Published online 25 February 2014 *Nucleic Acids Research, 2014, Vol. 42, No. 8* ***5361–5377***
doi:10.1093/nar/gku143

Selective unfolding of one Ribonuclease H domain of HIV reverse transcriptase is linked to homodimer formation

Xunhai Zheng, Lars C. Pedersen, Scott A. Gabel, Geoffrey A. Mueller, Matthew J. Cuneo, Eugene F. DeRose, Juno M. Krahn and Robert E. London*

Laboratory of Structural Biology, NIEHS, National Institutes of Health, Research Triangle Park, NC 27709, USA

Received October 28, 2013; Revised January 28, 2014; Accepted January 29, 2014

ABSTRACT

HIV-1 reverse transcriptase (RT), a critical enzyme of the HIV life cycle and an important drug target, undergoes complex and largely uncharacterized conformational rearrangements that underlie its asymmetric folding, dimerization and subunit-selective ribonuclease H domain (RH) proteolysis. In the present article we have used a combination of NMR spectroscopy, small angle X-ray scattering and X-ray crystallography to characterize the p51 and p66 monomers and the conformational maturation of the p66/p66′ homodimer. The p66 monomer exists as a loosely structured molecule in which the fingers/palm/connection, thumb and RH substructures are connected by flexible (disordered) linking segments. The initially observed homodimer is asymmetric and includes two fully folded RH domains, while exhibiting other conformational features similar to that of the RT heterodimer. The RH′ domain of the p66′ subunit undergoes selective unfolding with time constant ~6.5 h, consistent with destabilization due to residue transfer to the polymerase′ domain on the p66′ subunit. A simultaneous increase in the intensity of resonances near the random coil positions is characterized by a similar time constant. Consistent with the residue transfer hypothesis, a construct of the isolated RH domain lacking the two N-terminal residues is shown to exhibit reduced stability. These results demonstrate that RH′ unfolding is coupled to homodimer formation.

INTRODUCTION

HIV-1 reverse transcriptase (RT) is a multifunctional enzyme critical to the life cycle of the virus. The mature enzyme adopts a heterodimeric structure, such that both the polymerase and ribonuclease H active sites are located on the p66 subunit, while the p51 subunit also contains a polymerase domain which adopts an inactive fold and also includes a 14-residue fragment of the ribonuclease H domain (RH) (1,2). The structural basis for the asymmetric processing of HIV-1 RT to yield the mature RT p66/p51 heterodimer remains the major unresolved problem of RT assembly. For optimal activity of RT, dimer maturation requires that only one RH domain is removed from the homodimer. A viable model must also explain how the protease (PR) gains access to a cleavage site that is buried in the central β-strand of the folded RH. This lack of accessibility of the cleavage site is consistent with the proteolytic resistance of the isolated RH domain to HIV-1 PR (3,4). Based on these and other observations, it was initially proposed that the p66 homodimer exists as a conformational heterodimer in which one RH domain is partially unfolded in a manner sufficient to expose the cleavage site in that domain (3,5,6). Indirect support for this explanation was derived from the demonstration that the polymerase domain is metamorphic, adopting alternative folding patterns in each subunit (2) (Figure 1), and thus might represent a cleavage product of an asymmetrically folded p66/p66′ precursor. Furthermore, the significant polymerase activity of p66/p66′ (7,8) is consistent with a fold that is similar to that of the active heterodimer. However, little progress has been made toward characterizing the p66/p66′ structure or resolving the structural basis for the asymmetric cleavage. There is currently little information about the structure of the p66/p66′ homodimer, and studies supporting both a symmetric and an asymmetric structure have been reported (9,10). Recently other models have been considered, such as the possibility that formation of the two subunits is concerted rather than sequential, with proteolytic cleavage of RH occurring prior to dimer formation (11). These questions are of more than academic interest,

*To whom correspondence should be addressed. Tel: +1 919 541 4879; Fax: +1 919 541 5707; Email: London@niehs.nih.gov
Present address:
Matthew J. Cuneo, 1 Neutron Sciences Directorate, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA.

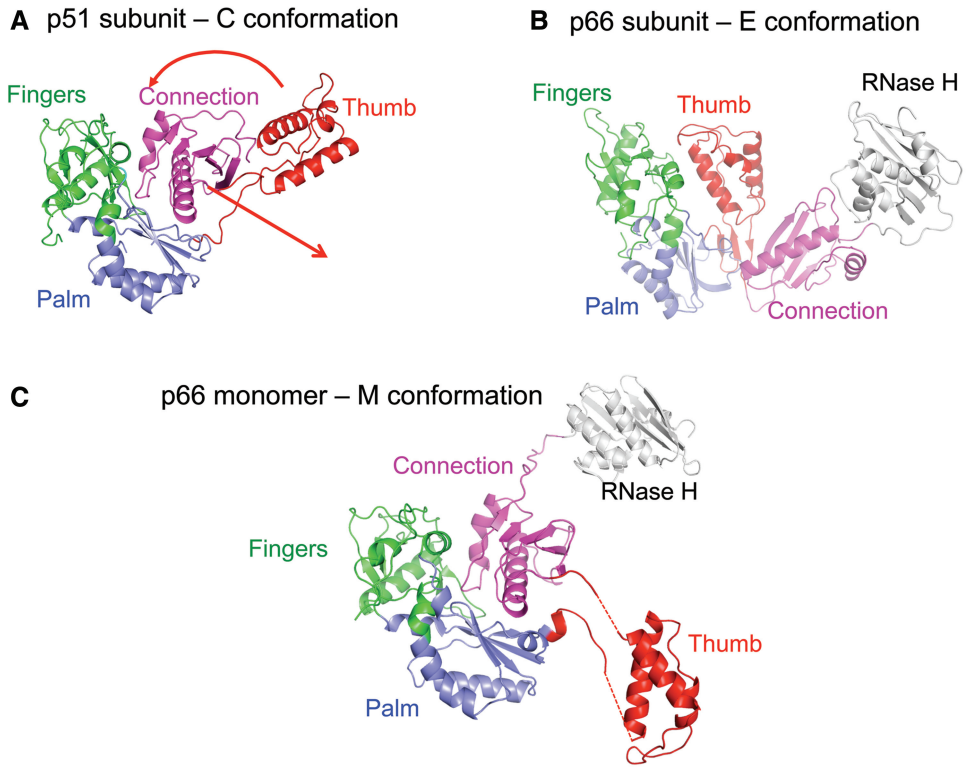

Figure 1. Domain/subdomain orientations in HIV-1 RT. The polymerase domain of HIV-1 RT is a metamorphic protein in which the sequence does not uniquely define the structure. The relative orientations of the domains/subdomains in the p51 subunit (**A**) and p66 subunit (**B**) are indicated. Color coding: fingers (green), palm (purple), thumb (red), connection (magenta) and RNase H (gray). Most of the secondary structure is preserved through the metamorphic transition, however some variation is observed, particularly involving the base of the thumb subdomain. The red arrows indicate the movements of the connection and thumb subdomains required to interconvert between the structures shown in panels A and B. (**C**) A model of the structure and domain orientations in the p66 monomer based on data obtained in the present article. In the monomer structure, the fingers/palm/connection, thumb and RNase H exist as three, flexibly linked domains. Linking segments derived from the thumb and connection domains are disordered. Note that in this figure, we have defined the thumb subdomain to include residues 230–319 as in our construct (Table 1).

since understanding the basis for RT formation and maturation could result in the identification of new targets for drug intervention.

Early kinetic studies of heterodimer formation indicated that a rapid dimerization process was followed by a much slower conformational reorganization, presumably reflecting an induced fit mechanism (12,13). In contrast, recent NMR and SAXS studies of the isolated polymerase domain are more consistent with a conformational selection model in which the predominant compact ('C') conformer, similar in structure to the p51 subunit of RT, infrequently isomerizes to the extended ('E') conformation that is similar to the active conformation in the p66 subunit (14). Dimerization then involves the interaction of monomers in the C and E conformation. Recent kinetic studies support the conclusion that both p66/p66′ and p66/p51 formation occur by a similar conformational selection process (15,16). Although the unique characteristics of metamorphic proteins have recently received enhanced attention (17–19), the implications of the metamorphic nature of the polymerase domain of RT for its conformational maturation have received little consideration.

The sequence redundancy of the two subunits represents a significant challenge to the analysis of this system, however NMR provides a potential approach for the analysis of the conformational transitions during p66/p66′ maturation and for characterization of the monomer species. In the present article, we have utilized an isoleucine labeling approach in order to achieve a compromise between sensitivity, detailed structural information, and complexity (Supplementary Figure S1). The behavior and conformational features of both the p51 and p66 monomers have been characterized, and the time-dependent spectral changes following dimerization provide insights into the dominant conformational changes of individual domains/subdomains during this process. Additional insight has been obtained with a deletion mutation that restricts conformational isomerization of the metamorphic polymerase domain. The energetic basis for the required unfolding of the supernumerary RNase H domain is found to result from the competition for several residues to which the inactively folded polymerase and RNase H domains both lay claim.

Experimental section

Construction, expression and purification of proteins

The constructs used in the current article are summarized in Table 1. All the p51 and p66 constructs contain the C280S mutation to provide greater stability against formation of oxidation products. All single mutations were

Table 1. Constructs used in this article

Abbreviations	Definition	Description	Vector
p66	HXB2 strain, residues 1–560		pet21(+)a
p66(L289K)	HXB2 strain, residues 1–560	Mutation L289K	pet21(+)a
p66ΔPL (palm loop deletion)	p66(Δ219–230)	Deleted residues 219–230	pet30(+)a
p51	P66(1–426)	Derived from p66	pet30(+)a
p51ΔPL (palm loop deletion)	RT430(Δ219–230)	Deleted residues 219–230 stop at position 430	pet30(+)a
RT216	p66 residues 1–216	Palm and fingers	pet30(+)a
RT319	p66 residues 1–319	Palm, fingers and thumb	pet30(+)a
Thumb subdomain	p66 residues 230–319		pet30(+)a
RH	p66(427–560)	Construct contains leader sequence MNEL prior to Y427	pet30(+)a
RHΔNT	p66(429–560)(L429M)	Replaces initial MNELYQL residues in RH with methionine	pet30(+)a

introduced using the QuikChange XL site directed mutagenesis kit (Agilent). Constructs were expressed in the *Escherichia coli* BL21(DE3) codon plus RIL strain at 25°C, and induced by addition of IPTG. The expression and purification of p66, p51, RT216 and the mutated analogs followed the same protocols as described previously (14,20). Expression and purification of the RH and RHΔNT domains were identical to that used in the previous study (21). We note that both constructs appeared as monomers and dimers eluting from the size exclusion column. Only the monomer fraction was retained.

Assignments of RT216 isoleucine methyl resonances were made by first assigning the backbone resonances in U-[^{2}H,^{13}C,^{15}N]RT216, which was expressed using M9 deuterated (99% D_2O) medium containing U-[^{2}H, ^{13}C] glucose and $^{15}NH_4Cl$ as the sole carbon and nitrogen sources. For assignment of the δ-$^{13}CH_3$ resonances, U-[1,2,3,4-$^{13}C_4$-Ile,^{2}H,^{12}C,^{15}N]RT216 was prepared by supplementing the medium with 50 mg of [1,2,3,4-$^{13}C_4$,3,3-$^{2}H_2$]2-oxobutyrate per liter, and unlabeled glucose. The Ile δ-$^{13}CH_3$ resonances of the isolated thumb subdomain were assigned using U-[^{13}C,^{15}N] RT(230-319), derived by growth in M9 medium containing U-[^{13}C] glucose and $^{15}NH_4Cl$ as the sole carbon and nitrogen sources. Assignments of the isoleucine δ-$^{13}CH_3$ resonances of the isolated RH domain were obtained from previous studies (21,22). pD values are uncorrected pH meter readings plus 0.42.

The first five samples in Table 1 utilized the [δ-$^{13}CH_3$-Ile] labeling pattern in a highly deuterated background for maximum sensitivity. They were prepared as described by Tugarinov and Kay (23) by growth on M9 minimal medium in 99% D_2O, supplemented with 50 mg/L [4-^{13}C,3,3-$^{2}H_2$]2-oxobutyrate 1 h prior to induction. Unless otherwise noted this labeling is identified as [$^{13}CH_3$-Ile] for simplicity.

