npg

ORIGINAL ARTICLE

Intermolecular recognition revealed by the complex structure of human CLOCK-BMAL1 basic helix-loop-helix domains with E-box DNA

Zixi Wang[1], Yaling Wu[2], Lanfen Li[1], Xiao-Dong Su[1]

[1]State Key Laboratory of Protein and Plant Gene Research, and Biodynamic Optical Imaging Center (BIOPIC), Peking University, Beijing 100871, China; [2]National Institute of Biological Sciences, Beijing 102206, China

CLOCK (circadian locomotor output cycles kaput) and BMAL1 (brain and muscle ARNT-like 1) are both transcription factors of the circadian core loop in mammals. Recently published mouse CLOCK-BMAL1 bHLH (basic helix-loop-helix)-PAS (period-ARNT-single-minded) complex structure sheds light on the mechanism for heterodimer formation, but the structural details of the protein-DNA recognition mechanisms remain elusive. Here we have elucidated the crystal structure of human CLOCK-BMAL1 bHLH domains bound to a canonical E-box DNA. We demonstrate that CLOCK and BMAL1 bHLH domains can be mutually selected, and that hydrogen-bonding networks mediate their E-box recognition. We identified a hydrophobic contact between BMAL1 Ile80 and a flanking thymine nucleotide, suggesting that CLOCK-BMAL1 actually reads 7-bp DNA and not the previously believed 6-bp DNA. To find potential non-canonical E-boxes that could be recognized by CLOCK-BMAL1, we constructed systematic single-nucleotide mutations on the E-box and measured their relevant affinities. We defined two non-canonical E-box patterns with high affinities, AACGTGA and CATGTGA, in which the flanking A7-T7′ base pair is indispensable for recognition. These results will help us to identify functional CLOCK-BMAL1-binding sites *in vivo* and to search for clock-controlled genes. Furthermore, we assessed the inhibitory role of potential phosphorylation sites in bHLH regions. We found that the phospho-mimicking mutation on BMAL1 Ser78 could efficiently block DNA binding as well as abolish normal circadian oscillation in cells. We propose that BMAL1 Ser78 should be a key residue mediating input signal-regulated transcriptional inhibition for external cues to entrain the circadian clock by kinase cascade.

Keywords: crystal structure; circadian; CLOCK; BMAL1; non-canonical E-box; phosphorylation
Cell Research (2013) **23**:213-224. doi:10.1038/cr.2012.170; published online 11 December 2012

Introduction

Circadian rhythms are adaptive mechanisms essential for almost all organisms in nature. They synchronize the behavior and physiological status to periodic changes of environment, ensuring that the biochemical processes are executed accurately and efficiently [1-4]. At molecular level, circadian clocks in mammals are mainly fulfilled through a core transcriptional negative feedback loop and several other auxiliary feedback loops [1, 2, 5, 6]. The

transcription control of the core loop is governed by two bHLH (basic helix-loop-helix) and PAS (period-ARNT-single-minded) domain-containing transcription factors, namely CLOCK (circadian locomotor output cycles kaput) and BMAL1 (brain and muscle ARNT-like 1) [1, 2, 5, 6]. It has been established that they form a heterodimer, bind E-box DNA elements and activate the transcription of clock-controlled genes [7, 8]. Two groups of resultant proteins, PERIOD (PER, sum of PER1-3) and CRYPTO-CHROME (CRY, or CRY1, 2), gradually accumulate and inhibit the activity of CLOCK-BMAL1 complex, constituting the negative feedback limb of the core loop [1, 2, 5, 6].

E-box sites are defined as mainly 6-bp DNA elements recognized by bHLH family transcription factors [9-11]. They facilitate the transcription of various genes

involved in cell proliferation, muscular and neural differentiation, immunoglobulin generation as well as circadian rhythms [9, 10, 12]. The E-box is distinguished by having a consensus sequence of CANNTG [13-15], with a palindromic canonical form of CACGTG [16]. By site selection and amplification method, the MOP4-BMAL1 (MOP4, also named NPAS2, a homolog of CLOCK) heterodimer was demonstrated to bind the canonical E-box, which is also considered to be a high-affinity binding site for CLOCK-BMAL1 [7]. Further large-scale studies on cycling genes in suprachiasmatic nuclei, the liver and the heart have shown that a number of non-canonical E-box elements also regulate rhythmic gene expression [17-19]. One representative example is the E'-box (CACGTT), which participates in mammalian *Per*2 gene oscillation [20]. Recently, genome-wide analyses of BMAL1-binding sites by ChIP array or deep sequencing techniques also identified many E-box-like and non-canonical E-box elements such as CCAATG, CATTGG, CATGTG, AACGTG, which extensively regulate clock-controlled genes [21, 22]. These E-boxes can further form tandem repeats to regulate gene expression [22, 23]. Despite these advances, a missing link exists in the direct binding assays between non-canonical E-boxes and CLOCK-BMAL1, and a comprehensive comparison of the affinity difference between various E-boxes.

The recently published heterodimeric structure of the mouse CLOCK-BMAL1 bHLH-PAS domains has shed light on how these two proteins interact with each other [24], but the structural details of the protein-DNA recognition mechanisms remain unclear. In this work, we have determined the ternary complex of human CLOCK-BMAL1 bHLH domains with an E-box DNA. We demonstrate that CLOCK His84 and BMAL1 Leu125 are the key residues for the mutual recognition between CLOCK and BMAL1 bHLH domains. By measuring the affinities of a series of single-mutated canonical E-box DNA by ITC (isothermal titration calorimetry) and by comparing their thermodynamic differences, we have identified two non-canonical E-box-binding patterns, AACGTGA and CATGTGA. We showed that the flanking thymine nucleotide positioned at the 7′ site, which is recognized by BMAL1 Ile80, is indispensable for the recognition of the non-canonical E-boxes AACGTG and CATGTG. Furthermore, we assessed the effects of phosphorylation on four serine residues in protein basic regions by mutating them to the phospho-mimicking glutamate residue individually. We found that only BMAL1 S78E could strongly inhibit DNA binding. This mutation also diminished the transcriptional activation in the luciferase assay *in vivo* and abolished the normal circadian oscillation in cells. Therefore, BMAL1 Ser78 should be a key residue

mediating input signal-regulated transcriptional inhibition.

Results

Crystal structure of CLOCK-BMAL1 bHLH domains and DNA complex

To explore the intermolecular recognition mechanism, we crystallized the complex structure of human CLOCK-BMAL1 bHLH domains bound to a canonical E-box DNA fragment. The crystals diffract to 2.4 Å resolution and belong to space group $P3_121$. The structure was determined by molecular replacement using modified Mad-Max-DNA structure (PDB: 1NLW) [25] as a search model. The final model was refined to R_{work}/R_{free} values of 19.5%/24.4% (Table 1). The electron density of the DNA chain was clearly observed after the initial refinement of the molecular replacement solution (Supplementary information, Figure S1). Each asymmetric unit contains one protein heterodimer and one double helical DNA chain (Figure 1A).

Resembling the previously solved protein-DNA complex structures in bHLH family [25-32], the basic helical regions of CLOCK and BMAL1 insert into the major groove of DNA. The extending parts following basic regions together with the second helices form a left-handed four-helical bundle, which is required for the heterodimer interaction and recognition. Notably, the intervening loops of both proteins are relatively small and do not appear to form any interactions with DNA bases. Hence, the intervening loops are not expected to participate in the recognition of flanking nucleotides beside E-box. Compared with the recently published mouse CLOCK-BMAL1 bHLH-PAS structure [33], the N-terminal fragment of human CLOCK bHLH domain in our structure is extended by three helical turns and moved outward by five degrees (Figure 1B), suggesting a DNA-induced conformational change.

