Molecular basis of host-adaptation interactions between influenza virus polymerase PB2 subunit and ANP32A

Aldo R. Camacho-Zarco1, Sissy Kalayil2, Damien Maurin1, Nicola Salvi1, Elise Delaforge1, Sigrid Milles1, Malene Ringkjøbing Jensen1, Darren J. Hart1, Stephen Cusack2 & Martin Blackledge1

Avian influenza polymerase undergoes host adaptation in order to efficiently replicate in human cells. Adaptive mutants are localised on the C-terminal (627-NLS) domains of the PB2 subunit. In particular, mutation of PB2 residue 627 from E to K rescues polymerase activity in mammalian cells. A host transcription regulator ANP32A, comprising a long C-terminal intrinsically disordered domain (IDD), is responsible for this adaptation. Human ANP32A IDD lacks a 33 residue insertion compared to avian ANP32A, and this deletion restricts avian influenza polymerase activity. We used NMR to determine conformational ensembles of E627 and K627 forms of 627-NLS of PB2 in complex with avian and human ANP32A. Human ANP32A IDD transiently binds to the 627 domain, exploiting multivalency to maximise affinity. E627 interrupts the polyvalency of the interaction, an effect compensated by an avian-unique motif in the IDD. The observed binding mode is maintained in the context of heterotrimeric influenza polymerase, placing ANP32A in the immediate vicinity of known host-adaptive PB2 mutants.
Influenza A virus (IAV) is responsible for 3–5 million severe cases every year, resulting in 250–500,000 deaths. Most influenza strains evolve exclusively in the large reservoir of water birds, but some highly pathogenic avian strains (e.g., H5N1, H5N8 and H7N9) can infect humans with lethal consequences (up to 60% mortality) and are potential pandemic threats for humanity if they develop human-to-human transmissibility. However, for these avian (av) viruses to efficiently replicate in mammalian cells, host adaptation of the viral polymerase is necessary. Replication of IAV is carried out by the RNA-dependent RNA viral polymerase that functions as a heterotrimeric complex, formed from separate components PA, PB1, and PB2. Few mutations are required for avian to human adaptation (up to 60% mortality) and are potential pandemic threats for water birds, but some highly pathogenic avian strains (e.g., H5N1, H7N9) can infect humans with lethal consequences.

Importantly, we show that the interaction between ANP32A and 627-NLS, the complex is polyvalent, and in all cases weak interaction sites until the nuclear localisation sequence (KRKR), situated to 78% and 58%, respectively, revealing that both ANP32A:627-NLS(E) and ANP32A:627-NLS(K) are indeed both intrinsically disordered (Supplementary Table 1, Supplementary Fig. 3), with none of the resonance in the IDD of ANP32A IDD. Backbone resonance assignment was completed and their interactions with 627-NLS of ANP32A and their interactions with 627-NLS differ principally due to a 33 amino acid insert in avANP32A (176–209) comprising an av-unique hexapeptide,176VLSLVK181, followed by a duplication of 27 amino acids also present in hANP32A IDD. Backbone resonance assignment was completed to 78% and 58%, respectively, revealing that h and avANP32A IDs are indeed both intrinsically disordered (Supplementary Figs. 1 and 2), with a slight tendency (20%) towards helical conformation for the hexapeptide, but negligible secondary structural tendency elsewhere.

Upon addition of the 627(K) domain, 1H and 15N chemical shift perturbations (CSPs) are seen for a large number of resonances in the IDD of hANP32A (Fig. 1b, Supplementary Fig. 1d). NMR relaxation rates measured at increasing titration mixtures (Fig. 1c) show maximal effects for the acidic strand 180DEDA183, while the largest CSPs are seen for the adjacent hydrophobic residues 184QVV186 (Fig. 1b, Supplementary Fig. 1d). Additional interactions are seen throughout the chain, in particular at 164YV165 and 214YND216. Comparison of backbone 13C shifts in the free and fully bound states reveals that only 179YDE182 shows any evidence of folding upon binding (Supplementary Fig. 1d), in this case into an extended β-sheet conformation, while the remainder of the chain remains highly flexible in the complex, retaining its random coil nature (Supplementary Fig. 1b). Similar evidence of multiple interaction sites is seen for the IDD of avANP32A in complex with the 627(E) domain (Supplementary Fig. 1d, Fig. 1d). Relaxation properties (Fig. 1c, h) and more specifically CSPs (Fig. 1e) of the hexapeptide are strongly influenced by the presence of both domains of 627-NLS, compared to constructs each comprising only the 627 or NLS domains and the linker. This implies that the specific conformational behaviour of the linker in integral 627-NLS, or the relative position of the two domains, are essential for the interaction with the hexapeptide of avANP32A IDD.

Interaction of 627-NLS and ANP32A reveals similar profiles to those measured for the IDs alone, with additional interactions involving the LRR for both h and avANP32A (Fig. 1f–i). In both h and avANP32A, the largest CSPs in the LRR are centred on 120LFN122, with additional shifts induced following the spine of the beta-helix (Fig. 1g). Notably, the cross-interaction, between hANP32A and 627-NLS(E) shows much smaller shifts than hANP32A:627-NLS(K), while comparison with avANP32A:627-NLS(E) clearly identifies increased shifts centred on the hexapeptide (Fig. 1i).

Although most of the IDD is involved in the interaction, the strongest binding or highest populations of binding interactions, occur at a distance of 25–35 amino acids from the LRR. In the case of avANP32A, this concerns the avian-unique hydrophobic hexapeptide, and in hANP32A the sequence 180DEDAQVV186. Further interactions are observed downstream of these interaction sites until the nuclear localisation sequence (KKKR), situated ~15 amino acids from the end of the chain, beyond that point no significant interactions are observed.

