4764–4773 *Nucleic Acids Research, 2009, Vol. 37, No. 14* *Published online 10 June 2009*
doi:10.1093/nar/gkp485

DNA binding and synapsis by the large C-terminal domain of ϕC31 integrase

Andrew R. McEwan, Paul A. Rowley and Margaret C. M. Smith*

Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2BX, UK

Received February 18, 2009; Revised May 16, 2009; Accepted May 19, 2009

ABSTRACT

The integrase (Int) from phage ϕC31 acts on the phage and host-attachment sites, *attP* and *attB*, to form an integrated prophage flanked by *attL* and *attR*. Excision (*attL* × *attR* recombination) is prevented, in the absence of accessory factors, by a putative coiled-coil motif in the C-terminal domain (CTD). Int has a serine recombinase N-terminal domain, required for synapsis of recombination substrates and catalysis. We show here that the coiled-coil motif mediates protein–protein interactions between CTDs, but only when bound to DNA. Although the histidine-tagged CTD (hCTD) was monomeric in solution, hCTD bound cooperatively to three of the recombination substrates (*attB, attL* and *attR*). Furthermore, when provided with *attP* and *attB*, hCTD brought these substrates together in a synaptic complex. Substitutions in the coiled-coil motif that greatly reduce Int integration activity, L460P and Y475H, prevented CTD–CTD interactions and led to defective DNA binding and no detectable DNA synapsis. A substitution, E449K, in full length Int confers the ability to perform excision in addition to integration as it has gained the ability to synapse *attL* × *attR*. hCTD^{E449K} was similar to hCTD in DNA binding but unable to form the CTD synapse suggesting that the CTD synapse is not essential but could be part of the mechanism that controls directionality.

INTRODUCTION

The integrase (Int) from the *Streptomyces* temperate phage, ϕC31, is widely used as a tool for genome engineering in model eukaryotes but its mechanism of action is still poorly understood (1). ϕC31 Int and many of its relatives offer an advantage over other site-specific recombinases in genome engineering applications as they are unidirectional in the absence of any accessory factors (2–4). Under these conditions Int can only recombine the integration substrates, *attP* and *attB* located normally on the phage and host chromosomes, respectively, to form *attL* and *attR*. For excision of the prophage, Int causes recombination between *attL* and *attR* to regenerate *attP* and *attB* in the presence of an accessory protein (5).

An early step in ϕC31 Int mediated recombination is the formation of the synapse, a nucleoprotein complex, which contains two recombination substrates held together by a presumed tetramer of recombinase subunits (6–8). We have shown that purified Int only synapses *attP* and *attB* and cannot synapse any other pair of substrates (7,8). This selectivity is the basis for its unidirectionality and is thought to be determined by different conformations of Int bound to its substrates (6–9).

ϕC31 Int is a large serine recombinase and thus possesses an N-terminal domain (NTD) that is conserved in all the serine recombinases (10). The NTDs provide the catalytic activity for DNA cleavage, strand exchange and joining of the recombination products and mediate protein–protein interactions that are essential for DNA synapsis (11,12). The activity of the large C-terminal domains (CTDs) of the large serine recombinases is less well understood although it is clear that this domain is required for DNA binding and for the control of integration and excision (6,13). Mutations in ϕC31 Int within a putative coiled-coil subdomain of the CTD gained the ability to recombine *attL* × *attR* in the absence of an accessory factor, although several of these mutants are also strongly defective in recombination activity generally (13). The hyperactive nature of these mutants was thought to be due to a defect in the ability of Int to inhibit *attL* × *attR* synapsis and recombination. In this report we present evidence that the coiled-coil motif mediates protein–protein interactions when the CTDs are bound to DNA. CTD–CTD interactions are required for stable DNA binding and are likely to be part of the mechanism that controls directionality.

MATERIALS AND METHODS

Strains, cultures and DNA manipulations

Escherichia coli DH5α was used for DNA manipulation and plasmid preparations. Plasmid pARM014, which

*To whom correspondence should be addressed. Tel: +44 1224 555739; Fax: +44 1224 555844; Email: maggie.smith@abdn.ac.uk

expresses residues 155–605 of ϕC31 Int, fused to a 6 × histidine tag and a TEV protease sequence at the N-terminus, was constructed by ligating a polymerase chain reaction (PCR) amplified DNA fragment (using pHS62 as the template and primers ARM6; 5′GCG AACATGTTCGACACGAAGAACCTTCAGCG, and ARM5; 5′GGCTTCTCGAGTGTCTCGCTACGCCGC) cut with PciI/XhoI to pEHISTEV cut with NcoI/XhoI (4,15,16). DNA fragments encoding the putative coiled-coil region (amino acids 445–519) were amplified by PCR (using pHS62 as the template and primers PRXCOIL5; 5′GCCATGGTGGGCAAGCTCACTGAG GCG, PRXCOIL2; 5′GCAAGCTTCCTAAACGGGTC CGTCGTACGC, PRXCOIL6; 5′GCCATGGTGGCAG GCGCGTACGACGGA, PRXCOIL4; 5′GCAAGCTTC TAGTCAAGGGGAAGTTTCGG) cut with NcoI and HindIII and ligated to NcoI/HindIII cut pLOU3 (New England Biolabs; Dr Louise Major) to generate pPARα (PRXCOIL5 and PRXCOIL2), pPARβ (PRXCOIL6 and PRXCOIL4) and pPARγ (PRXCOIL5 and PRXCOIL4). Amino-acid substitutions were introduced into pARM014 by site-directed mutagenesis (Quikchange, Stratagene) and pPARγ3, pPARγ5, pPARγ6 and pPARγ7 were constructed as for pPARγ except that the templates for PCR were replaced with pPAR57, pPAR25, pPAR12 and pPAR101 respectively (13).

