNAs.

In the present work, we constructed two thioether-macrocyclic peptide libraries initiated by either L or D isomers of *N*-(2-chloroacetyl)-tyrosine (ClAc^{L}Y or ClAc^{D}Y, respectively). To construct each peptide library, mRNA template libraries were designed to have AUG-$(NNK)_n$-UGC, where AUG and UGC assign ClAc^{L}Y or ClAc^{D}Y and C, respectively, which would undergo the thioether bond formation for cyclization; also where $(NNK)_n$ (N and K represent any of four bases and U or G, respectively) assign all possible 20 amino acids with a range of lengths between 4 and 12, *i.e.*, a mixture of peptides with a

Table 1. Selected Peptides with Frequency of Appearance and Their IC_{50} Values against Akt Isoforms[b]

Peptide	Sequence	Frequency	IC_{50} [nM]		
			Akt2	Akt1	Akt3
Pakti-L1	Ac-^{L}YILVRNRLLRVDCG-NH_2	28/37	110	>25,000	4,200
Pakti-L2	Ac-^{L}YWILITWPLVRRKCG-NH_2	2/37	120	~1,000[a]	~1,000[a]
Pakti-L3	Ac-^{L}YWIVLTWPIVTRRCG-NH_2	2/37	92	~1,000[a]	~1,000[a]
Pakti-L4	Ac-^{L}YTYWFVSMICG-NH_2	1/37	inactive	N.D.	N.D.
Pakti-L5	Ac-^{L}YIRRPWVPIMYLGCG-NH_2	3/37	active	N.D.	N.D.
Pakti-L6	Ac-^{L}YILVRNRPLRVDCG-NH_2	1/37	active	N.D.	N.D.
Pakti-D1	Ac-^{D}YAVRILGHYLQVGCG-NH_2	35/37	active	N.D.	N.D.
Pakti-D2	Ac-^{D}YLSRRHGLLFLIRCG-NH_2	1/37	inactive	N.D.	N.D.
Pakti-D3	Ac-^{D}YLSREFNLLFLVRCG-NH_2	1/37	active	N.D.	N.D.

[a] IC_{50} values were estimated by the graph shown in Figure.2b and c [b] Peptides shown in this table were synthesized by standard solid-phase synthesis and applied to the inhibition assay of kinase activities. IC_{50} values were determined by the plot shown in Figure 3. Pakti-L5, L6, D1, and D3 inhibited Akt2 kinase activity (the accurate IC_{50} value was undetermined), whereas Pakti-L4 and D2 were inactive by a preliminary assay. N.D. denotes "not determined" in this study.

random region of these lengths would be expressed. It should be noted that the thioether-macrocyclic peptide library initiated with ClAc^{D}Y would cover different conformational sequence space from that initiated with ClAc^{L}Y since a single D-isomer in the macrocyclic scaffold would potentially alter the tertiary structural conformers from those in all-L-macrocyclic peptides. We referred the respective libraries to as ^{L}Y- and ^{D}Y-libraries for our convenience in discussion below.

To perform the selection of active species, these libraries were integrated with an *in vitro* display method, so-called mRNA display,[38,39] where puromycin (Pu) was ligated to the 3′-end of a synthetic hybrid oligonucleotide (DNA-PEG-CCPu; the DNA sequence is complementary to the 3′-end sequence of mRNA, followed by a poly(ethylene glycol) linker, deoxyribonucleotides CC, and puromycin), which allows for fusion with the peptide C-terminus (Supplementary Table 2). To facilitate the integration, we designed the above mRNA libraries to have three repeats of (GGC)(AGC) encoding $(GS)_3$ (G, glycine; S, serine) after UGC (encoding a cysteine residue), UAG stop codon, followed by a short stretch of RNA that hybridizes the DNA-PEG-CCPu. It should be noted that our FIT system using this display experiment did not contain release factor-1 (RF1), which usually terminates the elongation at UAG stop codon; instead, ribosome would be stalled at UAG codon and the efficiency of puromycin-peptide fusion should be enhanced. By means of this system, the estimated complexity generated by the $(NNK)_n$ library transcribed from the corresponding DNA library was approximately 10^{14}, and the efficiency of the peptide-puromycin fusion was generally more than 30% in our hands for representative model peptides with the range of lengths and degenerated sequences used in the library (data not shown). Therefore, we confidently estimated the diversity of the initial display library to be 10^{12} or more. We referred to this system enabling display of nonstandard macrocyclic peptide libraries as the RaPID (Random nonstandard Peptide Integrated Discovery) system (Yamagishi *et al. Chem. Biol.*, in press). Significantly, the C-terminal end of each thioether-macrocyclic peptide in the library is covalently linked to its encoding mRNA *via* Pu, and thus active species of peptide based on their binding capability against a therapeutic target could be enriched by RT-PCR (reverse transcription polymerase chain reaction). Therefore, the RaPID selection process can be repeated until an active population emerges in the selected pool.

Selection of Anti-Akt2 Macrocyclic Peptides. We performed *in vitro* selection of the nonstandard macrocyclic peptides against Akt2 (Figure 1b). The ^{L}Y- and ^{D}Y-libraries were independently applied to the RaPID selection against a full-length Akt2 that was immobilized on Ni^{2+}-NTA magnetic beads *via* the interaction with hexa-His residues at the Akt2's N-terminus. Prior to the selection against Akt2-beads, the respective libraries were treated with the Ni^{2+}-NTA magnetic beads (up to 12 times) to remove undesired background-binding peptide species in the pool, and the peptide fraction unbound to the beads was then applied to the selection against Akt2-immobilized beads. At the sixth round, we observed an appreciable enrichment of active population in the both pools monitored by the recovery amount of selected cDNA by RT-PCR, while the background binding to the beads was significantly suppressed (Supplementary Figure 1).

The enriched pool was cloned and individual colonies were arbitrarily picked for sequencing, yielding a total of 37 DNA sequences from each of the ^{L}Y- and ^{D}Y-libraries (Table 1). The most abundant ^{L}Y-clone, referred to as Pakti-L1, was found 28 times in 37 clones, and shared the common sequence with Pakti-L6 found once. Pakti-L1 consists of 11 residues originating from the random sequence region, giving a total body length of a 13-mer macrocyclic peptide without the linker peptide region. Two clones, Pakti-L2 and 3, were each found twice, and shared a common sequence motif, consisting of 12 residues originating from the random region (a total of 14-mer body length). Pakti-L4 and L5 were independently found once and thrice, respectively. On the other hand, the ^{D}Y-library was dominated by a single kind of a peptide; Pakti-D1 was found 35 times. Pakti-D2 and D3 were found once, respectively, and shared the common sequence motif. Interestingly, no Pakti-D peptides had sequence similarity to any Pakti-L peptides. This implied that, even though the random region originated from

dx.doi.org/10.1021/cb200388k | *ACS Chem. Biol.* 2012, 7, 607–613

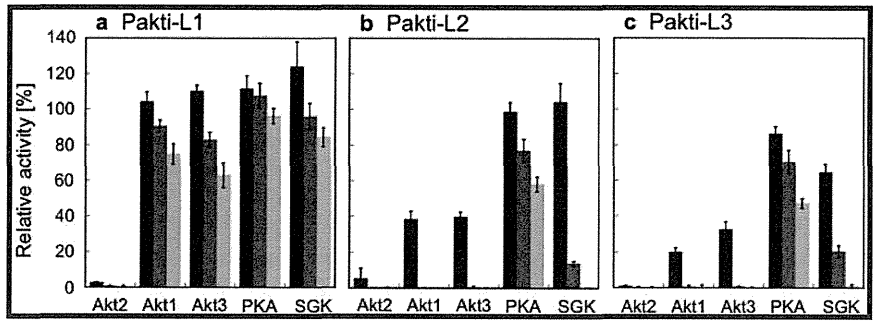


Figure 2. Inhibitory activities of Pakti-L1–L3 (a–c) against Akt isoforms and other kinases. (a) Akt- and isoform-selectivity of Pakti-L1. Black, gray, and light gray bars show observed activities of the kinase at the peptide concentrations of 1, 5, and 10 μM, respectively. The respective activity was defined as an average of activities of the kinase in triplicate in the presence of inhibitor relative to those in the absence of inhibitor. Error bars are shown as the standard deviations of the relative activities in triplicate. (b) Akt- and isoform-selectivity of Pakti-L2. (c) Akt- and isoform-selectivity of Pakti-L3.

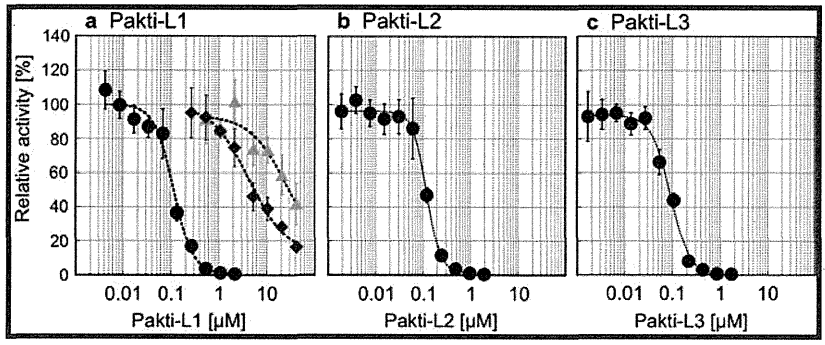


Figure 3. Titration of kinase activities as a function of Pakti-L inhibitor concentrations. (a) Inhibitory titration of Pakti-L1. Data plot in black circles are against Akt2. Those in light gray triangles and gray diamonds are against Akt1 and Akt3, respectively. IC_{50} values were determined by a curve fitting to a Hill equation generated by KaleidaGraph (Hulinks Inc.). (b) Inhibitory titration of Pakti-L2. (c) Inhibitory titration of Pakti-L3.

the same pool of mRNA sequences, the library composed of ^{L}Y-initiator covered different conformational sequence space from that of ^{D}Y-initiator.