Despite the evidence of clear interaction between ANP32A and 627-NLS, the complex is polyvalent, and in all cases weak (Supplementary Table 1, Supplementary Fig. 3), with none of the individual interactions exhibiting a stronger affinity than 800 μM

Results

IDDs of h and avANP32A and their interactions with 627-NLS. We have compared the two complexes using solution state NMR, initially from the side of ANP32A. The IDs of h and avANP32A comprise 63/96 and 79/129 Asp or Glu residues, respectively, leading to extensive spectral overlap. h and avANP32A IDs differ principally due to a 33 amino acid insert in avANP32A (176–209) comprising an av-unique hexapeptide, 176VLSLVK181, followed by a duplication of 27 amino acids also present in hANP32A IDD. Backbone resonance assignment was completed to 78% and 58%, respectively, revealing that h and avANP32A IDs are indeed both intrinsically disordered (Supplementary Figs. 1 and 2), with a slight tendency (20%) towards helical conformation for the hexapeptide, but negligible secondary structural tendency elsewhere.

Upon addition of the 627(K) domain, 1H and 15N chemical shift perturbations (CSPs) are seen for a large number of resonances in the IDD of hANP32A (Fig. 1b, Supplementary Fig. 1d). NMR relaxation rates measured at increasing titration mixtures (Fig. 1c) show maximal effects for the acidic strand 180DEDA183, while the largest CSPs are seen for the adjacent hydrophobic residues 184QVV186 (Fig. 1b, Supplementary Fig. 1d). Additional interactions are seen throughout the chain, in particular at 164YV165 and 214YND216. Comparison of backbone 13C shifts in the free and fully bound states reveals that only 179YDE182 shows any evidence of folding upon binding (Supplementary Fig. 1d), in this case into an extended β-sheet conformation, while the remainder of the chain remains highly flexible in the complex, retaining its random coil nature (Supplementary Fig. 1b). Similar evidence of multiple interaction sites is seen for the IDD of avANP32A in complex with the 627(E) domain (Supplementary Fig. 1d, Fig. 1d). Relaxation properties (Fig. 1c, h) and more specifically CSPs (Fig. 1e) of the hexapeptide are strongly influenced by the presence of both domains of 627-NLS, compared to constructs each comprising only the 627 or NLS domains and the linker. This implies that the specific conformational behaviour of the linker in integral 627-NLS, or the relative position of the two domains, are essential for the interaction with the hexapeptide of avANP32A IDD.

Interaction of 627-NLS and ANP32A reveals similar profiles to those measured for the IDs alone, with additional interactions involving the LRR for both h and avANP32A (Fig. 1f–i). In both h and avANP32A, the largest CSPs in the LRR are centred on 120LFN122, with additional shifts induced following the spine of the beta-helix (Fig. 1g). Notably, the cross-interaction, between hANP32A and 627-NLS(E) shows much smaller shifts than hANP32A:627-NLS(K), while comparison with avANP32A:627-NLS(E) clearly identifies increased shifts centred on the hexapeptide (Fig. 1i).

Although most of the IDD is involved in the interaction, the strongest binding or highest populations of binding interactions, occur at a distance of 25–35 amino acids from the LRR. In the case of avANP32A, this concerns the avian-unique hydrophobic hexapeptide, and in hANP32A the sequence 180DEDAQVV186. Further interactions are observed downstream of these interaction sites until the nuclear localisation sequence (KKKR), situated ~15 amino acids from the end of the chain, beyond that point no significant interactions are observed.

Despite the evidence of clear interaction between ANP32A and 627-NLS, the complex is polyvalent, and in all cases weak (Supplementary Table 1, Supplementary Fig. 3), with none of the individual interactions exhibiting a stronger affinity than 800 μM.
**Fig. 1 NMR of ANP32A with 627-NLS reveals a highly dynamic polyvalent complex.**

**a** Representation of the domains of ANP32A. Blue surface shows the LRR domain, pale blue ribbon-disordered IDD (100 or 132 amino acids). Sequence distribution of IDDs shown on right. Red = acidic, blue = basic and beige = hydrophobic or polar. AvIDD contains a 33 amino acids insert (horizontal line) comprising an avian-unique hexapeptide (H).

**b** Chemical shift titration of 627(K) into 15N-labelled IDD of hANP32A. hANP32A was 25 μM throughout. 627(K) concentrations were 0 (blue), 25 (green), 50 (grey-blue), 100 (orange) and 200 (red) μM. Measurement at 850 MHz, 293 K (see Methods section).

**c** 15N rotating frame relaxation ($R_1\rho$) of free (blue) 15N-labelled IDD of hANP32A at 300 μM, and upon mixing with 627(K) at 1:1 (orange) and 1:4 (red) ratio. Measurement at 850 MHz, 293 K.

**d** 15N rotating frame relaxation ($R_1\rho$) of free (blue) 15N-labelled IDD of avANP32A at 300 μM, and upon mixing with 627(E) at 1:1 (orange) and 1:4 (red) ratio. Maximum effect is seen at the equivalent motif, 33 amino acids downstream, from hANP32A.

**e** Chemical shift perturbation (CSP) of full-length avANP32A upon addition of 627(E) (orange), NLS (red) and 627-NLS(E) (blue). Concentration of avANP32A was 25 μM and the partner 50 μM throughout. Resonances from the LRR were not observable at this low concentration.