NMR experiments

^{1}H–^{13}C HMQC spectra were obtained at 25 or 35°C using Varian's gChmqc sequence on UNITY INOVA 500, 600 or 800 MHz spectrometers equipped with 5 mm Varian $^{1}H\{^{13}C,^{15}N\}$ triple-resonance cryogenically cooled probe, with actively shielded *z*-gradients. Typical acquisition parameters were 64 ms acquisition of 1024 points with 1.0 or 2.0s delay between scans and a ^{1}H sweep width of 14 ppm. In the indirect, ^{13}C-dimension, 96 points were acquired covering 10 ppm and the ^{13}C offset was set to 13 ppm.

All NMR experiments for backbone assignments of the subdomains were acquired at 25°C on Varian Inova 600 spectrometer, equipped with a triple resonance *z*-axis gradient Cold Probe. The backbone assignments for triply labeled RT216 were obtained from the following TROSY-based 3D triple resonance experiments: HNCA, HN(CO)CA, HN(CA)CB, HN(COCA)CB and HNCO (24). Chemical shift assignments of the δ-methyl groups of Ile were then accomplished from Ile,Leu-(HM)CM (CGCBCA)NH and Ile,Leu-HM(CMCGCBCA)NH experiments (23). All NMR data were processed using NMRPipe (25) and analyzed with NMRviewJ (26).

Dimerization and denaturation studies

All dimerization studies started with the monomeric forms, which were stored at 4°C as dilute (1–3 μM) samples in storage buffer: 50 mM Tris–HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA. For the dimerization studies, the p66 samples were initially concentrated to 80–100 μM and transferred to the NMR buffer: 25 mM Tris–d11 in D_2O, *pD* 7.5, 100 mM KCl, 0.02 % NaN_3. During these procedures, which required ∼1 h, sample temperature was maintained at 4°C. Undoubtedly, some dimerization occurs during the preparative stage prior to the NMR observations, and these changes contribute to the initial spectra.

The urea-induced unfolding of isolated RNase H domain and RHΔNT were performed at 25 °C on 50 μM RH or RHΔNT in 25 mM Tris–HCl–d11 8% D_2O, pH 6.8, containing 0.02% NaN_3, 20 uM DSS, using the Agilent gNfhsqc $^{1}H/^{15}N$ HSQC sequence with 3919 WATERGATE water suppression (27) on an INOVA 500 spectrometer with a cryogenically cooled probe. To insure similar protein concentrations at each urea concentration, premixed buffers were added to aliquots of lyophilized protein (28). The intensities for all well-dispersed resonances were recorded as a function of [urea] and fit to a sigmoid function to determine ΔΔG (28).

Data analysis

The rate of unfolding of the RH′ domain in the p66/p66′ homodimer was determined from intensity measurements

of the three resolved isoleucine resonances arising from this domain: Ile434′, Ile495′ and Ile521′. Since for each of these signals, the resonances of the p66′ subunit overlap with the resonances of the monomer, the intensities were corrected based on the Ile393 resonance that could be uniquely assigned to the monomer, as indicated below for Ile434:

$$I^{C}_{434}(t) = I^{M+C}_{434}(t) - k\ I^{M}_{393}(t) \quad (1)$$

The adjustable constant k was included to correct for differences in the relaxation parameters between Ile393 and each observed Ile resonance, and to adjust for variable fractions of monomer that were present in each individual enzyme preparation. Operationally, k was selected so that the asymptotic behavior of the RH′ domain-resonance intensities approached 0, since no unfolding was observed in any of the monomer species studied or in the isolated RH domain. Although in principle the parameter k was expected to vary for each Ile resonance in each different study, the variation in k was limited to 0.6–1.0 in order to minimize possible bias in the analysis of the data. Constraining k = 0.8 for the entire dataset reduced the mean time constant by only 8.5%, and does not significantly alter the conclusions of the study. The resulting decay curves were then analyzed using a two parameter exponential function:

$$I^{C}(t) = c_1 e^{-t/c_2} \quad (2)$$

where $c1$ is a variable pre-exponential factor and c2 the time constant characterizing the decay.

Unfolding of the RH′ domain was also found to be accompanied by a time-dependent increase in the intensity of multiple resonances near the random coil shift values of 0.86 ppm (^{1}H) and 12.9 ppm (^{13}C) (29). The time constants for the increasing intensity of the random coil peak and for several resonances in the p66 subunit of the homodimer resonances were determined by fitting to a single exponential model using the relation:

$$I(t) = c_1(1 - e^{t/c_2}) \quad (3)$$

The validity of the RH′ unfolding analysis is supported by the consistency of the time constants obtained form analysis of the Ile434′, Ile495′, and Ile521′ resonances using this approach and the time constants derived from the intensity measurements at the random coil position.

SAXS Experiments

SAXS data for p66(L289K) were recorded at National Synchrotron Light Source beamlines X21 and X9 (Brookhaven National Laboratory, Upton, NY), following procedures described previously (14). The effects of sample concentration on the calculated radius of gyration for the p66, monomer, homodimer and heterodimer are summarized in Supplementary Table S1.

X-ray crystallography

Crystals of the p51ΔPL monomer were grown at 4°C using the hanging drop vapor diffusion technique by mixing 1.5 µl of protein solution at 14.5 mg/ml in 20 mM Tris pH 8.0 and 100 mM NaCl with 1.5 µl of reservoir solution containing 100 mM imidazole pH 7.5, 4% galactose and 10% PEG8000. For data collection a crystal was transferred to a cryo-condition made up of the reservoir plus 20% ethylene glycol and flash frozen in liquid nitrogen. Data were collected on the SER-CAT ID beamline at Argonne National Laboratory (APS). The structure was solved by molecular replacement in PHENIX (30) using the p51 subunit of RT from pdb ID code 1QE1 (residues 219–230 deleted) as a search model. Residues 242–316 of the thumb domain were deleted after inspection of the electron density map. The model was further refined and modified using multiple rounds of refinement in PHENIX and model building including manual placement of the thumb domain (residues 245–311) followed by further refinement. The final model consists of residues 6–91, 94–245, 253–311 and 315–430. The X-ray parameters are summarized in the Supplementary Table S2, and a pdb file has been submitted to the protein data bank under the accession number 4KSE.

Nomenclature

In order to describe the complex behavior of this sequentially degenerate system, we have utilized the following nomenclature: The subunit and associated domains that become committed to developing into p51 and the supernumerary RH domain are indicated by primes, e.g. p66′, thumb′, RH′, etc. In some instances, we also have used the conformationally specific labeling introduced previously (14) in which we designate p66*M* as the monomer conformation; p66*E* corresponds to the more extended p66 conformation observed in the RT heterodimer; p66*C* corresponds to the p66 subunit that contains the compact and inactively folded polymerase domain, p51*C*. The individual resonances can then be identified as *M*, *E* or *C* indicating the conformational species to which they correspond. As presented below, the HMQC spectrum of the initially formed p66/p66′ homodimer indicates that p66 is in a conformation 'similar to, but not identical with' p66*E*. Since the conformation is evolving, we will label the corresponding resonances that show time-dependent changes as *E*i if observed initially, and *E* if observed at long times. Similarly, p66′ is close to, but not identical with the *C* conformation, so we identify the resonances as *C*i for the initially observed resonance. The constructs used in this study are summarized in Table 1.

RESULTS

Solution characterization of the p51 and p66 monomers

Conformational characterization of the p51 and p66 monomers is central to understanding both the dimerization and conformational maturation processes of RT. NMR studies were performed on isolated domains/subdomains summarized in Table 1. Backbone assignments of the fingers/palm (RT216), thumb (RT230–319) and RH domain were made on the U-[^{13}C,^{15}N] and U-[^{2}H,^{13}C,^{15}N] labeled molecules, and the isoleucine δ-methyl resonances assigned as described in the

Experimental section (Supplementary Figure S2 and Table S3). Methyl assignments were transferred to larger constructs when resonances were sufficiently isolated, and in some cases, additional assignments were made using site-specific mutants.

Earlier studies showed that low concentration, low salt and low temperature stabilize the monomeric form of p51 (14). A comparison of the ^{1}H-^{13}C HMQC spectra of [^{13}CH$_3$-Ile]p51 with the spectra obtained for other constructs is shown in Figure 2. A single resonance is observed for each resolved isoleucine residue, consistent with the monomeric nature of the subunit observed under these conditions (Figure 2). In contrast, the dimer spectra, discussed below, yield multiple resonances for most residues as a consequence of the distinct folds of the two subunits. A comparison of the ^{1}H-^{13}C HMQC spectrum of the [^{13}CH$_3$-Ile]p51 monomer with the assigned spectra for the fingers/palm subdomain [^{13}CH$_3$-Ile]RT216 (20), and for the isolated [^{13}CH$_3$-Ile]thumb (Table 1 and Figure 2A and B; Supplementary Figure S2) shows close agreement for nearly all of the well-resolved methyl resonances. Based on these comparisons, most of the better

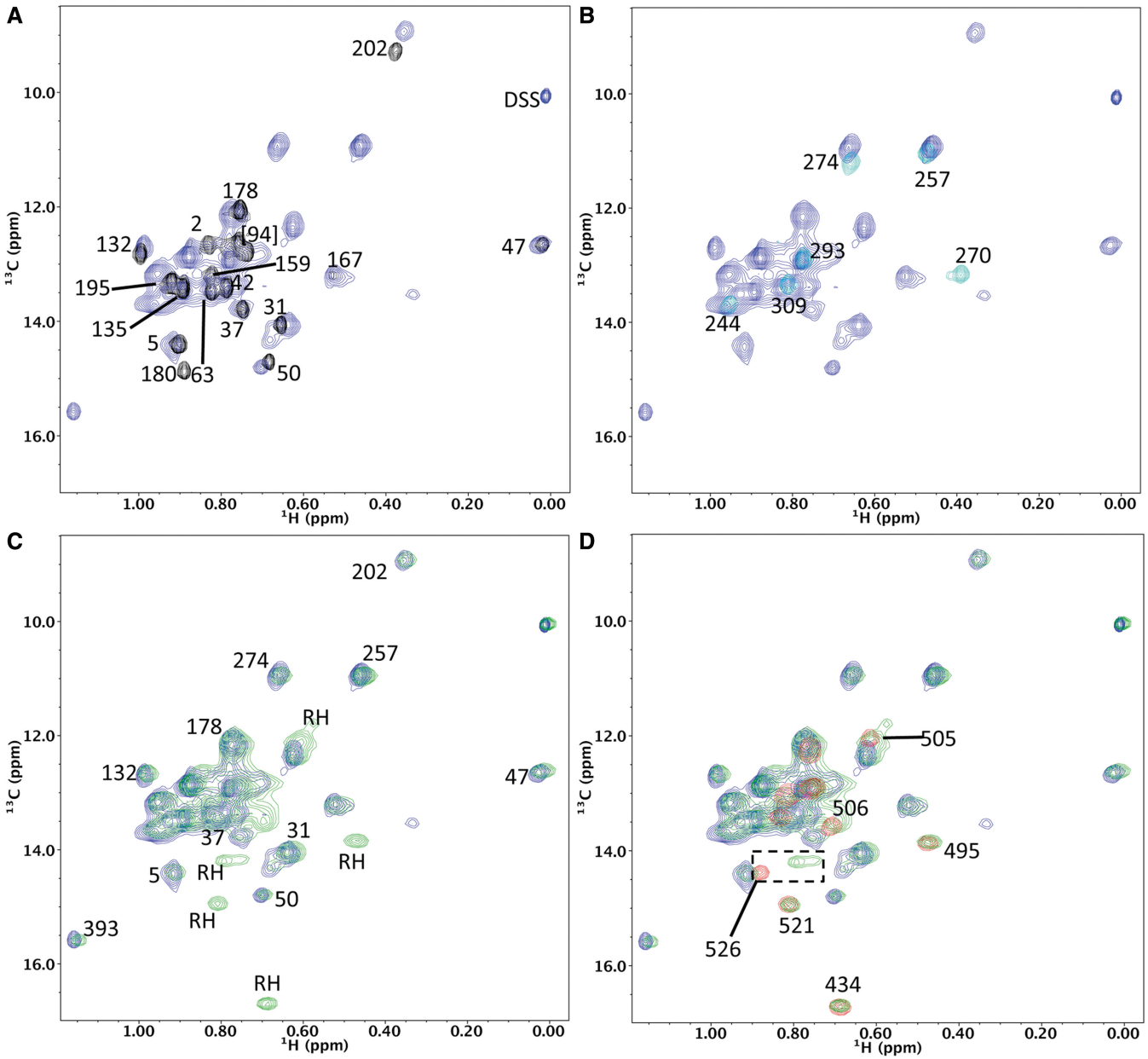

Figure 2. ^{1}H-^{13}C HMQC spectra of several [^{13}CH$_3$-Ile]-labeled constructs. The four panels show comparisons of various constructs to the spectrum of the isolated [^{13}CH$_3$-Ile]p51 monomer (20 μM, blue). The overlayed spectra shown correspond to: (**A**) [^{13}CH$_3$-Ile]RT216 (50 μM, black); (**B**) [^{13}CH$_3$-Ile]thumb (20 μM, cyan); (**C**) [^{13}CH$_3$-Ile,U-^{2}H]p66(L289K) (20 μM, green); (**D**) [^{13}CH$_3$-Ile,U-^{2}H]p66(L289K) (green) plus [^{13}CH$_3$-Ile]RH (15 μM, red). Note in particular how closely the spectra of the p51 monomer plus the isolated RH domain mimic the spectrum of the p66(L289K) monomer. The apparent shift of the Ile526 resonance, enclosed by the rectangle in panel (D), has a trivial explanation related to the sequence differences of the constructs. All samples were in 25 mM Tris–d11, *pD* = 7.52, 0.02 % NaN_3, 20 μM DSS and $T = 25^{\circ}$C.

resolved Ile methyl resonances of the p51 monomer are assigned by comparison with the isolated domains (Supplementary Table S3). These shifts are generally consistent with chemical shift predictions using higher resolution crystal structures (21,31–33), demonstrating that the overall folds of the isolated domains are preserved in the p51 monomer, and supporting the conclusion that the domain interactions within the monomer are insufficient to significantly perturb most of the resonances. The main exceptions to this generalization are the Ile270 and Ile274 resonances, corresponding to residues located at the base of the thumb in a conformationally hypervariable region of the protein.