Anti-Akt2 Macrocyclic Peptides Are Inhibitory Active. All peptides from the Pakti-L and -D families were chemically synthesized using a standard solid-phase method with an additional glycine-carboxamide at the C-terminus. The individual peptides were then subjected to kinase inhibitory screening using a universal S/T kinase assay system composed of biotin-labeled peptide substrates (HTRF KinEASE STK2 and STK3, the appropriate one of which was treated with the corresponding kinase). After the phosphorylation, the biotin in the substrate was captured with XL665-labeled streptavidin, and a phospho-specific monoclonal antibody labeled with Eu^{3+}-cryptate was used to detect the signal of time-resolved-fluorescence resonance energy transfer (TR-FRET). In this study, we used enzymatically active GST-Akt#-ΔPH (Carna Biosciences Inc.) in which GST, #, and ΔPH denoted glutathione *S*-transferase as a N-terminal tag, isoform number, and deletion of PH domain, respectively, and the respective Akt isoforms were phosphorylated at threonine/serine residues essential for the activation of kinase domain. Our preliminary activity screening for inhibition using Pakti-L1–L6 and Pakti-D1–D3 revealed that the Pakti peptides isolated in this study, except for Pakti-L4 and D2, were active inhibitors against Akt2 (Table 1). On the basis of this preliminary data, we chose three peptides from the Pakti-L family, Pakti-L1–L3, for further isoform-selectivity studies. In parallel, other representative S/T kinases, PKA (protein kinase A) and SGK (serum- and glucocorticoid-regulated protein kinase), were also tested for inhibition by these peptides (Figure 2a–c).

Akt- and Isoform-Selectivity of Pakti-L Peptides. The concentrations of each peptide were set at 1, 5, and 10 μM for the inhibitory test against Akt1–3, PKA, and SGK. Remarkably, under any of the above concentrations, Pakti-L1 completely shut down the kinase activity of Akt2, but all other kinases tested in this study retained activity over 50% even at 10 μM (Figure 2a). This result clearly shows a high isoform selectivity of Pakti-L1. Likewise, Pakti-L2 and L3 shut down the kinase activity of Akt2, but they also inhibited the kinase activity of both Akt1 and Akt3 to some extent (Figure 2b and c); at their 1 μM concentration, both Akt's gave an approximately 40% activity, and thus their IC_{50} values were estimated to be ~1 μM (Table 1). Similar to Pakti-L1, Pakti-L2 and L3 displayed very weak inhibitory activities against PKA and SGK (IC_{50} values could be <10 μM). Thus, Pakti-L2 and L3 are Akt2-selective but less isoform- and Akt-selective compared with Pakti-L1. Even though the *in vitro* selection of nonstandard macrocyclic peptides against Akt2 from the ^{L}Y- and ^{D}Y-libraries was based on their binding ability leading to enrichment, it has successfully yielded potent inhibitors from both libraries.

To assess more accurate IC_{50} values of the Akt2-inhibitors, we directly monitored the kinase activity of Akt2 by a classical filter-binding assay of ^{32}P-phosphorylated substrate peptide on a phosphocellulose paper.[40] For this experiment, we modified a known cross-reactive substrate, so-called crosstide, to generate a crosstide-KK in which two additional lysine residues were added to its C-terminus.[40,41] The phosphorylation activity was titrated in the presence of various concentrations of each peptide ranging from 2 to 2,100 nM against Akt2 (Figure 3a–c). All Pakti-L1–L3 displayed nearly the same inhibitory potencies against Akt2 with IC_{50} values of 100 nM (Table 1).

Because Pakti-L1 appears to exhibit a high selectivity to Akt2 according to the screening result (Figure 2a), we also attempted to determine the IC_{50} values against Akt1 and Akt3 (Figure 3a). The titration of Pakti-L1 in a range of 270−42,000 nM against Akt1 and Akt3 has revealed estimated IC_{50} values of 25 and 4.2 μM (Figure 3a and Table 1). It should be noted that the inhibitory activities of Pakti-L1 against Akt1 and Akt3 were intrinsically very weak, and therefore we were unable to completely titrate the inhibitory activity with 42 μM Pakti-L1, meaning that the above IC_{50} values were yet estimates. Nevertheless, the isoform selectivity of Pakti-L1 to Akt2 over Akt1 and Akt3 was 250- and 40-fold, respectively.

In summary, Pakti-L1, generated from ^{L}Y-library and studied in depth in the present work, displays an isoform-selective inhibitory activity against Akt2 over Akt1 and Akt3. It also exhibits nearly no activity against other families of S/T kinases. Thus, Pakti-L1 is a remarkable Akt-selective and Akt2 isoform-selective inhibitor. The other class of thioether-macrocyclic peptides, Pakti-L2 and L3, display potent inhibitory activity against Akt2, modest activities against Akt1/3, and very weak activity against other families of S/T kinases. Although they are weakly isoform-selective, they are highly Akt-selective inhibitors.

Discussion. The full-length of Akt2 used for the selection was an enzymatically inactive construct, *i.e.*, a nonphosphorylated form. Therefore, it is quite surprising that the majority of selected peptides found in ^{L}Y-library were able to show inhibitory potencies with a range of 100 nM IC_{50} against active Akt2. Moreover, all inhibitors found in this study inhibit the kinase activity of Akt2-ΔPH, suggesting that their binding site(s) should reside in the kinase domain of Akt2. The most intriguing question is how these macrocyclic peptides inhibit the kinase activity of Akt. Particularly, Pakti-L1 exhibits a high isoform-selective activity, so the questions where it binds in the Akt kinase domain and how it discriminates Akt2 over Akt1/3 are of greatest interest.

Since we have tertiary structural information of neither Pakti-L1 nor its complex with Akt2, we are only able to provide a speculative discussion for possible mechanisms of its isoform-selectivity, but it would be worthy to discuss such at this point. We propose two hypotheses as follows. The first hypothesis is that Pakti-L1 interacts with the substrate-binding domain and competitively inhibits Akt2 activity. It is known that a generic substrate sequence, such as crosstide, contains a RXRXX(S/T) motif (the phosphorylating S/T is assigned to position 0, and its N-terminal region is assigned to −1 to −5) where the arginine residues at positions −3 and −5, interact with the conserved glutamate residues (E236, E279, and E342) in Akt (Supplementary Figure 2a).[42] It turns out that Pakti-L1 has a RNR motif embedded in the middle of sequence (Table 1), and therefore this motif may be able to position Pakti-L1 into the active site. On the other hand, all Akt kinases share the same amino acid residues that interact with the substrate backbone amides or side chains, and unique residues in Akt2 differing from those in Akt1/3 are scattered outside of active site in the structure (see Supplementary Figure 2b). Therefore, if this hypothesis were correct, Pakti-L1 could use a unique mechanism, such as induced fit, to gain specific interactions, enabling the discrimination of Akt2 against Akt1/3.

The second hypothesis is that Pakti-L1 acts as an allosteric inhibitor. It may have a mechanism similar to Akti-1/2 where Pakti-L1 stabilizes the interaction between the kinase and PH domains, resulting in stabilization of the kinase-inactive "PH-in" form.[26] However, since Pakti-L1 is structurally bigger than Akti-1/2, the binding mode of Pakti-L1 should be largely different from that seen in Akti-1/2. Alternatively, Pakti-L1 interacts with a unique short sequence clustered in the region of 455−466 and allosterically inhibits the kinase activity (Supplementary Figure 2a and b). Unfortunately, this region is invisible in any of available X-ray structures of Akt, suggesting that this region may be largely unstructured. It is known that the downstream sequence, F470−Y475, docks on the "N-lobe" region when S474 is phosphorylated, resulting in activation of the kinase function.[42] Thus, if Pakti-L1 binds to this unique region and prevents the F470−Y475 region from docking, the kinase activity can be allosterically inhibited.

Since we do not have any data that enables us to rule out either hypothesis, we would like to leave open questions for the future investigations. Presumably, structural studies of the binary complex of the kinase domain with Pakti-L1 or the ternary complex with PH domain will be critical to reveal the exact inhibitory mechanism of Pakti-L1.

Conclusion. Here, we have reported thioether-macrocyclic peptide inhibitors with high Akt-selectivity generated by RaPID system. One of the inhibitors, Pakti-L1, was studied in depth and also showed a high isoform-selectivity against Akt2 over other isoforms. It is quite remarkable that even though the selection was performed against only Akt2, the selected peptides possessed isoform-selective properties. This suggests that the RaPID system could have a potential to yield isoform-selective inhibitors against various therapeutic targets of which isoform-selectivity is a strong requisite for the development of therapeutic agents. More extensive attempts by means of RaPID system against other therapeutic targets are currently underway in our laboratory.

METHODS

Preparation of ClAc^{L}Y and ^{D}Y Peptide Libraries. The mRNAs with the 4−12 repeated NNK sequences for the random region of amino acid sequence were prepared by *in vitro* transcription of their template cDNA amplified by PCR, respectively, using the primers P1, PNNK4−PNNK12, and P2 (Supplementary Table 2). The mRNA library was prepared by mixing the respective mRNAs with the 4−12 repeated NNK sequences with the following ratio, $(NNK)_4$:$(NNK)_5$:$(NNK)_6$:$(NNK)_7$:$(NNK)_8$:$(NNK)_9$:$(NNK)_{10}$:$(NNK)_{11}$:$(NNK)_{12}$ = 20^{-3}:20^{-2}:20^{-1}:1:10:10:10:10:10, adjusted to 10 μM total concentration of mRNA and denoted as a mRNA library. Two samples of 1 μM mRNA library were ligated with 1.5 μM of DNA-PEG-CCPu (Supplementary Table 2) by T4 RNA ligase at a scale of 200 μL total volume and incubated at RT for 30 min. The ligated libraries were purified by phenol-chloroform treatment and ethanol precipitation. The mRNA libraries fused with DNA-PEG-CCPu were translated in the Met-deficient FIT system at a scale of 150 μL total volume containing 50 μM ClAc^{L}Y-tRNA$^{fMet}_{CAU}$ or ClAc^{D}Y-tRNA$^{fMet}_{CAU}$,[36,37] prepared as described in Supporting Information, for 30 min at 37 °C and an additional 12 min at RT to enhance the mRNA-peptide conjugation efficiency. To quench the reaction, the samples were mixed with 15 μL of 200 mM EDTA (pH 7.5) and incubated at 37 °C for 30 min. For desalting, the samples were gel-filtered through 700 μL of fresh cross-linked dextran polymer beads, Sephadex G-25 Fine (GE Healthcare), equilibrated with PBST (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, and 0.05% (v/v) Tween20 adjusted to pH 7.6 with HCl) twice. Preparation of the two libraries, ClAc^{L}Y (^{L}Y) library and ClAc^{D}Y (^{D}Y) library, was completed by adding an equivalent volume of 2× blocking buffer (1 M NaCl and 0.2% (w/v) acetyl-BSA in PBST) into the filtrates.