Protein expression and purification

To express hCTD, *E. coli* BL21(DE3), pLysS, pARM014 was grown in LB supplemented with 30 μg ml^{-1} kanamycin at 20°C until A_{600} reached 0.6. Isopropyl β-D-thiogalactopyranoside (IPTG) was added (final concentration of 125 μM) and incubation continued overnight. The cell pellet was washed with phosphate buffered saline (PBS) and repelleted. Cells were sonicated in lysis buffer [25 mM Tris–HCl pH 7.75, 0.6 M NaCl, 10 mM imidazole, 20 μM lysozyme, 0.5 mM 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF), 10 μM *N*-(*trans*-epoxysuccinyl)-L-leucine 4-guanidinobutylamide (E-64), 2 mM Benzamidine]. The supernatant from the crude lysate was filtered (0.22 μm; Millipore), loaded onto a 5 ml His–Trap™ FF column, (GE Healthcare) and washed with wash buffer (25 mM Tris–HCl pH 7.75, 0.6 M NaCl, 50 mM imidazole). The protein was eluted with elution buffer (25 mM Tris–HCl pH 7.75, 0.6 M NaCl, 250 mM imidazole), equilibrated in size exclusion buffer (20 mM Tris–HCl pH 8.0, 500 mM NaCl, 10 mM β-mercaptoethanol) and applied onto a HiPrep 16/60 Sephacryl S-200 size exclusion chromatography column (GE Healthcare). Peak fractions were combined, diluted into storage buffer [20 mM Tris–HCl pH 8.0, 500 mM NaCl, 2 mM *tris*(2-carboxyethyl)phosphine hydrochroride (TCEP)] and kept at −20°C after addition of 50% glycerol.

The hMBP-Int fusions derived from pPARα, pPARβ pPARγ, pPARγ 3,5,6,7 were expressed as described for the hCTD except that the cultures contained 30 μg ml^{-1} carbenicillin and chloramphenicol, were incubated at 30°C and the final IPTG concentration was 300 μM. Cell lysis was in PBS pH 7.4, 10 mM imidazole and protease inhibitor (Complete™, Roche). The supernatant from the crude lysate was filtered (0.22 μm filter; Millipore) and loaded onto a 1 ml His–Trap™ HP, (GE Healthcare) pre-equilibrated with SK buffer (20 mM Sodium phosphate, 500 mM KCl, pH 7.4) with 10 mM imidazole. The column was washed with the same buffer and eluted with SK buffer containing 250 mM imidazole. Peak (A_{280}) fractions were collected and stored at 4°C.

Analytical size-exclusion chromatography (ASEC)

ASEC was performed on an AKTAprime™ (GE Healthcare) coupled to a UV detector. The purified hCTD and mutants (10–20 μM) were loaded onto a Superose 6 HR 10/300 column (GE Healthcare) pre-equilibrated with storage buffer (as for purification). The apparent molecular weight was calculated from the peaks obtained at A_{280} compared to molecular weight standards (Bio-Rad). ASEC of the purified hMBP-coiled-coil fusion proteins was performed as described for the hCTDs except that protein (19–56 μM) was loaded onto a Superdex 200 column (GE Healthcare) that had been pre-equilibrated and cooled with ice cold 20 mM sodium phosphate buffer (pH 7.4).

Wild-type ϕC31 Int and the catalytically inactive derivative IntS12A were purified from *E. coli* BL21 (DE3) pLysS containing pHS62 or pMSX6 as described previously (8).

Protein crosslinking

Protein samples were diluted in 20 mM phosphate buffer (pH 7.4) to a concentration of 2.5 μM and incubated with 0.1% glutaraldehyde for up to 5 min. Reactions were stopped with the addition of a 1:1 ratio of 5% β-mercaptoethanol and SDS–PAGE sample buffer, followed by incubation for 5 min at 95°C. Products were analysed by SDS–PAGE.

DNA-binding assays

DNA-binding assays were performed as describe previously (13). Briefly, DNA fragments or annealed oligonucleotides were labelled by incorporation of [α-^{32}P] dCTP with DNA polymerase (Klenow exo$^-$ fragment). Proteins were added to the labelled probe (~1.5 nM) and incubated at 30°C for 30 min. For synapse assays an unlabelled fragment (~32 nM) prepared by PCR was added to the reactions and incubated at 30°C for 2 h. After incubation the sample was separated by electrophoresis in a non-denaturing 4% polyacrylamide gel. Radioactive bands were detected from the dried gels after exposing to a Fuji phosphor image plate and scanning with a Fuji FLA-3000 phosphorimager.