Several resonances in the connection domain were assigned by site-directed mutagenesis, since the isolated connection domain proved to be intractable. Most importantly, the well-resolved Ile393 resonance (Figure 2 and Supplementary Figure S3) is of particular value for dimerization analysis since, as demonstrated below, it is significantly shifted in the monomer relative to its positions in both subunits of the p66/p66′ and p51/p51′ homodimers.

In order to obtain spectra for the p66 monomer we utilized the L289K mutation that interferes with dimerization (14,34). A comparison of the spectra of the [$^{13}CH_3$-Ile]-labeled p51 and p66 monomers shows nearly perfect consistency (Figure 2C). Nearly all of the p66(L289K) isoleucine methyl resonances that do not overlap with p51 arise from peaks that co-resonate with the isoleucine methyl resonances in isolated RH, as shown by superposition with the HSQC spectrum of the isolated [$^{13}CH_3$-Ile] RH domain (Figure 2D). The results summarized above are consistent with the prediction of Wang *et al.* (35), that the p51 and p66 monomers would adopt similar conformations in solution, and that this conformation would also be similar to that of the p51 subunit of RT as a consequence of the much lower surface exposure of hydrophobic residues in this conformation. This result is also similar to the conclusion of our previous studies of the isolated polymerase domain (14), which concluded that the compact, p51*C* conformation of the monomer is strongly favored over the extended, p51*E* conformation. Additionally, the excellent agreement between the resonances in the isolated RH domain and the p66 monomer supports the conclusion that the RH domain is only weakly coupled to the remaining structure, and in particular does not interact strongly with the thumb domain, although an interface between these domains exists in the RT heterodimer. One exception to this generalization is the Ile526 resonance, and this difference has a trivial explanation resulting from differences in the sequences of the constructs used (Supplementary Figure S4).

A key feature of the spectrum of the p66(L289K) monomer is the close agreement of the Ile methyl resonance shifts with those observed in the isolated RH domain (Figure 2D). Of the RH domain Ile resonances, Ile434 is of particular interest, since it is located in the proteolytically sensitive segment near the PR cleavage site (residues 440–441) (36). Thus, on the basis of the Ile methyl resonances in the RT domain of the p66 monomer, it is predicted that the RH domain of the p66 monomer retains its folded structure and the cleavage site should be no more accessible than it is in the isolated RH domain.

The conformation of p66(L289K) was also evaluated using small angle X-ray scattering (SAXS). Data from the SAXS study is consistent with the conclusion that the p66(L289K) is predominantly monomeric and monodisperse (Figure 3 and Supplementary Table S1).

Simulated SAXS-scattering curves were generated using the p66 subunit from an RT crystal structure (pdb: 1DLO), or using a model of the polymerase domain in a p51-type conformation linked to a variably positioned RH domain by residues derived from partial unfolding of the C-terminus of the connection domain. The fit of the experimental scattering intensity to the p66 subunit derived from a RT crystal structure is substantially poorer ($\chi^2 = 1.5$) than the fit to a linked, two-domain p51∼RH model ($\chi^2 = 0.58$), in which the position of the RH domain was arbitrarily selected (Figure 3A). Multiple orientations of the RH domain were generated using the program EOM (37) and the EOM-generated ensemble of structures produced only a small improvement in the data fit, reducing χ^2 = from 1.5 to 1.3 ± 0.5. Nevertheless, some of the component structures of the ensemble exhibited substantially better fits than the ensemble (Figure 3B). We therefore conclude that the SAXS data are consistent with the multidomain structure determined from crystallographic and NMR data, but are not sufficient to discriminate among various alternatives.

Crystallographic characterization of an isomerization-deficient polymerase domain

No crystallographic data is currently available for the p66 or p51 monomer. To capture this state, we engineered a mutation that interferes with the interconversion of the metamorphic polymerase domain between the *E* and *C* conformations (Figure 1A and B). Since the data summarized above indicate that the monomers significantly favor the *C* conformation, we used the p51 RT subunit as a model, and introduced a deletion mutation consistent with the *C* but not the *E* conformation. The mutated form allows us to more completely characterize the monomer state, and provides a basis for subunit-selective labeling of p66/p66′. The deletion was motivated by the observation that a loop in the palm subdomain of the p51 subunit, corresponding to residues 219–230, is typically disordered and frequently not observed, and residues 218 and 231 are in sufficient proximity to allow direct linkage (Figure 4A). Alternatively, the corresponding residues in the 'E' conformation adopted by the polymerase domain in the p66 subunit are ordered and residues 218 and 231 are separated by ∼20 Å (Figure 4B). The palm loop deletion (ΔPL) mutants consistently elute as monomers on size-exclusion columns in comparison to their wild-type counterparts (data not shown).

The crystal structure of monomeric p51ΔPL reveals a domain orientation pattern that is similar to that of the p51 subunit of RT, while exhibiting two significant differences (Figure 4C): (i) The C-terminal residues of the connection domain corresponding to residues ∼421–429,

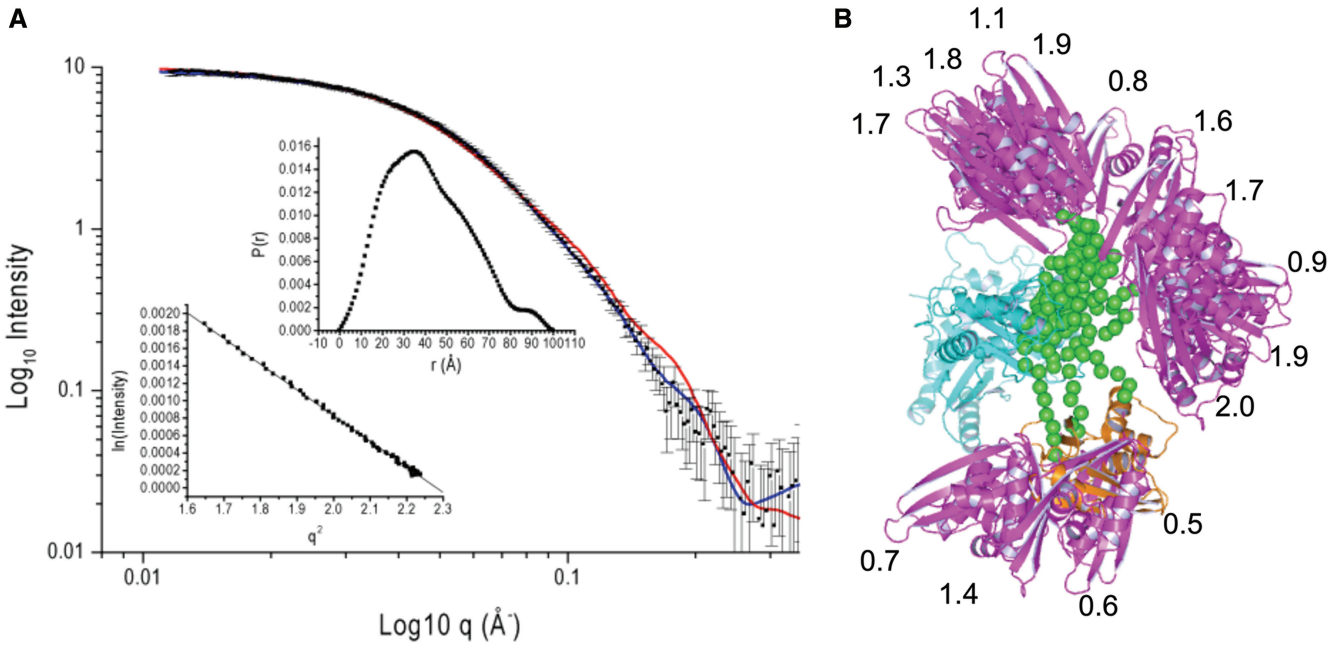

Figure 3. Characterization of the p66(L289K) monomer with SAXS. (**A**) SAXS intensity I(q) data inset with pairwise vector length distribution curve and Guinier fit of the I(q) scattering data. The red solid line is the calculated scattering curve based on the p66 subunit, the blue solid line corresponds to the two domain polymerase∼RH structure in which the polymerase domain adopts the compact C conformation and using the RH position shown in orange in panel (B). (**B**) Fifteen structures showing the variably positioned RH domain corresponding to an initial, arbitrarily selected position of the RH domain (orange) and to 14 structures from the optimal ensemble generated by the EOM analysis (magenta RH domains); the corresponding χ^2 errors for each structure are also shown. The ribbon diagrams show the fingers/palm/thumb/connection in cyan and the variably positioned RH domain in magenta/orange. The linking residues derived from the C-terminus of the connection domain are shown as green spheres.

which form helix αM′ in the p51 subunit of RT, are unfolded, and (ii) the three-helix bundle that forms the core of the thumb remains intact, however, the domain has rotated around the two hinge residues (Trp239 and Val317) and segments 246–252 and 312–314, connecting the hinge region with the three-helix core, are disordered (Figure 4C). Interestingly, these are the same hinge residues that allow conformational flexibility of the thumb in the p66E subunit (39). Observation of both the C-terminal residues and the thumb in the crystal structure likely results from stabilization by lattice contacts. The two major conformational changes involving the thumb orientation and the connection domain C-terminal residues are expected to be cooperative since in the p51 subunit of the RT heterodimer, residues in helix αM′ interact directly with residues in the thumb′ subdomain (Figure 4C, blue structure). The comparison shown in Figure 4C indicates that the monomer, while adopting a conformation similar to the 'C' form observed in RT, has a more variable thumb subdomain orientation.

In combination, the NMR results presented above for the isoleucine-labeled p51 and p66 monomers and the crystal structure of the isolated p51ΔPL subunit allow characterization of the solution conformation of the p66 monomer. The p51ΔPL structure contains an unfolded C-terminus which can function as a linker with the folded RH domain, so that an additional RH domain can be added to form a p66 monomer without any further modification of the p51ΔPL structure. The resulting structure, shown in Figure 1C, satisfies both the NMR and SAXS constraints requiring that the monomer contain linked p51 and RH domains that do not interact significantly. The p66 monomer thus has a loose structural organization characterized by three variably positioned substructures: the fingers/palm/connection, the thumb and the RH domains. The thumb and RH are connected to the fingers/palm/connection by flexible linking segments. The close agreement between the isoleucine methyl shifts observed in the monomer and for the isolated domains (Figure 2) is consistent with weak interface contacts in the p66 monomer. This is an important feature of the monomer; if the monomer adopted the *C* rather than the *M* conformation, additional interface contacts would need to be broken to allow the domain rearrangements necessary for the transition to the *E* conformation. As a result, the p66 monomer is primed for the domain rearrangements required for conversion to the extended 'E' conformation to a much greater extent than would be the case if the polymerase monomer were initially in a p51-like 'C' conformation.