Selection Procedure by RaPID System. The following enrichment process was independently performed with the respective ^{L}Y and ^{D}Y libraries. At the first round, the ^{L}Y and ^{D}Y libraries were applied to

dx.doi.org/10.1021/cb200388k | ACS Chem. Biol. 2012, 7, 607−613

no protein-immobilized Ni^{2+}-NTA magnetic beads (Life Technologies, Cat. No. 10104D) twice to remove all the components tagged with the hexahistidine in the Met-deficient FIT system and undesired background beads binders from the libraries (this process is referred to as preclear below), then applied to $(His)_6$-Akt2-immobilized Ni^{2+}-NTA magnetic beads, and incubated at 4 °C for 30 min (this process is referred to as positive selection). After the incubation, the beads were washed with 500 μL of PBST four times. The washed beads were resuspended in 40 μL of reverse-transcription reaction mixture. The mRNAs remaining on the Akt2-immobilized beads were reverse-transcribed by MMLV reverse-transcriptase (Promega, Cat. No. M1705) at 42 °C for 1 h using the primer P2. The resultant cDNAs on the beads were eluted by being mixed with 200 μL of 1× PCR reaction buffer and heated at 95 °C for 5 min, with immediate separation of the supernatant from the beads. The amounts of the eluted cDNAs from the beads in the respective libraries were measured by qPCR, and the eluted cDNAs were amplified by PCR using the primer P1 and P2 for the preparation of the mRNA libraries for the next round. The aforementioned processes of (selected) library preparation, preclear, and positive selection constitute one cycle of the enrichment process.

From the second round, in the enrichment process, the reverse-transcription of mRNA to prepare the cDNAs was performed prior to the selection and the selection was performed by autodispenser machine (Nikkyo Technos Co., Ltd.). At the second round, the individual peptide libraries were prepared by *in vitro* transcription of the amplified cDNAs, ligation of the mRNA libraries with DNA-PEG-CCPu, *in vitro* translation, and reverse transcription. The libraries prepared in 5 μL of the translation scale and diluted to 100 μL with blocking buffer and PBST, the fresh magnetic beads, the wash buffer (PBST), inactive Akt2 protein, and H_2O were set in the appropriate wells of the 96-well plate. The machine was programmed to run the following processes, immobilization of Akt2 protein to Ni^{2+}-NTA magnetic beads, washing the extra Akt2 protein from the beads, and applying the respective library to the preclear and selection processes. After iterative preclear processes up to 6–12 times depending upon the rounds, the respective libraries were applied to the Akt2-immobilized beads for the selection and incubated at about 10 °C for 15–30 min. The beads were washed with 100 μL of PBST three times and resuspended with 40 μL of H_2O, and then the machine program was executed to completion. The beads were mixed with 50 μL of 2× PCR buffer and adjusted to 100 μL total volume with H_2O. The eluted cDNAs were recovered by heating the sample at 95 °C for 5 min and separating the supernatant from the beads. The recovered cDNAs were amplified by PCR and then transcribed to prepare mRNA libraries for the next round. The enrichment process was completed at the sixth round, where appreciable enrichment of the recovery rate was observed.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

This material is available free of charge *via* the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: hsuga@chem.s.u-tokyo.ac.jp.

■ ACKNOWLEDGMENTS

We thank Y. Aoki for early contribution to this work, M. Hayashi for assistance with the structural analysis of Akt, and C. J. Hipolito for proofreading. We also thank M. Noguchi and F. Suizu (Hokkaido University) for the gift of PH domain in our earlier work, and H. Murakami, Y. Gotoh, and M. Higuchi (The University of Tokyo) for thoughtful discussions. This work was supported by a JSPS Grant-in-Aid for the Specially Promoted Research (21000005), a research and development projects of the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization (NEDO) to H.S. and Grants-in-Aid for JSPS fellows to J.M. in part (21-9079).

■ REFERENCES

(1) Bellacosa, A., Kumar, C. C., Di Cristofano, A., and Testa, J. R. (2005) Activation of AKT kinases in cancer: implications for therapeutic targeting. *Adv. Cancer Res. 94*, 29–86.

(2) Romashkova, J. A., and Makarov, S. S. (1999) NF-kappaB is a target of AKT in anti-apoptotic PDGF signalling. *Nature 401*, 86–90.

(3) Staal, S. P. (1987) Molecular cloning of the akt oncogene and its human homologues AKT1 and AKT2: amplification of AKT1 in a primary human gastric adenocarcinoma. *Proc. Natl. Acad. Sci. U.S.A. 84*, 5034–5037.

(4) Sun, M., Wang, G., Paciga, J. E., Feldman, R. I., Yuan, Z. Q., Ma, X. L., Shelley, S. A., Jove, R., Tsichlis, P. N., Nicosia, S. V., and Cheng, J. Q. (2001) AKT1/PKBalpha kinase is frequently elevated in human cancers and its constitutive activation is required for oncogenic transformation in NIH3T3 cells. *Am. J. Pathol. 159*, 431–437.

(5) Bellacosa, A., de Feo, D., Godwin, A. K., Bell, D. W., Cheng, J. Q., Altomare, D. A., Wan, M., Dubeau, L., Scambia, G., Masciullo, V., Ferrandina, G., Benedetti Panici, P., Mancuso, S., Neri, G., and Testa, J. R. (1995) Molecular alterations of the AKT2 oncogene in ovarian and breast carcinomas. *Int. J. Cancer 64*, 280–285.

(6) Cheng, J. Q., Godwin, A. K., Bellacosa, A., Taguchi, T., Franke, T. F., Hamilton, T. C., Tsichlis, P. N., and Testa, J. R. (1992) AKT2, a putative oncogene encoding a member of a subfamily of protein-serine/threonine kinases, is amplified in human ovarian carcinomas. *Proc. Natl. Acad. Sci. U.S.A. 89*, 9267–9271.

(7) Yuan, Z. Q., Sun, M., Feldman, R. I., Wang, G., Ma, X., Jiang, C., Coppola, D., Nicosia, S. V., and Cheng, J. Q. (2000) Frequent activation of AKT2 and induction of apoptosis by inhibition of phosphoinositide-3-OH kinase/Akt pathway in human ovarian cancer. *Oncogene 19*, 2324–2330.

(8) Cho, H., Mu, J., Kim, J. K., Thorvaldsen, J. L., Chu, Q., Crenshaw, E. B. 3rd, Kaestner, K. H., Bartolomei, M. S., Shulman, G. I., and Birnbaum, M. J. (2001) Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB beta). *Science 292*, 1728–1731.

(9) Brodbeck, D., Cron, P., and Hemmings, B. A. (1999) A human protein kinase Bgamma with regulatory phosphorylation sites in the activation loop and in the C-terminal hydrophobic domain. *J. Biol. Chem. 274*, 9133–9136.

(10) Tschopp, O., Yang, Z. Z., Brodbeck, D., Dummler, B. A., Hemmings-Mieszczak, M., Watanabe, T., Michaelis, T., Frahm, J., and Hemmings, B. A. (2005) Essential role of protein kinase B gamma (PKB gamma/Akt3) in postnatal brain development but not in glucose homeostasis. *Development 132*, 2943–2954.

(11) Luo, Y., Shoemaker, A. R., Liu, X., Woods, K. W., Thomas, S. A., de Jong, R., Han, E. K., Li, T., Stoll, V. S., Powlas, J. A., Oleksijew, A., Mitten, M. J., Shi, Y., Guan, R., McGonigal, T. P., Klinghofer, V., Johnson, E. F., Leverson, J. D., Bouska, J. J., Mamo, M., Smith, R. A., Gramling-Evans, E. E., Zinker, B. A., Mika, A. K., Nguyen, P. T., Oltersdorf, T., Rosenberg, S. H., Li, Q., and Giranda, V. L. (2005) Potent and selective inhibitors of Akt kinases slow the progress of tumors in vivo. *Mol. Cancer Ther. 4*, 977–986.

(12) Okuzumi, T., Fiedler, D., Zhang, C., Gray, D. C., Aizenstein, B., Hoffman, R., and Shokat, K. M. (2009) Inhibitor hijacking of Akt activation. *Nat. Chem. Biol. 5*, 484–493.

