HCF-1 self-association via an interdigitated Fn3 structure facilitates transcriptional regulatory complex formation

Jihye Park[a], Fabienne Lammers[b], Winship Herr[b], and Ji-Joon Song[a,1]

[a]Department of Biological Sciences and Graduate School of Nanoscience and Technology (World Class University), KAIST (Korea Advanced Institute of Science and Technology) Institute for the BioCentury, KAIST, Daejeon 305-701, Korea; [b]Center for Integrative Genomics, University of Lausanne, Génopode, 1015 Lausanne, Switzerland

Host-cell factor 1 (HCF-1) is an unusual transcriptional regulator that undergoes a process of proteolytic maturation to generate N-(HCF-1$_N$) and C- (HCF-1$_C$) terminal subunits noncovalently associated via self-association sequence elements. Here, we present the crystal structure of the self-association sequence 1 (SAS1) including the adjacent C-terminal HCF-1 nuclear localization signal (NLS). SAS1 elements from each of the HCF-1$_N$ and HCF-1$_C$ subunits form an interdigitated fibronectin type 3 (Fn3) tandem repeat structure. We show that the C-terminal NLS recruited by the interdigitated SAS1 structure is required for effective formation of a transcriptional regulatory complex: the herpes simplex virus VP16-induced complex. Thus, HCF-1$_N$–HCF-1$_C$ association via an integrated Fn3 structure permits an NLS to facilitate formation of a transcriptional regulatory complex.

crystallography | chromatin | nuclear localization sequence

Host cell factor-1 (HCF-1; also known as HCFC1) is a metazoan transcriptional regulator. HCF-1 was initially identified as a human coactivator for immediate-early gene expression of herpes simplex virus (HSV) by forming a complex (VP16-induced complex, VIC) with the HSV protein VP16 and the cellular transcriptional regulator Oct-1 (1). HCF-1 regulates cell-cycle progression by mediating associations between DNA-binding transcription factors and chromatin modifying complexes (2–12).

Many HCF-1 proteins, including all known vertebrate HCF-1s, undergo a process of proteolytic maturation. Thus, human HCF-1 is synthesized as a 2035-aa precursor, which is cleaved and modified by O-GlcNAC transferase (13) to generate HCF-1$_N$ and HCF-1$_C$ subunits possessing different roles in, for example, cell-cycle regulation (14). After cleavage, the two resulting HCF-1$_N$ and HCF-1$_C$ subunits remain noncovalently associated via two "self-association sequences" called SAS1 and SAS2: SAS1 is the primary association element and its sequence is conserved in HCF-1, whereas SAS2 is secondary, less well conserved, and not considered in this study (15).

SAS1 is composed of a short 43-aa HCF-1$_N$ segment called SAS1N and a 197-aa HCF-1$_C$ segment called SAS1C (Fig. 1A). Wilson et al. (15) suggested that SAS1C has two tandem fibronectin type 3 (Fn3) repeat structures forming a binding site for the 43-aa SAS1N to promote HCF-1$_N$–HCF-1$_C$ association.

To understand the mechanism of HCF-1 self-association, we have determined the crystal structure of SAS1 with the neighboring nuclear localization signal (called here SAS1–NLS). Contrary to expectation, SAS1N and SAS1C together—not SAS1C alone—form an integrated tandem Fn3 structure. Furthermore, we show that the self-association tethers the basic NLS to the HCF-1$_N$ Kelch domain to facilitate VP16-induced complex formation.

Results

Crystal Structure of the HCF-1 SAS1 Self-Association Domain. The SAS1–NLS crystal structure (Fig. 1B) was determined to 2.7 Å resolution using the Se single-wavelength anomalous dispersion (SAD) method (Table S1 and Fig. S1). Fig. S2 details schematically the different molecules used in this study as well as their nomenclature (Fig. S2). In the SAS1–NLS crystal structure, there are two molecules per asymmetric unit, each possessing a different conformation (Fig. 1C): in one conformation (shown in Fig. 1B and in gray in Fig. 1C), a part of the NLS (HCF-1 residues 2014–2020) is well structured; in the other conformation (shown in green in Fig. 1C), the NLS is disordered. Unless otherwise specifically noted, we refer to the conformation with a structured NLS (Fig. 1B).

The SAS1–NLS crystal structure reveals several interesting features (Fig. 1B). It is composed of two structural domains: one domain (Fn3-1) is composed of SAS1N (HCF-1 residues 360–402) and the N-terminal portion of SAS1C (1806–1885 amino acids), and the other domain (Fn3-2) consists of the C-terminal portion of SAS1C (1892–2002 amino acids). These two domains are connected through a structured and highly conserved connecting loop (1886–1891 amino acids) (Fig. 1C and Figs. S3A and S4B). The relative orientation between the two domains in the two SAS1–NLS conformations differs, with the conformation possessing a disordered NLS displaying a wider angle between the two domains (Fig. 1C).

Interestingly, the structured NLS is located between the two domains (Fig. 1B and Fig. S4). Electrostatic surface presentations of the two SAS1–NLS conformations reveal highly charged surfaces in which the basic NLS in the NLS-containing conformation forms a highly basic patch (circled in Fig. 1D, Left) that is largely absent in the disordered NLS conformation (Fig. 1D, Right). Thus, NLS binding seems to be coordinated with an SAS1 conformational change in which the two SAS1 domains grasp the NLS by closing onto each other, thus generating a highly basic surface.