Evaluation of the perturbation introduced by the palm loop deletion

As noted above, deletion of the palm loop was engineered to block the C⟷E interconversion without perturbing the C conformation. The validity of this approach is supported by the data presented above and in our previous study (14) indicating that the p51 and p66 monomers both adopt a conformation similar to that observed in the p51

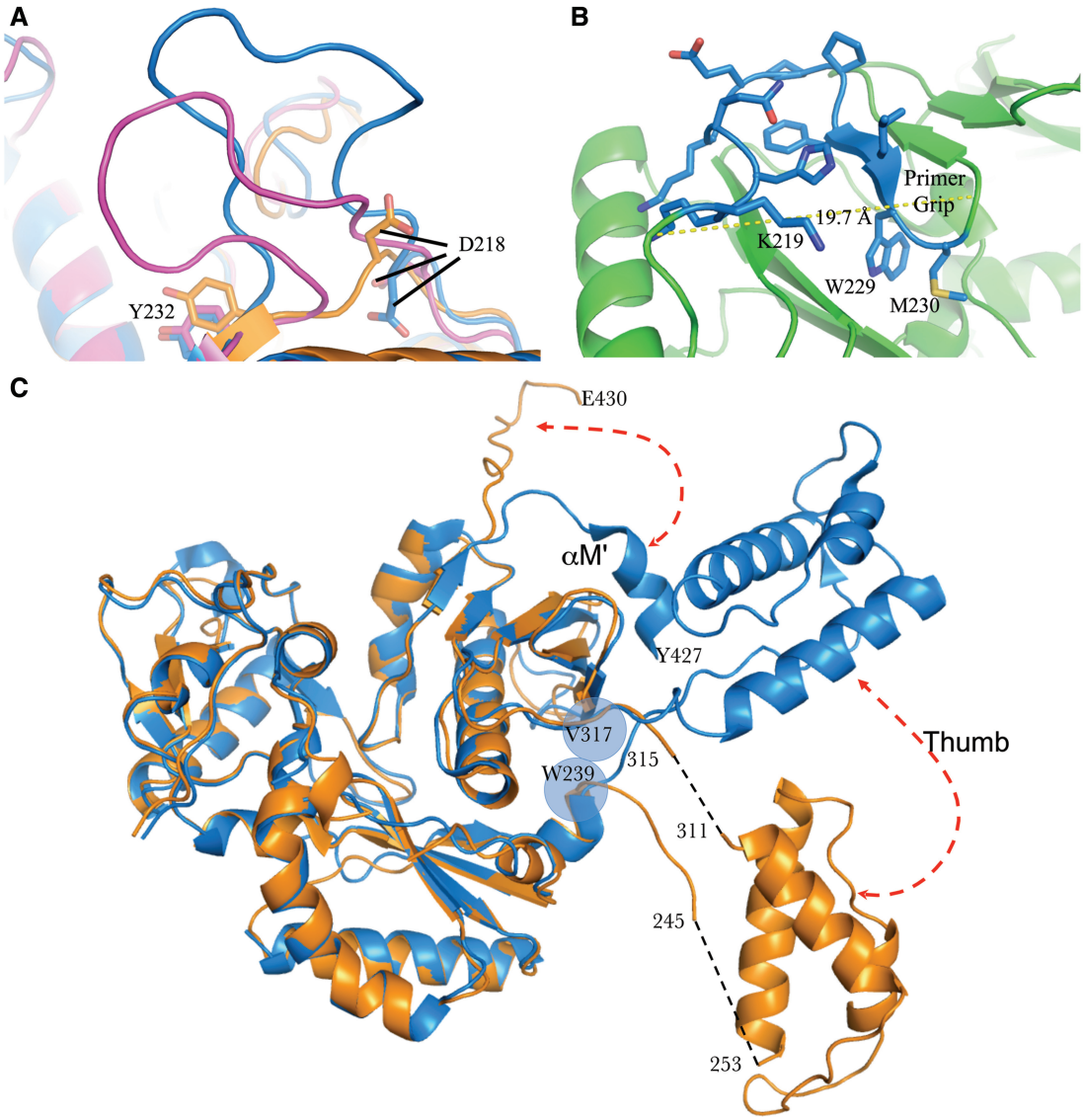

Figure 4. Design and crystal structure of the p51 monomer. (**A**) Overlayed ribbon diagrams of structures showing the disordered loop, residues 219–230, in two p51 subunits of RT (pdb files 1S6P, blue; 2BE2, magenta), along with the corresponding region in the p51ΔPL monomer structure, lacking this loop (orange). Residues Asp218 and Tyr232 are also included to orient the view. We note that in many RT structures, the electron density for the loop residues is sufficiently weak so that they are not included in the structure. (**B**) Ribbon diagram of the region of the p66 subunit of RT (green) containing residues 219–230 (blue). Deletion of these residues would eliminate multiple interactions and require substantial movement of segments near residues 218 and 231 to close the resulting gap. The dotted line shows the distance between the α-carbons of residues 218 and 231. (**C**) Ribbon diagram overlay corresponding to the p51ΔPL monomer crystal structure (orange) and the p51 subunit of RT (pdb code: 3DLK, (38)) (blue). Dashed red arrows indicate the major structural differences. The positions of the two thumb subdomain hinge resides previously identified in the p66E subunit (W239, V317) (39) are indicated by transparent blue circles. Residues adjacent to the disordered (missing) segments of the thumb, 246–252 and 312–314, are annotated.

subunit of RT, and by the analysis of Wang *et al.* (35) also concluding that the polymerase monomer would adopt a conformation similar to that of the p51 subunit of RT, i.e. the C conformation. In principle, this deletion may introduce additional perturbations to the solution conformations of the p51ΔPL and p66ΔPL monomers. In order to evaluate the significance of this perturbation, we compared the spectra of p66 containing two different mutations that both favor the monomer. The ^{1}H-^{13}C HMQC spectra of the [$^{13}CH_3$-Ile]p66ΔPL and the [$^{13}CH_3$-Ile]p66(L289K) monomers are extremely similar (Figure 5A). A few small shift perturbations can be assigned to Ile residues 178 and 202 that are located near the site of the deletion (Figure 5B). The basis for the monomer selectivity supported by these two mutations is different. The L289K mutation biases the conformational equilibrium of the polymerase domain toward the C conformation, perhaps as a result of replacing a solvent-exposed leucine residue with a hydrophilic lysine residue, but does not preclude adoption of the E conformation (14), while as discussed above, the constraint imposed by the palm loop deletion on the conformational

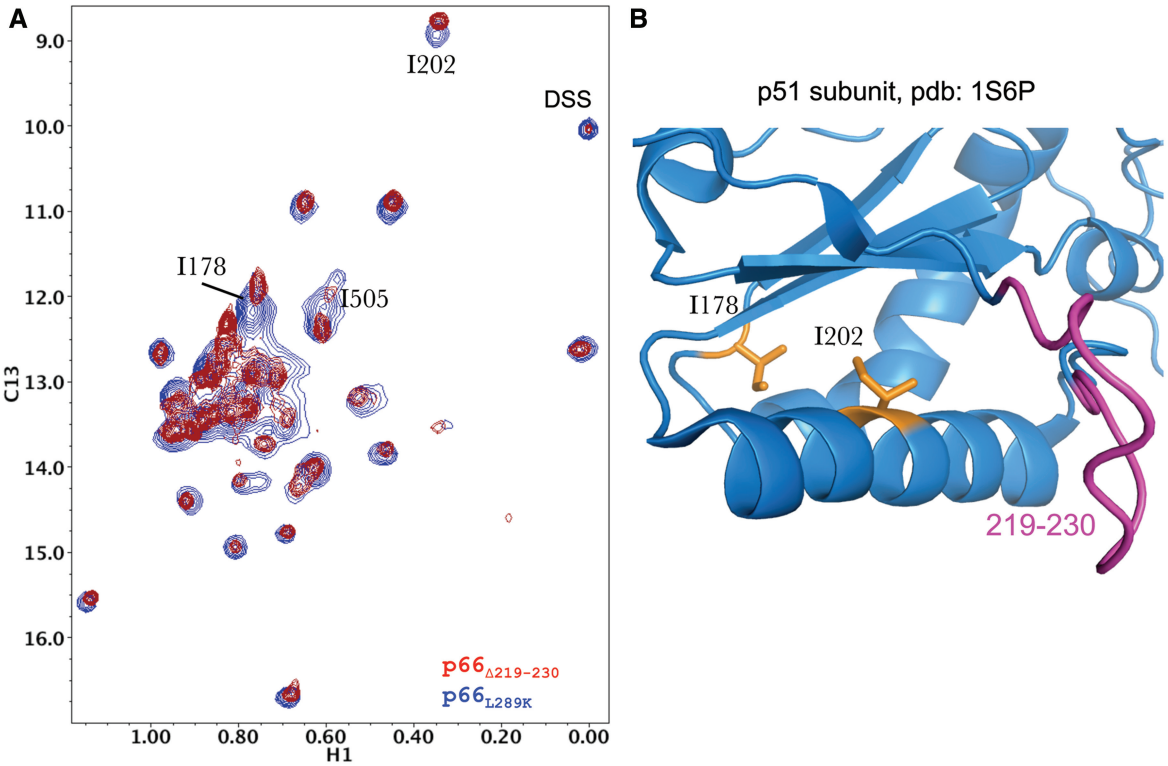

Figure 5. Spectral comparison of two p66 monomers. (**A**) ^{1}H-^{13}C HMQC spectral overlay corresponding to [^{13}CH$_3$-Ile,U-^{2}H]p66ΔPL (red) and to [^{13}CH$_3$-Ile,U-^{2}H]p66(L289K) (blue). (**B**) Ribbon diagram illustrating the relative positions of the palm loop (pdb: 1S6P), and residues Ile178 and Ile202. The small shift differences of the Ile178 and Ile202 resonances are consistent with their proximity to the palm loop.

equilibrium is considerably more substantial, significantly interfering with the E conformation. Based on this comparison, we conclude that the palm-loop-deletion produces only small, localized perturbations, and does not lead to any dramatic conformational perturbations beyond this region of the protein.

Chemical shift perturbations related to heterodimer formation

Altering concentration, ionic strength and temperature to values favoring dimer formation resulted in many spectral changes that are apparent in the first time period after the experimental conditions are changed, as well as additional changes that transpire with much slower rates. Since any of the shift changes observed in the spectrum of the homodimer relative to that of the monomer are similar to those observed in the RT heterodimer, we first evaluated the shift perturbations resulting from heterodimer formation (Figure 6A). In general, the shift perturbations result from two factors: (i) the conformation differences between the two polymerase domains in each subunit, and (ii) perturbations that result directly from formation of the p66:p51 interface. These factors result in spectral inequivalence for many resonances in the heterodimer that arise from the pair of corresponding Ile residues in the two subunits.

Interestingly, nearly all of the RH domain Ile methyl resonances exhibit significant shifts in the heterodimer relative to their positions in the isolated domain or the p66 monomer (Figure 6B). This result was unexpected since, in contrast with the thumb and connection subdomains, the structural differences between the isolated and RT-incorporated RH domain are minimal. The RMSD for residues 427–550 between the isolated RH domain (pdb: 3K2P) and RT (pdb: 3DLK, p66 subunit), is 0.38Å. The chemical shift changes presumably result from the high sensitivity of NMR shifts to the microenvironment of each nucleus. The need to satisfy the constraints of multiple interfaces—an intrasubunit interface with the p66 connection subdomain, and an inter-subunit interface with the connection and thumb subdomains on p51, may result in subtle conformational changes. Regardless of the basis for these shifts, the perturbations of the RH domain Ile resonances in the p66 subunit are extremely useful for characterizing the conformational maturation of the p66/p66′ homodimer.

A significant correlation between dimer shift perturbations and the involvement of the domains in interface contacts was also observed. Thus, the Ile resonances in the p66-fingers and p51-palm show few shift perturbations, while the resonances in the p66-palm and the p51-fingers, which form part of the heterodimer interface, show more significant shift perturbations. The most readily observed examples are Ile47 (fingers) and Ile178 (palm) (Supplementary Figure S5).