RESULTS

A motif in the CTD that controls the directionality of recombination mediates protein–protein interactions

The region 445–524 is predicted to form two α helices H1 and H2 separated by a short linker (Figure 1) (13). Based on a COILS prediction the first of these helices, H1, has a high probability of forming coiled-coil whereas the

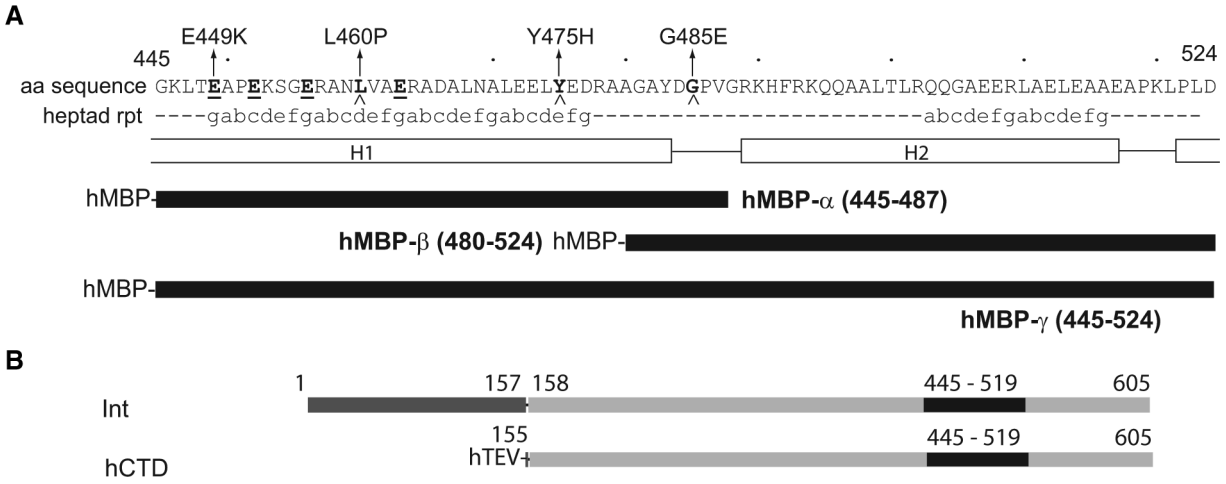

Figure 1. Constructs containing the putative coiled-coil motif and the CTD of ϕC31 Int. (**A**) The sequence of Int from amino acids 445–524 is shown with the heptad repeat (positions a–g) of the predicted coiled-coil underneath. The coiled-coil prediction was described previously (13) and was performed using the COILS programme (17,22). The extent of predicted α-helix is shown by the boxes H1 and H2 (23). Underlined residues and those indicate by arrowheads indicate where mutation can give hyperactive or defective Ints, respectively. Solid black lines show the amino acid sequence fused to the hMBP with the name of each construct (and amino acid coordinates) in bold. (**B**) The domain structure of ϕC31 Int and the hCTD construct. The NTD (dark grey) ends with a protease sensitive site at K157. The CTD (light grey) contains the putative coiled-coil motif (black). hTEV is the histidine tag and TEV protease site fused at the N-terminus to Int residues 155–605.

probability of H2 being coiled-coil was much lower (13,17). Several substitutions in this motif (E449K, E452K, E456K, E463K) were shown previously to cause hyperactivity, allowing these mutant Ints to recombine *attL* × *attR* in addition to *attP* × *attB* (13). Other mutations (L460P and Y475H) in this putative coiled-coil region were strongly defective in *attB* × *attP* recombination (13). Coiled-coil secondary structures often mediate protein–protein interactions and so the ability of the 445–524 regions to oligomerize was investigated. Three constructs were made, hMBP-α, hMBP-β and hMBP-γ, that respectively contained H1, H2 and H1 and H2 fused to a histidine-tagged maltose-binding protein (hMBP). hMBP, hMBP-α and hMBP-β all behaved as monomers by size exclusion chromatography, whereas hMBP-γ behaved as a dimer (Figure 2A). These observations were confirmed by glutaraldehyde crosslinking (Figure 2C). Thus both the predicted H1 and H2 helical regions are required for the efficient dimerization of the motif. The mutations from IntE449K, IntL460P and IntY475H were introduced into hMBP-γ. hMBP-γ^{E449K} behaved in a similar manner to hMBP-γ in size exclusion chromatography but hMBP-γ^{Y475H} and hMBP-γ^{L460P} both lost the ability to dimerize (Figure 2B and D). A substitution, G485E, in the putative linker between H1 and H2 was also tested (Figure 1A) as IntG485E was defective for *attP* × *attB* recombination but able to recombine *attL* × *attR* at very low levels in *in vivo* recombination assays (data not shown). hMBPγ^{G485E} was also unable to dimerize (Figure 2B and D). These data show that, in isolation, the putative coiled-coil motif mediates protein–protein interactions and this interaction is required for Int function.