The most striking feature of SAS1–NLS is the binding mode between SAS1N and SAS1C, in which an interdigitated SAS1N/SAS1C structure is formed (Fig. 1B). The entire SAS1N and the N-terminal portion of SAS1C together form a single Fn3 domain (Fn3-1), whereas the C-terminal portion of SAS1C adopts a canonical Fn3 structure (Fn3-2) (Fig. 1E and Fig. S3B). Thus, instead of SAS1C containing an autonomous tandem Fn3 structure as suggested (15), a tandem Fn3 structure is created by SAS1N and SAS1C association. Given this interdigitated structure, we refer
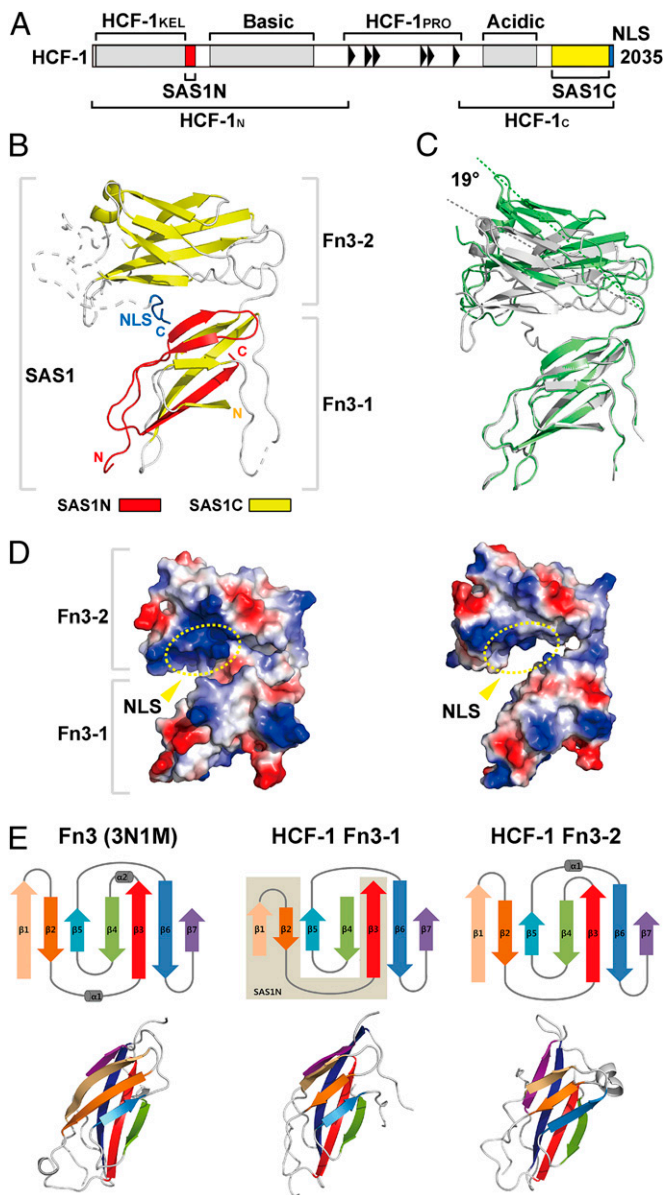
Fig. 1. Crystal structure of the HCF-1 SAS1 and NLS. (*A*) A schematic diagram showing conserved regions of HCF-1 including HCF-1$_{KEL}$, SAS1N, basic, HCF-1$_{PRO}$, acidic, SAS1C, and NLS. The regions used for this study are shown in red for SAS1N, yellow for SAS1C, and blue for NLS. (*B*) A ribbon diagram of SAS1–NLS. SAS1–NLS is composed of two Fn3 domains, Fn3-1 and Fn3-2. SAS1N of HCF-1$_N$ (red) and SAS1C of HCF-1$_C$ (yellow; as indicated in *A*) form an interdigitated Fn3-1 structure. A part of the NLS (blue) is located between the two Fn3 domains and interacts with the β1 strand of Fn3-1 and the β7 strand of Fn3-2. Disordered regions are shown as dashed lines. (*C*) Comparison between two SAS1–NLS conformations in an asymmetric unit. One conformation without a structured NLS (green) is more open than the other with a structured NLS (gray). (*D*) Electrostatic surface representation of the two SAS1–NLS conformations: with (*Left*) or without (*Right*) structured NLS. Left: the NLS located between Fn3-1 and Fn3-2 (yellow dashed circle) creates a basic patch (indicated by an arrow). Right: The basic patch is absent in the molecule where the NLS is disordered. (*E*) Fn3-1 and Fn3-2 adopt a canonical Fn3 topology. Topology and structure diagrams of a canonical Fn3 domain (PDB ID code 3N1M, the Fn3 domain of Brother of CDO [BOC]) (*Left*), and Fn3-1 (*Middle*) and Fn3-2 (*Right*) domains of HCF-1. Fn3-1 is an interdigitated Fn3 structure composed of both SAS1N (β1–β3 strands; shaded gray) and SAS1C (β4–β7 strands; unshaded) component β strands.

to the region of Fn3-1 originating from SAS1N (i.e., the entirety of SAS1N) as "Fn3-1N" and that originating from SAS1C as "Fn3-1C" (Fig. 1*B* and Fig. S2). Thus, the crystal structure of SAS1–NLS shows that HCF-1$_N$ and HCF-1$_C$ associate in an unexpected binding mode forming a single interdigitated structure from two distant parts, SAS1N and SAS1C.