Mutual recognition mechanism between CLOCK and BMAL1 bHLH domains

Since CLOCK and BMAL1 bHLH domains share high sequence homology (35% identity for the 60 residues compared), individual CLOCK or BMAL1 bHLH domain can form a homodimer structure in solution indicated by gel filtration and ultracentrifugation analysis (Supplementary information, Figure S2A and S2B). However, when both bHLH domains are present, they can discriminate one another to form a stable heterodimer (Figure 1A). To uncover the mechanism of mutual recognition between bHLH domains, we examined the heterodimer interface. The helical bundle responsible for

Table 1 Data collection and refinement statistics

| Data collection	
Space group	P3$_1$21
Cell dimensions	
a, b, c (Å)	98.74, 98.74, 62.88
α, β, γ (°)	90, 90, 120
Resolution (Å)	38.8-2.40
	(2.46-2.40)
R_{sym}	9.4 (63.7)
$I/\sigma I$	33.5 (6.7)
Completeness (%)	99.9 (100)
Redundancy	19.8 (19.8)
Refinement	
Resolution (Å)	38.8-2.40
No. of reflections	14045
R_{work}/R_{free} (%)	19.5/24.4
No. of atoms	
Protein	958
DNA	509
Water	106
B-factors (Å^2)	
Protein	47.7
DNA	52.3
Water	44.6
R.m.s deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.3
Ramachandran plot	
Favored (%)	97.4
Allowed (%)	2.6
Generously allowed (%)	0
Disallowed (%)	0

Values in parentheses are for highest-resolution shell.

the heterodimer interactions can be divided into five layers (Figure 2A). Of these, layers 2 and 3 have symmetric residue arrangement (except CLOCK Ser77 and BMAL1 Ala118, but both are small side-chain residues), indicating that these two central layers cannot render the selectivity for heterodimer.

To find the crucial residues involved in mutual recognition, we built the models of the CLOCK and BMAL1 homodimer through structural alignment. In both modeled structures, we observed serious side-chain clash, including CLOCK Phe50 in layer 1 and CLOCK His84, BMAL1 Leu125 in layer 5 (Figure 2A, right). Thus, these incompatible residues are expected to render the preference for heterodimer formation, as they will destabilize the homodimeric structure. To test this hypothesis,

we measured their relevant binding affinities by ITC. CLOCK H84L and BMAL1 L125H mutations, which mimic BMAL1 and CLOCK homodimer structures of layer 5, respectively, lead to decreased affinities by ~100-200 folds (Figure 2B). In contrast, BMAL1 M88F and CLOCK F50M mutations, which mimic CLOCK and BMAL1 homodimer structures of layer 1, respectively, do not cause observable effect on mutual recognition (Figure 2B and Supplementary information, Table S1). These results indicate that CLOCK His84 and BMAL1 Leu125 in layer 5 are the key residues involved in mutual recognition.

We thus proposed a model to explain the recognition process between CLOCK and BMAL1 bHLH domains (Figure 2C). Individual CLOCK and BMAL1 form less stable homodimers due to the steric hindrance in layer
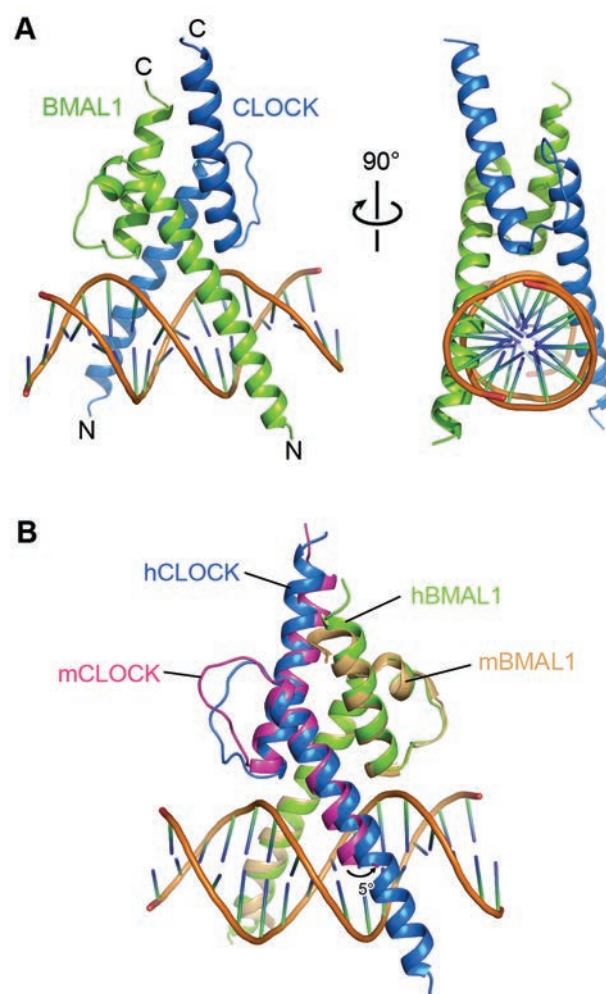
Figure 1 Overall structure. **(A)** Heterodimeric structure of human CLOCK-BMAL1 bHLH domains with DNA. CLOCK is colored blue and BMAL1 is colored green. **(B)** Structural alignment of human CLOCK-BMAL1 bHLH domains and the counterpart of mouse (PDB: 4F3L).
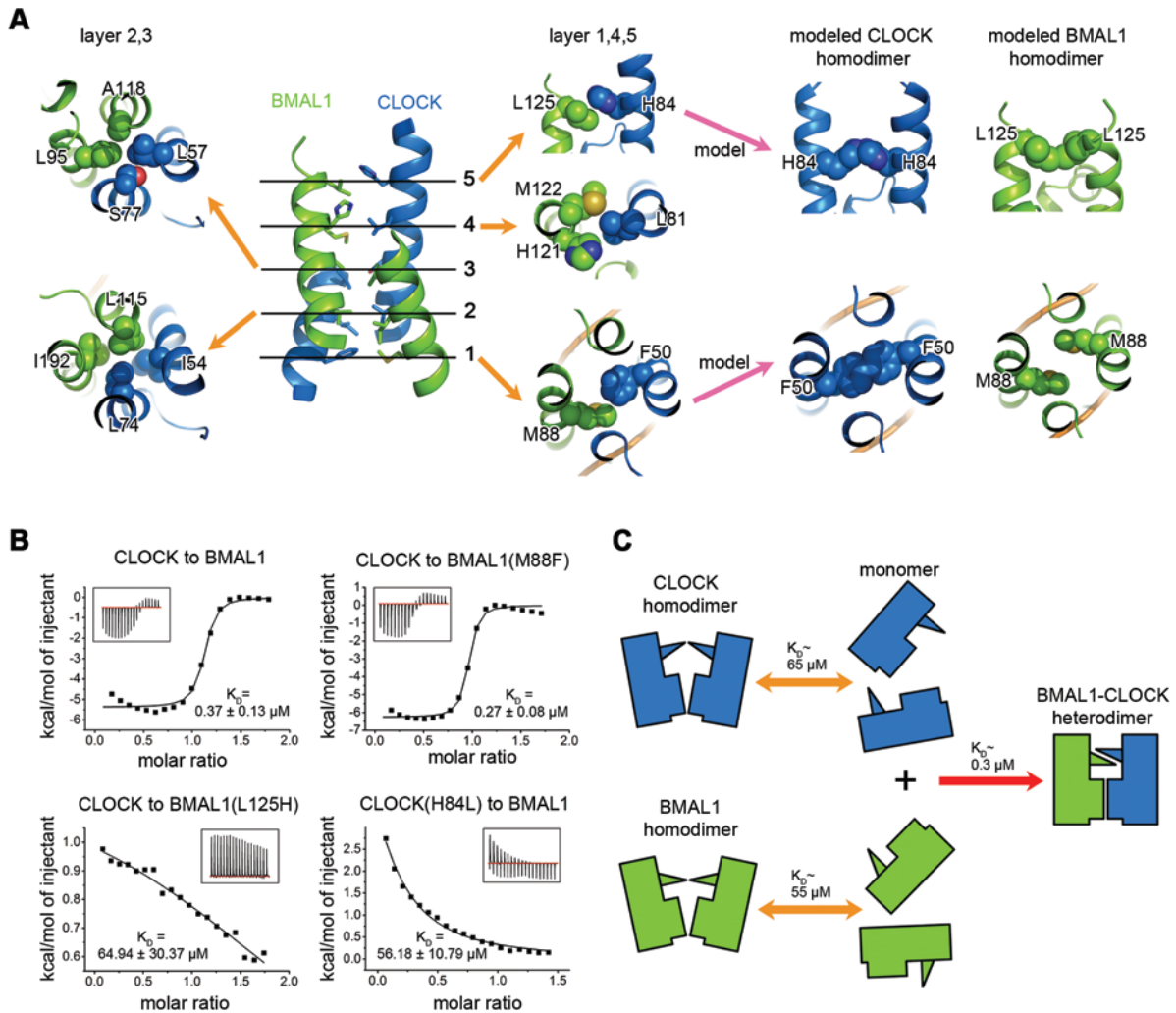
Figure 2 The mutual recognition mechanism between CLOCK and BMAL1 bHLH domains. **(A)** The four-helix bundle composing the heterodimer interface can be divided into five layers. In layers 2 and 3, the residues in symmetric positions are the same (except CLOCK S77 and BMAL1 A118), so they do not confer the selectivity for the heterodimer. In the modeled homodimeric structures, CLOCK Phe50 in layer 1 and CLOCK His84, BMAL1 Leu125 in layer 5 have steric clash. **(B)** ITC measurements for the titration of the CLOCK bHLH domain into the BMAL1 bHLH domain. Mutated residues are indicated in subtitles. K_D ± s.d. is shown. Homodimer-mimicking mutations in layer 5 (BMAL1 L125H and CLOCK H84L) greatly diminish the interaction. Homodimer-mimicking mutation in layer 1 (BMAL1 M88F) has no observable effect on mutual selection. **(C)** A proposed mechanistic model for mutual recognition. In the heterodimeric structure, the two recognition residues (CLOCK His84 and BMAL1 Leu125) are well stacked, thus leading a shift of the equilibrium to the heterodimer formation.