**f** 15N rotating frame relaxation ($R_1\rho$) of free (red) 15N-labelled full-length hANP32A at 300 μM, and upon mixing with 627-NLS(K) at 1:1 ratio (blue). Measurement at 850 MHz, 293 K. g CSP of full-length hANP32A upon addition of 627-NLS(K) (blue). Concentration of hANP32A was 300 μM and 627-NLS(K) 600 μM. The cross-interaction between hANP32A and 627-NLS(K) is shown for comparison at the same stoichiometric ratio (orange). h 15N rotating frame relaxation ($R_1\rho$) of free (red) 15N-labelled full-length avANP32A at 300 μM, and upon mixing with 627-NLS(E) at 1:1 ratio (blue). Measurement at 850 MHz, 293 K. i CSP of full-length avANP32A upon addition of 627-NLS(E) (blue). Concentration of avANP32A was 300 μM and 627-NLS(E) 600 μM. The cross-interaction between hANP32A and 627-NLS(E) is shown for comparison at the same stoichiometric ratio (orange). The avian-specific insert is highlighted by shifting data from hANP32A from residues following the beginning of the insert by 33 amino acids.
for the interaction of hANP32A with 627(K) (1700 μM for αvANP32A with 627(E)). Notably, the interactions between hANP32A and 627(E) are also weaker (>1400 μM), suggesting that the single E627K mutation plays an important role. Interaction with integral 627-NLS is weaker due to the open-closed equilibrium reducing the population of available binding states. Interaction of the LRR of ANP32A alone with the 627 domain reveals weaker affinity (>1500 μM). Although these interactions are weak, the extended interaction surface involving 80–100 disordered amino acids in h and αvANP32A nevertheless results in tighter binding as experienced by 627-NLS.

Interactions of αv and h-adapted 627-NLS with αv and hANP32A. The interaction was also investigated from the side of 627-NLS domains of PB2. The two domains exhibit an open-closed equilibrium (Fig. 2a) that is populated ~40:60 at 293 K leading to two sets of resonances for the majority of the protein. Addition of hANP32A to 627-NLS(K) resulted in CSPs throughout the protein (Supplementary Fig. 4). The largest shifts are observed in the open form, suggesting that ANP32A interacts preferentially with this conformation. The closed form of 627-NLS(K) is stabilised by a tripartite salt bridge, and can be removed from the equilibrium by mutation of the implicated amino acids (R650, D730/E687)27. CSPs induced upon interaction with the ANP32A IDD are illustrated for clarity using an open-only mutant (Fig. 2b), and the distribution of CSPs as measured on the wild-type proteins (Fig. 2c). Figure 2d illustrates the major shifts observed for 627-NLS(K) upon addition of hANP32A IDD. Although basic sidechains are found on both sides of the domain (Fig. 2e), CSPs are observed mainly on one face of the 627 domain, again suggesting that the interaction is not uniquely driven by electrostatic attraction with the acidic IDD. Despite strong similarity of the CSP profiles, clear additional shifts are measurable for 627-NLS(K):hANP32A compared to 627-NLS(E):αvANP32A (Fig. 2c, Supplementary Fig. 4), particularly in the vicinity of R630, L631 and 587–591, forming a continuous interaction surface that is strongly enhanced in 627-NLS(K) (Fig. 2f). It is interesting to note that residues 590 and 591 provide an alternative pathway to host adaptation, as evidenced by the comparison of the CSP-ensemble of conformers that best fit the experimental data. The experimental results show strong PREs, reporting on weak, or sparsely populated contacts between the two proteins, providing long-range positional constraints that are complementary to the short-range modulation of the dynamics of each site as measured by spin relaxation. These orthogonal experimental probes reveal different aspects of the same interfaces, while providing a high level of confirmatory experimental validation (vide infra).

The experimental results show strong PREs, reporting on tighter or more populated interactions, distributed over long stretches of the IDD for certain spin-label positions, and little broadening for other positions. Such profiles again suggest a polyvalent interaction, whereby distinct sites dispersed along the IDD visit the same sites on 627-NLS. Interestingly, some labels (in particular 587 and 631) induce a well-defined pattern over the long β-sheet on the concave face of the LRR of ANP32A, allowing the determination of its orientation with respect to the 627 domain (Fig. 3d, see ‘Methods’ section), despite the weak interaction between the two domains (Supplementary Table 1). This orientation is in full agreement with the observed CSP on the surface of the LRR of ANP32A (Fig. 1), where the largest chemical shifts are seen for F121, Y122 and C123. These residues are positioned in closest proximity to the 627 domain in the ensemble of conformers that best fit the entire set of experimental data. The interaction is apparently stabilised by hydrophobic interactions, involving residues on the surface of 627 and ANP32A, and an electrostatic interaction between D130 (ANP32A) and R646 (627) (Fig. 3e). It seems likely that the observed interaction relates to observations that have recently implicated D130 in host adaptation20,21,23. The optimal poses determined for 627-NLS(K):hANP32A and 627-NLS(E):αvANP32A complexes are very similar. The interface between 627 and ANP32A is bordered by a nearly continuous ridge of solvent accessible basic sidechains, including K589, R630, R641
and R650, that is completed by the presence of K627 in the case of 627-NLS(K) (Fig. 3f). Inspection of the PRE data from the IDD regions of h and avANP32A reveals weaker effects in the region immediately following the LRR for avANP32A, and more contacts with the dislocated NLS domain (Fig. 3a, b), in particular