Interactions Between SAS1N and SAS1C. Because SAS1N and SAS1C are highly conserved (Fig. S3*A*), to understand how the sequence conservation is correlated with the SAS1–NLS structure, we colored the highly conserved residues on the surface of the SAS1N structure in cyan and of the SAS1C structure in green as shown in Fig. 2, revealing a striking distribution of conserved residues. In Fn3-1, the conserved residues are mostly located on the interface between Fn3-1N (compare Fig. 2 *A* and *B*) and Fn3-1C (compare Fig. 2 *C* and *E*), whereas in Fn3-2, conserved residues are located on the surface (Fig. 2 *B* and *D*–*F*) as well as in the interior. At the Fn3-1N and Fn3-1C interface, the conserved residues form a large hydrophobic core (Fig. 3*A*, *I*). There are a few highly conserved residues located on the surface of Fn3-1 and these residues also appear to be involved in holding Fn3-1N and Fn3-1C together (Fig. 3*A*, *II*).

To examine the contributions of conserved Fn3-1 interface residues to SAS1 assembly, we mutated Fn3-1 residues and assayed SAS1 assembly by bacterial cosynthesis and GST pull-down assay between wild-type and mutant SAS1N and SAS1C proteins (Fig. 3 *B* and *C*). To disrupt the hydrophobic interface, we generated mutations V382A, W384A and Y393A, L395A in Fn3-1N, and mutations V1815A, F1864A, V1866A, and V1866E in Fn3-1C. Although none of the mutations fully disrupted SAS1N and SAS1C–NLS association, W384A in Fn3-1N and V1866E in Fn3-1C, two interacting residues, had the greatest effect (Fig. 3*B*). The relatively marginal effects of the alanine substitutions may reflect the hydrophobic properties of alanine side chains (Fig. 3*B*).

To examine the contributions of conserved interacting residues located on the surface of Fn3-1, we generated the Q396A and W1812A mutants. W1812A shows a moderate effect on association, whereas Q396A shows little if any effect (Fig. 3*B*).

Given the relatively modest effects of the mutations on SAS1 assembly, we next asked if there might be additive effects of combined mutations on the hydrophobic Fn3-1N and Fn3-1C interface and the interacting Fn3-1 surface. Because the W1812A-interacting Fn3-1 surface mutation shows a moderate decrease in SAS1 assembly, we combined this mutation with both Fn3-1N interface (V382A, W384A, and Y393A) and interacting surface (Q396A) mutants. Combined, the Fn3-1C W1812A and both classes of Fn3-1N alanine mutants showed severe defects in SAS1 assembly (Fig. 3*C*). These data show that Fn3-1N and Fn3-1C associate through the hydrophobic interface as well as the interacting surface of Fn3-1.

Association Between HCF-1$_N$ and HCF-1$_C$ via Fn3-1 Promotes Transcriptional Regulatory Complex Formation. To investigate the function of HCF-1$_N$ and HCF-1$_C$ association through forming an interdigitated Fn3-1 structure, we examined its role in formation of the VIC transcriptional regulatory complex. HCF-1 forms a VIC with VP16 and Oct-1 on a so-called TAATGARAT DNA target sequence (1). VP16 consists of a core region required for VIC formation and a potent C-terminal transcriptional activation domain enriched in acidic amino acids (also known as acidic activation domain) (Fig. S2*C*). The N-terminal Kelch domain of HCF-1$_N$ (Fig. 1*A* and Fig. S2) is responsible for binding the VP16 core (called VP16$_{ΔAD}$) and stabilizing a VIC containing VP16$_{ΔAD}$ (16), but when a VIC is formed with the full-length acidic activation domain–containing VP16 (VP16$_{fl}$), HCF-1$_C$ sequences (residues 1495–2035) are also important for VIC formation (17). Thus, because VP16$_{fl}$ VIC formation involves both HCF-1$_N$ (the Kelch domain) and HCF-1$_C$ sequences, it provides a functional assay for
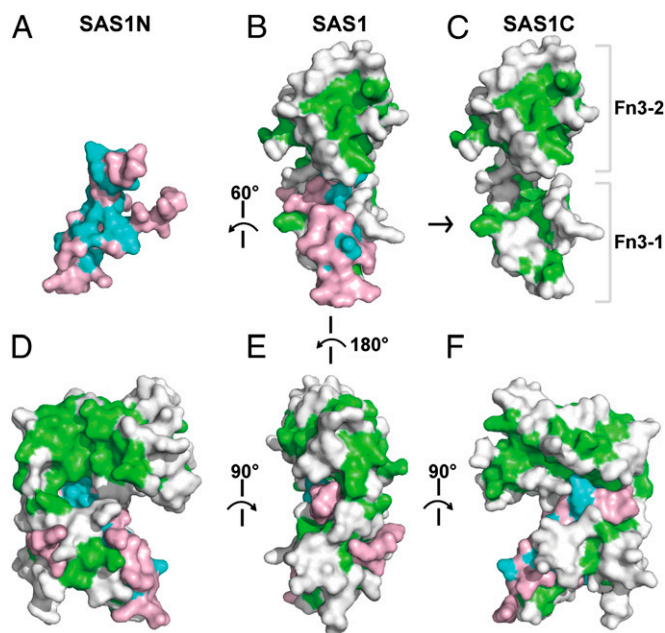
Fig. 2. Surface representations of SAS1–NLS showing highly conserved residues. Highly conserved residues are shown on the surface of SAS1–NLS. SAS1N is shown in pink with highly conserved residues indicated in cyan and SAS1C is shown in gray with conserved residues indicated in green. Most of conserved residues of Fn3-1 are located at the associating interface between SAS1N and SAS1C (A–C). Conserved residues of Fn3-2 are distributed throughout the surface of Fn3-2 (D–F).

SAS1 assembly in an important process—transcriptional regulatory complex assembly.