5. When both proteins are present, the equilibrium will rapidly shift to the heterodimer in which the residues are well stacked. This model suggests that bHLH domains alone can be mutually selected and contribute to the mutual recognition of full-length CLOCK and BMAL1 molecules.

Recognition of canonical E-box by CLOCK-BMAL1

By site selection and amplification methods, the canonical form of E-box CACGTG was previously iden-

tified as a high-affinity DNA-binding site for MOP4-BMAL1, as well as a binding site for CLOCK-BMAL1 [7]. Since the resulting consensus binding sequence contains no degenerate site, it is predicted that each base pair of the E-box is recognized. As expected, CLOCK-BMAL1 heterodimer interacts with each site of the E-box in the structure, with Arg39, Glu43, Arg47 in CLOCK and His77, Glu81, Arg85 in BMAL1 as the sequence-specific recognition residues (Figure 3A and 3B). Although the base-reading residues in both proteins have

a similar conformation, they have unique features. First, CLOCK Arg39 forms a pair of hydrogen bonds with one end of guanine of the E-box, whereas BMAL1 His77 forms one hydrogen bond with the other end of guanine; Second, CLOCK Arg47 lacks direct hydrogen bond interaction with the center guanine, thus making a looser contact than that of BMAL1 Arg85 (Figure 3A and 3B).

The precision of the interaction between protein and DNA is further illustrated by quantification of the DNA-binding affinities of the heterodimeric and homodimeric proteins. The heterodimeric protein binds to canonical E-box DNA (marked as wild-type (WT) DNA) with a K_D value of 1.52 ± 0.10 μM, ~5-10 folds lower than that of both homodimeric proteins (Figure 3C and Supplementary information, Table S1). These results indicate that the non-functional CLOCK or BMAL1 homodimers, if coexisted with heterodimer *in vivo*, should have little competitive influence on DNA binding.

Previous biochemical data has shown that the redox states of NAD cofactors have a significant influence on DNA binding by regulating CLOCK or NPAS2 bHLH domains directly [34]. It was reported that the oxidized NAD^+ and $NADP^+$ inhibit CLOCK-BMAL1 heterodimer binding to DNA, whereas the reduced NADH and NA-DPH have enhancing effects. These effects are probably fulfilled by directly binding of NAD cofactors to the bHLH domain, thus resulting in a conformational change [34]. To quantify the effects of the NAD cofactors on DNA binding, we measured the protein-DNA binding affinities in the presence of 10 mM NAD cofactors by ITC. Upon addition of NAD^+, NADH, $NADP^+$ and NADPH, the K_D values are 1.18 ± 0.07, 2.13 ± 0.17, 2.54 ± 0.11 and 2.74 ± 0.16 μM, respectively (Supplementary information, Figure S3 and Table S1), all of which are quite similar to that without ligand addition. Furthermore, we have crystallized the protein-DNA complex in the presence of NADH or NADPH and solved the crystal structure, but no interpretable electron density for NAD cofactors is seen. These results suggest that none of the NAD cofactors has a direct influence on DNA binding through
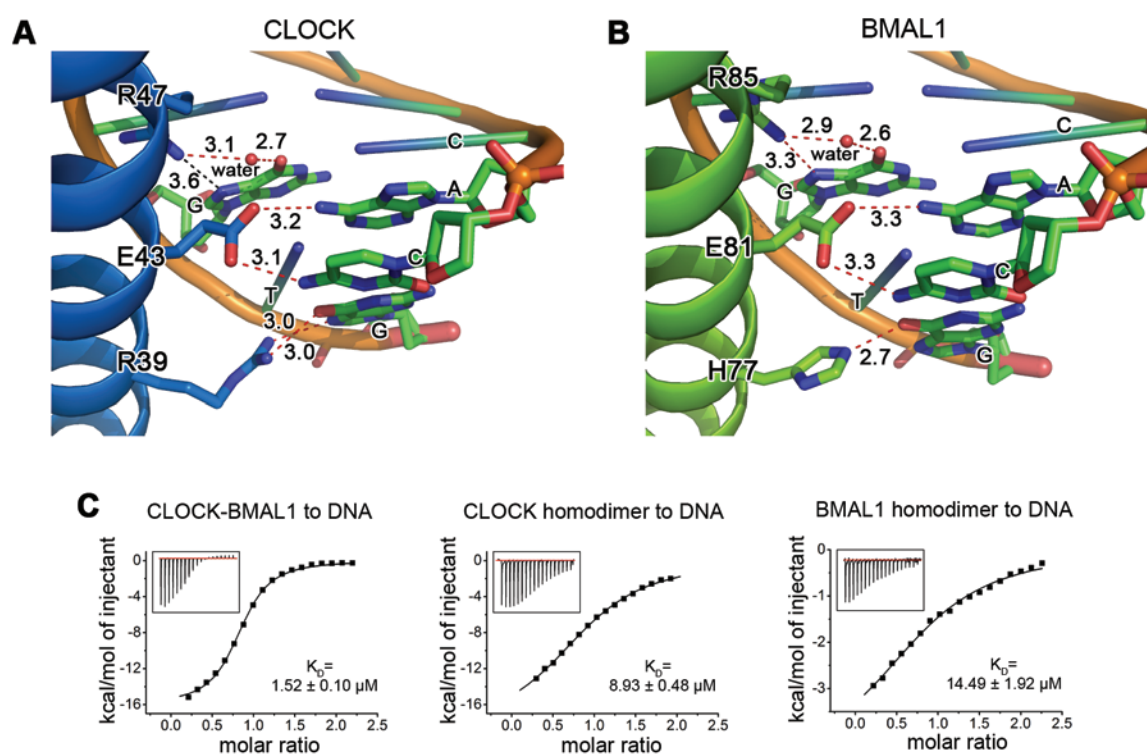
Figure 3 The recognition mechanism of CLOCK-BMAL1 bHLH domains to canonical E-box. **(A)** Detailed interactions between CLOCK and E-box. **(B)** Detailed interactions between BMAL1 and E-box. The red dashed lines represent the hydrogen-bonding contacts and the black dashed line represents the distance between two atoms. The bond lengths or distance are shown in Angstroms (Å). **(C)** ITC measurements of the binding affinities between hetero-/homodimeric proteins and the E-box DNA. bHLH domain proteins were titrated into a 16-bp blunt-ended double-stranded DNA that contains a canonical E-box in the center (Sequence: AGGAACACGTGACCCA. It is designated as WT DNA. E-box is underlined). Although the CLOCK bHLH homodimer has a higher affinity than BMAL1, both homodimers have a much lower affinity than heterodimer. K_D ± s.d. is shown.