The isolated hCTD is a monomer

Serine recombinases have a conserved dimerization motif in the NTDs (6,12,18). Protein–protein interactions between CTDs will be masked by this interface so a fragment of Int containing residues 155–605 fused to an N-terminal histidine tag and a TEV protease site (hCTD) that lacks the NTD dimer interface was purified.

In size exclusion chromatography the hCTD behaved as a monomer, eluting as a 52-kDa protein (Figure 3). The monomeric nature of the hCTD is consistent with the observation that the isolated CTD from the related serine Int from phage Bxb1 is also monomeric (6). Amino-acid substitutions E449K, L460P and Y475H were introduced into hCTD. hCTD^{Y475H} behaved in a similar manner to the wild-type hCTD in size exclusion chromatography, whilst hCTD^{E449K} and hCTD^{L460P} had predicted molecular weights of 79 and 84 kDa, respectively (Figure 3). The lower retention volumes observed with hCTD^{E449K} and hCTD^{L460P} most likely reflect differences in protein conformation rather than the acquisition of an oligomerization interface as full length, IntL460P and IntY475H were indistinguishable from native Int in size exclusion chromatography (13). These data suggest that the putative coiled-coil motif is either buried or sequestered when Int is free in solution.

Mutations in the putative coiled-coil motif disrupt DNA binding by hCTD

When ϕC31 Int binds to its attachment sites two complexes are observed (7) (Figures 4G, 5A and D). As free Int is a dimer, we propose that the more abundant and lower mobility complex contains a DNA bound dimer, with each Int subunit contacting one of the two-half sites of each attachment site (7) (Figure 6). Complex I, which is always in much lower abundance and has a higher mobility than complex II, most likely contains a monomer of Int bound to one of the two-half sites (7). When the hCTD was used in binding assays with the attachment sites low (complex II) and high (complex I)

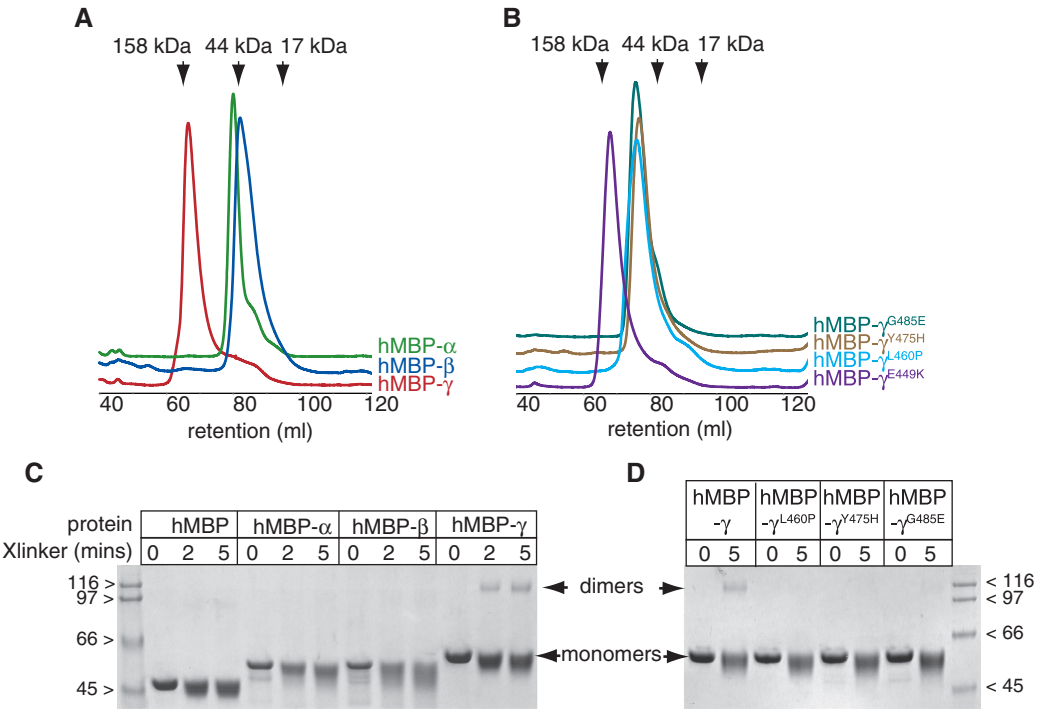

Figure 2. The putative coiled-coil motif mediates protein–protein interactions. (**A**) Gel filtration traces of the hMBP fusions to fragments of Int containing the putative coiled-coil motif. Standards (BioRad) were run before and after the samples. The calculated monomeric molecular weights for hMBP-α, hMBP-β and hMBP-γ were 50 kDa, 50 kDa and 54 kDa, respectively, and the apparent molecular weights were 52 kDa, 36 kDa and 125 kDa, respectively. (**B**) Gel filtration of hMBP-γ containing mutations which had no effect (hMBP-γ^{E449K}) or prevented oligomerization (hMBP-γ^{L460P}, hMBP-γ^{Y475H} and hMBP-γ^{G485E}). The apparent molecular weights of these constructs were 125 kDa, 63 kDa, 58 kDa and 67 kDa, respectively. (**C** and **D**) Glutaraldehyde crosslinking of purified hMBP and hMBP fusions to fragments containing the putative coiled-coil motif. After treatment with glutaraldehyde the proteins were separated by SDS–PAGE and stained with Coomassie blue. Molecular weight standards were from BioRad.