The HCF-1$_C$ sequences necessary for VP16$_{fl}$ VIC formation were largely not defined (17). We therefore first tested whether the SAS1C–NLS used in the structural analysis (residues 1806–2035) is sufficient to form a VIC with VP16$_{fl}$. We cotranslated in vitro an HCF-1$_N$ construct containing the Kelch domain and SAS1N (called HCF-1$_{Kelch–SAS1N}$) with SAS1C–NLS and performed a VIC formation assay with VP16$_{\Delta AD}$ (Fig. 4A, lanes 3–7, and Fig. S5) and VP16$_{fl}$ (lanes 8–20). Whereas, HCF-1$_C$ sequences are more important for VP16$_{fl}$ than VP16$_{\Delta AD}$ VIC formation (Fig. 4A, compare lanes 4 and 9), the HCF-1$_C$ SAS1C–NLS with cotranslated HCF-1$_{Kelch–SAS1N}$ is sufficient to promote effective VP16$_{fl}$ VIC formation (compare lanes 9 and 12; note the retarded VIC mobility in lane 12 from the added presence of SAS1C–NLS in the VIC). These data suggest that SAS1C–NLS in HCF-1$_C$ is sufficient to promote effective VIC formation with VP16$_{fl}$.

We next asked whether HCF-1$_N$/HCF-1$_C$ association via the Fn3-1N and Fn3-1C interaction observed in the crystal structure is critical for VIC formation. We assayed VIC formation with the HCF-1$_{Kelch–SAS1N}$ (N) mutants N$_{V382A}$, N$_{W384A}$, and N$_{Q396A}$, and the SAS1C–NLS (C) mutants C$_{W1812A}$ and C$_{V1866E}$, and with combined mutants of N$_{V382A}$+C$_{W1812A}$, N$_{V384A}$+C$_{W1812A}$, and N$_{Q396A}$+C$_{W1812A}$. The two mutations, W384A (N$_{W384A}$) and V1866E (N$_{V1866E}$), which showed the greatest defects in SAS1 assembly in the GST pull-down experiments (Fig. 3B), greatly impaired the ability of VP16$_{fl}$ to form VICs (Fig. 4A, lanes 14 and 17). Even mutations that exhibited little or moderate effects in the GST pull-down experiments (i.e., V382A, Q396A, and W1812A) displayed more pronounced effects in the VIC-formation assay (Fig. 4A, lanes 13, 15, and 16), suggesting that the VIC-formation assay is a more sensitive assay for SAS1 assembly than the GST pull-down assay. Consistent with the GST pull-down assays, combined mutants were not able to form a VIC (Fig. 4A, lanes 18–20; note the residual VIC formation observed is the result of

SAS1C-independent VIC formation as evidenced by the comigration of this VIC with the N-only VIC in lane 9). These data show that HCF-1$_N$ and HCF-1$_C$ association through Fn3-1 is critical for wild-type VP16 VIC formation.

SAS1 is created via the association of SAS1N and SAS1C segments that form the interdigitated Fn3-1. Supporting the importance of this interaction, the conserved amino acids in Fn3-1 are mostly located at the SAS1N and SAS1C interface (Fig. 2). These observations led us to hypothesize that the principal function of Fn3-1 is to hold HCF-1$_N$ and HCF-1$_C$ together. To test this hypothesis, we constructed a fusion protein (called N·C) with the C terminus of HCF-1$_{Kelch–SAS1N}$ covalently linked to the N terminus of the HCF-1$_{C\_SAS1C–NLS}$ via a six-glycine linker. We then introduced mutations into the N·C fusion protein, hypothesizing that if the function of Fn3-1 in VIC formation is limited to holding HCF-1$_N$ and HCF-1$_C$ together, then the covalent linkage should rescue the defects of the SAS1N/SAS1C association mutants. Consistent with this hypothesis, the fusion protein rescued the VP16$_{fl}$ VIC formation activity of the SAS1N (W384A) and SAS1C (W1812A) mutations either separately or combined (Fig. 4B, compare lanes 6–8 with lanes 9–11).

We next asked whether the Fn3-1 domain is required at all for VIC formation by deleting one or the other or both of the Fn3-1N and Fn3-1C segments in the N·C fusion protein. Although these deletions cause a change in VIC mobility, they do not affect the level of VIC formation (Fig. 4B, lanes 12–14). These results
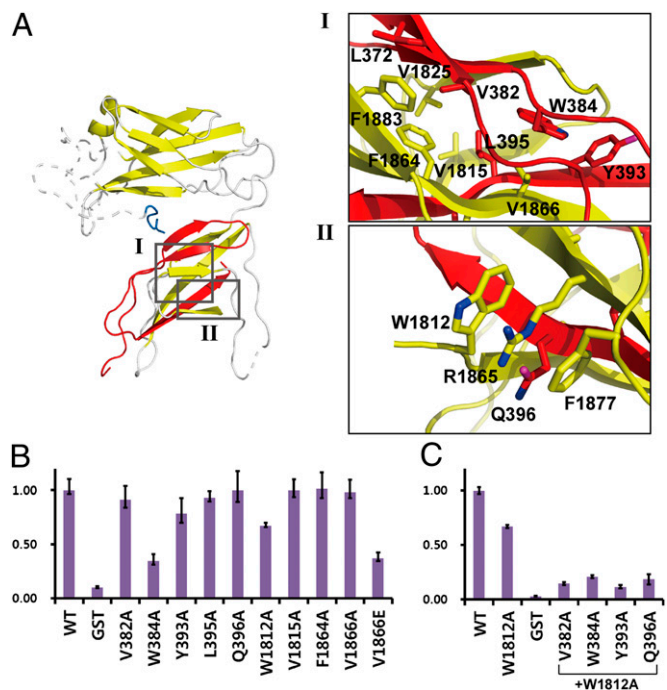