bHLH domain. It is important to explore the detailed role of NAD cofactors in circadian regulation, as the changes of the concentration of NAD cofactors attributed to food intake or neuronal activity can be an input cue to entrain circadian clock, thus linking the cellular metabolism and circadian regulation together [34]. It is more likely that, as recently reported, NAD^+ regulates circadian rhythm by promoting SIRT1-dependent histone deacetylation, thus downregulating the CLOCK-BMAL1-dependent transcription [35-37], but not by regulating DNA binding through bHLH domain.

BMAL1 Ile80 recognize flanking thymine nucleotide of E-box

With exception to the canonical E-box region, the DNA used in crystallization is not strictly palindromic. However, in the crystals, the DNA chains have uniform direction to the asymmetric heterodimeric proteins (Figure 4A). This observation suggests that additional interactions exist between the protein and DNA bases to determine a specific DNA orientation. After examination, we observed a hydrophobic contact between BMAL1 Ile80 and a flanking thymine nucleotide (Figure 4A and 4B). Thus, the DNA bases recognized by CLOCK-BMAL1 become non-palindromic and can be divided into CLOCK half-site (CAC) and BMAL1 half-site (GTGA) (Figure 4A). To verify whether the BMAL1-mediated hydrophobic contact contributes to DNA binding, we titrated the heterodimeric proteins into site 7-mutated DNA 7ATd (ACACGTGT, the E-box is underlined and flanking nucleotides at both sides are shown, hereinafter) (Figure 4C and Supplementary information, Table S1). As expected, this DNA mutation lead to a decrement of binding affinity by ~2-fold compared with WT DNA (Figure 3C, left).

The preference for the flanking base pair A7-T7′ was first identified by the previously reported site selection and amplification experiment, and was further verified by a competition assay [7]. We provided a molecular explanation for these results and quantified the contribution of this hydrophobic contact in base reading. In addition, we did not find any flanking interaction in the CLOCK half-site, which was supported by the finding that the flanking bases at the CLOCK half-site were not conserved in the site selection result [7].

The flanking thymine nucleotide determines the specific binding to non-canonical E-boxes

To explore the non-canonical E-boxes that could potentially be recognized by CLOCK-BMAL1, we generated systematic single-nucleotide mutations on each site of the canonical E-box and measured their relevant binding affinities by ITC (Supplementary information, Table S1). We found two sequences that have obvious higher affinities than the others. They are 1CAd (AAACGTGA) and 3CTz (ACATGTGA), with K_D values of ~4 to 5 μM (Figure 4D and 4E, left). Unexpectedly, when we titrated CLOCK-BMAL1 bHLH heterodimeric protein into 6GTd (ACACGTTA), a DNA fragment sharing the same E-box as 1CAd (AAACGTGA), the binding affinity could not be measured (Figure 4D, right). This observation indicates that the interaction with the flanking nucleotide is required and crucial for recognition of the non-canonical E-box. If the flanking hydrophobic interaction is involved in recognition of the non-canonical E-box, then to bind 1CAd DNA, CLOCK Arg39 should interact with T1′, whereas to bind 6GTd DNA, BMAL1 His77 should interact with T6. The observation that only 1CAd DNA is recognized suggests that CLOCK can tolerate a C1 to A1 (or G1′ to T1′) mutation in DNA molecule, thus binding to 1CAd DNA, whereas BMAL1 cannot tolerate a G6 to T6 mutation, resulting in non-recognition to 6GTd DNA.

Similar to the case discussed above, binding to 4GAz (ACACATGA) was undetectable although it has the same E-box as 3CTz (ACATGTGA) (Figure 4E, right). This result can be explained by the different DNA-binding features between CLOCK and BMAL1. The interaction between CLOCK Arg47 and G3′ is looser than that between BMAL1 Arg85 and G4 (Figure 3A and 3B). Therefore, CLOCK can tolerate a G3′ to A3′ mutation in DNA molecule, but BMAL1 cannot tolerate a G4 to A4 mutation. In addition, we titrated heterodimeric proteins into two tandem E-box DNA fragments named DbpI2wt and DbpI2sp10 [22], in which binding to non-canonical E-box (GCACATTC) was undetectable (Supplementary information, Figure S4 and Table S1).