the spin-label at position 699 in the NLS broadens the IDD maximally at the position of the hexapeptide of avANP32A. This again confirms that the interaction with the immediately proximal basic face of 627, defined by the differential CSPs (Fig. 2f), is weaker in the case of avANP32A.
**Fig. 2 Interaction of 627-NLS and ANP32A observed from the perspective of 627-NLS.** a Representation of the domains of 627-NLS. The C-terminal domains of PB2 comprise two sub-domains (627—orange and NLS—grey) connected by a linker (red) and terminated by an NLS peptide (NLS). The molecule exists in equilibrium between closed and open forms that exchange at 50 s$^{-1}$ at room temperature. The position of the E627K adative mutation is shown in blue. b HSQC of the open-only form of 2D, 13C and 15N-labelled 627-NLS(K) (D730A/E687A; 200 μM; red), indicating chemical shifts of selected sites upon addition of hANP32A IDD (400 μM blue). c CSP of 627-NLS(L) (300 μM) upon addition of hANP32A IDD at a ratio of 1:4 (red) and 627-NLS(E) (300 μM) upon addition of avANP32A IDD at a ratio of 1:4 (blue). Spectra recorded on 13D, 13C and 15N-labelled 627-NLS at 293 K and 850 MHz. Only shifts affecting resonances corresponding to the open form are shown for clarity. d CSP of 627-NLS(K) (red) induced by addition of hANP32A derived from b above the threshold of 0.25 (dashed line) for 627 (grey) and NLS (yellow) domains. Two orientations of the protein are shown. e Distribution of basic sidechains (blue) on the surface of the 627 and NLS domains. By comparison with f, it is evident that one of the two faces preferentially interacts with the IDD. f Representation of the largest differential shifts on the surface of the 627 domain upon addition of ANP32A as shown in c. Amino acids showing the largest difference in CSP (>0.15) in the 627-NLS(K):hANP32A IDD complex compared to the 627-NLS(E):avANP32A IDD are highlighted in red (position of 627 shown in blue). g Interaction of full-length 627-NLS with ANP32A induces a change in the open-closed equilibrium. Red—peaks reporting on the open and closed forms of 627-NLS(E) or 627-NLS(K) before the addition (blue) of h or avANP32A, respectively. The interaction with ANP32A potentiates the equilibrium in both cases, fully removing the closed form in the case of the avian pair. Residues distal from the main interaction site were chosen to reduce the risk of peak disappearance due to direct interaction.

**Fig. 3 Experimental characterisation of the dynamic 627-NLS:ANP32A interaction complexes.** a Experimental (orange) and calculated (blue) paramagnetic relaxation enhancements measured on hANP32A in the presence of paramagnetically labelled 627-NLS. Intensity ratios compare spectra recorded in the presence of oxidised and reduced forms of TEMPO-maleimide for complex admixtures of hANP32A (300 μM) and 627-NLS(K) (150 μM). Calculated values result from representative ensembles selected using the ASTEROIDS approach. b Experimental paramagnetic relaxation enhancements measured on avANP32A in the presence of paramagnetically labelled 627-NLS. Intensity ratios compare spectra recorded in the presence of oxidised and reduced forms of TEMPO-maleimide for complex admixtures of avANP32A (220 μM) and av627-NLS (400 μM). This admixture was chosen to replicate the population of bound state estimated in the case of 627-NLS(K):hANP32A. c Position of the eight cysteine mutations used to label 627-NLS. Five mutations were selected over the surface of the 627 domain and three on NLS. One mutant protein was expressed and purified for each site on both 627-NLS(K) and 627-NLS(E), and labelled with TEMPO-maleimide (see ‘Methods’ section). d Optimisation of the position of the folded domains of ANP32A and 627-NLS(K). Two thousand starting poses generated by the programme Haddock were optimised with respect to experimental PREs measured on the folded domain. The ten best fitting poses are shown. The same procedure was used for PREs measured on the avANP32A: 627-NLS(E) complex, resulting in similar best fitting structures. e Localisation of intermolecular interactions possibly stabilising the interface between the two proteins, involving hydrophobic (F121 and Y144 in ANP32A, and Y592 and F633 in 627-NLS) and electrostatic interactions (D130 in ANP32A and R646 in 627-NLS). f Position of positively charged ridge of solvent-exposed basic sidechains in the vicinity of the interface between the folded domains.

Ensemble analysis of the PRE data using the ASTEROIDS approach$^{31}$, accounting for the flexibility in the 627-NLS linker and IDD domains of ANP32A (Supplementary Fig. 8), allows us to propose a molecular description of the conformational sampling of the entire complex. Representative ensembles of conformations of the multi-domain complex (Fig. 4) reveal the sampling of conformational space of the IDD and NLS domains relative to the position of 627 and the LRR of ANP32A. Comparison of the position of the IDDs over the ensembles (Fig. 4d, e) reveals more restricted sampling for hANP32A, localising residues 175−200 in the vicinity of the basic ridge on the surface of 627. Sampling of avANP32A is more dispersed.
Fig. 4 Comparison of 627-NLS(K):hANP32A and 627-NLS(E):avANP32A interaction complexes. a Average distance difference matrix, showing the average difference \( d_{hh} - d_{avav} \) in distance between amino acid positions in the IDD of ANP32A (x-axis) and the 627-NLS domains (y-axis) over the two ensembles. The colour code shown on the right is measured in Å.

b Representation of the key interactions between hANP32A and 627-NLS(K). Polyvalent interactions between 627 and the IDD localise the disordered domain in the vicinity of the basic patch on the surface of 627 (see Fig. 3). The different IDD chains are shown to represent different binding modes and are truncated at residue 200 for clarity. Red positions in the IDD indicate the acidic sidechains. Note that this representation illustrates the tendency over the entire ensemble that is highly disperse (see d and e).

c Representation of the key interactions between avANP32A and 627-NLS(E). In this case, the average distance between the IDD and the surface of 627 is larger, but in general closer to the NLS domain, in particular the hexapeptide of ANP32A, and the linker between 627 and NLS. The two IDD chains represent reduced polyvalency compared to hANP32A and 627-NLS(K), and are again truncated at residue 200 for clarity.

d Representative ensemble of conformations describing the conformational space sampled by the hANP32A:627-NLS(K) complex. The NLS domain has been removed for clarity, and the IDD truncated at residue 210. The position of residue 190 is indicated as a red sphere.

e Representative ensemble of conformations describing the conformational space sampled by the avANP32A:627-NLS(E) complex. Representation as in d.

(Fig. 4b–e), as shown quantitatively in the average distance map (Fig. 4a), that identifies closer contacts in the 627-NLS(K):hANP32A complex for residues 160, 180 and 205 with 570, 586 and 627, and closer contacts between the linker and NLS regions, and avANP32A region 150–200 (Fig. 4a).