(13) Heerding, D. A., Rhodes, N., Leber, J. D., Clark, T. J., Keenan, R. M., Lafrance, L. V., Li, M., Safonov, I. G., Takata, D. T., Venslavsky, J. W., Yamashita, D. S., Choudhry, A. E., Copeland, R. A., Lai, Z., Schaber, M. D., Tummino, P. J., Strum, S. L., Wood, E. R., Duckett, D. R., Eberwein, D., Knick, V. B., Lansing, T. J., McConnell, R. T., Zhang, S., Minthorn, E. A., Concha, N. O., Warren, G. L., and Kumar, R. (2008) Identification of 4-(2-(4-amino-1,2,5-oxadiazol-3-yl)-1-ethyl-7-{[(3*S*)-3-piperidinylmethyl]oxy}-1*H*-imidazo[4,5-c]pyridin-4-yl)-2-

methyl-3-butyn-2-ol (GSK690693), a novel inhibitor of AKT kinase. *J. Med. Chem. 51*, 5663–5679.
(14) Yap, T. A., Walton, M. I., Hunter, L. J., Valenti, M., de Haven Brandon, A., Eve, P. D., Ruddle, R., Heaton, S. P., Henley, A., Pickard, L., Vijayaraghavan, G., Caldwell, J. J., Thompson, N. T., Aherne, W., Raynaud, F. I., Eccles, S. A., Workman, P., Collins, I., and Garrett, M. D. (2011) Preclinical pharmacology, antitumor activity, and development of pharmacodynamic markers for the novel, potent AKT inhibitor CCT128930. *Mol. Cancer Ther. 10*, 360–371.
(15) Grimshaw, K. M., Hunter, L. J., Yap, T. A., Heaton, S. P., Walton, M. I., Woodhead, S. J., Fazal, L., Reule, M., Davies, T. G., Seavers, L. C., Lock, V., Lyons, J. F., Thompson, N. T., Workman, P., and Garrett, M. D. (2010) AT7867 is a potent and oral inhibitor of AKT and p70 S6 kinase that induces pharmacodynamic changes and inhibits human tumor xenograft growth. *Mol. Cancer Ther. 9*, 1100–1110.
(16) Kondapaka, S. B., Singh, S. S., Dasmahapatra, G. P., Sausville, E. A., and Roy, K. K. (2003) Perifosine, a novel alkylphospholipid, inhibits protein kinase B activation. *Mol. Cancer Ther. 2*, 1093–1103.
(17) Hiromura, M., Okada, F., Obata, T., Auguin, D., Shibata, T., Roumestand, C., and Noguchi, M. (2004) Inhibition of Akt kinase activity by a peptide spanning the betaA strand of the proto-oncogene TCL1. *J. Biol. Chem. 279*, 53407–53418.
(18) Kim, D., Sun, M., He, L., Zhou, Q. H., Chen, J., Sun, X. M., Bepler, G., Sebti, S. M., and Cheng, J. Q. (2010) A small molecule inhibits Akt through direct binding to Akt and preventing Akt membrane translocation. *J. Biol. Chem. 285*, 8383–8394.
(19) Yang, L., Dan, H. C., Sun, M., Liu, Q., Sun, X. M., Feldman, R. I., Hamilton, A. D., Polokoff, M., Nicosia, S. V., Herlyn, M., Sebti, S. M., and Cheng, J. Q. (2004) Akt/protein kinase B signaling inhibitor-2, a selective small molecule inhibitor of Akt signaling with antitumor activity in cancer cells overexpressing Akt. *Cancer Res. 64*, 4394–4399.
(20) Berndt, N., Yang, H., Trinczek, B., Betzi, S., Zhang, Z., Wu, B., Lawrence, N. J., Pellecchia, M., Schonbrunn, E., Cheng, J. Q., and Sebti, S. M. (2010) The Akt activation inhibitor TCN-P inhibits Akt phosphorylation by binding to the PH domain of Akt and blocking its recruitment to the plasma membrane. *Cell Death Differ. 17*, 1795–1804.
(21) Miao, B., Skidan, I., Yang, J., Lugovskoy, A., Reibarkh, M., Long, K., Brazell, T., Durugkar, K. A., Maki, J., Ramana, C. V., Schaffhausen, B., Wagner, G., Torchilin, V., Yuan, J., and Degterev, A. (2010) Small molecule inhibition of phosphatidylinositol-3,4,5-triphosphate (PIP3) binding to pleckstrin homology domains. *Proc. Natl. Acad. Sci. U.S.A. 107*, 20126–20131.
(22) Yan, L. (2009) MK-2206: A potent oral allosteric AKT inhibitor, *AACR Meeting Abstracts 2009*, Abstract DDT01-1.
(23) Liu, R., Liu, D., Trink, E., Bojdani, E., Ning, G., and Xing, M. (2011) The Akt-specific inhibitor MK2206 selectively inhibits thyroid cancer cells harboring mutations that can activate the PI3K/Akt pathway. *J. Clin. Endocrinol. Metab. 96*, E577–585.
(24) Lindsley, C. W., Zhao, Z., Leister, W. H., Robinson, R. G., Barnett, S. F., Defeo-Jones, D., Jones, R. E., Hartman, G. D., Huff, J. R., Huber, H. E., and Duggan, M. E. (2005) Allosteric Akt (PKB) inhibitors: discovery and SAR of isozyme selective inhibitors. *Bioorg. Med. Chem. Lett. 15*, 761–764.
(25) Gilot, D., Giudicelli, F., Lagadic-Gossmann, D., and Fardel, O. (2010) Akti-1/2, an allosteric inhibitor of Akt 1 and 2, efficiently inhibits CaMKIalpha activity and aryl hydrocarbon receptor pathway. *Chem. Biol. Interact. 188*, 546–552.
(26) Wu, W. I., Voegtli, W. C., Sturgis, H. L., Dizon, F. P., Vigers, G. P., and Brandhuber, B. J. (2010) Crystal structure of human AKT1 with an allosteric inhibitor reveals a new mode of kinase inhibition. *PLoS One 5*, e12913.
(27) Litman, P., Ohne, O., Ben-Yaakov, S., Shemesh-Darvish, L., Yechezkel, T., Salitra, Y., Rubnov, S., Cohen, I., Senderowitz, H., Kidron, D., Livnah, O., Levitzki, A., and Livnah, N. (2007) A novel substrate mimetic inhibitor of PKB/Akt inhibits prostate cancer tumor growth in mice by blocking the PKB pathway. *Biochemistry 46*, 4716–4724.
(28) Obata, T., Yaffe, M. B., Leparc, G. G., Piro, E. T., Maegawa, H., Kashiwagi, A., Kikkawa, R., and Cantley, L. C. (2000) Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. *J. Biol. Chem. 275*, 36108–36115.
(29) Luo, Y., Smith, R. A., Guan, R., Liu, X., Klinghofer, V., Shen, J., Hutchins, C., Richardson, P., Holzman, T., Rosenberg, S. H., and Giranda, V. L. (2004) Pseudosubstrate peptides inhibit Akt and induce cell growth inhibition. *Biochemistry 43*, 1254–1263.
(30) Shin, I., Edl, J., Biswas, S., Lin, P. C., Mernaugh, R., and Arteaga, C. L. (2005) Proapoptotic activity of cell-permeable anti-Akt single-chain antibodies. *Cancer Res. 65*, 2815–2824.
(31) Levy, D. S., Kahana, J. A., and Kumar, R. (2009) AKT inhibitor, GSK690693, induces growth inhibition and apoptosis in acute lymphoblastic leukemia cell lines. *Blood 113*, 1723–1729.
(32) Murakami, H., Ohta, A., Ashigai, H., and Suga, H. (2006) A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat. Methods 3*, 357–359.
(33) Ohta, A., Murakami, H., Higashimura, E., and Suga, H. (2007) Synthesis of polyester by means of genetic code reprogramming. *Chem. Biol. 14*, 1315–1322.
(34) Kawakami, T., Murakami, H., and Suga, H. (2008) Messenger RNA-programmed incorporation of multiple N-methyl-amino acids into linear and cyclic peptides. *Chem. Biol. 15*, 32–42.
(35) Kawakami, T., Murakami, H., and Suga, H. (2008) Ribosomal synthesis of polypeptoids and peptoid-peptide hybrids. *J. Am. Chem. Soc. 130*, 16861–16863.
(36) Goto, Y., Ohta, A., Sako, Y., Yamagishi, Y., Murakami, H., and Suga, H. (2008) Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. *ACS Chem. Biol. 3*, 120–129.
(37) Goto, Y., Katoh, T., and Suga, H. (2011) Flexizymes for genetic code reprogramming. *Nat. Protoc. 6*, 779–790.
(38) Nemoto, N., Miyamoto-Sato, E., Husimi, Y., and Yanagawa, H. (1997) In vitro virus: bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett. 414*, 405–408.
(39) Roberts, R. W., and Szostak, J. W. (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. U.S.A. 94*, 12297–12302.
(40) Hastie, C. J., McLauchlan, H. J., and Cohen, P. (2006) Assay of protein kinases using radiolabeled ATP: a protocol. *Nat. Protoc. 1*, 968–971.
(41) Cross, D. A., Alessi, D. R., Cohen, P., Andjelkovich, M., and Hemmings, B. A. (1995) Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature 378*, 785–789.
(42) Yang, J., Cron, P., Good, V. M., Thompson, V., Hemmings, B. A., and Barford, D. (2002) Crystal structure of an activated Akt/protein kinase B ternary complex with GSK3-peptide and AMP-PNP. *Nat. Struct. Biol. 9*, 940–944.

Organic & Biomolecular Chemistry

Dynamic Article Links
View Article Online

Cite this: *Org. Biomol. Chem.*, 2012, **10**, 5783

www.rsc.org/obc

COMMUNICATION

Selective thioether macrocyclization of peptides having the N-terminal 2-chloroacetyl group and competing two or three cysteine residues in translation†‡

Kazuhiro Iwasaki,[a] Yuki Goto,[b] Takayuki Katoh[b] and Hiroaki Suga*[b]

Received 11th February 2012, Accepted 29th February 2012
DOI: 10.1039/c2ob25306b

Published on 01 March 2012 on http://pubs.rsc.org | doi:10.1039/C2OB25306B

The mode of thioether macrocyclization of peptides containing an N-terminal 2-chloroacetyl group and two or three competing cysteine residues at downstream positions has been extensively studied, leading to a strategy for designated formation of overlapping-bicyclic peptides or dumbbell-type bicyclic peptides.