Selective unfolding of the RH′ domain in the p66/p66′ homodimer

In order to elucidate the conformational processes underlying homodimer formation, we adjusted the

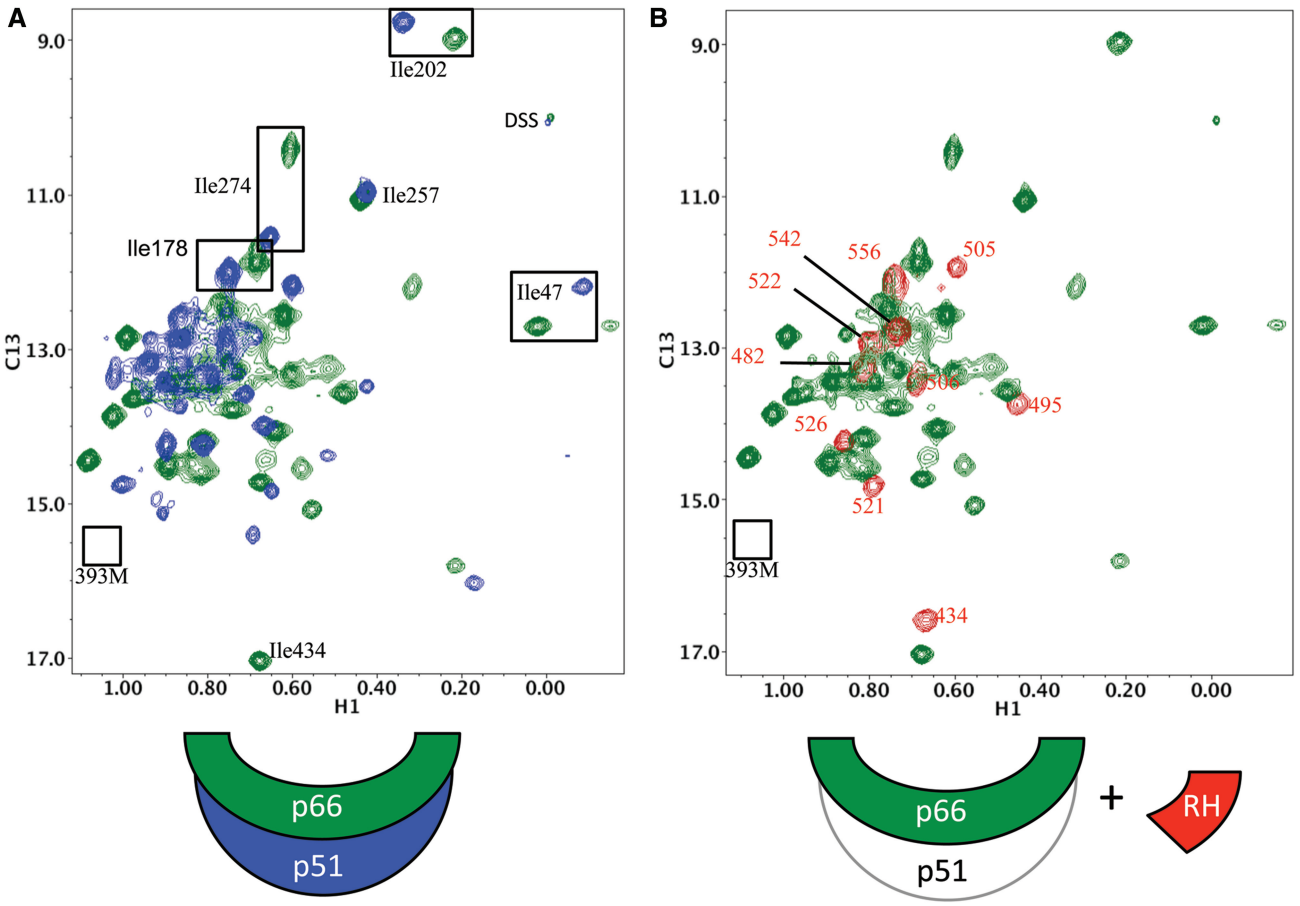

Figure 6. ^{1}H-^{13}C HMQC spectra of isoleucine labeled RT subunits. (**A**) Spectral overlay of [$^{13}CH_3$-Ile,U-$^{2}H]_{66}$RT (50 μM, green) with [$^{13}CH_3$-Ile,U-$^{2}H]_{51}$RT (60 μM, blue) and (**B**) [$^{13}CH_3$-Ile,U-$^{2}H]_{66}$RT (green) with isolated [$^{13}CH_3$-Ile]RNase H (15 μM, red). Schematic diagrams illustrating the labeling pattern used in each panel are shown at the bottom of the figure. Several of the better resolved Ile resonances are annotated, and the RH domain Ile resonances are annotated in red. The position of the missing Ile393 monomer resonance is also indicated. The labeled RT spectrum was obtained in 25 mM Tris–d11, D_2O, *pD* = 7.52, 150 mM KCl, 0.02% NaN_3, 20 μM DSS and the RH-domain spectrum was obtained in the same buffer but without the KCl. All spectra were obtained at 25°C.

experimental conditions to favor dimeric species, as described in Materials and methods section. The approach followed in these studies provides no information on changes occurring on a time scale shorter than ~2 h. However, most dimerization studies have indicated that there are some very slow processes involved, making these transitions accessible to NMR analysis (7,13,16). Although changes in ionic strength can lead to resonance perturbations not directly resulting from dimerization, these are generally small and differ from the shift changes discussed above that result from dimer formation (Supplementary Figure S6). Figure 7A shows the ^{1}H-^{13}C HMQC spectrum of isoleucine-labeled p66 acquired during the first 5½ h acquisition period after adjusting experimental conditions to favor dimerization, as described in the Experimental section. The spectrum is overlayed with that of isoleucine-labeled p51/p51′, which also forms a homodimer that exhibits a low level of polymerase activity (40). The ^{1}H-^{13}C HMQC spectra obtained for both the [$^{13}CH_3$-Ile]p51 and [$^{13}CH_3$-Ile]p66 samples exhibit many of the shift perturbations that are observed for the two conformationally distinct subunits of the heterodimer, supporting the conclusion that the p51/p51′ and p66/p66′ species that predominate during the first several hours adopt many of the conformational features of the heterodimer. For example, the isolated Ile202 and Ile47 resonances exhibit shifts characteristic of both the E and the C conformations observed in the heterodimer (Figure 6A). For overlapping resonances such as Ile47M and Ile47E, Ile202M and Ile202C, and Ile274M and Ile274C, the peaks in the p51 sample are assigned to both the monomer and dimer species, while the same resonances in the p66 sample arise primarily from the dimer. These spectra demonstrate that during the first spectral accumulation period after conditions favoring dimer formation are initiated, both homodimers exist as conformational heterodimers.

It is also apparent that the intensity ratios for the *E* and *C* resonances differ between the p51/p51′ and the p66/p66′. In principle, the monomer/dimer ratios for the two samples can be determined by evaluating the intensities of overlapping *E*, *C* and *M* resonances for multiple residues.

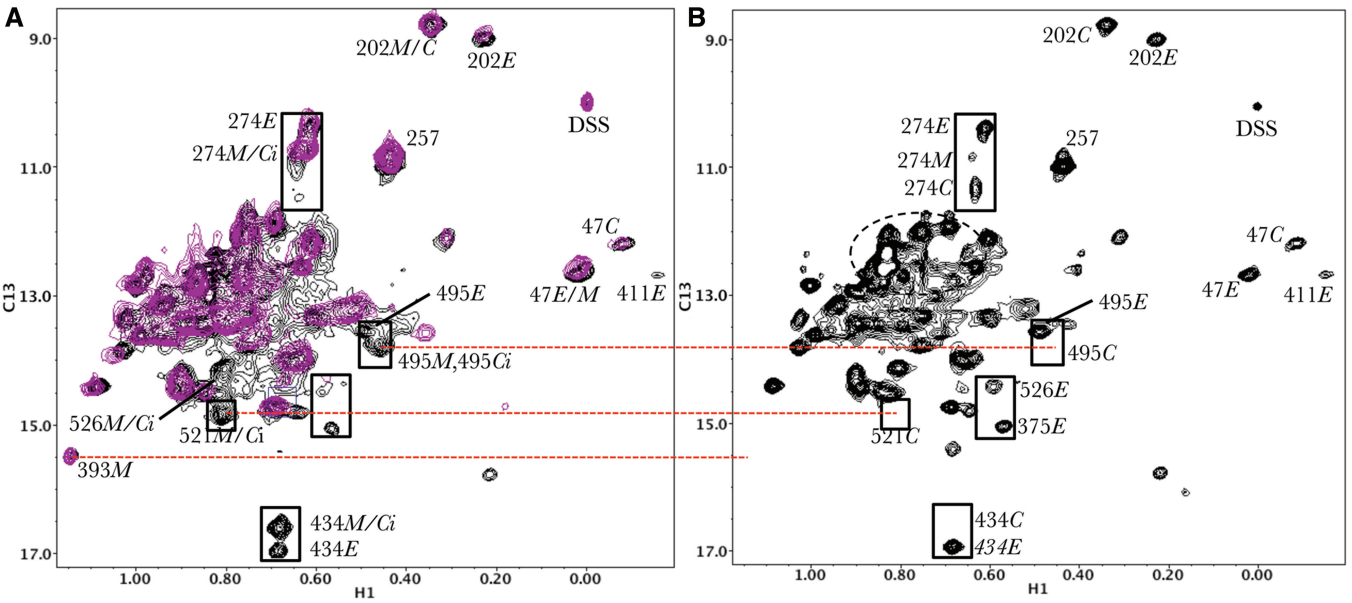

Figure 7. Time-dependent changes in the ^{1}H-^{13}C HMQC spectra of the isoleucine-labeled p66 homodimer. (**A**) Overlay of the initial spectrum obtained for 100 μM of [$^{13}CH_3$-Ile,U-^{2}H]p66 during the first 5.5 h after changing buffer to favor dimerization (black) with the spectrum of 100 μM [$^{13}CH_3$-Ile,U-^{2}H]p51 (magenta); (**B**) ^{1}H-^{13}C HMQC spectrum obtained during the eighth accumulation period (38.5–44 h after the onset of dimerization). In panels (A) and (B), several groups of resolved resonances that exhibit slow time-dependent changes are indicated in boxes. For a number of resonances, the shift of the monomer overlaps that of either the E or *C* forms of the dimer. As indicated by the Ile393M resonance, there is a substantial monomer concentration in the p51/p51′ homodimer, but less in the p66/p66′ homodimer. Since a number of resonances exhibit slow time dependent changes, these are labeled as *Ci*, since they correspond to the initial but not the ultimate shifts for the resonances of the p66′ subunit. The intensities of the RH′ domain *Ci* resonances are all very low by the eighth accumulation period, while the Ile274 *Ci* resonance is replaced by a second resonance. Dimerization buffer was 25 mM Tris–d11 in D_2O, pD = 7.52, 100 mM KCl, 0.02 % NaN_3, 20 uM DSS and T = 35°C. The spectrum of p51 was obtained at 25°C.

However, it is more convenient to determine the monomer concentration from an analysis of the Ile393*M* resonance at $\delta^{1}H,^{13}C$ = 1.16, 15.5 ppm, since this resonance is observed exclusively in the monomer (compare Figures 2 and 6). Figure 7A shows that a strong Ile393 resonance is present at the monomer position in the p51 sample, while the corresponding resonance in the p66 sample is considerably weaker. Thus, during this initial accumulation period, the p51 sample corresponds to a monomer/dimer mixture, while the p66 sample is mostly dimeric. This result is consistent with the much lower dimerization equilibrium constant for p51/p51′ relative to p66/p66′ (41). Similar conclusions are supported by analysis of other resonances, e.g. the relative intensities of the 202*E* and 202*C*/*M* resonances for the p51 and p66 samples, however the analysis using the Ile393*M* resonance is simpler, as this resonance is assigned exclusively to the monomeric species.

A comparison of the initial spectrum of the p66/p66′ homodimer (Figure 7A) with the spectrum of the heterodimer (Figure 6A) reveals a number of differences. Among the most significant is the appearance of multiple additional resonances that align well with those of the isolated RH domain. These resonances are shifted relative to their positions in the RT heterodimer (Figure 6B). Thus, we are observing two sets of RH domain resonances in the homodimer, and these are assigned to the RH and RH′ domains of the two subunits. We therefore conclude that the p66/p66′ homodimer present during the first accumulation period has two-folded RH domains; the resonances of one (RH) exhibit shifts similar to those observed in the RT heterodimer, while the second set of resonances (RH′) exhibit shifts similar to those of the monomer or, equivalently, the isolated RH domain. This characteristic is most readily apparent for the two isolated Ile434 resonances, labeled Ile434*E* and Ile434*M*/*Ci* (Figure 7), but can also be seen for other RH domain resonances. These two sets of resonances are thus assigned to the RH domain in p66 and to the RH′ domain in p66′.