In addition, A2 or T5 site mutated DNA molecules also have detectable binding, but their affinities are much lower, with K_D values generally between 10 to 20 μM (Figure 4F). Furthermore, compared with high-affinity DNA molecules (WT, 7ATd, 1CAd, 3CTz), the A2- or T5-mutated DNA molecules have clearly different thermodynamic profiles during binding. As shown in Figure 4F, binding to high-affinity DNA molecules was mainly driven by enthalpy changes, an indication of specific hydrogen bond interactions. However, binding to A2- or T5-mutated DNA molecules was mainly driven by entropy changes, suggesting a loss of specificity. These low-affinity A2- or T5-mutated E-boxes appear largely in tandem form in front of some clock-controlled genes such as *Per3*, *Ptma*, *Qk*, *Trfp*, *Rexo2* [23]. However, their detailed roles *in vivo* require further investigation due to their low affinities.
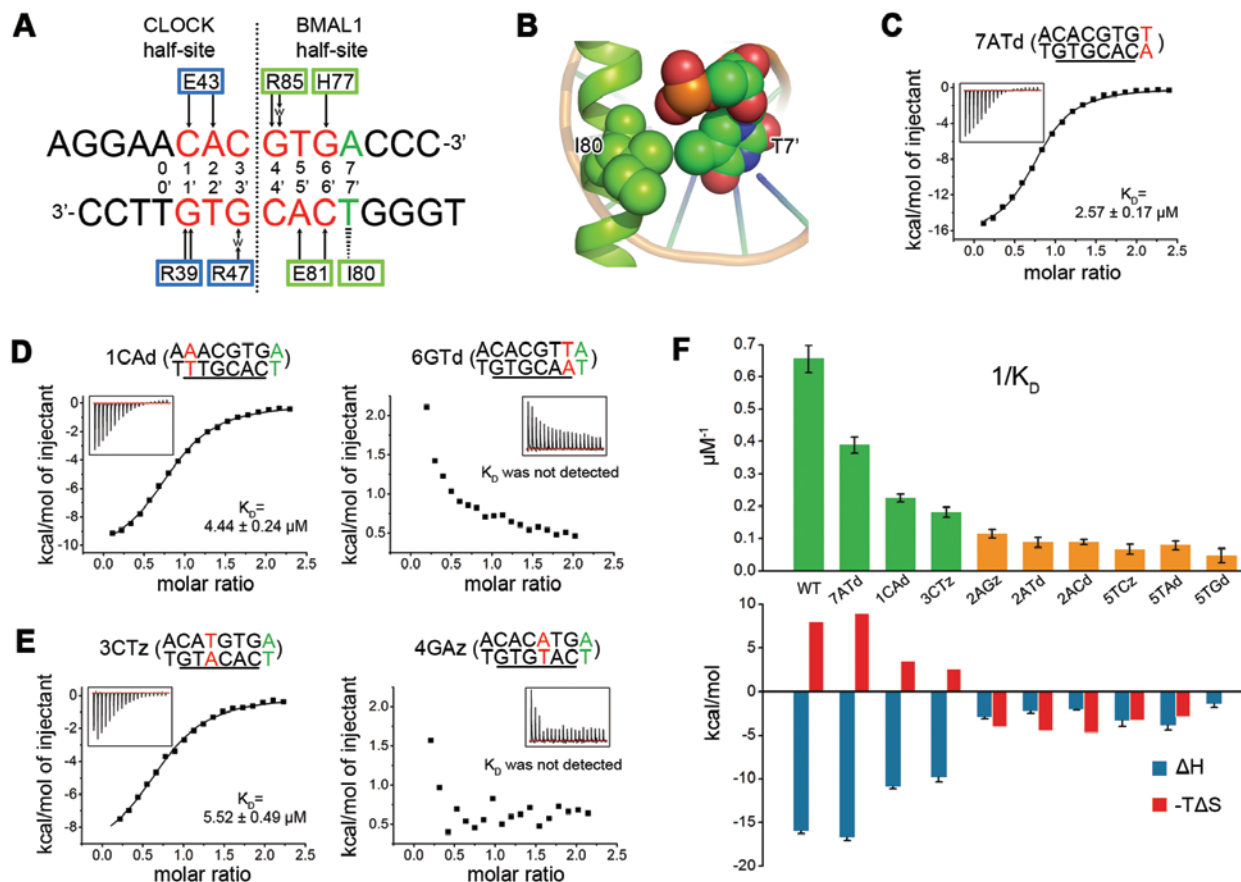
Figure 4 CLOCK and BMAL1 read DNA bases asymmetrically. **(A)** A schematic recognition diagram of CLOCK-BMAL1 to DNA in the crystal structure. Residues that participate in DNA recognition are labeled by a blue rectangle in CLOCK and a green rectangle in BMAL1. E-box is in red. Flanking base pairs that are recognized by BMAL1 is colored green. Black arrows represent hydrogen bonds. Lower-case 'w' means water molecule. Zebra triangle represents hydrophobic contact. DNA is divided into the CLOCK half-site and the BMAL1 half-site by a dashed line. **(B)** A close view of the hydrophobic contact between BMAL1 Ile80 and flanking T7′. **(C)** Titration of CLOCK-BMAL1 bHLH heterodimer into a 16 bp canonical E-box DNA with flanking T7′ mutated (7ATd, ACACGTGT). **(D)** CLOCK-BMAL1 binds to 1CAd (AAACGTGA) but not 6GTd (ACACGTTA). **(E)** CLOCK-BMAL1 binds to 3CTz (ACATGTGA) but not 4GAz (ACACATGA). The core DNA sequences used in titration are shown in subtitles and the E-box is underlined. Mutated base pairs are colored red, and the K_D ± s.d. is shown. **(F)** The A2 or T5 site single-mutated E-box DNA molecules have much lower binding affinities and different thermodynamic profiles. $1/K_D$ ± s.d. is shown. The core sequences of the mutated DNA are 2AGz (ACGCGTGA), 2ATd (ACTCGTGA), 2ACd (ACCCGTGA), 5TCz (ACACGCGA), 5TAd (ACACGAGA), 5TGd (ACACGGGA). See Supplementary information, Table S1 for full-length DNA used in ITC measurements.

In summary, we have defined two non-canonical E-box binding patterns with high affinities, AACGTGA and CATGTGA, both of which have a similar thermodynamic profile to canonical E-box during interactions with CLOCK-BMAL1. From a structural perspective and based on the results obtained from ITC, we conclude that when E-box is in two non-canonical forms (AACGTG and CATGTG), the specific binding to E-box will be determined by the existence of a hydrophobic contact between flanking thymine nucleotide and BMAL1 Ile80.
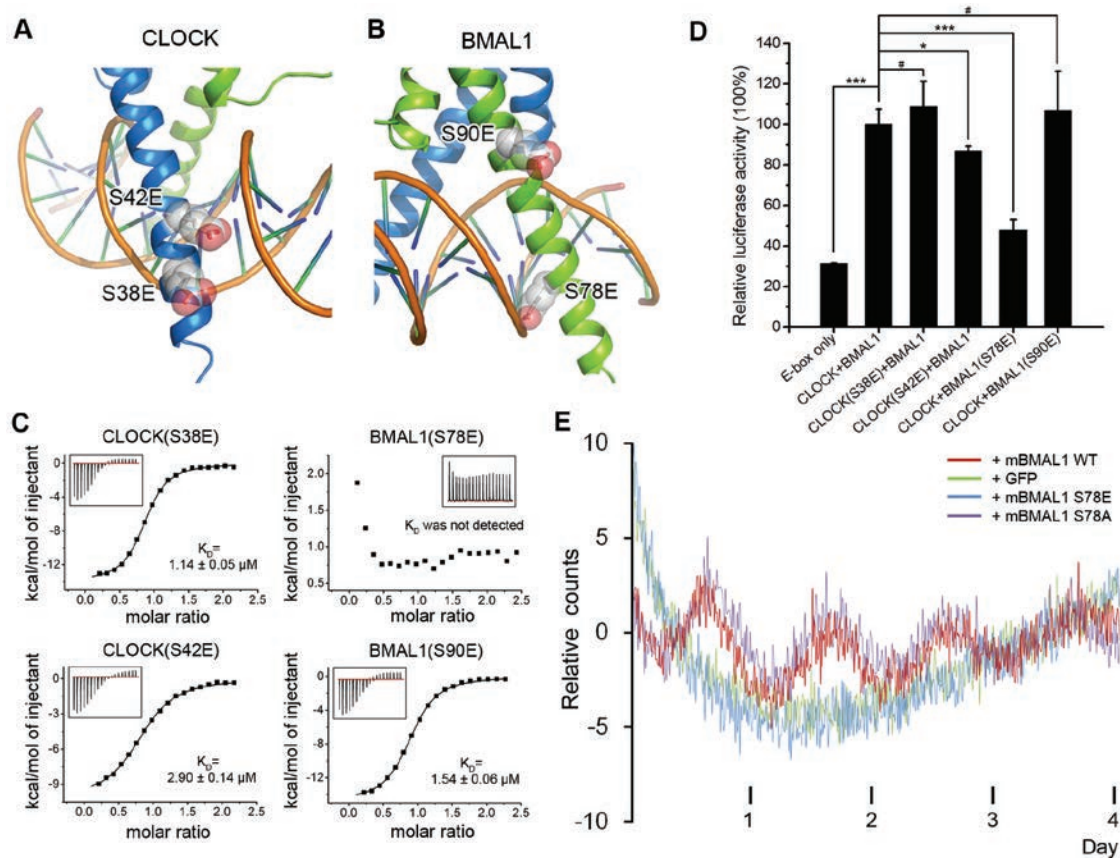
Potential phosphorylation sites in basic regions

To ensure the robustness and precision of the clock, circadian components are extensively regulated. Phosphorylation of CLOCK and BMAL1 basic regions can inhibit their binding to DNA, thus downregulating their transcriptional activity. CLOCK and BMAL1 each has two serine residues in their basic regions (CLOCK Ser38, Ser42 and BMAL1 Ser78, Ser90) (Figure 5A and 5B). It has been reported that three of the four serine residues (CLOCK Ser38, Ser42 and BMAL1 Ser90) can

be phosphorylated *in vivo* [38, 39]. However, BMAL1 Ser78 is also very likely to undergo phosphorylation predicted by NetPhos 2.0 Server [40], which is supported by the finding of a mass spectrometry record of phosphorylated BMAL1 Ser78 in PhosphoSite database [41]. To compare the effects of the four serine residues on phosphorylation-mediated DNA-binding inhibition, we mutated them to phospho-mimicking glutamate residues individually and measured the binding affinities of corresponding heterodimeric proteins to WT DNA. Of the four mutations, CLOCK S38E, S42E and BMAL1 S90E have no significant influence on DNA binding, whereas the BMAL1 S78E mutation almost completely abolishes DNA binding (Figure 5C and Supplementary information, Table S1).