Fig. 3. Interactions between SAS1N and SAS1C. (A) Detailed views of the interactions between SAS1N and SAS1C. The several hydrophobic residues (Leu372, Val382, Trp384, Tyr393, and Leu395 from Fn3-1N, and Val1815, Val1825, Phe1864, Val1866, and Phe1883 from Fn3-1C) form a hydrophobic core at the associating interface (I). Highly conserved Trp (W1812), Arg (R1865), and Phe (F1877) residues of SAS1C stack onto each other, and wrap around the β3 strand of SAS1N. The side chain of Gln396 interacts with the guanidino group of Arg1865 to apparently stabilize the interaction between Trp1812 and Arg1865 (II). (B and C) Mutating residues involved in SAS1N/SAS1C interaction can affect the association between SAS1N and SAS1C–NLS. Residues involved in the interaction between SAS1N and SAS1C were mutated and SAS1N/SAS1C association assessed by GST-pull-down with GST–SAS1N. Amounts of bound SAS1C–NLS were measured in three replicate experiments.
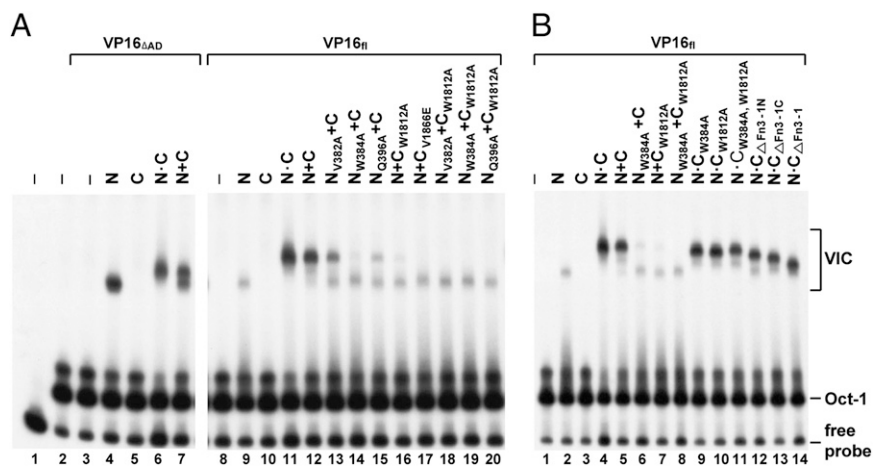
Fig. 4. The SAS1N/SAS1C association via Fn3-1 is critical for VP16-induced complex formation. (*A*) VIC formation assay with wild-type and SAS1N/SAS1C association surface mutants with DNA probe alone (lane 1) added Oct-1 POU DNA-binding domain (lanes 2–20), VP16$_{\Delta AD}$ (lanes 3–7), and VP16$_{fl}$ (lanes 8–20) proteins. HCF-1 proteins are N in lanes 4 and 9, C in lanes 5 and 10, N·C in lanes 6 and 11, N+C in lanes 7 and 12, and N+C mutants as indicated in lanes 13–20. N: HCF-1$_{Kelch–SAS1N}$, C: HCF-1$_{C\_SAS1C–NLS}$, N·C: a covalently linked HCF-1$_{Kelch–SAS1N}$ and HCF-1$_{C\_SAS1C–NLS}$ protein, N+C: cotranslated HCF-1$_{Kelch–SAS1N}$ and HCF-1$_{C\_SAS1C–NLS}$. (*B*) VIC formation assay of covalently linked SAS1N and SAS1C mutants with VP16$_{fl}$ and Oct-1 POU DNA-binding domain. HCF-1 proteins are none in lane 1, N in lane 2, C in lane 3, N·C in lane 4, N+C in lane 5, N+C mutants in lanes 6–8, and N·C mutants in lanes 9–14.

indicate that the sole function of Fn3-1 in VIC formation is in bringing the HCF-1$_N$ and HCF-1$_C$ subunits together, which is consistent with nearly all conserved residues in Fn3-1 being situated on the interface between Fn3-1N and Fn3-1C. Our data show that the association of HCF-1$_N$ and HCF-1$_C$ via an interdigitated Fn3-1 domain is critical to form a transcription regulatory complex, and that the principal function of Fn3-1 in this role is to hold the HCF-1$_N$ and HCF-1$_C$ subunits together.

Nuclear Localization Signal Promotes VP16-Induced Complex Formation.
The finding that HCF-1$_{C\_SAS1C–NLS}$ is required for VP16$_{fl}$ VIC formation and that the function of Fn3-1 is limited to holding the HCF-1$_N$ and HCF-1$_C$ subunits together suggests that the ability to promote VP16$_{fl}$ VIC formation resides in the Fn3-2–NLS segment. To dissect the contribution of Fn3-2–NLS to VIC formation, we generated a series of deletion constructs in the N·C fusion to exclude the possibility that the deletions destabilize HCF-1$_N$ and HCF-1$_C$ association and, as shown in Fig. 5*A* and summarized in Fig. 5*B*, assayed their activities in VIC formation with VP16$_{\Delta AD}$ (Fig. 5*A*, *Upper*) or VP16$_{fl}$ (Fig. 5*A*, *Lower*). First, we deleted the entire Fn3-2–NLS region (N·C$_{\Delta Fn3-2\Delta2003–2035}$, lane 4) or its

components: the Fn3-2 domain itself (N·C$_{\Delta Fn3-2}$, lane 2) or the region containing the C-terminal NLS (N·C$_{\Delta2003–2035}$, lane 3). Fn3-2–NLS or NLS region alone deletion, but not Fn3-2 deletion, interfered with VP16$_{fl}$ but not VP16$_{\Delta AD}$ VIC formation (Fig. 5*A Upper* and *Lower*, compare lanes 2–4), suggesting that the NLS region alone is involved in VP16$_{fl}$ VIC formation.