A comparison of the spectra obtained during the first and 8th accumulation periods (Figure 7A and B) reveals multiple changes that transpire over a time course of many hours (Supplementary Figure S7). Well-resolved resonances exhibiting slow, time-dependent changes correspond to Ile375 and Ile526 on the p66 subunit, and to Ile274′ and each of the resolved RH′ domain resonances on p66′. The first three resonances above are shifted to different positions, all of which are similar to their positions in the p66/p51′ heterodimer, while each of the RH′ resonances gradually loses intensity (Figure 8A). In parallel with the decay of the RH′ resonances, the intensities of resonances located near the expected random coil shift for the Ile Cδ methyl group increase sharply and then level off (Figure 8B). These data clearly demonstrate that the initially formed p66/p66′ homodimer contains two folded RH domains, and that the RH′ domain undergoes domain-selective unfolding.

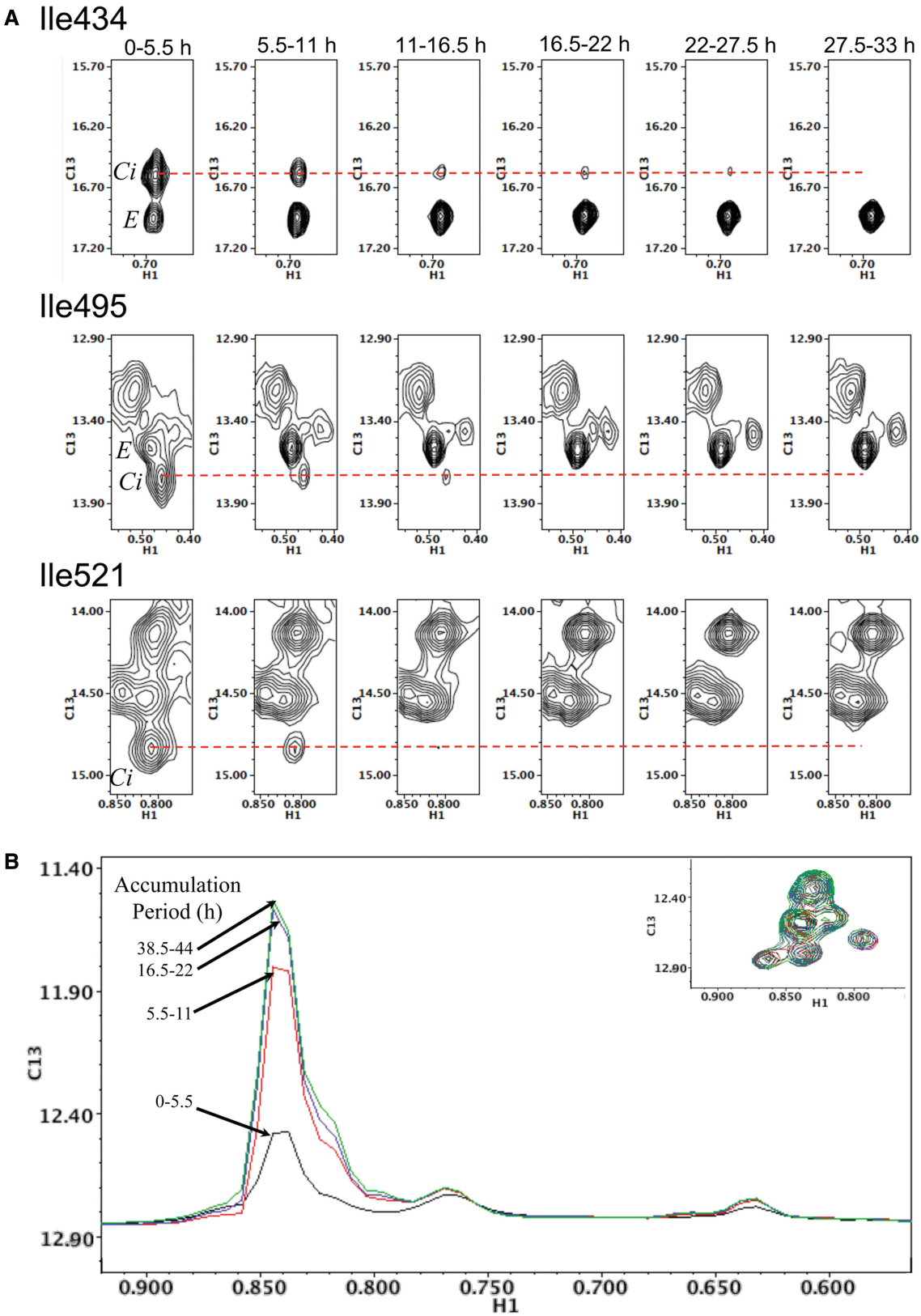

Figure 8. Time-dependent behavior of RH domain resonances in [$^{13}CH_3$-Ile,U-^{2}H]p66/[$^{13}CH_3$-Ile,U-^{2}H]p66′. Each time period corresponds to a sequential 5.5 h accumulation. (**A**) Spectral regions containing the Ile434, Ile495 and Ile521 resonances. For Ile434 and Ile495, the resonances from the RH domains in both subunits are assigned, while for Ile521 only the resonance in the RH′ domain was assigned. (**B**) 1D ^{1}H slices through the random coil ^{13}C chemical shift of the ^{1}H-^{13}C HMQC spectra corresponding to the accumulation periods indicated. First accumulation period (0–5.5 h, black), second period (5.5–11 h, red), fourth period (16.5–22 h, blue), eighth period (38.5–44 h, green). The 2D inset shows that the intensity is divided among several resonances near the random coil shift position. Other conditions as in Figure 7.

A more quantitative estimate of the RH′ decay is limited by overlap between the resolved Ile434′, Ile495′ and Ile521′ resonances assigned to the RH′ domain in the p66′ subunit of the homodimer and the corresponding resonances of the monomer (Figure 2D). Analysis of the time-dependent intensities for each of these resolved resonances required a bi-exponential fit, with slow components characterized by fractional intensities of 10–20% of the total and time constants of ~40 h. Since we observed no time-dependent decay for the RH resonances of either the monomer or the isolated-RH domain, RH′ unfolding was uniquely attributed to formation of the p66/p66′ homodimer. As summarized in the Data analysis section and in Supplementary Figures S8 and S9, the monomer component contribution to the three resolved RH′ resonances listed above was selectively removed by subtraction of a weighted fraction of the Ile393 monomer resonance intensity. Subtraction of the monomer resonance intensity removes this slow component, facilitating a simplified analysis using a single exponential function (Supplementary Figure S8 and Table S4). The basis for the persistence of small levels of monomer species in our studies is unclear. One possibility is that the mature p66/p66′ homodimer that contains an unfolded RH′ domain is characterized by a somewhat larger dissociation constant due to interference of the unfolded RH′ with the dimer interface. In the presence of HIV protease, the unfolded RH′ domain will be rapidly degraded (4), so this effect would not be present in the virion. The mean time constant describing the decay of the Ile434′, Ile495′ and Ile521′ resonance intensities was 6.5 ± 0.4 h (standard error), similar to the mean time constant of 6.6 ± 0.3 for the increase of the random coil resonance intensity (Supplementary Table S4).

The intensity of the Ile434E resonance arising from the p66 subunit of the homodimer has typically reached ~50% of its asymptotic value during the first 5.5 h accumulation period, and reached equilibrium with a mean time constant of 5.4 ± 0.7 h. This value is probably sufficiently close to the RH′ unfolding rate to indicate that global structural perturbations resulting from the extra RH′ domain are eliminated as the RH′ domain unfolds.

To summarize, the time-dependent NMR study of the Ile-labeled p66/p66′ homodimer reveals many of the features of the RT heterodimer, but shows additional resonances attributable to a second RH′ domain on the p66′ subunit. These resonances decay slowly over the course of the study. In addition to the slow decay of the RH′ resonances, resonances assigned to 274C, 375E and 526E exhibit slow, time-dependent changes, and are shifted to positions that characterize the RT heterodimer (compare Figures 6A and 7). Thus the maturation process is not limited to RH′ domain unfolding. A more complete discussion of the time-dependent conformational changes is beyond the scope of the present article, and will be presented elsewhere.

Destabilization of the RH′ domain by residue transfer

The basis for the RH′ domain unfolding is undoubtedly complex. One of the proposals for RH′ unfolding involves a transfer of residues from the RH′ domain to the connection′ subdomain of the p66′ subunit (2,3). As shown schematically in Figure 9A, the extended (E) conformation of the polymerase domain is compatible with a folded RH domain, while the compact, inactive (*C*) conformation of the polymerase domain, observed in the p51 subunit of the RT heterodimer, is not. This incompatibility is most clearly illustrated by considering the two alternative environments of Tyr427′ in p51′ and in RH′ (Figure 9B and C). Since it is not possible for Tyr427′ to be in two places at once, either the connection′ or the RH′ domain can be fully folded, but not both simultaneously. Thus, in order to form the p51′ subunit that is observed in numerous crystal structures of the RT heterodimer, at least several residues must be transferred from RH′ to the connection′ subdomain, with the exact number of residues being model dependent. In contrast, the monomeric structure of p66 is characterized by a partially unraveled connection domain that forms a linker to the RH domain (Figure 1C). Thus, dimerization is correlated with formation of a p66′-structure that ultimately forms a p51′-polymerase domain, requiring the residue transfer outlined above, leading to selective destabilization of the RH′ domain.

Effect of N-terminal residues on RH-domain stability

In order to evaluate the extent of RH domain destabilization resulting from the transfer of N-terminal RH-domain residues to the polymerase domain, we compared the behavior of the unmodified domain with that of a truncated construct, 'RHΔNT' in which the two N-terminal Tyr427 and Gln428 residues were removed, and residue Leu429 was mutated to methionine (Table 1). The ^{1}H-^{15}N HSQC spectrum of RHΔNT is qualitatively similar to that of wt RH (Figure 10A). Thus, the first two residues of the RH domain are not required to obtain a conformation similar to that of the wt domain. The resonance perturbations in the HSQC spectrum involve the region around Tyr427, and as well as residues of helices α_R2 and α_R4 and β_R5 (nomenclature as in Wang *et al.* (35)). These perturbations are consistent with the structure of the domain. Tyr427 is positioned against helix 4 and the sidechain hydroxyl is in H-bonding distance with Ile506 carbonyl on helix 2. The N-terminal loop lies against the edge of the β_R5 of the central β-sheet and helps to anchor it in the structure. Hence, the N-terminal segment appears to play a significant role in stabilizing the structure of the RH domain.

In order to more quantitatively evaluate the contribution of the N-terminal residues to the RH domain stability, urea denaturation studies were performed, as described by Yang *et al.* (28). The intensities of 24 well-resolved amide resonances were analyzed as a function of urea concentration, and the results indicate that the free energy of RHΔNT is reduced by 2.0 kcal/mol relative to RH ($\Delta\Delta G = 2.0 \pm 0.4$ kcal/mol). Intensity plots for six amide resonances in RH and RHΔNT are shown in Figure 10B. While this is a significant reduction in stability, unfolding of the RH domain thus appears to require more than removal of the two N-terminal residues, since it is still capable of forming a well-folded structure.