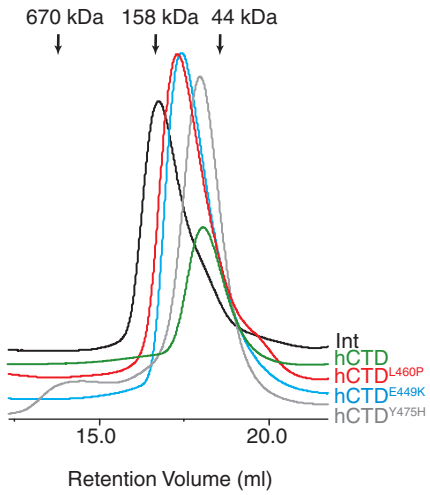

Figure 3. Size exclusion chromatography of the isolated hCTD and mutant derivatives. Gel filtration traces of hCTD (green), hCTD^{Y475H} (grey), hCTD^{L460P} (red) and hCTD^{E449K} (light blue) and wild-type Int (black). The calculated momomeric form of the hCTD is 54 kDa. The apparent molecular weights for hCTD, hCTD^{E449K}, hCTD^{L460P} and hCTD^{Y475H} were 52 kDa, 79 kDa, 84 kDa and 55 kDa, respectively. Full length Int eluted with an apparent molecular weight of 121 kDa (predicted monomer is 67 kDa).

mobility complexes were also observed (Figure 4). The binding affinities for the four attachment sites by the hCTD were similar to those obtained with wild-type Int (Figure 4) (7,13). As hCTD was monomeric in solution, we expected the hCTD to occupy the two half sites independently on the basis of concentration and affinity. This appeared to be the case for *attP* binding where complex I is the most abundant complex up to 83 nM (Figure 4C and D). In contrast the hCTD appeared to bind to *attB*, *attL* and *attR* cooperatively, where even at low concentrations of protein the majority of bound probe was in complex II (Figure 4A, B, E and F). Indeed, binding by the hCTD to *attB* appeared to be dependent on co-operativity as when a probe (BX^{50}), containing an *attB* site in which one arm was heavily mutated to prevent binding, was used in a binding assay, the binding affinity by hCTD was severely reduced compared to an *attB*^{50} probe (Figure 4G). Under the same conditions the affinity of native Int for BX^{50} was only mildly affected (Figure 4G). The cooperative binding by hCTD to *attB*, *attL* and *attR* could be explained if protein–protein interactions occurred between adjacently bound hCTDs or if binding to a half site by hCTD altered the DNA conformation to favour binding of a second molecule to the adjacent half site.

If DNA binding depends on CTD–CTD interactions mediated by the putative coiled-coil region, mutations in this motif should affect the DNA-binding properties of the CTDs. hCTD^{L460P} bound to the attachment sites with either reduced co-operativity and/or reduced affinity (Figure 4A, C, E and F) and hCTD^{Y475H} binding displayed greatly reduced affinities (Figure 4A and C and data not shown). The simplest explanation of these data