Importantly, comparison of PRE profiles of hANP32A:627-NLS(E) demonstrates that the cross-interaction, that represents the case encountered when a non-adapted avian IAV infects human cells, shows less extensive and in general weaker contacts compared to hANP32A:627-NLS(K) and avANP32A:627-NLS(E) (Fig. 5). The absence of K627 is again seen to diminish the numerous polyvalent contacts present in hANP32A:627-NLS(K) as illustrated by the profiles induced by probes attached to 587 and 631, while the shorter IDD and lack of the hexapeptide abrogates the extensive interaction surface and the linker-specific contact present in avANP32A:627-NLS(E) (compare profiles induced by probe attached at 699). The required components for interaction modes specific to either 627-NLS(K):hANP32A or 627-NLS(E):avANP32A are therefore both weakened when avian polymerase interacts with hANP32A.

hANP32A:627-NLS in the context of integral FluB polymerase. To determine whether the interactions characterised here are relevant in the context of a human-adapted full-length influenza polymerase, we initially repeated the chemical shift titrations using the heterotrimeric influenza B (FluB) polymerase (bound to vRNA), with the IDD of hANP32A (Fig. 6a). The interaction sites...
Given the spectroscopic similarities of the interaction of ANP32A are closest to the LRR domain due to slow tumbling in solution. Particle results in extreme line broadening for the residues that are found to be the same as for 627-NLS from IAV (Fig. 6b). In this study, we describe and compare the molecular complexes formed by the human-adapted or avian-adapted 627-NLS domains with the respective ANP32A host proteins, in order to understand the nature and specificity of these interactions. All of the implicated proteins exhibit extensive intrinsic disorder. The elaboration of atomic resolution descriptions of such highly disordered complexes requires methodologies that can account for the ensemble of structures sampled by the IDDs of av and hANP32A within the 627-NLS:ANP32A complexes, indicating a broader capture radius for ANP32A. Figure 6c, f illustrates the expected conformational space sampled by the IDDs of av and hANP32A within the 627-NLS:ANP32A complexes, indicating a broader capture radius for the avian complex. A similar procedure was applied to the ensemble of conformationally very distinct transcriptionally inactive IAV polymerase bound to the cRNA 5′ terminus, where the 627 domain sits on the surface, and is displaced and rotated relative to the polymerase core by ~70°. In this case, the 627-NLS:ANP32A complex can be easily accommodated (Fig. 6g, h). The position of NLS on the surface of the polymerase, and its observed positional variability in existing structures, suggests that the open form of 627-NLS is sampled in both transcriptionally active and inactive polymerases. Finally, the recently determined dimeric structure of apo IAV polymerase accommodates the ensemble equally well (Supplementary Fig. 9).

Discussion
In this study, we describe and compare the molecular complexes formed by the human-adapted or avian-adapted 627-NLS domains with the respective ANP32A host proteins, in order to understand the nature and specificity of these interactions. All of the implicated proteins exhibit extensive intrinsic disorder. The elaboration of atomic resolution descriptions of such highly disordered complexes requires methodologies that can account for the ensemble of structures sampled by the two proteins in solution. Investigation of the hANP32A:627-NLS(K) and hANP32A:627-NLS(E) complexes using complementary structural data from NMR chemical shifts, spin relaxation and paramagnetic relaxation combined with quantitative ensemble modelling reveals the existence of highly dynamic molecular assemblies that exhibit very different interaction modes.

The LRR of ANP32A interfaces with the 627 domain of PB2 at the C-terminal end of its concave surface, apparently stabilised via hydrophobic and electrostatic interactions. This interaction is
intrinsically very weak, but experimental NMR, including CSP and PREs, accurately report on the relative position of the two domains. The disordered domain interacts transiently with a basic patch on the surface of the 627 domain, preferentially in the case of hANP32A:627-NLS(K), involving weak, polyvalent interactions of acidic and hydrophobic residues of the IDD. The presence of multiple low-affinity interactions (in the millimolar range individually) distributed throughout 80 amino acids of the...
IDD of hANP32A results in an effective increase in affinity to ~50 μM for 627-NLS(K), an effect of avidity that has been observed in a number of systems exhibiting extensive intrinsic disorder38,39. The critical E627K mutation completes a continuous ridge of solvent-exposed positively charged residues6 that are available for interaction with the highly dynamic acidic IDD. This ridge colocalizes interacting residues of the IDD in the vicinity of the surface, that differentially implicates residues 589 and 591, 629, 631, 635 and 637 when the E627K mutation is present. By contrast the IDD of avANP32A interacting with 627-NLS(E) populates fewer conformations in the vicinity of the surface of 627. This lack of interaction, due to the interruption of the positively charged surface by the presence of E627, is compensated by an even broader conformational sampling of the IDD that exploits a more extensive interaction surface, implicating the hexapeptide motif specific to avANP32A and the NLS domain, and in particular the linker region of 627-NLS. It is again interesting to note that two adaptive mutations (V683T and A684S)40 have been identified in this linker region. By binding predominantly to the open form of 627-NLS, ANP32A potentiates the equilibrium of open and closed forms of 627-NLS, an effect that is more efficient in the case of the 627-NLS(E):avANP32A interaction, likely as a result of the interaction of the hexapeptide. Investigating the interaction of individual peptides isolated from the IDD, we are able to confirm the multivalent nature of the interaction of the hIDD with 627-NLS, and to strongly substantiate the differential binding modes of the two complexes.

Notably, the ‘cross-interaction’ between hANP32A and 627-NLS (E) exhibits neither the stabilisation properties mediated by the aIDD in the avANP32A:627-NLS(E) complex, nor the polyvalent binding specific to 627(K) as observed in the hANP32A:627-NLS (K) interaction. This lack of adapted molecular mechanisms results in fewer and weaker contacts between human ANP32A and avian-adapted 627-NLS. In combination, these effects may explain the inefficiency of avian polymerase in human cells in the absence of avANP32A or 627-NLS(K).