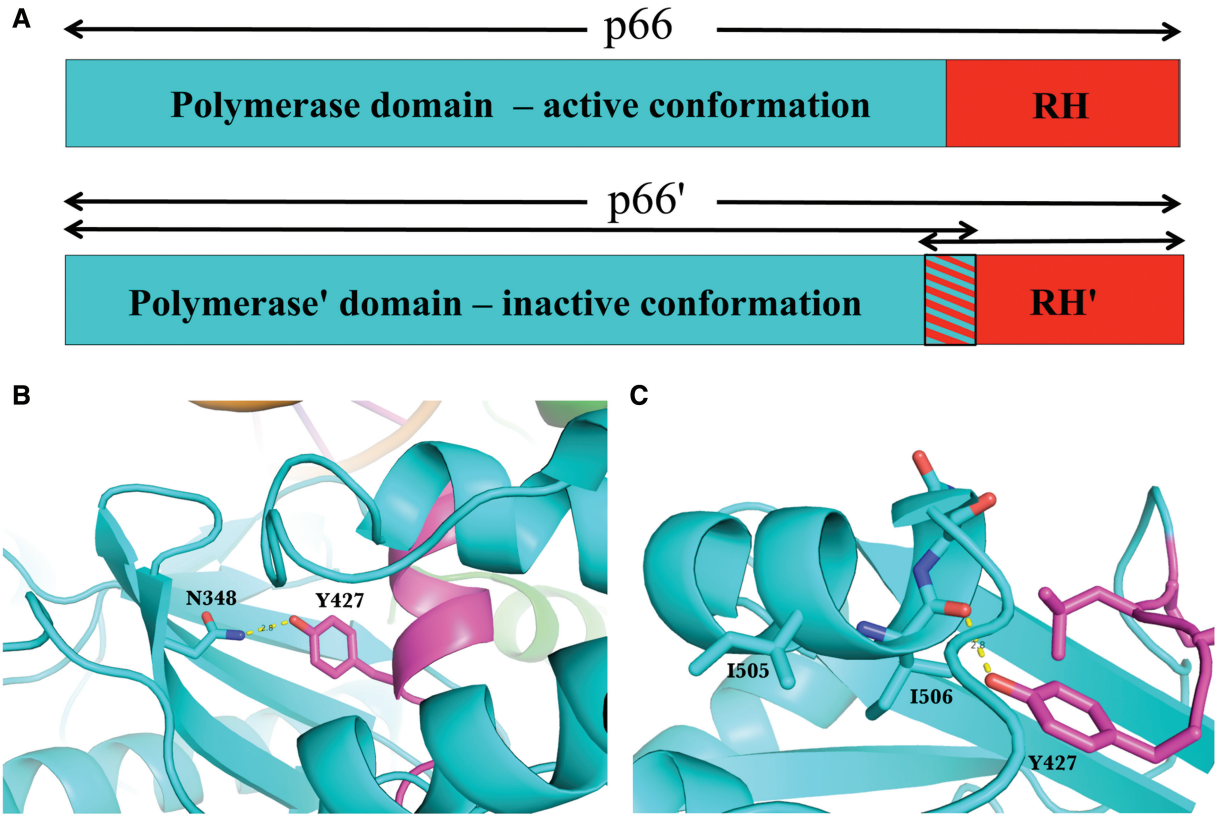

Figure 9. Schematic basis for selective RH′ domain unfolding. (**A**) Domain organization of p66 and p66′. In the p66 subunit, the polymerase domain in its extended, active conformation and the RH domain coexist. In p66′, the polymerase′ domain in the compact, inactive conformation that is observed in the p51 subunit of RT competes with the RH′ domain for common residues near Tyr427′. In panels (**B**) and (**C**), the two alternate environments of Tyr427′ in the inactively folded polymerase′ domain and the RH′ domain are shown.

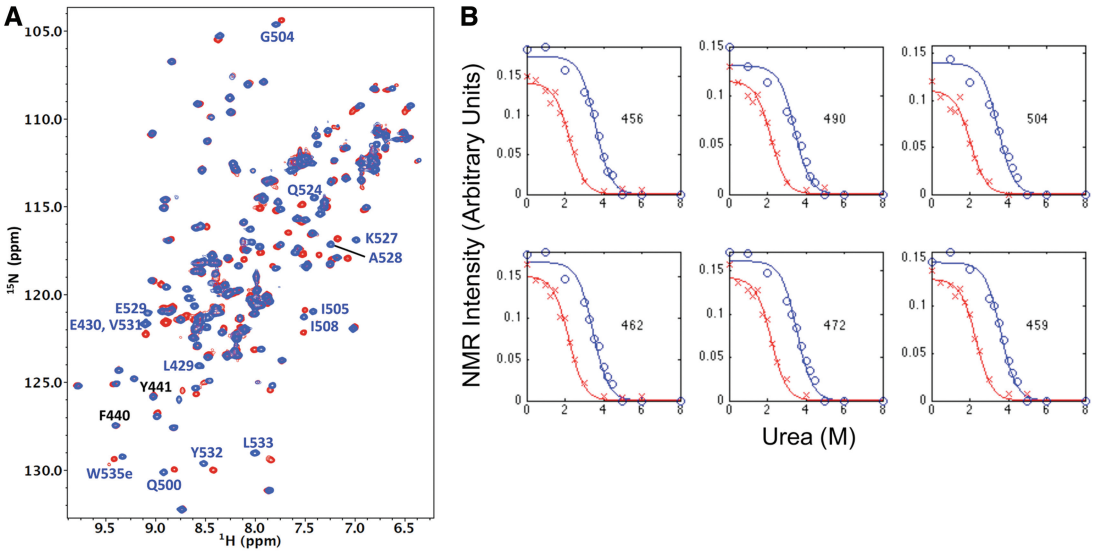

Figure 10. Effect on N-terminal residues on RH domain stability. (**A**) Overlay of the ^{1}H-^{15}N HSQC spectra of U-[^{15}N]RH (blue) with U-[^{15}N]RHΔNT (red), in which the N-terminal MNELYEL residues are replaced by M. Some of the well-resolved resonances that exhibit more significant shift changes are annotated in blue. The amide resonances at the cleavage site corresponding to F440 and Y441, are not directly affected by the deletion. (**B**) Intensities of amide resonances for the residues 456, 490, 504, 462, 472 and 459 as a function of urea concentration (RH domain, blue; RHΔNT, red). These studies indicate a significantly reduced stability of the truncated construct.

Additional destabilization likely results from further transfer of residues and from steric clashes between the destabilized RH′ and the remaining dimer structure.

DISCUSSION

It has been proposed that the maturation of the reverse transcription complex within the virion involves formation of a p66/p66′ homodimer with subsequent partial or complete unfolding of a single RH′ domain sufficient to expose the PR cleavage site, followed by formation of the p66/p51′ heterodimer. This pathway is supported by data obtained on cell lysates and virions using dimerization-deficient gag-pol mutants that contain a p66(W401A) mutation (42) and by *in vitro* studies of a dimerization deficient p66(L234A) mutant (43). However, no *in vitro* studies have directly demonstrated RH′ unfolding and the results of *in vitro* studies have generally been consistent with alternate pathways for heterodimer formation (9,11,36,44).

The present study article provides the first detailed characterization of the substrate for homodimerization: the p66 monomer. Although as predicted previously, the monomer exhibits many of the features of the p51 subunit of RT (35), it adopts a somewhat looser organization that involves three substructures: the fingers/palm/connection, the thumb and the RH domains, connected by flexible linking segments (Figure 1C). Since the interconversion of the metamorphic polymerase domain between its alternative conformations involves primarily domain rearrangements, the looser monomer structure is primed for such interconversion; only one interface—between the fingers/palm and connection subdomains, needs to be broken in order to adopt the active, E conformation.

Domain-selective unfolding was monitored using the three resolved isoleucine resonances assigned to the RH′ domain, Ile434′, Ile495′ and Ile521′. Quantitative analysis of the unfolding rate is limited by the overlap of these resonances with the corresponding resonances of the p66 monomer, however a correction based on the intensity of the Ile393 monomer resonance resulted in a mean-time constant of 6.5 h, which is similar to the time constant obtained for the time-dependent increase in the intensity of a resonance near the random coil position. This result is thus consistent with the dimerization-dependent RH′-unfolding model. We propose that this subunit-specific RH′ unfolding is a consequence of the metamorphic polymerase domain. The incompatibility of the polymerase domain in its compact, inactive conformation with a folded RH′ domain results in a tug-of-war for residues common to both structures (Figure 9). A tug-of-war induced unfolding of a model protein has been demonstrated by Peng and Li (45), and RT maturation represents a realization of this process in a critically important HIV protein. Analogous destabilization of the protein ubiquitin resulting from loss of the N-terminal methionine has previously been demonstrated (46). The spectra shown in Figure 7A unequivocally indicate that conformational asymmetry of the p66/p66′ homodimer, similar to that observed in the p66/p51 heterodimer, precedes RH′-domain unfolding, thus eliminating a class of previously proposed models in which RH′ hydrolysis precedes the establishment of the asymmetric structure observed in the heterodimer (9,36,44).

The mean lifetime of the HIV virion has been estimated as ∼7 h (47,48), and recent simulations by Konnyu *et al.* (49) show that the formation of p51 and p15 (the isolated RH domain, since the model does not include the intradomain cleavages that occur) occurs over several hours. The rates obtained in the present article are thus marginal, since for optimal infectivity the viral proteins need to be fully active at the time of infection. However, several factors present in the virion are likely to accelerate this process. First, the higher protein concentrations in the virion will favor dimer formation and hence dimerization-dependent RH′ unfolding. In the studies of Wapling *et al.* (42), mutations that interfere with p66 homodimer formation were found to reduce processing of p66 to p51 in cell lysates, but the effect was less pronounced in the virion. This variability was attributed to the ability of the higher effective concentration of p66 monomers in the virion to overcome the effects of the mutation, forming loose or transient dimers. Another catalyst for dimerization and the resulting RH′ unfolding within the virion may be the formation of the reverse transcription complex that includes tRNALys3 annealed with the viral primer binding sequence (PBS) in complex with the p66/p66′ homodimer (50–52). Specifically, it has been established that the virion accumulates tRNALys3 (51), which forms an 18-bp complex with the viral RNA PBS (50). Substrate binding typically favors protein folding and consistent with this expectation, Cabodevilla *et al.* (7) found that addition of template/primer to p66 samples resulted in a 30-fold decrease of the equilibrium dissociation constant of the homodimer. Preliminary NMR studies similar to those reported here indicate that addition of dsDNA significantly accelerates RH′ unfolding (data not shown). Since, there are only two copies of the viral PBS per virion, effective catalysis of RH′ unfolding for more than two homodimers requires equilibration of the PBS with multiple homodimers—a process that is likely to occur over the virion lifespan. A third potential contributor to homodimerization of p66 is the possible formation a dimeric structure of the uncleaved or incompletely cleaved polyprotein, as has been suggested by some research (53). Further studies of these effects are currently in progress.

At the present time, all clinically used RT inhibitors target the heterodimer. Limitations on existing drugs result from the significant toxicity associated with nucleoside RT inhibitors (NRTIs) (54) as well as the ability of the virus to develop drug-resistant mutations (55). The monomer structure and its interconversion between the *M*, *C* and *E* conformations represent additional potential targets for drug intervention. Previous studies of the polymerase domain demonstrate that a mutation can interfere with dimerization via its effect on the isomerization step (14). Several groups have proposed RT dimerization as a significant target for drug development (11,13,56). The data obtained in the present study provide additional

insights into the RT conformational maturation process that should help to guide future work in this area.

ACCESSION NUMBERS

pdb accession No.: 4KSE.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Lin Yang of the X9 beamline, at the National Synchrotron Light Source at Brookhaven National Laboratory, for assistance with data collection. Use of the X9 beamline is supported by the United States Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract DE-AC02-98CH10886. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

FUNDING

National Institute of Health (Intramural Research Program); National Institute of Environmental Health Sciences [Research Project Number Z01-ES050147 to R.E.L., and Research Project Number Z1A ES102645 to L.C.P.]; National Institutes of Health, NIEHS (Delivery Order HHSN273200700046U to E.F.D.). Funding for Open Access charge: Funding from Intramural Research Project [Number Z01-ES050147].

Conflict of interest statement. None declared.