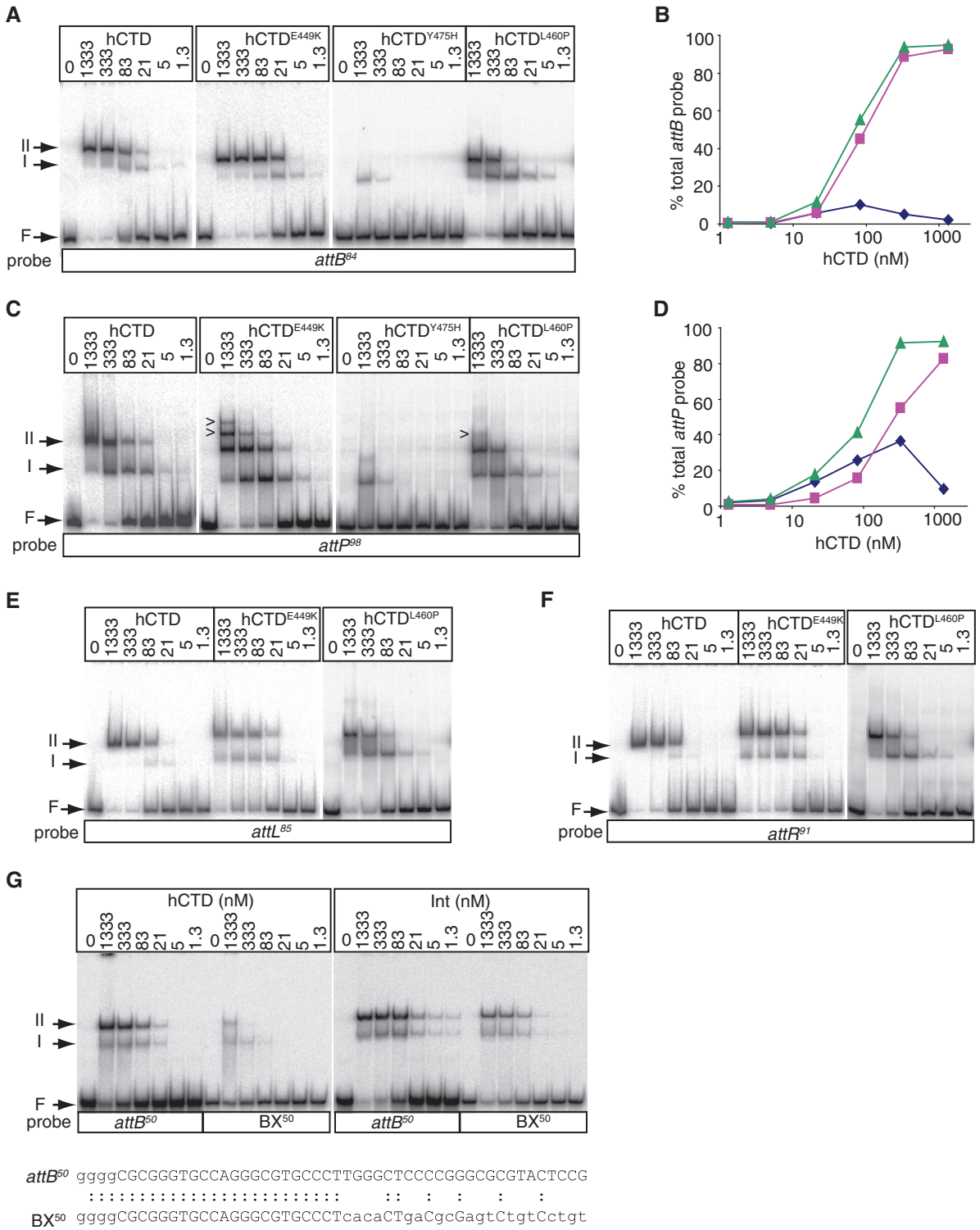

Figure 4. Binding by hCTD and mutant derivatives to DNA substrates. DNA binding by decreasing concentrations (nM) of hCTD and mutant derivatives to *attB* (**A**), *attP* (**C**), *attL* (**E**) and *attR* (**F**) was assayed by band shift assays in non-denaturing polyacrylamide gels. The numbers in superscript indicate the length of the DNA probe. F is the free probe. The positions of complex I (I) and complex II (II) are indicated. The arrowheads indicate possible trimers and tetramers of protein bound to the *attP* probe. (**B** and **D**) The amount of radioactivity in complex I (closed diamonds), complex II (closed squares) and complexes I + II (closed triangles) as a percentage of the total probe in the lane was calculated and plotted against the concentration of hCTD in the binding assay. (**G**) Binding by hCTD and Int to a 50 bp labelled *attB* probe ($attB^{50}$) or a 50 bp labelled probe in which the right half site (B′) has been extensively mutated (BX^{50}). The sequences of the top strands of each probe are shown as an alignment. The colon indicates identical base pair in both probes.