To assess the contribution of the NLS to VIC formation, we created N·C molecules (Fig. 5*B*) that lacked either the NLS (N·C$_{\Delta2014–2035}$, lane 8) or the region tethering the NLS to Fn3-2 (N·C$_{\Delta2003–2013}$, lane 5). Deleting the NLS but not the tethering sequence (Fig. 5*A*, compare lanes 5 and 8) prevented VP16$_{fl}$ VIC formation. These data suggest that the ability to form a VIC with VP16$_{fl}$ resides exclusively within the NLS segment. To dissect the NLS region further, we deleted the structured region (N·C$_{\Delta2014–2020}$, lane 6) and the rest (N·C$_{\Delta2021–2035}$, lane 7) of the NLS separately. Interestingly, in these two constructs, although the level of VP16$_{fl}$ VIC formation is decreased, the ability to form a VIC was not abolished, indicating that the N-terminal and C-terminal portions of the NLS function redundantly in VP16$_{fl}$ VIC formation (Fig. 5*A*, lanes 6 and 7).
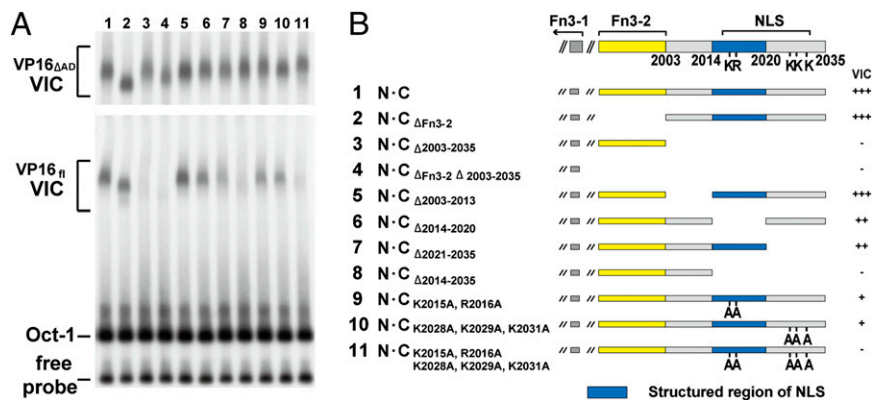
Fig. 5. The NLS is required for VP16-induced complex formation with full-length VP16. (*A*) VIC formation with Oct-1 POU DNA-binding domain, and VP16$_{\Delta AD}$ (*Upper*) or VP16$_{fl}$ (*Lower*) of a series of mutants of N·C near Fn3-2 and NLS. (*B*) A schematic illustrating mutant constructs and their VIC formation abilities (*Left*). Fn3-2 is shown in yellow and the structured NLS is shown in deep blue. Basic residues (K2015, R2016, K2028, K2029, and K2031) in the NLS are indicated as K or R, and their alanine mutations indicated as A.
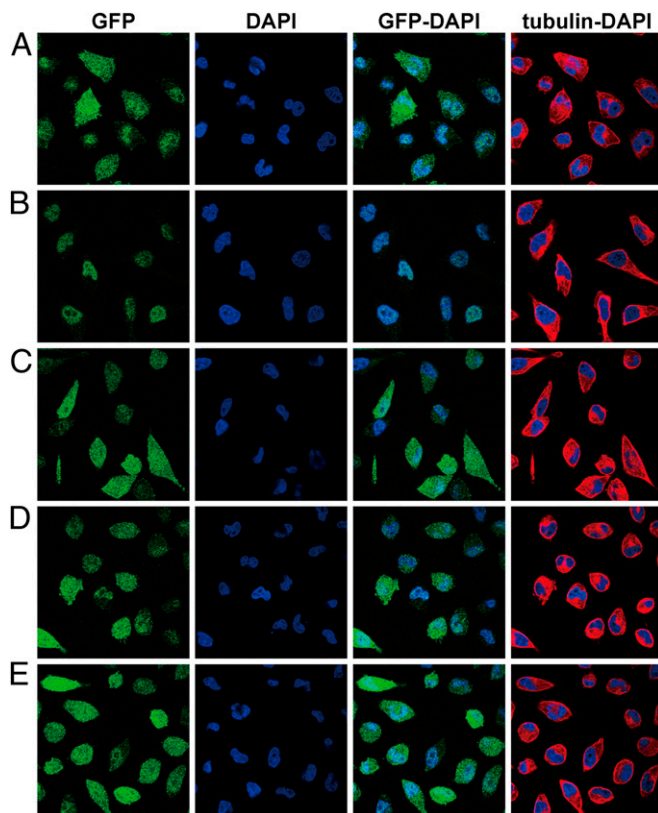
Fig. 6. Nuclear localization potential of the NLS. Nuclear localization of GST-GFP alone (*A*), GST-GFP-HCF-1_NLS (*B*), and the basic residue mutants: GST-GFP-HCF-1_NLS$_{K2015A/R2016A}$ (*C*), GST-GFP-HCF-1_NLS $_{K2028A/K2029A/K2031A}$ (*D*), and GST-GFP-HCF-1_NLS$_{K2015A/R2016A/K2028A/K2029A/K2031A}$ (*E*). GFP signal, DAPI staining, GFP-DAPI merged, and tubulin-DAPI merged are labeled above the figure sections.