These results can be well explained from a structural perspective. Side chains of CLOCK S38E and BMAL1 S90E protrude away from the DNA molecule (Figure 5A and 5B). Although CLOCK S42E forms charge repulsion with the DNA backbone, the heterodimeric proteins carrying this mutation can still bind DNA strongly (Figure 5C). Notably, the inhibitory role played by BMAL1 S78E in DNA binding is more than a consequence of charge repulsion, but a result of steric hindrance (Figure 5B).

Phospho-mimicking BMAL1 S78E mutation compromises circadian oscillation in cells

To determine whether the *in vitro* measured binding affinities of different mutants depict physiological rel-

Figure 5 Potential phosphorylation sites in basic regions and their effects on DNA binding. **(A)** Modeled phospho-mimicking mutations in CLOCK basic region. **(B)** Modeled phospho-mimicking mutations in BMAL1 basic region. **(C)** Effects of phospho-mimicking mutations on DNA binding. CLOCK-BMAL1 bHLH domains carrying the indicated mutations were titrated into WT E-box DNA. K_D ± s.d. is shown. Only the BMAL1 S78E mutation effectively blocks DNA binding. **(D)** Luciferase reporter assay demonstrates that BMAL1 S78E can inhibit DNA binding *in vivo*, thus diminishing the transcriptional activity of CLOCK-BMAL1. Data were represented as mean ± s.d. of three repeats in one assay. P value was calculated by Student's t-test. ***P < 0.001; *P < 0.05; #P > 0.1. **(E)** Real-time whole-cell luciferase assay. *mBMAL1* WT, *mBMAL1* S78A, *mBMAL1* S78E and *GFP* genes were transfected into mouse mBMAL1$^{-/-}$ mPer2^{Luciferase} fibroblast cells, respectively. Real-time expression of the fusion protein mPER2-luciferase was monitored for 4 days. Red, purple, blue and green lines represent the luminescent signal using *mBMAL1* WT, *mBMAL1* S78A, *mBMAL1* S78E and *GFP* as rescue genes, respectively.

evance, we used a luciferase reporter system to test the transcriptional activities of full-length human CLOCK-BMAL1 carrying these structure-based mutations. When non-tagged human WT *CLOCK* and *BMAL1* genes were co-transfected with the luciferase reporter gene into HEK293T cells, they increased the luminescent signal up to 4 folds. The signal strength in CLOCK S38E, S42E and BMAL1 S90E transfected cells is similar to WT, indicating that these phospho-mimicking mutations have no significant effect on DNA binding and transcriptional activation. In contrast, the BMAL1 S78E transfected cells have an obvious decrement of the luminescent signal compared with WT, which is in accordance with ITC result (Figure 5D).

To further investigate the functional role of the BMAL1 mutants, we introduced *mBMAL1* WT, *mBMAL1* S78E, *mBMAL1* S78A or *GFP* genes into the mouse mBMAL1$^{-/-}$ mPer2^{Luciferase} [42] fibroblast cell line by lentivirus-based transfection, and tested whether they could rescue the mPER2 oscillation in cells. Real-time monitoring of the fusion protein mPER2-luciferase expression showed that mBMAL1 S78E, similar to the GFP control, could not restore the mPER2 oscillation in cells as WT (Figure 5E). Notably, the mBMAL1 S78A mutant retained the ability to induce the oscillated signal of mPER2, suggesting that dephosphorylation on BMAL1 Ser78 is critical for maintaining the clock oscillation.

Discussion

Hydrophobic contacts are mainly employed to read pyrimidines in DNA binding proteins [43]. In our structure, we observed a hydrophobic contact between BMAL1 and the flanking thymine, which suggests that CLOCK-BMAL1 heterodimer actually reads 7-bp DNA and not the previously believed 6-bp DNA. The additional base pair appears to be not so important when the E-box has a canonical form. However, to recognize non-canonical E-boxes, the flanking interaction is indispensable. Through systematic DNA mutations and ITC measurements, we have identified two non-canonical E-box patterns with high affinities, AACGTGA and CATGTGA. The former pattern composes a previously demonstrated functional E'-box (AACGTG or CACGTT), and a flanking A7-T7′ nucleotide that is rather conserved in mammalian *Per*2 promoters [20, 44]. Notably, compared with the functional binding pattern AACGTGA, the pattern CATGTGA, has a similar affinity to CLOCK-BMAL1 and similar thermodynamic profiles during binding, indicating it is also functional *in vivo*. However, if the conserved flanking A7-T7′ is substituted, both E-boxes are very likely to lose their function. In conclusion, we suggest the pattern

AACGTGA and CATGTGA should be used instead of AACGTG and CATGTG for non-canonical binding site searching.

Circadian rhythms are mainly synchrotronized by day-night cycles. In addition, cellular environment changes attributed to metabolism also affect the circadian rhythm profoundly [45]. In this study, we identified a possible phosphorylation site in BMAL1 basic regions that may be one of the cutting points for the external cues to regulate the circadian core transcriptional process. We have shown that the S78E mutant of BMAL1 significantly compromises the ability of DNA binding as well as the transcriptional activity of the complex in cells. Compared with the previously reported inhibitory phosphorylation sites in CLOCK basic regions (Ser38 and Ser42) [38], phosphorylation on BMAL1 Ser78 will be more crucial and efficient in downregulating the transcriptional activity of CLOCK-BMAL1. Considering that many kinases can phosphorylate and regulate the activity of BMAL1 at multiple sites [6, 46, 47], we propose that BMAL1 Ser78 should be a key residue mediating input signal-regulated transcriptional inhibition for external cues to entrain the circadian clock by kinase cascade.

In general terms, the association and dissociation of transcription factors with DNA are key aspects to determine the transcriptional activities. In the fruit fly, circadian transcription by CLK-CYC is such a rhythmic DNA binding process that is regulated by interactions with PER-TIM [48]. In mammals, it is previously reported that CLOCK-BMAL1 heterodimer remains bound to DNA throughout the circadian cycle [49]. However, a contrary report showed mouse CLOCK-BMAL1 could undergo rhythmic binding to multiple E-boxes in regulating the oscillation of *Dbp* gene [50]. We have demonstrated that phospho-mimicking S78E mutant of BMAL1 efficiently blocks DNA binding, which provides a molecular rationale for the possibility of rhythmic binding of CLOCK-BMAL1 during circadian cycle. In addition, phosphorylation on BMAL1 Ser78 may also help to promote protein recycling by disassociating the complex from DNA when the circadian transcriptional process is terminated.