Naturally occurring peptides isolated from various organisms often have macrocyclic structures.[1] These structural features rigidify their tertiary structure, often granting them high affinities to the binding partners or targets. Unlike inter-sidechain disulfide bonds mostly found in ribosomally synthesized proteins, macrocyclization in such peptides takes place in various ways, *i.e.* head-to-tail, head-to-sidechain, sidechain-to-tail, inter-sidechain and their combinations,[2] giving not only their structural diversities but also greater physiological stability.[3] Since the formation of these unique bonds does not occur spontaneously, post-translational modification enzymes or enzymes in a part of non-ribosomal peptide synthetase are involved in the transformation.[4] On the other hand, as such enzymes generally have a preference toward certain sets of peptide sequences and/or particular recognition motifs upstream and/or downstream of the target sequences, their versatility for the synthesis of diverse and artificial macrocyclic peptides could be limited.

To circumvent this limitation, we recently devised a methodology[5] for spontaneous thioether macrocyclization of ribosomally expressed peptides using Flexible *In-vitro* Translation (FIT) system.[6] In this system, an artificial amino acid bearing a 2-chloroacetyl (ClAc) group on the α-amino group was incorporated by initiation codon reprogramming at the N-terminus, and reacts with sulfhydryl group in a cysteine (C) residue at a downstream position (Fig. 1a). Although this head-to-sidechain macrocyclization chemistry itself has been known for the chemical synthesis of peptides,[7] the above method using the FIT system represents the first example of thioether macrocyclization of ribosomally expressed peptides. Most importantly, this methodology allows us to readily construct various sequences and lengths of thioether macrocyclic peptides by simply designing their mRNA sequences which are transcribed from the corresponding synthetic DNAs *in vitro*.[5] Moreover, we have recently constructed such genetically encoded libraries of thioether macrocyclic peptides with a complexity of over 10^{12} and applied them to the discovery of inhibitors against various enzyme targets, *e.g.* E6AP,[8] Akt2,[9] SIRT2,[10] by integration with the mRNA display technique,[11] referred to as RaPID (Random non-standard Peptide Integrated Discovery) system as a whole.

A typical design of the above libraries is to have an *N*-ClAc-amino acid encoded by AUG as an initiator, random elongator amino acid sequences encoded by $(NNK)_n$ (K = U or G, n = 7–15), $(NNU)_n$, or $(NNC)_n$, followed by a cysteine (C) encoded

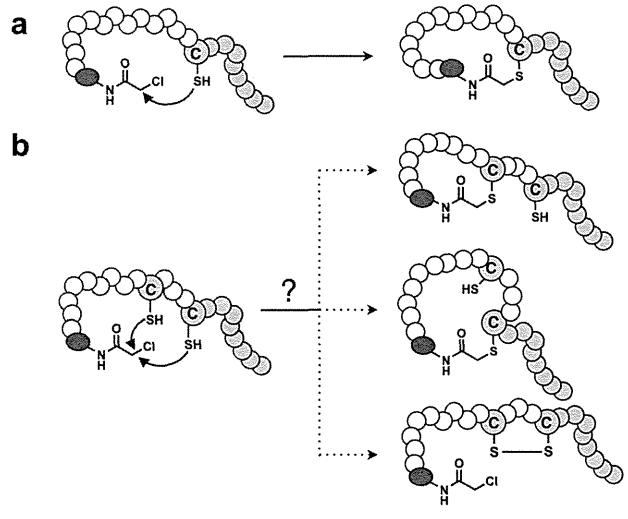


Fig. 1 Schematic representation of peptide macrocyclization closed by a thioether bond. (**a**) Peptide with N-terminal ClAc group that reacts with a cysteine (C) residue. (**b**) Three possible cyclization pathways in *N*-ClAc-peptides with two cysteine residues.

[a]*Department of Chemistry and Biotechnology, Graduate School of Engineering, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
[b]*Department of Chemistry, Graduate School of Science, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan. E-mail: hsuga@chem.s.u-tokyo.ac.jp*

† This article is part of the *Organic & Biomolecular Chemistry* 10th Anniversary issue.

‡ Electronic supplementary information (ESI) available: Materials, Experimental procedures, MS data, and MS/MS data. See DOI: 10.1039/c2ob25306b

This journal is © The Royal Society of Chemistry 2012

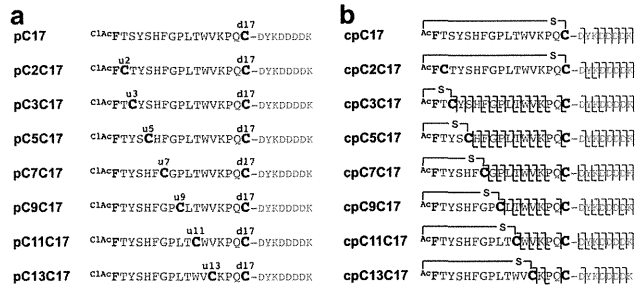


Fig. 2 Peptides and their fragmentation profiles observed in MALDI-TOF/TOF mass spectrometry. (**a**) Sequences of linear peptides expressed in this study. (**b**) Observed fragment peaks in each cyclized peptide and putative thioether bond formation based on the fragmentation profile. L shape and rotated L shape denote observations of y-ion and b-ion peaks, respectively.

Published on 01 March 2012 on http://pubs.rsc.org | doi:10.1039/C2OB25306B
Downloaded by Tokyo Daigaku on 28 January 2013

by UGC. As a result, the major fraction of expressed peptides would cyclize between the N-terminus and the designated C residue at the C-terminus. However, during the course of such selection studies, we encountered active peptide sequences containing two (or possibly more) cysteine residues, one of which appeared in the random region encoded by UGU or UGC. In such peptides (Fig. 1b), both C residues could potentially react with the N-terminal ClAc group to form a thioether bond,[8] but it was unclear which bond would be preferentially or even selectively formed in such peptide sequences (Fig. 1b, top *vs.* middle). Alternatively, a disulfide bond between two C residues could possibly compete with the thioether bond formation (Fig. 1b, bottom). To understand a general rule of macrocyclization in such peptides, here we have designed experiments to express model peptides varying the position of an 'upstream Cys' (^{u}C) residue in addition to the designated 'downstream Cys' (^{d}C) residue.

We used a FIT system customized for the initiation codon reprogramming, where an initiator tRNA charged with *N*-ClAc-phenylalanine (^{ClAc}Phe) by means of flexizyme was supplemented to a Met-deficient cell-free translation system. We designed seven mRNA sequences encoding 25-mer model peptides (Fig. 2a), each of which was initiated with ^{ClAc}Phe, elongated with 15 residues containing ^{u}C and ^{d}C, and further elongated with a Flag tag (DYKDDDDK, where D, Y, K are aspartate, tyrosine, and lysine; it is used for not only an affinity purification tag but also for facilitating ionization of expressed peptides in mass spectrometry). In these peptides, ^{u}C was located in the 2nd position or an odd position from the 3rd to the 13th position along with a fixed ^{d}C at the 17th position (pC2C17–pC13C17, where the former and latter number denotes the position of ^{u}C and ^{d}C residue and p strands for peptide). Thus, each peptide would possibly produce three macrocyclic structures with a thioether bond between the N-terminus and ^{u}C or ^{d}C, or an ^{u}C–^{d}C disulfide bond. As a control, an mRNA sequence encoding a simple thioether macrocyclic peptide was also prepared (pC17).

All mRNA sequences expressed the encoded peptide with a loss of approximately 36 Da from the molecular mass of the full-length peptide, *i.e.* a loss of HCl, as a major product peak confirmed by MALDI-TOF (Fig. S1‡). This result indicated that the thioether bond formation preferentially occurred over the disulfide bond formation. The respective peptides were then subjected to the fragmentation by MALDI-TOF/TOF, of which the individual peaks were assigned to the expected fragments. Interestingly, two peptides, cpC17 and cpC2C17 (c stands for cyclic) were poorly fragmented at the region between F1 and ^{d}C17 (Fig. S2‡). On the other hand, cpC3C17–cpC13C17 afforded various peaks of fragments, supposedly originating from the linear region when macrocyclization took place between the N-terminus and ^{u}C; whereas, fragment of the thioether macrocyclic region was not found similar to that observed for cpC17 (Fig. S3‡). These results consistently suggested that macrocyclization of pC3C17–pC13C17 took place between the N-terminus and ^{u}C, giving the respective structures consisting of a macrocyclic head and a linear tail with a free sulfhydryl on ^{d}C17. In contrast, macrocyclization of pC2C17 predominantly took place between the N-terminus and ^{d}C17, giving the macrocyclic structure with a free sulfhydryl on ^{u}C2 (Fig. 2b). This selective macrocyclization could be attributed to the fact that the nucleophilic attack of the sulfhydryl group of ^{u}C2 on the α-carbon of the *N*-ClAc group required the formation of a constrained 9-membered ring, which might be unfavourable over macrocyclization between the N-terminus and ^{d}C17.

To solidify the above hypothesis, we expressed cpC3C17 peptide containing a single ester bond (cpC3e12C17; e denoting the ester bond) where tryptophan at the 12th position (W12) was replaced with phenyllactic acid (^{HO}F) by reprogramming of its codon (Fig. S4a‡).[12] As a control, we also expressed cpC17 with the same ester substitution at the 12th position (Fig. S4b‡). MALDI-TOF analysis of each peptide confirmed the production of macrocyclic peptide-ester hybrid as a single major product (Fig. S4c and d‡). Upon subjecting to alkaline hydrolysis, the ester bond was selectively cleaved to yield a product or products as follows: in cpC3e12C17, two fragment peaks were observed, each of which was assigned to the thioether macrocyclic peptide with the carboxylate tail (**1**) and the linear tail peptide (**2**) cleaved at ^{HO}F, respectively (Fig. S4c‡); in a control peptide cpe12C17, on the other hand, a single peak with the molecular weight corresponding to the hydrolysis product at the 12th position (**3**) was detected as a single piece of sequence (Fig. S4d‡). These results were consistent with our hypothesis that the macrocyclization of pC3C17–pC13C17 took place between the N-terminus and ^{u}C.