The HCF-1 NLS (KRPMSSPEMKSAPKKSKADGQ, 2015–2035 amino acids) is predicted to be of a bipartite form (18) in which it is the N-terminal part (SKRPMSS, 2014–2020 amino acids) that is well-structured in one of the conformations in the crystal. Because the NLS is required for VIC formation only with VP16$_{fl}$ and the VP16 activation domain is highly acidic, the basic sequences within the NLS (Fig. S3*A*) could play a critical role in forming a VIC with VP16$_{fl}$. To test this hypothesis, we examined whether the basic residues in the NLS are critical for VIC formation.

We therefore generated three mutants in which the basic residues in the N-terminal (Lys2015, Arg2016) or C-terminal (Lys2028, Lys2029, Lys2031) parts of the NLS are mutated separately or in combination and measured the VIC formation activities with these mutants. Whereas the separate N-terminal or C-terminal NLS mutants displayed slightly decreased VIC formation (Fig. 5, lanes 9 and 10), the combined mutations strongly affected VIC formation (Fig. 5, lane 11). These results suggest that the two NLS basic regions contribute to VIC formation with VP16$_{fl}$.

We next asked whether these basic residues required for efficient VIC formation are also critical for the nuclear localization potential of the NLS. For this purpose, we fused the wild-type HCF-1_NLS (2003–2035 amino acids) or the mutant HCF-1_NLS (K2015A-R2016A, K2028A-K2029A-K2031A, and K2015A-R2016A-K2028A-K2029A-K2031A) to a GST-GFP fusion and examined subcellular localization after stable synthesis in human HeLa Flp-In cells as shown in Fig. 6. In this assay, GST-GFP was localized to the cytoplasm and nucleus (Fig. 6*A*), whereas the wild-type GST-GFP-NLS was localized to the nucleus alone (Fig. 6*B*). In contrast, the GST-GFP-NLS mutants displayed similar combined cytoplasmic/nuclear staining as wild-type GST-GFP (compare Fig. 6 *C–E* with *A*) indicating that the basic residues required for efficient VP16$_{fl}$ VIC formation are also critical for the nuclear localization. These results indicate that the HCF-1 NLS possesses a dual function in localizing HCF-1 to the nucleus and promoting VIC formation.

Discussion

This study shows that the HCF-1$_N$ and HCF-1$_C$ subunits associate by forming an integrated SAS1N–SAS1C (i.e., SAS1) structure composed of tandem Fn3 domains, one Fn3 being an integrated composite of SAS1N and SAS1C sequences. Furthermore, our studies reveal that this stable association permits the NLS to promote formation of a transcription regulatory complex: the VP16$_{fl}$-containing VIC.

Although the interaction between two proteins through "strand complementation" has been observed in other proteins (19), the SAS1 structure was unexpected because it had been hypothesized that the SAS1C region alone could form two Fn3 domains and that these would provide a dynamic surface for reversible interaction with SAS1N making it possible to regulate HCF-1$_N$ and HCF-1$_C$ subunit association (15). Given the evident stability of SAS1N–SAS1C structure—as evidenced, for example, by its resistance to numerous different alanine substitutions (Fig. 3)—we suggest that it is unlikely that the HCF-1$_N$ and HCF-1$_C$ subunits dissociate and reassociate once they form the single hybrid Fn3-1 domain, which we imagine occurs cotranslationally (or soon thereafter). Thus, SAS1 forms a stable structure, which, in those species in which HCF-1 proteins undergo proteolysis, probably stably holds the two generated subunits together.

This understanding leads us to a view in which a progenitor HCF-1 molecule containing what we imagine was a relatively common tandem Fn3 domain structure had residues added between the β3 and β4 strands of Fn3-1 (Fig. 1*E*) and that this addition led over time to the acquisition of multiple elements including the basic and acidic segments and proteolytic processing region as seen in human HCF-1 (Fig. 1*A*). Thus, HCF-1 subunit association probably did not evolve because there was proteolytic maturation, but rather that proteolytic maturation evolved because there was an existing structure—SAS1—to hold the two molecules resulting from proteolytic maturation together.

The stable intersubunit structure revealed here may also exist in other heterodimeric proteins generated by proteolysis (e.g., the mixed-lineage leukemia protein) (20). In combination, HCF-1 and the mixed-lineage leukemia protein apparently establish an evolutionarily attractive mechanism—formation of a single stable structural domain composed of two distant regions in a large precursor protein—for stable association of protein subunits generated by proteolysis.

Although the HCF-1$_N$ Kelch domain is sufficient to bind VP16 and to form a VIC with a VP16 molecule lacking its acidic activation domain, the HCF-1$_C$ subunit is additionally required for robust formation of a VIC with full-length VP16 (17). Unexpectedly, within HCF-1$_C$, it is the NLS region alone that plays an important role in forming a VIC with full-length VP16. The HCF-1 NLS is of a bipartite form such that there are two basic segments. Interestingly, each of these can promote full-length VP16 VIC formation (Fig. 5). Thus, it is a basic region of the HCF-1$_C$ subunit that promotes VIC formation when there is an acidic activation domain present in VP16. This observation suggests that basic patches in the NLS (but not necessarily the basic patch observed in the NLS-containing SAS1–NLS conformation; Fig. 1*D*) function to overcome an inhibitory effect of the acidic activation domain of VP16 in forming a VIC. Such an inhibitory effect could be repulsion between the acidic VP16 activation domain and the acidic DNA in a VIC. Whichever the mechanism, however, these results show that sequences within an NLS involved in nuclear localization can also possess a function other than nuclear localization.