Materials and Methods

Protein preparation

The bHLH of *Homo sapiens* CLOCK (residues 29-89) and BMAL1 (residues 66-128) were cloned into pET-21b (Novagen) vector using *Nde* I and *Xho* I restriction sites. A tryptophan codon was inserted in front of *Xho* I site for monitoring the absorption of both proteins at 280 nm. The plasmids were transformed into *E. coli* strain Rosetta and the proteins were overexpressed in LB medium at 30 °C for 5 h by induction with 0.5 mM β-D-

thiogalactopyranoside (IPTG). The cells were harvested by centrifugation, resuspended in buffer A (1 M NaCl, 20 mM Tris-HCl, pH 7.8) and lysed by sonication. After centrifugation at 25 000 rpm for 2 h, the soluble fraction of the cell lysate was loaded on a Ni^{2+} affinity column in buffer A, washed and eluted by 20% and 80% buffer B (500 mM NaCl, 500 mM imidazole, 20 mM Tris-HCl, pH 7.8), respectively. Equal moles of CLOCK and BMAL1 proteins were thoroughly mixed and incubated on ice for 10 min, followed by gel filtration using a Superdex 75 column in buffer C (500 mM NaCl, 20 mM Tris-HCl, pH 7.8). For subsequent crystallization analysis, the proteins were desalted into buffer D (200 mM NaCl, 20 mM Tris-HCl, pH 7.8), concentrated to 25 mg/ml and stored at −80 °C.

Crystallization and data collection

DNA primers used in crystallization are: forward, AGGAA-CACGTGACCC; reverse, TGGGTCACGTGTTCC. This DNA sequence was derived from a previously reported site selection and amplification experiment [7]. Equal moles of 15 bp forward and reverse primers were annealed to form a sticky-ended double-stranded DNA in buffer D (200 mM NaCl, 20 mM Tris-HCl, pH 7.8). To get the complex crystals, the DNA and heterodimeric proteins were mixed at a molar ratio of 1.1:1, and incubated on ice for 10 min. The final concentration of proteins for crystallization is ~17 mg/ml. Crystal screening was carried out using the sitting drop vapor diffusion method at 16 °C by mixing 1 μl DNA-protein complex and 1 μl reservoir solution containing 20% (w/v) PEG 3350 and 0.2 M magnesium formate. Crystals were grown to full size in 1 week. Before flash frozen, the crystals were soaked in the cryoprotectant containing 20% (w/v) PEG 3350, 0.2 M magnesium formate, 20% glycerol for ~2 min. The data were collected at KEK, Photon Factory beamline BL17A, Tsukuba, Japan, using a Quantum 315 CCD detector (Area Detector Systems Corporation). The diffraction data were integrated and scaled using HKL-2000 program [51].

Structure determination and refinement

Due to the strong anisotropic nature of the data, ellipsoidal truncation and anisotropic scaling of the structure factors were performed using Diffraction Anisotropy Server [52]. Corrected structure factors were used in phasing and refinement. Initial phases were determined by molecular replacement using the modified PDB 1NLW as a search model. The structure was manually built using Coot [53] combined with refinement using *Phenix.refine* [54]. The atomic coordinates and diffraction data have been deposited to Protein Data Bank with the accession code 4H10.

ITC measurements

See Supplementary information, Table S1 for DNA sequences used in ITC measurements. ITC measurements were performed using ITC200 (GE Healthcare) at 25 °C. All annealed double-stranded blunt-ended DNA used in ITC experiments were purified by gel filtration using a Superdex 75 column. The main peak eluted was collected and the DNA concentration was determined by measuring the absorbance at 260 nm using Nanodrop (Thermo Fisher Scientific). Both proteins and DNA were in buffer C (500 mM NaCl, 20 mM Tris-HCl, pH 7.8). For determination of the affinities between two proteins, typically 1 mM (monomer concentration) WT or mutated CLOCK bHLH protein was titrated into

0.12 mM (monomer concentration) BMAL1 bHLH protein. For determination of the affinities between protein and DNA, typically 0.4 mM protein (heterodimer or homodimer concentration) was titrated into 0.036 mM double-stranded DNA. The thermograms were fitted to a single binding site model using Origin software.

Analytical ultracentrifugation

Analytical ultracentrifugation measurements were performed using ProteomeLab XL-I (Beckman Coulter) analytical ultracentrifuge with an An-60 Ti rotor. bHLH domains of CLOCK, BMAL1 and their complexes were analyzed in buffer C (500 mM NaCl, 20 mM Tris-HCl, pH 7.8) with a speed of 60 000 rpm (262 000× *g*) rotating for at least 6 h at 25 °C.

Luciferase reporter assay

Luciferase reporter assay was performed in a 96-well plate and each sample was performed in triplicate wells. HEK293T cells were seeded in DMEM supplemented with 10% FBS and cultured to 80% confluence at 37 °C with 5% CO_2. Before transfection, the cells were harvested by centrifugation and resuspended to 1.6×10^6 cells/ml. 50 μl suspended cells were dispensed into individual wells to ~8×10^4 cells per well. For preparing the transfection sample of each well, the following amounts of plasmids were diluted in 25 μl Opti-MEM I reduced serum medium: 10 ng pGL3-3 × E-box, 40 ng pcDNA3.1-hBMAL1, 120 ng pcDNA3.1-hCLOCK and 30 ng pcDNA3.1 for normalization (total 200 ng per well). Then 0.5 μl Lipofectamine 2000 (Invitrogen) was diluted in 25 μl Opti-MEM I reduced serum medium and incubated for 5 min. The diluted plasmids and Lipofactamine were then combined and incubated for 20 min, followed by transferring the mixture (total 50 μl) into one well containing cells. After transfection, the plate was incubated at 37 °C with 5% CO_2. 24 h after transfection, 50 μl medium was removed from each well. The cells were lysed and luciferase activity was measured using a Bright-Glo Luciferase Assay System (Promega) according to the manufacturer's protocol.

Real-time whole-cell luciferase reporter assay

mBMAL1 WT, *mBMAL1* S78E, *mBMAL1* S78A and *GFP* genes were constructed into the lentivirus vector using the ViraPower Promoterless Lentiviral Gateway Kits (Invitrogen) according to the manufacturer's protocol. The four kinds of lentiviruses were produced as previously described [55]. Mouse $mBMAL1^{-/-}$ mPer-$2^{Luciferase}$ fibroblast cells were cultured in DMEM supplemented with 10% FBS until 80% confluence in a six-well plate. To infect the cells, the medium in each well was replaced with 1 ml virus suspension, 1 ml fresh DMEM supplemented with 10% FBS, antibiotics (100 units/ml penicillin, 100 μg/ml streptomycin) and Polybrene (a final concentration of 8 μg/ml). After 18 h, the medium containing virus was replaced with fresh DMEM supplemented with 10% FBS and antibiotics. Approximately 48 h after infection, the infected cells were seeded in 10-cm dishes, and blasticidin was added to the medium to a final concentration of 5 μg/ml. The dishes were then incubated for 3 days to select stably transduced cells. To monitor real-time whole-cell luciferase activity, the blasticidin-resistant $mBMAL1^{-/-}$ mPer2^{Luciferase} fibroblast cells were plated into 3.5 cm dishes and cultured to 90% confluence. The cells were then changed into HEPES-buffered DMEM containing 1 μM luciferin (Promega) and B-27 supplements (Invitrogen), followed by sealing the dishes with the coverslip and vacuum grease. The real-time

luminescence was recorded by Lumicycle (Actimetrics) for 4 days in a 36 °C incubator.

References

1   Panda S, Hogenesch JB, Kay SA. Circadian rhythms from flies to human. *Nature* 2002; **417**:329-335.

2   Reppert SM, Weaver DR. Coordination of circadian timing in mammals. *Nature* 2002; **418**:935-941.

3   Green CB, Takahashi JS, Bass J. The meter of metabolism. *Cell* 2008; **134**:728-742.

4   Maury E, Ramsey KM, Bass J. Circadian rhythms and metabolic syndrome: from experimental genetics to human disease. *Circ Res* 2010; **106**:447-462.