It is interesting to compare the interaction results measured by different NMR-based techniques with existing studies, where results measured using different techniques seem to paint a slightly different picture. While the individual interaction sites reported in our study indicate a weaker binding between avANP32A and 627-NLS (E) as compared to hANP32A and 627-NLS(K), some biochemical and cell biology studies report on a higher affinity of avANP32A to avian-adapted polymerases17,24. This apparent contradiction may result from different tagging techniques used in the reported pull-down assays, but we also note that overall affinities between proteins are not necessarily comparable to individual multivalent interactions in terms of affinities, and it will be interesting to resolve the dependencies in the future.

Importantly, NMR indicates that the complex characterised for the minimal 627-NLS:ANP32A interaction is maintained in the context of the integral FluB polymerase in its transcriptionally active form, showing very similar NMR-binding characteristics to hANP32A. Superposition of the binding pose of ANP32A with respect to 627 onto the 627 domain in the associated transcriptionally active polymerase structure26 would place the LRR domain in a similarly dimensioned pocket bounded by PB2 domains (Fig. 6d). Intriguingly, recently characterised host-adaptive PB2 mutants33 lie in the immediate vicinity of the surface of the folded domain of ANP32A in this binding conformation. Influenza polymerase exhibits extensive plasticity in solution, as demonstrated by recent electron microscopy and X-ray crystallographic studies describing multiple states of the polymerase26,28,29,34,35,41,42, and it appears highly likely that many of these states will be in conformational exchange in solution. It is also possible that the highly dynamic and transient nature of the interaction with ANP32A enables or enhances exchange between these forms during the viral cycle. It is therefore of interest to investigate the possible impact of our observed binding mode on other known states of influenza polymerase. The proximity of ANP32A IDD to the interface between the 627 and mid-domains of PB2, both of which undergo large-scale reorientations and dislocation between apo-, rRNA-bound and vRNA-bound polymerase, also raises the possibility that ANP32A interaction is associated with these conformational changes. In the transcriptionally inactive structure, the 627 domain appears almost dislocated from the core, a conformation that can easily accommodate the binding pose of ANP32A determined for the minimal complex.

The information contained in our ensemble descriptions of the ANP32A:627-NLS interactions therefore allow us to speculate further on its role in viral function. Recent observations have established that dimerisation of influenza polymerase is essential for the initiation of vRNA synthesis during replication35. The apo-polymerase structure of IAV polymerase that was recently solved in its dimeric form is also found to be compatible with the binding pose determined here (Supplementary Fig. 9), again with the 627 domain dislocated and sitting on the surface of the catalytic core domains. It has been suggested that ANP32A plays a role in assembly or regulation of this dimerisation process, for example, by recruitment of a second apo `packaging' polymerase to a replicating polymerase to initiate formation of the progeny viral RNP43. In this context, the polyvalent nature of the interaction between ANP32A and 627-NLS may be of functional relevance, allowing for more than one polymerase to simultaneously bind to ANP32A thereby colocalizing two polymersases to facilitate viral replication. In all of these aspects, the more extensive effective capture radius of the IDD of avANP32A may be important.

It is known that the intrinsically disordered, phosphorylated CTD of host RNA polymerase II (Pol II) binds to the surface of PA of influenza polymerase to facilitate the cap-snatching mechanism44,45. Given the highly negatively charged nature of the IDD of ANP32A it seems possible that it may play a regulatory role in this interaction, for example, by competing with the phosphorylated CTD to inhibit the interaction with Pol II. Notably, the extremely long IDDs, allied to the fact that ANP32A binds to the phosphorylated CTD to inhibit the interaction with Pol II. In all of these aspects, the more extensive effective capture radius of the IDD of avANP32A may be important.

In summary, the description of these highly dynamic species-specific assemblies reveals unique mechanistic insight into the role of the ANP32 family in host adaptation of avian influenza polymerase to the human cells, and provides a molecular framework for understanding the considerable volume of experimental observation measured on this complex system, as well as informing the identification of novel targets for IAV inhibition.

Methods

Constructs. A codon optimised 627-NLS construct from PB2 subunit was synthesised encoding amino acids 538–759 from avian H5N1 A/duck/Shantou/4610/2003 for expression in Escherichia coli (Geneart, Regensburg, Germany)27. In addition, constructs containing just the 627 domain (amino acids 538–693) or the NLS domain (amino acids 678–759) were generated. Avian ANP32A (Geneart plasmid # 413932.3) was synthesised and codon optimised for expression in E. coli (GenScript, New Jersey, USA). Plasmids containing just the intrinsically disordered region of avian ANP32A (aIVDD, amino acids 149–281) or human ANP32A (hIVDD, amino acids 144–249) were generated. All constructs were cloned into a plasmid derived from pET9a with an N-terminal His-tag and a TEV cleavage site (MGEHHHHHHDDDPITENLYFQQ). pQE70-ANP32A was a gift from Konrad Buessow (Addgene plasmid # 31563; http://n2t.net/addgene:31563; RRID: Addgene_31563)47.
Protein expression and purification. Plasmids were transformed into E. coli Rosetta cells, and the cultures were grown in LB and induced with IPTG for 16 h at 18°C. Cultures were harvested by centrifugation, resuspended in buffer A (50 mM Tris-HCl pH 7.5 and 200 mM NaCl) with protease inhibitors (complete, Roche) and bacterial lysis was performed by sonication. All proteins were purified by affinity chromatography on Ni-NTA agarose (Qiagen), followed by incubation with TEV protease at 4°C coupled with dialysis into buffer A. A second affinity column with Ni-NTA agarose was performed and the following orthogonally coupled to a Superdex 75 column (GE Healthcare) for size-exclusion chromatography in buffer A.