With all knowledge together, we designed bicyclic peptides consisting of a head-to-sidechain thioether bond and an inter-sidechain disulfide bond. We first demonstrated expression of a peptide, referred to as pC2C10C17, where leucine at the 10th position (L10) in pC2C17 was replaced with a 'middle Cys' (^{m}C10) residue. Because ^{u}C2 should not react with the N-terminal ClAc group, we expected that the sulfhydryl group of ^{m}C10 would spontaneously react with the N-terminus, leaving two thiol-free cysteine residues, ^{u}C2 and ^{d}C17; then these two residues would form a disulfide bond under mild oxidation conditions, yielding an overlapping-bicyclic backbone structure (Fig. 3a). In fact, the mass spectrum of the expressed pC2C10C17 after air-oxidization showed a peak corresponding to the expected bicyclic structure (Fig. 3b, pC2C10C17-oxi). Furthermore, the tandem mass analysis of cpC2C10C17-oxi by MALDI-TOF/TOF showed fragment peaks only corresponding

This journal is © The Royal Society of Chemistry 2012

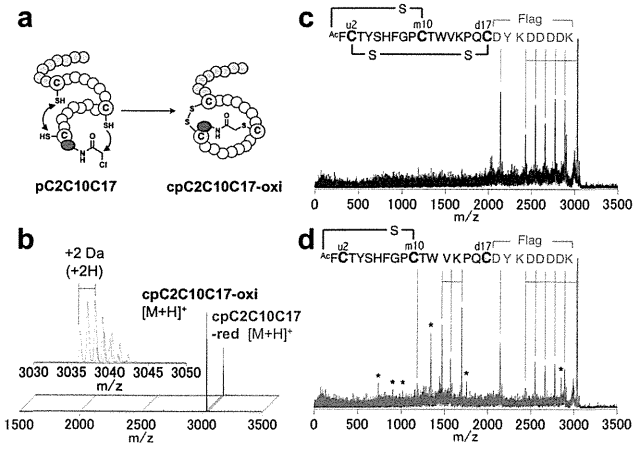


Fig. 3 Expression of an overlapping-bicyclic peptide consisting of a thioether bond and a disulfide bond. (a) Schematic representation of macrobicyclic peptides based on the pC2C10C17 scaffold. Because C2 is less susceptible to the reaction with the *N*-ClAc group, C10 predominantly forms a thioether bond and the remaining C2 and C17 form a disulfide bond. (b) MALDI-TOF spectra of cpC2C10C17 with two forms. Mild air-oxidation of the expressed peptide yielded the peptide, cpC2C10C17-oxi, that was consistent with the molecular mass containing a thioether bond and a disulfide bond (black spectrum). Its TCEP treatment yielded a mass with an additional 2 Da (cpC2C10C17-red), suggesting that the disulfide bond was reduced (red spectrum). The inset figure shows an expanded 3030–3050 *m/z* area of the superimposed spectra of cpC2C10C17-oxi (Cal. 3036.23, Obs. 3036.08) and cpC2C10C17-red (Cal. 3038.24, Obs. 3038.32). (c) MALDI-TOF/TOF spectrum of cpC2C10C17-oxi. (d) MALDI-TOF/TOF spectrum of cpC2C10C17-red. Assigned b-ion peaks are shown in the figure. Peaks labeled with asterisks are assigned as y-ion peaks.

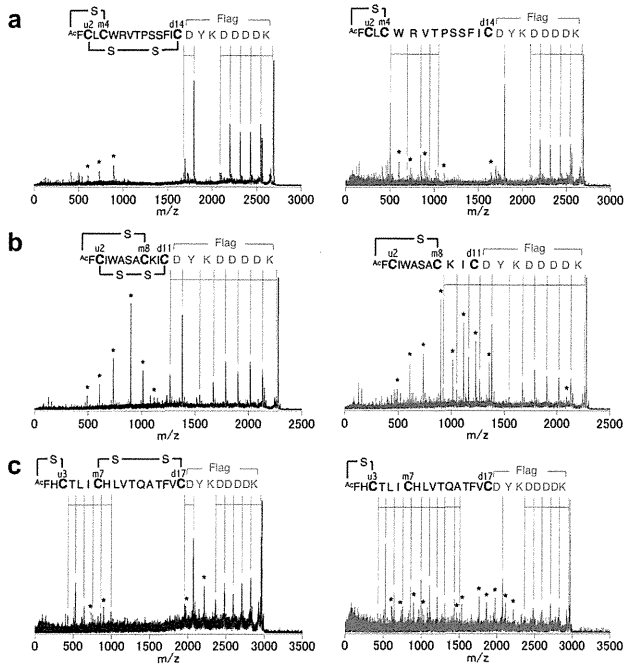


Fig. 4 MALDI-TOF/TOF spectra of overlapping- or dumbbell-type macrocyclic peptides. (a) cpC2C4C14, (b) cpC2C8C11, and (c) cpC3C7C17. Mild air-oxidation of the respective expressed peptide yielded the bicyclic structure that was consistent with the molecular mass containing a thioether bond and a disulfide bond (spectra of peptide-oxi shown in black). Upon treatment with TCEP, the molecular mass of each peptide-oxi increased by 2 Da, suggesting that the disulfide bond was reduced to yield monocyclic peptide (spectra of peptide-red shown in red). Assigned b-ion peaks are shown in the figure. Peaks labeled with asterisks are assigned as y-ion peaks.

to the linear Flag region (Fig. 3c), indicating that the desired overlapping-bicyclic structure was formed. Upon reduction treatment of pC2C10C17-oxi with tris(2-carboxyethyl)phosphine (TCEP), the molecular mass of pC2C10C17-oxi increased by approximately 2 Da, implying that the disulfide bond was reduced to form pC2C10C17-red (Fig. 3b, pC2C10C17-red). Tandem mass analysis of pC2C10C17-red yielded additional peaks corresponding to some fragmentations in the region between ^{m}C10 and ^{d}C17 (Fig. 3d). This result gave solid evidence that the designed overlapping-bicyclic peptide consisting of the head-to-sidechain and inter-sidechain bonds was produced.

To demonstrate the designer generality of this approach, we also expressed ^{u}C2-peptides varying in the position of ^{m}CX (X = 4, 5, 6, or 8) in the combination of ^{d}CY (Y = 7, 10, 11, or 14) with arbitrary sequence compositions and lengths. In all cases, production of the desired bicyclic peptide was confirmed by MALDI-TOF/TOF (Fig. 4a for X = 4 and Y = 14, and b for X = 8 and Y = 11; see also Fig. S5a, b, c, and d‡ for other combinations). Moreover, ^{u}C3-peptides having two different positions of ^{m}CX (X = 7, and 13) were also expressed, yielding dumbbell-type peptides (Fig. 4c for X = 7; see also Fig. S5e‡ for X = 13). Taken together, this designer bicyclization strategy enables us to construct various bicyclic peptides with desired sequence compositions.

Conclusions

Here we have shown selective macrocyclization of peptides having the N-terminal ClAc group with two competing C residues expressed in translation machinery under the reprogrammed genetic code. The nearest ^{u}C (upstream C) residue generally dictates the spontaneous thioether bond formation *via* thiol-nucleophilic attack on the α-carbon of the ClAc group, thus forming a macrocyclic structure and leaving a ^{d}C (downstream C) residue unreacted. However, there is one exception where ^{u}C embedded at position 2 (^{u}C2) does not react with the ClAc group and instead ^{d}C residue predominantly reacts with the ClAc group, leaving the ^{u}C residue thiol-free. Based on this knowledge, we have devised a strategy for the synthesis of bicyclic peptides in which an additional middle C (^{m}C) residue between the ^{u}C2 and ^{d}C residues is embedded in the sequence, cyclizing the peptide by a thioether bond between ClAc and ^{m}C followed by a disulfide bond formation of ^{u}C2 and ^{d}C. This bicyclization is selective and fairly general independent of sequence compositions. Thus, this strategy is applicable for constructing a library consisting of such constrained bicyclic peptide scaffolds. Particularly, the integration of this library with RaPID system will allow us to rapidly select bicyclic peptides against a chosen protein target. This opens a new avenue to search a unique peptide 3-dimensional sequence space for bioactive peptides.

Published on 01 March 2012 on http://pubs.rsc.org | doi:10.1039/C2OB25306B

Acknowledgements

We thank H. Murakami for the discussion and development of the FIT system. This work was supported by a Grants-in-Aid of Japan Society for Promotion of Science (JSPS), the Specially Promoted Research (21000005) and the Industrial Science and Technology Program in the New Energy and Industrial Technology Development Organization to H.S.; Grants-in-Aid for JSPS Fellows (22–9695) to K.I.; Grants-in-Aid for Young Scientists (B) (22750145) from JSPS and PRESTO, Japan Science and Technology Agency (10206) to Y.G.; and Grants-in-Aid for Young Scientists (B) (22710210) from JSPS to T.K.

Published on 01 March 2012 on http://pubs.rsc.org | doi:10.1039/C2OB25306B
Downloaded by Tokyo Daigaku on 28 January 2013

Notes and references

1 D. P. Fairlie, G. Abbenante and D. R. March, *Curr. Med. Chem.*, 1995, **2**, 654–686.

2 (*a*) A. Paladini and L. C. Craig, *J. Am. Chem. Soc.*, 1954, **76**, 688–692; (*b*) M. Dreyfuss, E. Harri, H. Hofmann, H. Kobel, W. Pache and H. Tscherter, *Eur. J. Appl. Microbiol.*, 1976, **3**, 125–133; (*c*) A. C. Whyte, B. K. Joshi, J. B. Gloer, D. T. Wicklow and P. F. Dowd, *J. Nat. Prod.*, 2000, **63**, 1006–1009; (*d*) T. Tanaka, E. Tsukuda, M. Nozawa, H. Nonaka, T. Ohno, H. Kase, K. Yamada and Y. Matsuda, *Mol. Pharmacol.*, 1994, **45**, 724–730; (*e*) B. A. Johnson, H. Anker and F. L. Meleney, *Science (Washington, D. C.)*, 1945, **102**, 376–377.