5   Dunlap JC. Molecular bases for circadian clocks. *Cell* 1999; **96**:271-290.

6   Zhang EE, Kay SA. Clocks not winding down: unravelling circadian networks. *Nat Rev Mol Cell Biol* 2010; **11**:764-776.

7   Hogenesch JB, Gu YZ, Jain S, Bradfield CA. The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors. *Proc Natl Acad Sci USA* 1998; **95**:5474-5479.

8   Gekakis N, Staknis D, Nguyen HB, *et al.* Role of the CLOCK protein in the mammalian circadian mechanism. *Science* 1998; **280**:1564-1569.

9   Massari ME, Murre C. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol* 2000; **20**:429-440.

10   Murre C, Bain G, van Dijk MA, *et al.* Structure and function of helix-loop-helix proteins. *Biochim Biophys Acta* 1994; **1218**:129-135.

11   Ledent V, Vervoort M. The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. *Genome Res* 2001; **11**:754-770.

12   Munoz E, Baler R. The circadian E-box: when perfect is not good enough. *Chronobiol Int* 2003; **20**:371-388.

13   Ephrussi A, Church GM, Tonegawa S, Gilbert W. B lineage-specific interactions of an immunoglobulin enhancer with cellular factors *in vivo*. *Science* 1985; **227**:134-140.

14   Murre C, McCaw PS, Baltimore D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 1989; **56**:777-783.

15   Murre C, McCaw PS, Vaessin H, *et al.* Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* 1989; **58**:537-544.

16   Swanson HI, Chan WK, Bradfield CA. DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins. *J Biol Chem* 1995; **270**:26292-26302.

17   Ueda HR, Chen W, Adachi A, *et al.* A transcription factor response element for gene expression during circadian night. *Nature* 2002; **418**:534-539.

18   Storch KF, Lipan O, Leykin I, *et al.* Extensive and divergent circadian gene expression in liver and heart. *Nature* 2002; **417**:78-83.

19   Panda S, Antoch MP, Miller BH, *et al.* Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* 2002; **109**:307-320.

20   Yoo SH, Ko CH, Lowrey PL, *et al.* A noncanonical E-box enhancer drives mouse Period2 circadian oscillations *in vivo*. *Proc Natl Acad Sci USA* 2005; **102**:2608-2613.

21   Hatanaka F, Matsubara C, Myung J, *et al.* Genome-wide profiling of the core clock protein BMAL1 targets reveals a strict relationship with metabolism. *Mol Cell Biol* 2010; **30**:5636-5648.

22   Rey G, Cesbron F, Rougemont J, Reinke H, Brunner M, Naef F. Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. *PLoS Biol* 2011; **9**:e1000595.

23   Nakahata Y, Yoshida M, Takano A, *et al.* A direct repeat of E-box-like elements is required for cell-autonomous circadian rhythm of clock genes. *BMC Mol Biol* 2008; **9**:1.

24   Huang N, Chelliah Y, Shan Y, *et al.* Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science* 2012; **337**:189-194.

25   Nair SK, Burley SK. X-ray structures of Myc-Max and Mad-Max recognizing DNA: molecular bases of regulation by proto-oncogenic transcription factors. *Cell* 2003; **112**:193-205.

26   Ferre-D'Amare AR, Pognonec P, Roeder RG, Burley SK. Structure and function of the b/HLH/Z domain of USF. *EMBO J* 1994; **13**:180-189.

27   Brownlie P, Ceska T, Lamers M, *et al.* The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. *Structure* 1997; **5**:509-520.

28   Parraga A, Bellsolell L, Ferre-D'Amare AR, Burley SK. Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 Å resolution. *Structure* 1998; **6**:661-672.

29   Ma PC, Rould MA, Weintraub H, Pabo CO. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* 1994; **77**:451-459.

30   Ferre-D'Amare AR, Prendergast GC, Ziff EB, Burley SK. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* 1993; **363**:38-45.

31   Shimizu T, Toumoto A, Ihara K, *et al.* Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. *EMBO J* 1997; **16**:4689-4697.

32   Longo A, Guanga GP, Rose RB. Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition. *Biochemistry* 2008; **47**:218-229.

33   Huang N, Chelliah Y, Shan Y, *et al.* Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator com-

34 Rutter J, Reick M, Wu LC, McKnight SL. Regulation of clock and NPAS2 DNA binding by the redox state of NAD cofactors. *Science* 2001; **293**:510-514.

35 Wijnen H. Circadian rhythms. A circadian loop asSIRTs itself. *Science* 2009; **324**:598-599.

36 Ramsey KM, Yoshino J, Brace CS, *et al*. Circadian clock feedback cycle through NAMPT-mediated NAD$^+$ biosynthesis. *Science* 2009; **324**:651-654.

37 Nakahata Y, Sahar S, Astarita G, Kaluzova M, Sassone-Corsi P. Circadian control of the NAD$^+$ salvage pathway by CLOCK-SIRT1. *Science* 2009; **324**:654-657.

38 Yoshitane H, Takao T, Satomi Y, Du NH, Okano T, Fukada Y. Roles of CLOCK phosphorylation in suppression of E-box-dependent transcription. *Mol Cell Biol* 2009; **29**:3675-3686.

39 Tamaru T, Hirayama J, Isojima Y, *et al*. CK2alpha phosphorylates BMAL1 to regulate the mammalian clock. *Nat Struct Mol Biol* 2009; **16**:446-448.

40 Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 1999; **294**:1351-1362.

41 Hornbeck PV, Chabra I, Kornhauser JM, Skrzypek E, Zhang B. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* 2004; **4**:1551-1561.

42 Yoo SH, Yamazaki S, Lowrey PL, *et al*. PERIOD2:: LU-CIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *Proc Natl Acad Sci USA* 2004; **101**:5339-5346.

43 Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS. Origins of specificity in protein-DNA recognition. *Annu Rev Biochem* 2010; **79**:233-269.

44 Akashi M, Ichise T, Mamine T, Takumi T. Molecular mechanism of cell-autonomous circadian gene expression of Period2, a crucial regulator of the mammalian circadian clock. *Mol Biol Cell* 2006; **17**:555-565.

45 Bass J, Takahashi JS. Circadian integration of metabolism and energetics. *Science* 2010; **330**:1349-1354.

46 Sanada K, Okano T, Fukada Y. Mitogen-activated protein kinase phosphorylates and negatively regulates basic helix-loop-helix-PAS transcription factor BMAL1. *J Biol Chem* 2002; **277**:267-271.

47 Gallego M, Virshup DM. Post-translational modifications regulate the ticking of the circadian clock. *Nat Rev Mol Cell Biol* 2007; **8**:139-148.

48 Yu W, Zheng H, Houl JH, Dauwalder B, Hardin PE. PER-dependent rhythms in CLK phosphorylation and E-box binding regulate circadian transcription. *Genes Dev* 2006; **20**:723-733.

49 Lee C, Etchegaray JP, Cagampang FR, Loudon AS, Reppert SM. Posttranslational mechanisms regulate the mammalian circadian clock. *Cell* 2001; **107**:855-867.

50 Ripperger JA, Schibler U. Rhythmic CLOCK-BMAL1 binding to multiple E-box motifs drives circadian *Dbp* transcription and chromatin transitions. *Nat Genet* 2006; **38**:369-374.

51 Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997; **276**:307-326.

52 Strong M, Sawaya MR, Wang S, Phillips M, Cascio D, Eisenberg D. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 2006; **103**:8060-8065.

53 Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004; **60**:2126-2132.

54 Adams PD, Grosse-Kunstleve RW, Hung LW, *et al*. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 2002; **58**:1948-1954.

55 Tiscornia G, Singer O, Verma IM. Production and purification of lentiviral vectors. *Nat Protoc* 2006; **1**:241-245.

(**Supplementary information** is linked to the online version of the paper on the *Cell Research* website.)