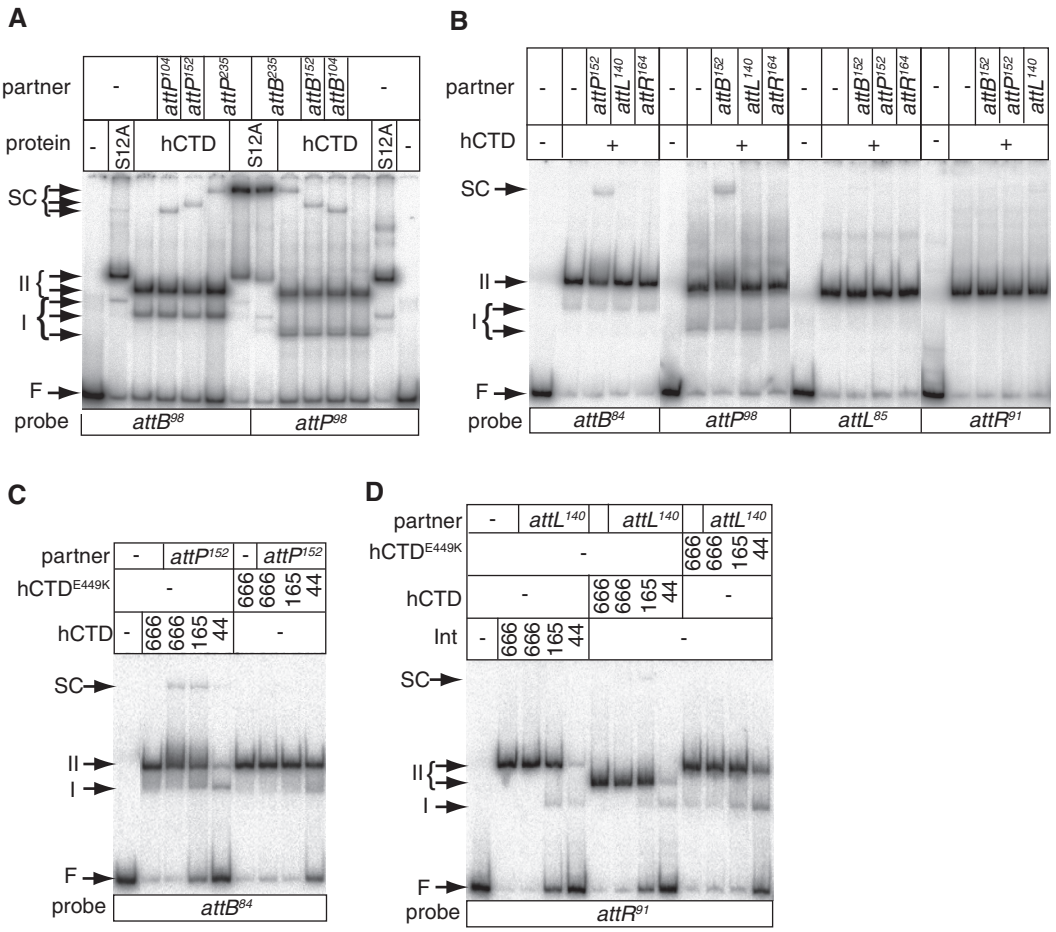

Figure 5. Synapsis of *attP* and *attB* by hCTD. Formation of the synaptic complexes were assayed by band shift assays in non-denaturing polyacrylamide gels. (**A**) Comparison of synaptic complexes obtained with full length IntS12A (S12A) and hCTD. Addition of IntS12A or hCTD to *attB* or *attP* resulted in complexes I and II and addition of an unlabelled partner *att* sites resulted in the formation of supershifts or synaptic complexes (SC) (8). (**B**) The hCTD synaptic complexes were only obtained when provided with *attP* and *attB*; other pairs of recombination sites did not supershift the probe. (**C** and **D**) Inability of $hCTD^{E449K}$ to form synaptic complexes with an *attP* and *attB* pair (C) or with an *attL* and *attR* pair (D). Three concentrations of the hCTD, $hCTD^{E449K}$ and, in D only, Int (666, 165 and 44 nM) were used. The numbers in superscript indicate the lengths of the DNA fragments used as probes or as partners.

is that interactions mediated by the putative coiled-coil motif are required for DNA binding. IntY475H was shown previously to bind to *att* sites with equal or greater affinities than native Int (13). It follows that the NTD stabilizes DNA binding, either by a direct interaction between the CTD and NTD or because of the dimeric nature of IntY475H (13). DNA binding was not greatly affected in $hCTD^{E449K}$, although some additional complexes were observed with *attP* (see below). As introduction of the E449K mutation into the hMBP-γ did not affect oligomerization, this observation is consistent with the requirement for the coiled-coil interactions to stabilize DNA binding.

Synapsis of *attP* and *attB* by hCTD

We have previously shown that Int forms a synaptic complex with *attP* and *attB* and that if a catalytically inactive Int, IntS12A, is used the synaptic complexes accumulate to a high level (8,13). This complex occurs only if IntS12A is provided with *attP* and *attB*; no other pair of *att* sites form a synaptic complex (8). As the coiled-coil motif was proposed to have a role in synapsis (13), we tested whether the isolated hCTD could synapse *attP* with *attB* (Figure 5). In super-shift assays with a radiolabelled DNA probe (either *attB* or *attP*), an unlabelled (235 bp) partner site and the hCTD, complexes of similar mobility to the IntS12A synaptic complex were detected (Figure 5A). The presence of the unlabelled partner site in the putative CTD synapse was confirmed by observing decreasing mobility of the complex in response to the increasing size of the unlabelled site. The synaptic complex formed by hCTD did not accumulate to as high levels as with IntS12A, and this is consistent with the NTD contributing to the stability of a full synaptic interface. As seen with full length Int, hCTD only brought DNA encoding *attP* and *attB* together in a synapse; other pairs of sites did not give rise to stable supershifted complexes (Figure 5B).

We have shown previously that IntS12A,L460P and IntS12A,Y475H do not synapse *attP* and *attB*.

As expected hCTD^{L460P} and hCTD^{Y475H} were also unable to synapse *attP* and *attB* (data not shown). Thus mutations that prevent the protein–protein interaction by the putative coiled-coil motif also prevented hCTD synapsis.

Surprisingly E449K, a substitution that enables *attL* × *attR* synapsis in full length Int, also inhibited synapsis of the hCTDs; no supershifted complexes could be observed with isolated hCTD^{E449K} with *attP*/*attB* or *attL*/*attR* pairs (Figure 5C and D). However hCTD^{E449K} does appear to be able to oligomerize when bound to *attP*, possibly forming trimers and tetramers, even at low protein concentration (Figure 4C). Higher order complexes were just detectable at the highest protein concentration with native Int, hCTD and hCTD^{L460P} with *attP* (Figure 4C).