Among the domains or regions comprising the SAS1–NLS structure presented here, a function for Fn3-1 is HCF-1$_N$ and HCF-1$_C$ association, and a function for the NLS is nuclear localization and formation of a transcription regulatory complex. In contrast, although there is a high degree of surface conservation in the Fn3-2 domain (Fig. 2) suggesting a function conserved in evolution, no mechanistic activity has been attributed to it. Given that other Fn3 domains are well known to be involved in protein–protein interactions, it may be involved in interacting with other HCF-1 binding partners (21).

The Fn3-1 and Fn3-2 domains are connected by a very highly conserved connecting loop (Figs. S3A and S4B). This high-sequence conservation suggests that the connecting loop does more than just connect the two domains. Given that there are two conformations of the Fn3-1 and Fn3-2 domains depending on the presence of an ordered NLS, it is possible that the connecting loop has a role in regulating the opening and closing of the two Fn3 domains on binding of other molecules such as VP16 to release the NLS to facilitate nuclear localization and/or transcription complex formation. Nevertheless, it cannot be ruled out that one or the other or both of the conformations observed is due to crystal packing forces.

The NLS-containing SAS1 structure has the NLS positioned between the Fn3-1 and Fn3-2 domains. We imagine that, in this configuration, the NLS would need to be repositioned to function and that such a repositioning could be achieved by a transition to the open Fn3-1/Fn3-2 conformation. It would be interesting to test whether an HCF-1–binding protein such as VP16 might induce the two Fn3 domains to open and thus release and activate the NLS for either VIC formation or HCF-1–dependent nuclear import of VP16 (18).

Although the molecular and cellular biology of HCF-1 has been extensively studied, no atomic resolution structure of HCF-1 has been reported. This work provides a structural view into the function and mechanisms of HCF-1.

Materials and Methods

SAS1–NLS was expressed in *E.coli*, and purified by Ni affinity, anion exchange, and gel-filtration chromatography. The crystals were obtained by the hanging-drop vapor diffusion method and the structure was solved using the SAD method. The VP16-induced complex formation was measured by electrophoretic mobility retardation assay. Nuclear localization was performed in HeLa cells stably transformed via Flpase-induced recombination. The detailed materials and methods are described in *SI Materials and Methods*.

1. Wysocka J, Herr W (2003) The herpes simplex virus VP16-induced complex: The makings of a regulatory switch. *Trends Biochem Sci* 28(6):294–304.
2. Knez J, Piluso D, Bilan P, Capone JP (2006) Host cell factor-1 and E2F4 interact via multiple determinants in each protein. *Mol Cell Biochem* 288(1-2):79–90.
3. Tyagi S, Chabes AL, Wysocka J, Herr W (2007) E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. *Mol Cell* 27(1):107–119.
4. Vogel JL, Kristie TM (2000) The novel coactivator C1 (HCF) coordinates multiprotein enhancer formation and mediates transcription activation by GABP. *EMBO J* 19(4):683–690.
5. Vercauteren K, Gleyzer N, Scarpulla RC (2008) PGC-1-related coactivator complexes with HCF-1 and NRF-2beta in mediating NRF-2(GABP)-dependent respiratory gene expression. *J Biol Chem* 283(18):12102–12111.
6. Yu H, et al. (2010) The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Mol Cell Biol* 30(21):5071–5085.
7. Dejosez M, et al. (2010) Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. *Genes Dev* 24(14):1479–1484.
8. Cai Y, et al. (2010) Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. *J Biol Chem* 285(7):4268–4272.
9. Smith ER, et al. (2005) A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol Cell Biol* 25(21):9175–9188.
10. Wysocka J, Myers MP, Laherty CD, Eisenman RN, Herr W (2003) Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes Dev* 17(7):896–911.
11. Yokoyama A, et al. (2004) Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol Cell Biol* 24(13):5639–5649.
12. Liang Y, Vogel JL, Narayanan A, Peng H, Kristie TM (2009) Inhibition of the histone demethylase LSD1 blocks alpha-herpesvirus lytic replication and reactivation from latency. *Nat Med* 15(11):1312–1317.
13. Capotosti F, et al. (2011) O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. *Cell* 144(3):376–388.
14. Julien E, Herr W (2003) Proteolytic processing is necessary to separate and ensure proper cell growth and cytokinesis functions of HCF-1. *EMBO J* 22(10):2360–2369.
15. Wilson AC, Boutros M, Johnson KM, Herr W (2000) HCF-1 amino- and carboxy-terminal subunit association through two separate sets of interaction modules: Involvement of fibronectin type 3 repeats. *Mol Cell Biol* 20(18):6721–6730.
16. Wilson AC, Freiman RN, Goto H, Nishimoto T, Herr W (1997) VP16 targets an amino-terminal domain of HCF involved in cell cycle progression. *Mol Cell Biol* 17(10):6139–6146.
17. LaBoissière S, Walker S, O'Hare P (1997) Concerted activity of host cell factor subregions in promoting stable VP16 complex assembly and preventing interference by the acidic activation domain. *Mol Cell Biol* 17(12):7108–7118.
18. La Boissière S, Hughes T, O'Hare P (1999) HCF-dependent nuclear import of VP16. *EMBO J* 18(2):480–489.
19. Puorger C, et al. (2008) Infinite kinetic stability against dissociation of supramolecular protein complexes through donor strand complementation. *Structure* 16(4):631–642.
20. García-Alai MM, Allen MD, Joerger AC, Bycroft M (2010) The structure of the FYR domain of transforming growth factor beta regulator 1. *Protein Sci* 19(7):1432–1438.
21. Scarr RB, Sharp PA (2002) PDCD2 is a negative regulator of HCF-1 (C1). *Oncogene* 21(34):5245–5254.

BIOCHEMISTRY