3 D. R. March, G. Abbenante, D. A. Bergman, R. I. Brinkworth, W. Wickramasinghe, J. Begun, J. L. Martin and D. P. Fairlie, *J. Am. Chem. Soc.*, 1996, **118**, 3375–3379.

4 (*a*) D. Schwarzer, R. Finking and M. A. Marahiel, *Nat. Prod. Rep.*, 2003, **20**, 275–287; (*b*) R. Finking and M. A. Marahiel, *Annu. Rev. Microbiol.*, 2004, **58**, 453–488; (*c*) M. A. Fischbach and C. T. Walsh, *Chem. Rev.*, 2006, **106**, 3468–3496.

5 Y. Goto, A. Ohta, Y. Sako, Y. Yamagishi, H. Murakami and H. Suga, *ACS Chem. Biol.*, 2008, **3**, 120–129.

6 (*a*) H. Murakami, A. Ohta, H. Ashigai and H. Suga, *Nat. Methods*, 2006, **3**, 357–359; (*b*) Y. Goto, T. Katoh and H. Suga, *Nat. Protocols*, 2011, **6**, 779–790.

7 (*a*) L. Yu, Y. H. Lai, J. V. Wade and S. M. Coutts, *Tetrahedron Lett.*, 1998, **39**, 6633–6636; (*b*) K. D. Roberts, J. N. Lambert, N. J. Ede and A. M. Bray, *Tetrahedron Lett.*, 1998, **39**, 8357–8360; (*c*) F. D. T. Lung, C. R. King and P. P. Roller, *Lett. Pept. Sci.*, 1999, **6**, 45–49; (*d*) S. C. Pero, L. Oligino, R. J. Daly, A. L. Soden, C. Liu, P. P. Roller, P. Li and D. N. Krag, *J. Biol. Chem.*, 2002, **277**, 11918–11926; (*e*) P. Li, M. L. Peach, M. C. Zhang, H. P. Liu, D. J. Yang, M. Nicklaus and P. P. Roller, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 895–899; (*f*) F. M. Brunel and P. E. Dawson, *Chem. Commun.*, 2005, 2552–2554; (*g*) R. Tugyi, G. Mezo, E. Fellinger, D. Andreu and F. Hudecz, *J. Pept. Sci.*, 2005, **11**, 642–649; (*h*) K. D. Roberts, J. N. Lambert, J. Ede and A. M. Bray, *J. Pept. Sci.*, 2006, **12**, 525–532; (*i*) C. J. Porter and J. A. Wilce, *Biopolymers*, 2007, **88**, 174–181; (*j*) K. D. Roberts and N. J. Ede, *J. Pept. Sci.*, 2007, **13**, 811–821; (*k*) S. Jiang, C. Z. Liao, L. Bindu, B. L. Yin, K. W. Worthy, R. J. Fisher, T. R. Burke, M. C. Nicklaus and P. P. Roller, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 2693–2698.

8 Y. Yamagishi, I. Shoji, S. Miyagawa, T. Kawakami, T. Katoh, Y. Goto and H. Suga, *Chem. Biol.*, 2011, **18**, 1562–1570.

9 Y. Hayashi, J. Morimoto and H. Suga, *ACS Chem. Biol.*, 2012, DOI: 10.1021/cb200388k.

10 J. Morimoto, Y. Hayashi and H. Suga, *Angew. Chem., Int. Ed. Engl.*, 2012, DOI: 10.1002/anie.201108118.

11 (*a*) R. W. Roberts and J. W. Szostak, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 12297–12302; (*b*) N. Nemoto, E. MiyamotoSato, Y. Husimi and H. Yanagawa, *FEBS Lett.*, 1997, **414**, 405–408.

12 (*a*) In addition to genetic code reprogramming, a hydroxy acid has been incorporated into amber codon in the following papers: A. Ohta, H. Murakami, E. Higashimura and H. Suga, *Chem. Biol.*, 2007, **14**, 1315–1322; (*b*) J. A. Ellman, D. Mendel and P. G. Schultz, *Science (Washington, D. C.)*, 1992, **255**, 197–200; (*c*) P. M. England, H. A. Lester and D. A. Dougherty, *Biochemistry*, 1999, **38**, 14409–14415; (*d*) S. W. Millward, T. T. Takahashi and R. W. Roberts, *J. Am. Chem. Soc.*, 2005, **127**, 14142–14143.

 This journal is © The Royal Society of Chemistry 2012

OPEN ACCESS Freely available online

PLoS one

Dynamics of Hepatitis B Virus Quasispecies in Association with Nucleos(t)ide Analogue Treatment Determined by Ultra-Deep Sequencing

Norihiro Nishijima[1], Hiroyuki Marusawa[1]*, Yoshihide Ueda[1], Ken Takahashi[1], Akihiro Nasu[1], Yukio Osaki[2], Tadayuki Kou[3], Shujiro Yazumi[3], Takeshi Fujiwara[4], Soken Tsuchiya[4], Kazuharu Shimizu[4], Shinji Uemoto[5], Tsutomu Chiba[1]

1 Department of Gastroenterology and Hepatology, Graduate School of Medicine, Kyoto University, Kyoto, Japan, **2** Department of Gastroenterology and Hepatology, Osaka Red Cross Hospital, Osaka, Japan, **3** Department of Gastroenterology and Hepatology, Tazuke Kofukai Medical Research Institute, Kitano Hospital, Osaka, Japan, **4** Department of Nanobio Drug Discovery, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto, Japan, **5** Department of Surgery, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Abstract

Background and Aims: Although the advent of ultra-deep sequencing technology allows for the analysis of heretofore-undetectable minor viral mutants, a limited amount of information is currently available regarding the clinical implications of hepatitis B virus (HBV) genomic heterogeneity.

Methods: To characterize the HBV genetic heterogeneity in association with anti-viral therapy, we performed ultra-deep sequencing of full-genome HBV in the liver and serum of 19 patients with chronic viral infection, including 14 therapy-naïve and 5 nucleos(t)ide analogue(NA)-treated cases.

Results: Most genomic changes observed in viral variants were single base substitutions and were widely distributed throughout the HBV genome. Four of eight (50%) chronic therapy-naïve HBeAg-negative patients showed a relatively low prevalence of the G1896A pre-core (pre-C) mutant in the liver tissues, suggesting that other mutations were involved in their HBeAg seroconversion. Interestingly, liver tissues in 4 of 5 (80%) of the chronic NA-treated anti-HBe-positive cases had extremely low levels of the G1896A pre-C mutant (0.0%, 0.0%, 0.1%, and 1.1%), suggesting the high sensitivity of the G1896A pre-C mutant to NA. Moreover, various abundances of clones resistant to NA were common in both the liver and serum of treatment-naïve patients, and the proportion of M204VI mutants resistant to lamivudine and entecavir expanded in response to entecavir treatment in the serum of 35.7% (5/14) of patients, suggesting the putative risk of developing drug resistance to NA.

Conclusion: Our findings illustrate the strong advantage of deep sequencing on viral genome as a tool for dissecting the pathophysiology of HBV infection.

Citation: Nishijima N, Marusawa H, Ueda Y, Takahashi K, Nasu A, et al. (2012) Dynamics of Hepatitis B Virus Quasispecies in Association with Nucleos(t)ide Analogue Treatment Determined by Ultra-Deep Sequencing. PLoS ONE 7(4): e35052. doi:10.1371/journal.pone.0035052

Editor: Antonio Bertoletti, Singapore Institute for Clinical Sciences, Singapore

Received November 17, 2011; **Accepted** March 8, 2012; **Published** April 16, 2012

Copyright: © 2012 Nishijima et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by JSPS Grant-in-aid for Scientific Research 21229009, 23390196, and Health and Labor Science Research Grants (H22-08) and Research on Hepatitis from the Ministry of Health, Labor and Welfare, Japan. (http://mhlw-grants.niph.go.jp/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: maru@kuhp.kyoto-u.ac.jp

Introduction

Hepatitis B virus (HBV) is a non-cytopathic DNA virus that infects approximately 350 million people worldwide and is a main cause of liver-related morbidity and mortality [1–3]. The absence of viral-encoded RNA-dependent DNA polymerase proofreading capacity coupled with the extremely high rate of HBV replication yields the potential to rapidly generate mutations at each nucleotide position within the entire genome [4]. Accordingly, a highly characteristic nature of HBV infection is the remarkable genetic heterogeneity at the inter- and intra- patient level. The latter case of variability as a population of closely-related but nonidentical genomes is referred to as viral quasispecies [5,6]. It is well recognized that such mutations may have important implications regarding the pathogenesis of viral disease. For example, in chronic infection, G to A point mutation at nucleotide (nt) 1896 in the pre-core (pre-C) region as well as A1762T and G1764A mutations in the core-promoter region are highly associated with HBeAg seroconversion that in general results in the low levels of viremia and consequent clinical cure [7–9]. In contrast, acute infection with the G1896A pre-C mutant represents a high risk for fulminant hepatic failure [10,11]. Although these facts clearly illustrate the clinical implications of certain viral mutation, increasing evidence strongly suggests that

the viral genetic heterogeneity is more complicated than previously thought [12,13].

The major goals of antiviral therapy in patients with HBV infection are to prevent the progression of liver disease and inhibit the development of hepatocellular carcinoma [14]. Oral nucleos(t)ide analogue (NA) have revolutionized the management of HBV infection, and five such antiviral drugs, including lamivudine, adefovir, entecavir, tenofovir, and telbivudine, are currently approved medications [15,16]. These agents are well-tolerated, very effective at suppressing viral replication, and safe, but one of the major problems of NA therapy is that long-term use of these drugs frequently causes the emergence of antiviral drug-resistant HBV due to substitutions at specific sites in the viral genome sequences, which often negates the benefits of therapy and is associated with hepatitis flares and death [16,17]. It is unclear whether viral clones with antiviral resistance emerge after the administration of antiviral therapy or widely preexist among treatment-naïve patients.