REFERENCES

1. Jacobomolina,A., Ding,J.P., Nanni,R.G., Clark,A.D., Lu,X.D., Tantillo,C., Williams,R.L., Kamer,G., Ferris,A.L., Clark,P. *et al.* (1993) Crystal-structure of human-immunodeficiency-virus type-1 reverse-transcriptase complexed with double-stranded dna at 3.0 angstrom resolution shows bent DNA. *Proc. Natl Acad. Sci. USA*, **90**, 6320–6324.
2. Kohlstaedt,L.A., Wang,J., Friedman,J.M., Rice,P.A. and Steitz,T.A. (1992) Crystal-structure at 3.5 angstrom resolution of hiv-1 reverse-transcriptase complexed with an inhibitor. *Science*, **256**, 1783–1790.
3. Hostomska,Z., Matthews,D.A., Davies,J.F., Nodes,B.R. and Hostomsky,Z. (1991) Proteolytic release and crystallization of the rnase-h domain of human-immunodeficiency-virus type-1 reverse-transcriptase. *J. Biol. Chem.*, **266**, 14697–14702.
4. Tomasselli,A.G., Sarcich,J.L., Barrett,L.J., Reardon,I.M., Howe,W.J., Evans,D.B., Sharma,S.K. and Heinrikson,R.L. (1993) Human-immunodeficiency-virus type-1 reverse-transcriptase and ribonuclease-H as substrates of the viral protease. *Protein Sci.*, **2**, 2167–2176.
5. Davies,J.F., Hostomska,Z., Hostomsky,Z., Jordan,S.R. and Matthews,D.A. (1991) Crystal-structure of the ribonuclease-H domain of HIV-1 reverse-transcriptase. *Science*, **252**, 88–95.
6. Jacobomolina,A. and Arnold,E. (1991) HIV reverse-transcriptase structure-function-relationships. *Biochemistry*, **30**, 6351–6361.
7. Cabodevilla,J.F., Odriozola,L., Santiago,E. and Martinez-Irujo,J.J. (2001) Factors affecting the dimerization of the p66 form of HIV-1 reverse transcriptase. *Eur. J. Biochem.*, **268**, 1163–1172.
8. Beard,W.A. and Wilson,S.H. (1993) Kinetic-analysis of template.primer interactions with recombinant forms of HIV-1 reverse-transcriptase. *Biochemistry*, **32**, 9745–9753.
9. Anderson,S.F. and Coleman,J.E. (1992) Conformational-changes of HIV reverse-transcriptase subunits on formation of the heterodimer - correlation with k(Cat) and K(M). *Biochemistry*, **31**, 8221–8228.
10. Sharma,S.K., Fan,N.S. and Evans,D.B. (1994) Human-immunodeficiency-virus type-1 (HIV-1) recombinant reverse-transcriptase - asymmetry in P66 subunits of the P66/P66 homodimer. *Febs. Lett.*, **343**, 125–130.
11. Srivastava,S., Sluis-Cremer,N. and Tachedjian,G. (2006) Dimerization of human immunodeficiency virus type 1 reverse transcriptase as an antiviral target. *Curr. Pharm. Des.*, **12**, 1879–1894.
12. Divita,G., Rittinger,K., Geourjon,C., Deleage,G. and Goody,R.S. (1995) Dimerization kinetics of HIV-1 and HIV-2 reverse-transcriptase - a 2-step process. *J. Mol. Biol.*, **245**, 508–521.
13. Restle,T., Muller,B. and Goody,R.S. (1990) Dimerization of human-immunodeficiency-virus type-1 reverse-transcriptase - a target for chemotherapeutic intervention. *J. Biol. Chem.*, **265**, 8986–8988.
14. Zheng,X.H., Mueller,G.A., Cuneo,M.J., DeRose,E.F. and London,R.E. (2010) Homodimerization of the p51 subunit of HIV-1 reverse transcriptase. *Biochemistry*, **49**, 2821–2833.
15. Venezia,C.F., Meany,B.J., Braz,V.A. and Barkley,M.D. (2009) Kinetics of association and dissociation of HIV-1 reverse transcriptase subunits. *Biochemistry*, **48**, 9084–9093.
16. Braz,V.A., Holladay,L.A. and Barkley,M.D. (2010) Efavirenz binding to HIV-1 reverse transcriptase monomers and dimers. *Biochemistry*, **49**, 601–610.
17. Murzin,A.G. (2008) Biochemistry - metamorphic proteins. *Science*, **320**, 1725–1726.
18. Bryan,P.N. and Orban,J. (2010) Proteins that switch folds. *Curr. Opin. Struct. Biol.*, **20**, 482–488.
19. Kuloglu,E.S., McCaslin,D.R., Markley,J.L. and Volkman,B.F. (2002) Structural rearrangement of human lymphotactin, a C chemokine, under physiological solution conditions. *J. Biol. Chem.*, **277**, 17863–17870.
20. Unge,T., Knight,S., Bhikhabhai,R., Lovgren,S., Dauter,Z., Wilson,K. and Strandberg,B. (1994) 2.2 A resolution structure of the amino-terminal half of HIV-1 reverse transcriptase (fingers and palm subdomains). *Structure*, **2**, 953–961.
21. Zheng,X., Mueller,G.A., DeRose,E.F. and London,R.E. (2012) Metal and ligand binding to the HIV-RNase H active site are remotely monitored by Ile556. *Nucleic Acids Res.*, **40**, 10543–10553.
22. Pari,K., Mueller,G.A., DeRose,E.F., Kirby,T.W. and London,R.E. (2003) Solution structure of the RNase H domain of the HIV-1 reverse transcriptase in the presence of magnesium. *Biochemistry*, **42**, 639–650.
23. Tugarinov,V. and Kay,L.E. (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J. Am. Chem. Soc.*, **125**, 13868–13878.
24. Yang,D.W. and Kay,L.E. (1999) Improved (HN)-H-1-detected triple resonance TROSY-based experiments. *J. Biomol. NMR*, **13**, 3–10.
25. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) Nmrpipe - a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR*, **6**, 277–293.
26. Johnson,B.A. and Blevins,R.A. (1994) NMR view - a computer-program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
27. Sklenar,V., Piotto,M., Leppik,R. and Saudek,V. (1993) Gradient-tailored water suppression for H-1-N-15 HSQC experiments optimized to retain full sensitivity. *J. Magnet. Reson. Series A*, **102**, 241–245.
28. Yang,S.A., Noble,C.G. and Yang,D.W. (2009) Characterization of DLC1-SAM equilibrium unfolding at the amino acid residue level. *Biochemistry*, **48**, 4040–4049.

29. Wishart,D.S., Bigam,C.G., Holm,A., Hodges,R.S. and Sykes,B.D. (1995) H-1, C-13 and N-15 random coil NMR chemical-shifts of the common amino-acids .1. investigations of nearest-neighbor effects. *J. Biomol. NMR*, **5**, 67–81.
30. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sec. D-Biol. Crystallogr.*, **66**, 213–221.
31. Hansen,D.F., Neudecker,P. and Kay,L.E. (2010) Determination of isoleucine side-chain conformations in ground and excited states of proteins from chemical shifts. *J. Am. Chem. Soc.*, **132**, 7589–7591.
32. Han,B., Liu,Y.F., Ginzinger,S.W. and Wishart,D.S. (2011) SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR*, **50**, 43–57.
33. London,R.E., Wingad,B.D. and Mueller,G.A. (2008) Dependence of amino acid side chain 13C shifts on dihedral angle: application to conformational analysis. *J. Am. Chem. Soc.*, **130**, 11097–11105.
34. Goel,R., Beard,W.A., Kumar,A., Casasfinet,J.R., Strub,M.P., Stahl,S.J., Lewis,M.S., Bebenek,K., Becerra,S.P., Kunkel,T.A. *et al.* (1993) Structure-function studies of HIV-1(1) reverse-transcriptase - dimerization-defective mutant L289k. *Biochemistry*, **32**, 13012–13018.
35. Wang,J., Smerdon,S.J., Jager,J., Kohlstaedt,L.A., Rice,P.A., Friedman,J.M. and Steitz,T.A. (1994) Structural basis of asymmetry in the human-immunodeficiency-virus type-1 reverse-transcriptase heterodimer. *Proc. Natl Acad. Sci. USA*, **91**, 7242–7246.
36. Lowe,D.M., Aitken,A., Bradley,C., Darby,G.K., Larder,B.A., Powell,K.L., Purifoy,D.J.M., Tisdale,M. and Stammers,D.K. (1988) Hiv-1 reverse-transcriptase - crystallization and analysis of domain-structure by limited proteolysis. *Biochemistry*, **27**, 8884–8889.
37. Bernado,P., Mylonas,E., Petoukhov,M.V., Blackledge,M. and Svergun,D.I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.*, **129**, 5656–5664.
38. Bauman,J.D., Das,K., Ho,W.C., Baweja,M., Himmel,D.M., Clark,A.D., Oren,D.A., Boyer,P.L., Hughes,S.H., Shatkin,A.J. *et al.* (2008) Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design. *Nucleic Acids Res.*, **36**, 5083–5092.
39. Hsiou,Y., Ding,J., Das,K., Clark,A.D., Hughes,S.H. and Arnold,E. (1996) Structure of unliganded HIV-1 reverse transcriptase at 2.7 angstrom resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure* **4**, 853–860.
40. Dufour,E., El Dirani-Diab,R., Boulme,F., Fournier,M., Nevinsky,G., Tarrago-Litvak,L., Litvak,S. and Andreola,M.L. (1998) p66/p51 and p51/p51 recombinant forms of reverse transcriptase from human immunodeficiency virus type 1 - interactions with primer tRNA(Lys3), initiation of cDNA synthesis, and effect of inhibitors. *Eur. J. Biochem.*, **251**, 487–495.
41. Venezia,C.F., Howard,K.J., Ignatov,M.E., Holladay,L.A. and Barkley,M.D. (2006) Effects of efavirenz binding on the subunit equilibria of HIV-1 reverse transcriptase. *Biochemistry*, **45**, 2779–2789.
42. Wapling,J., Moore,K.L., Sonza,S., Mak,J. and Tachedjian,G. (2005) Mutations that abrogate human immunodeficiency virus type 1 reverse transcriptase dimerization affect maturation of the reverse transcriptase heterodimer. *J. Virol.*, **79**, 10247–10257.
43. Sluis-Cremer,N., Arion,D., Abram,M.E. and Parniak,M.A. (2004) Proteolytic processing of an HIV-1 pol polyprotein precursor: insights into the mechanism of reverse transcriptase p66/p51 heterodimer formation. *Int. J. Biochem. Cell Biol*, **36**, 1836–1847.
44. Abram,M.E. and Parniak,M.A. (2005) Virion instability of human immunodeficiency virus type 1 reverse transcriptase (RT) mutated in the protease cleavage site between RT p51 and the RT RNase H domain. *J. Virol*, **79**, 11952–11961.
45. Peng,Q. and Li,H.B. (2009) Direct observation of tug-of-war during the folding of a mutually exclusive protein. *J. Am. Chem. Soc.*, **131**, 13347–13354.
46. Bamezai,S., Banez,M.A.T. and Breslow,E. (1990) Structural and functional-changes associated with modification of the ubiquitin methionine. *Biochem*, **29**, 5389–5396.
47. Perelson,A.S., Neumann,A.U., Markowitz,M., Leonard,J.M. and Ho,D.D. (1996) HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time. *Science*, **271**, 1582–1586.
48. Perelson,A.S. and Nelson,P.W. (1999) Mathematical analysis of HIV-1 dynamics *in vivo*. *Siam Rev.*, **41**, 3–44.
49. Konnyu,B., Sadiq,S.K., Turanyi,T., Hirmondo,R., Muller,B., Krausslich,H.G., Coveney,P.V. and Muller,V. (2013) Gag-Pol processing during HIV-1 virion maturation: a systems biology approach. *Plos Comput. Biol.*, **9**, e1003103.
50. Derocquigny,H., Gabus,C., Vincent,A., Fourniezaluski,M.C., Roques,B. and Darlix,J.L. (1992) Viral-RNA annealing activities of human-immunodeficiency-virus type-1 nucleocapsid protein require only peptide domains outside the Zinc fingers. *Proc. Natl Acad. Sci. USA*, **89**, 6472–6476.
51. Jiang,M., Mak,J., Ladha,A., Cohen,E., Klein,M., Rovinski,B. and Kleiman,L. (1993) Identification of transfer-RNAs incorporated into wild-type and mutant human-immunodeficiency-virus type-1. *J. Virol.*, **67**, 3246–3253.
52. Warrilow,D., Tachedjian,G. and Harrich,D. (2009) Maturation of the HIV reverse transcription complex: putting the jigsaw together. *Rev. Medical Virol.*, **19**, 324–337.
53. Figueiredo,A., Moore,K.L., Mak,J., Sluis-Cremer,N., de Bethune,M.P. and Tachedjian,G. (2006) Potent nonnucleoside reverse transcriptase inhibitors target HIV-1 Gag-Pol. *Plos Pathog.*, **2**, 1051–1059.
54. Lewis,W. and Dalakas,M.C. (1995) Mitochondrial toxicity of antiviral drugs. *Nat. Med.*, **1**, 417–422.
55. Menendez-Arias,L. (2010) Molecular basis of human immunodeficiency virus drug resistance: an update. *Antiv. Res.*, **85**, 210–231.
56. Agopian,A., Gros,E., Aldrian-Herrada,G., Bosquet,N., Clayette,P. and Divita,G. (2009) A new generation of peptide-based inhibitors targeting HIV-1 reverse transcriptase conformational flexibility. *J. Biol. Chem.*, **284**, 254–264.