To produce 15N-labelled or 13C- and 15N-labelled proteins for NMR spectroscopy, bacteria were grown in M9 minimal medium containing MEM vitamins ( Gibco), supplemented with 1 g L−1 of 15NCl2 and 2 g L−1 of unlabelled or 13C-glucose. To produce additionally 2H-labelled proteins, the M9 minimal medium was prepared in D2O and 2 g L−1 of deuterated 15NCl2. Protein purity was checked by SDS−PAGE and mass spectrometry. Single-point mutations of 627-NLS were done using the Quick change method30, using Phusion high-fidelity DNA polymerase and DpnI (Thermo Scientific). Cysteine residues were purified as mentioned above for wild-type protein; however, 10 mM of dithiothreitol (DTT) was added after the second Ni-NTA column to keep proteins in a reduced state until labelling. The heterotrimeric human influenza polymerase from B/Memphis/13/03 (Flub) was expressed as a self-cleaving polyprotein and a molar excess of 4-maleimido-TEMPO dissolved in DMSO was added to the reduced 627-NLS cysteine mutants. The reaction was incubated for 12 h at 4°C and then injected into a Superdex S75 column to eliminate the excess of TEMPO labelled samples (700 μM of the 627-NLS (E627) and 150 μM of the 627-NLS (K627) mutants). Measurements in full-length ANP32A were performed with 300 MHz 15N-labelled protein and 400 MHz of the 627-NLS (E627) and 13C-labelled 627-NLS (E627) mutants.

Determination of the relative position of ANP32A and 627. Experimentally determined PREs were measured on h and avANP32A in the presence of different spin-labelled forms of h and av627-NLS, respectively, were used to determine the relative position of the folded domain of ANP32A with respect to the 627 domain. Two thousand different positions of the two domains were generated using the programme Haddock56,57, varying over a wide range of distances and orientations. Positions of spin-label-bearing sidechains were generated on the basis of rotameric libraries (see Supplementary Fig. 8)31,32, and an ensemble of sidechain positions was used to calculate expected PREs on ANP32A for a given position of the 627 domain for each label. Admixtures were adjusted to ensure a population of the bound state of 10% for both complexes. The position of each of the 2000 starting conformations was averaged over a range of ±10 Å along three orthogonal cartesian axes at a resolution of 0.1 Å, and the best fitting position retained. The ten best fitting structures are shown in Fig. 3d.

Ensemble descriptions of ANP32A:627-NLS complexes. Having determined the relative position of the two domains, the flexible parts were constructed onto this configuration. For both h and av complexes, the statistical coil model flexible meca- no27 was used to predict 10,000 conformations of the linker region of 627-NLS, the NLS domain, the NLS peptide that terminates 627-NLS, and the 96 or 128 amino acid IDD of h and avANP32A. Conformers were calculated using amino acid-specific potentials that reproduce the experimentally observed behaviour of the IDD domains, and were calculated to avoid steric overlap between any of the domains (Supplementary Fig. 8). PREs over the entire ANP32A molecule (folded and unfolded domains) were calculated for each of the conformers calculated for each complex, and these conformations were used as a basis set from which ensembles were selected using the ASTEROIDS approach36,58. Ensemble size was estimated on the basis of direct and cross-validated PRE profiles (60 conformers were used for both h and av complexes).

Distance matrices were calculated by calculating the average distance between Cα atoms between the two proteins in the selected ensembles from the two complexes, and the distance difference matrix shown in Fig. 5a by subtracting the matrix from hANP32A:627-NLS(K) from the avANP32A:627-NLS(E) matrix.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
All of the NMR data presented in the article are available from the authors upon request. The NMR chemical shift assignments, and PREs of avian and human ANP32A have been deposited in the Biological Magnetic Resonance Bank (BMRB; bmrbr.wisc.edu/) with accession codes 28134 and 28135, respectively.

Received: 13 March 2020; Accepted: 29 June 2020; Published online: 21 July 2020

References
1. Lozano, R. et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380, 2095–2128 (2012).
2. Nicholls, H. Pandemic Influenza: the Inside Story. PLoS Biol. 4, e50 (2006).
3. Almond, J. W. A single gene determines the host range of influenza virus. Nature 270, 617–618 (1977).
4. Tarendeau, F. et al. Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. PLoS Pathog. 4, e1000136 (2008).
5. Mehe, A. & Doudna, J. A. Adaptive strategies of the influenza virus polymerase for replication in humans. Proc. Natl Acad. Sci. USA 106, 21312–21316 (2009).
6. Gabriel, G., Czudai-Matwich, V. & Klenc, H.-D. Adaptive mutations in the H5N1 polymerase complex. Virus Res. 178, 53–62 (2013).
7. Tarendeau, F. et al. Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. Nat. Struct. Mol. Biol. 14, 225–233 (2007).
8. Subbarao, E. K., Kawoaka, Y. & Murphy, B. R. Rescue of an influenza A virus wild-type PB2 gene and a mutant derivative bearing a site-specific temperature-sensitive and attenuating mutation. J. Virol. 67, 7223–7228 (1993).
9. Massin, P., van der Werf, S. & Naffakh, N. Residue 627 of PB2 is a determinant of cold sensitivity in RNA replication of avian influenza viruses. J. Virol. 75, 5389–5404 (2001).
10. de Jong, M. D. et al. Fatal outcome of human influenza A (H5N1) strain is associated with high viral load and hypercytokinemia. Nat. Med. 12, 1203–1207 (2006).
11. Kiriu, J., Bucci, M. D., Poole, D. S. & Mehe, A. Conserved features of the PB2 domain impact influenza virus polymerase function and replication. J. Virol. 88, 5977–5986 (2014).
12. Reilly, P. T., Yu, Y., Hamiche, A. & Wang, L. Cracking the ANP32A:627 whip’s important functions, unequal requirement, and hints at disease implications. BioEssays 36, 1062–1071 (2014).
13. Long, J. S. et al. Species difference in ANP32A alters influenza A virus replication and restricts host infection. Nature 529, 101–104 (2016).
14. Sugiyama, K., Kawaguchi, A., Okuwaki, M. & Nagata, K. pp32 and APRIL are host cell-derived regulators of influenza virus RNA synthesis from cRNA. Life Sci. 84, e8939 (2015).
15. Lowen, A. C. Viology: Host protein clips bird flu’s wings in mammals. Nature 529, 30–31 (2016).
16. Mehe, A. The avian influenza virus polymerase brings ANP32A home to roost. Cell Host Microbe 19, 137–138 (2016).
17. Domingues, P. & Hale, B. G. Functional insights into ANP32A-dependent influenza virus polymerase host restriction. Cell Rep. 20, 2538–2546 (2017).
18. Baker, S. F., Ledwith, M. P. & Mehle, A. Differential splicing of ANP32A in birds alters its ability to stimulate RNA synthesis by restricted influenza polymerase. Cell 284, 2381–2388 (2016).