There has been a recent advance in DNA sequencing technology [18]. The ultra-deep sequencers allow for massively parallel amplification and detection of sequences of hundreds of thousands of individual molecules. We recently demonstrated the usefulness of ultra-deep sequencing technology to unveil the massive genetic heterogeneity of hepatitis C virus (HCV) in association with treatment response to antiviral therapy [19]. On the other hand, there are a few published studies in which this technology was used to characterize genetic HBV sequence variations [20–22]. Margeridon-Thermet et al reported that the 454 Life Science GS20 sequencing platform provided higher sensitivity for detecting drug-resistant HBV mutations in the serum of patients treated with nucleoside and nucleotide reverse-transcriptase inhibitors [20]. Solmone et al also reported the strong advantage conferred by the same platform to detect minor variants in the serum of patients with chronic HBV infection [21]. Although in these previous studies low-abundant drug-resistant variants were successfully detected, the analyses were focused on the reverse-transcriptase region of circulating HBV in the serum and thus the whole picture of HBV genetic heterogeneity as well as the *in vivo* dynamics of HBV drug resistant variants in response to anti-viral treatment remains to be clarified. Moreover, intrahepatic viral heterogeneity in patients that achieved the clearance of circulating HBV is largely unknown.

By taking the advantage of an abundance of genetic information obtained by utilizing the Illumina Genome Analyzer II (Illumina, San Diego, CA) as a platform of ultra-deep sequencing, we determined the whole HBV sequence in the liver and serum of patients with chronic HBV infection to evaluate viral quasispecies characteristics. Moreover, we investigated the prevalence of rare drug-resistant HBV variants as well as detailed dynamic changes in the viral genetic heterogeneity in association with NA administration. Based on the abundant genetic information obtained by ultra-deep sequencing, we clarified the precise prevalence of HBV clones with G1896A pre-C mutations in association with HBe serostatus in chronically infected patients with or without NA treatment. We also detected a variety of minor drug-resistant clones in treatment-naïve patients and their dynamic changes in response to entecavir administration, demonstrating the potential clinical significance of naturally-occurring drug-resistant mutations.

Materials and Methods

Ethics Statement

The Kyoto University ethics committee approved the study, and written informed consent for participation in this study was obtained from all patients. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Patients

The liver tissues of 19 Japanese patients that underwent living-donor liver transplantation at Kyoto University due to HBV-related liver disease were available for viral genome analyses. These individuals included 13 men and 6 women, aged 41 to 69 years (median, 55.2 years) and all but one were infected with genotype C viruses. Participants comprised 19 patients with liver cirrhosis caused by chronic HBV infection, including 14 antiviral therapy-naïve cases (chronic-naïve cases) and 5 cases receiving NA treatment, with either lamivudine or entecavir (chronic-NA cases) (Table 1). Serum HBV DNA levels were significantly higher in chronic-naïve cases than in chronic NA cases (median serum HBV DNA levels were 5.6, and <2.6 log copies/ml, respectively, Table 1). Liver tissue samples were obtained at the time of transplantation, frozen immediately, and stored at −80°C until use. Serologic analyses of HBV markers, including hepatitis B surface antigen (HBsAg), antibodies to HBsAg, anti-HBc, HBeAg, and antibodies to HBeAg, were determined by enzyme immunoassay kits as described previously [23]. HBV DNA in the serum before transplantation was examined using a polymerase chain reaction (PCR) assay (Amplicor HBV Monitor, Roche, Branchburg, NJ). To examine the dynamics of viral quasispecies in response to anti-HBV therapy, paired serum samples of 14 treatment-naïve patients before and after administration of daily entecavir (0.5 mg/day) were subjected to further analyses on viral genome.

Direct population Sanger sequencing

DNA was extracted from the liver tissue and serum using a DNeasy Blood & Tissue Kit (Qiagen, Tokyo, Japan). To define the consensus reference sequences of HBV in each clinical specimen, all samples were first subjected to direct population Sanger sequencing using the Applied Biosystems 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA). Oligonucleotide primers for the HBV genome were designed to specifically amplify whole viral sequences as two overlapping fragments using the sense primer 169_F and antisense primer 2847_R to yield a 2679-bp amplicon (amplicon 1), and the sense primer 685_F and antisense primer 443_R to yield a 2974-bp amplicon (amplicon 2; Table S1). HBV sequences were amplified using Phusion High-Fidelity DNA polymerase (FINZYMES, Espoo, Finland). All amplified PCR products were purified using the QIAquick Gel Extraction kit (Qiagen) after agarose gel electrophoresis and used for direct sequencing. The serum of a healthy HBV DNA-negative volunteer was used as a negative control.

Viral genome sequencing by massively-parallel sequencing

Massively-parallel sequencing with multiplexed tags was performed using the Illumina Genome Analyzer II as described [19]. The end-repair of DNA fragments, addition of adenine to the 3′ ends of DNA fragments, adaptor ligation, and PCR amplification by Illumina PCR primers were performed as described previously [24]. Briefly, the viral genome sequences were amplified by high-fidelity PCR using oligonucleotide primers as described above, sheared by nebulization using 32 psi N2 for 8 min, and then the sheared fragments were purified and concentrated using a QIAquick PCR purification Kit (Qiagen). Nucleotide overhangs resulting from fragmentation were then converted into blunt ends using T4 DNA polymerase and Klenow

Table 1. Characteristics of patients with chronic HBV infection analyzed in this study.

	Chronic-naïve (N = 14)	Chronic-NA (N = 5)
Age†	55.5 (41–69)	55.0 (49–68)
Sex (male/female)	9/5	4/1
Alanine aminotransaminase (IU/l)†	41 (10–74)	30 (15–65)
Total bilirubin (mg/dl)†	0.9 (0.5–31.1)	1.7 (0.6–4.5)
Platelet count ($\times 10^4$/mm^3)†	12.7 (3.3–27.6)	5.1 (3.6–11.3)
HBV genotype		
B	1	0
C	13	5
Viral load (log copies /ml)†	5.6 (<2.6–8.8)*	<2.6 (<2.6–5.3)*
HBe-serostatus (HBeAg+/HBeAb+)	8/6	0/5
Fibrosis		
F0–F2	6	0
F3–F4	8	5
Activity		
A0–A1	7	3
A2–A3	7	2

†Values are median (range).
*P = 0.042.
doi:10.1371/journal.pone.0035052.t001

enzymes, followed by the addition of terminal 3′ A-residues. An adaptor containing unique 6-bp tags, such as "ATCACG" and "CGATGT" (Multiplexing Sample Preparation Oligonucleotide Kit, Illumina), was then ligated to each fragment using DNA ligase. We then performed agarose gel electrophoresis of adaptor-ligated DNAs and excised bands from the gel to produce libraries with insert sizes ranging from 200 to 350 bp. These libraries were amplified independently using a minimal PCR amplification step of 18 cycles by Illumina PCR primers with Phusion High-Fidelity DNA polymerase. The DNA fragments were then purified with a MinElute PCR Purification Kit (Qiagen), followed by quantification using the NanoDrop 2000C (Thermo Fisher Scientific, Waltham, MA) to make a working concentration of 10 nM. Cluster generation and sequencing was performed for 64 cycles on the Illumina Genome Analyzer II according to the manufacturer's instructions. The obtained images were analyzed and base-called using GA pipeline software version 1.4 with the default settings provided by Illumina.

Genome Analyzer sequence data analysis

Using the high performance alignment software "NextGene" (SoftGenetics, State College, PA), the 64 base-pair reads obtained from the Genome Analyzer II were aligned with the reference sequences of 3215 bp that were determined by direct population Sanger sequencing of each clinical specimen. Reads with 90% or more bases matching a particular position of the reference sequences were aligned. Furthermore, two quality filters were used for sequencing reads: the reads with a median quality score of more than 30 and no more than 3 uncalled nucleotides were allowed anywhere in the 64 bases. Only sequences that passed the quality filters, rather than raw sequences, were analyzed and each position of the viral genome was assigned a coverage depth, representing the number of times the nucleotide position was sequenced.

Allele-specific quantitative real-time PCR and semiquantitative PCR to determine the relative proportion of G1896A pre-C mutant

To determine the relative proportion of the G1896A pre-C mutant, allele-specific quantitative real-time PCR was performed based on the previously described method [25,26]. Oligonucleotide primers were designed individually to amplify the pre-C region of wild-type and the G1896A pre-C mutant HBV. Three primers were used for this protocol, two allele-specific sense primers, 1896WT_F (for wild-type) and 1896MT_F (for the G1896A pre-C mutant), and one common antisense primer, 2037_R (Table S1). Quantification of wild-type and the G1896A pre-C mutant was individually performed by real-time PCR using a Light Cycler 480 and Fast Start Universal SYBR Master (Roche, Mannheim, Germany) [27]. The relative proportion of the G1896A pre-C mutant was determined to calculate the G1896A pre-C mutant/total HBV ratios. Performance of this assay was tested using mixtures of two previously described plasmids, pcDNA3-HBV-wt#1 and pcDNA3-HBV-G1896A pre-C mutant [28]. Semiquantitative PCR was performed using primers described above, then agarose gel electrophoresis was performed.

Statistical analysis

Results are expressed as mean or median, and range. Pretreatment values were compared using the Mann-Whitney U-test or the Kruskal Wallis H-test. *P* values less than 0.05 were considered statistically significant.

The viral quasispecies characteristics were evaluated by analyzing the genetic complexity based on the number of different sequences present in the population. Genetic complexity for each site was determined by calculating the Shannon entropy using the following formula:

$$Sn = -\frac{\sum_{i=1}^{n} fi(lnfi)}{N}$$