19. Long, J. S., Mistry, B., Haslam, S. M. & Barclay, W. S. Host and viral determinants of influenza A virus species specificity. Nat. Rev. Microbiol. 17, 67 (2019).

20. Stoller, E. et al. ANP32 proteins are essential for influenza virus replication in human cells. J. Virol. 93, e00217–e00219 (2019).

21. Zhang, H. et al. Fundamental contribution and host range determination of ANP32A and ANP32B in influenza A virus polymerase activity. J. Virol. 93, e0174–19 (2019).

22. Baker, S. F. & Mehle, A. ANP32B, or not to be, that is the question for influenza virus. eLife 8, e48084 (2019).

23. Long, J. S. et al. Species specific differences in use of ANP32B proteins by influenza A virus. eLife 8, e45066 (2019).

24. Mistry, B. et al. Elucidating the interactions between influenza virus polymerase and host factor ANP32A. J. Virol. 94, e01353–19 (2020).

25. Kuzuhara, T. et al. Structural basis of the influenza A virus RNA polymerase PB2 RNA-binding domain containing the pathogenicity-determinant lysine 627 residue. J. Biol. Chem. 284, 6855–6860 (2009).

26. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. Nature 516, 355–360 (2014).

27. Delaforge, E. et al. Large-scale conformational dynamics control H5N1 influenza polymerase PB2 binding to importin α. J. Am. Chem. Soc. 137, 15123–15134 (2015).

28. Hengrun, N. et al. Crystal structure of the RNA-dependent RNA polymerase from influenza C virus. Nature 527, 114–117 (2015).

29. Thierry, E. et al. Influenza polymerase can adopt an alternative configuration involving a radial packing of PB2 domains. Mol. Cell 61, 125–137 (2016).

30. Yamada, S. et al. Biological and structural characterization of a host-adapting avian influenza polymerase. J. Biol. Chem. 280, 11015–11018 (2005).

31. Salmon, L. et al. NMR characterization of long-range order in intrinsically disordered proteins using NMR spectroscopy. Chem. Rev. 114, 6632–6660 (2014).

32. Holm, L. et al. Comprehensive mapping of adaptation of the avian influenza polymerase protein PB2 to humans. eLife 8, e45079 (2019).

33. Wadziki, J. M., Koubia, T. & Cusack, S. Structure and function of influenza polymerase. Cold Spring Harb. Perspect. Med. 10, 438372 (2020).

34. Fan, H. et al. Structures of influenza A virus RNA polymerase offer insight into viral genome replication. Nature 573, 287–290 (2019).

35. Jensen, M. R., Zweckstetter, M., Huang, J. & Blackledge, M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. J. Mol. Biol. 418, 99 (2012).

36. Huang, J.-R. et al. Transient electrostatic interactions dominate the conformational equilibrium sampled by multidomain splicing factor U2AF65: a combined NMR and SAXS study. J. Am. Chem. Soc. 136, 7068–7077 (2014).

37. Millns, S. et al. Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. Cell 163, 734–745 (2015).

38. Borgia, A. et al. Extreme disorder in an ultrahigh-affinity protein complex. Nature 555, 61–66 (2018).

39. Hayashi, T., Wills, S., Bussey, K. A. & Takimoto, T. Identification of influenza A virus PB2 residues involved in enhanced polymerase activity and virus growth in mammalian cells at low temperatures. J. Virol. 89, 8042–8049 (2015).

40. Peng, Q. et al. Structural insight into RNA synthesis by influenza D polymerase. Nat. Microbiol. 4, 1750–1759 (2019).

41. Koubia, T., Drmcova, P. & Cusack, S. Structural snapshots of actively transcribing influenza polymerase. Nat. Struct. Mol. Biol. 26, 460–470 (2019).

42. Fedor, E. & Te Velthuis, A. J. W. Structure and function of the influenza virus transcription and replication machinery. Cold Spring Harb. Perspect. Med. 10, a038398 (2019).

43. Lukarska, M. et al. Structural basis of an essential interaction between influenza polymerase and Pol II CTD. Nature 541, 117–121 (2017).

44. Martin, I. S. et al. A mechanism for the activation of the influenza virus transcriptase. Mol. Cell 70, 1101–1110.e14 (2018).

45. Milles, S. et al. An ultraflex weak interaction in the intrinsically disordered replication machinery is essential for measles virus function. Sci. Adv. 4, eaat7778 (2018).

46. Bassew, K. et al. Structural genomics of human proteins-target selection and generation of a public catalogue of expression clones. Microb. Cell Factories 4, 21 (2005).

47. Braman, J., Papworth, C. & Greener, A. Site-directed mutagenesis using double-stranded plasmid DNA templates. Methods Mol. Biol. Clifton NJ 57, 31–44 (1996).

48. DELAGLIO, F. et al. NMRPipe - a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 6, 277–293 (1995).