These data show that the isolated hCTD can bring *attP* and *attB* specifically together in a synaptic complex and that the ability to synapse is disrupted by mutations in the coiled-coil motif.

DISCUSSION

The data presented here show that the putative coiled-coil motif in the large CTD of ϕC31 Int provides a protein interaction interface that is required for substrate recognition and, consequently, *attP* × *attB* synapsis. Two substitutions L460P and Y475H, shown previously to reduce recombination activity in full length Int, abolish oligomerization of the coiled-coil (13) (Figure 2). Both residues are predicted to be directly involved in the coiled-coil interactions (19). L460 occupies the 'd' position in the heptad repeat predicted by the COILS algorithm (17) (Figure 1A). In conjunction with 'a' residues, amino acids at 'a' and 'd' form a hydrophobic core in typical coiled-coil interactions (19). The substitution L460P would not only affect a hydrophobic interface, but would also disrupt the α-helix in H1. Y475H is likely to disrupt predicted polar or charged interactions between residues at the 'e' position of the heptad repeat that, along with 'g' residues where several strongly hyperactive mutations map, influence stability and oligomerization (Figure 1A) (19). Thus L460 and Y475 probably have direct roles in the observed coiled-coil interactions.

While the hMBP-γ construct containing the isolated coiled-coil motif fused to maltose-binding protein oligomerized, the isolated hCTD, which also contains the putative coiled-coil, behaved as a monomer in size exclusion chromatography (Figure 3). This implies that the coiled-coil motif, in the context of free hCTD and possibly also Int, is unable to mediate protein–protein interactions. We propose that the coiled-coil interactions can occur, however, between adjacently bound hCTDs on binding to *attB*, *attL* and *attR*. Binding by hCTD to these sites is cooperative and mutations in the coiled-coil motif (L460P and Y475H) reduce co-operativity and binding affinity. Binding by hCTD to *attP* was much less cooperative, suggesting that the coiled-coil domain does not interact between adjacent hCTDs bound to *attP*. Previously DNAseI footprinting showed the extent of protection by Int on *attP*, *attB*, *attL* and *attR* (7). Int bound to *attP* has a longer footprint than Int bound to *attB* and this is consistent with the reported minimal functional sites for *attP* (39 bp) and *attB* (34 bp) (20). Together these observations suggest that Int binds to *attP* and *attB* with different spacing relative to the crossover site. Possibly the coiled-coil motifs are too far apart in the *attP* bound subunits to interact with each other (Figure 6). These observations show that the hCTD complex with *attP* is clearly different from the hCTD complex with *attB* and this supports our proposition that it is the specific conformations of Int incurred by binding to *attP* and *attB* that determines whether the substrates can synapse and recombine.

Our model suggests that the putative coiled-coil becomes exposed when the hCTD binds to DNA such that, if the DNA contains *attB*, *attL* or *attR* interactions occur between adjacently bound CTDs of Int (Figure 6A(i)]. The coiled-coil motifs in the Int subunits bound to *attP* may be unpaired and these could initiate CTD interactions to form a tetramer, bringing the *attP* and *attB* sites together in a synapse formed by interacting CTDs [Figure 6A(ii)]. In other serine recombinases synapsis via the conserved NTDs leads to activation and catalysis (11). It seems likely that the productive or full ϕC31 Int synapse also requires a synaptic interface mediated through the NTDs [Figure 6A(iii)]. The CTD synapse could be an intermediate complex that is required to relieve a proposed block to the formation of the NTD synaptic interface (see below). The hCTD tetrameric interface only forms in the presence of hCTD bound to *attP* and *attB* and this specific selection of substrates for synapsis is reminiscent of full length Int. Thus the CTD synapse is likely to be part of the mechanism that discriminates against the use of other pairs of substrates for recombination.

The data presented here lead us to conclude that the CTD, possibly the coiled-coil itself, inhibits the formation of the tetrameric interface at the NTD unless both *attP*-bound and *attB*-bound Int are present. Previous data showed that introduction of E449K to a mutant in the NTD (IntV129A) that was defective in synapsis could partially rescue the activity, pointing towards an interaction between the CTD and the NTD which acts at synapsis (14). E449 lies at the start of the putative coiled-coil motif and IntE449K was shown previously to be as active as wild-type Int in *attP* × *attB* synapsis and can also synapse *attL* × *attR*, *attL* × *attL* and *attR* × *attR* (13). E449 is predicted to lie at the 'g' position within the coiled-coil motif and substitutions at other 'g' positions have similar hyperactive phenotypes (13). Unlike L460P and Y475H, E449K did not disrupt oligomerization of the hMBP-γ or DNA binding by the hCTD. Thus E449 probably does not contribute directly to the same protein–protein interface as that generated through L460 and Y475. It therefore seems likely that E449 and other acidic amino acids that are predicted to lie on the same face of the coiled-coil motif, affect how the coiled-coil interacts with other residues, motifs or domains within Int. hCTD^{E449K} differs from hCTD as it could not form a stable CTD tetramer (incidentally implicating a role for the coiled-coil motif directly or indirectly in the CTD interactions at this stage.) At the same time IntE449K

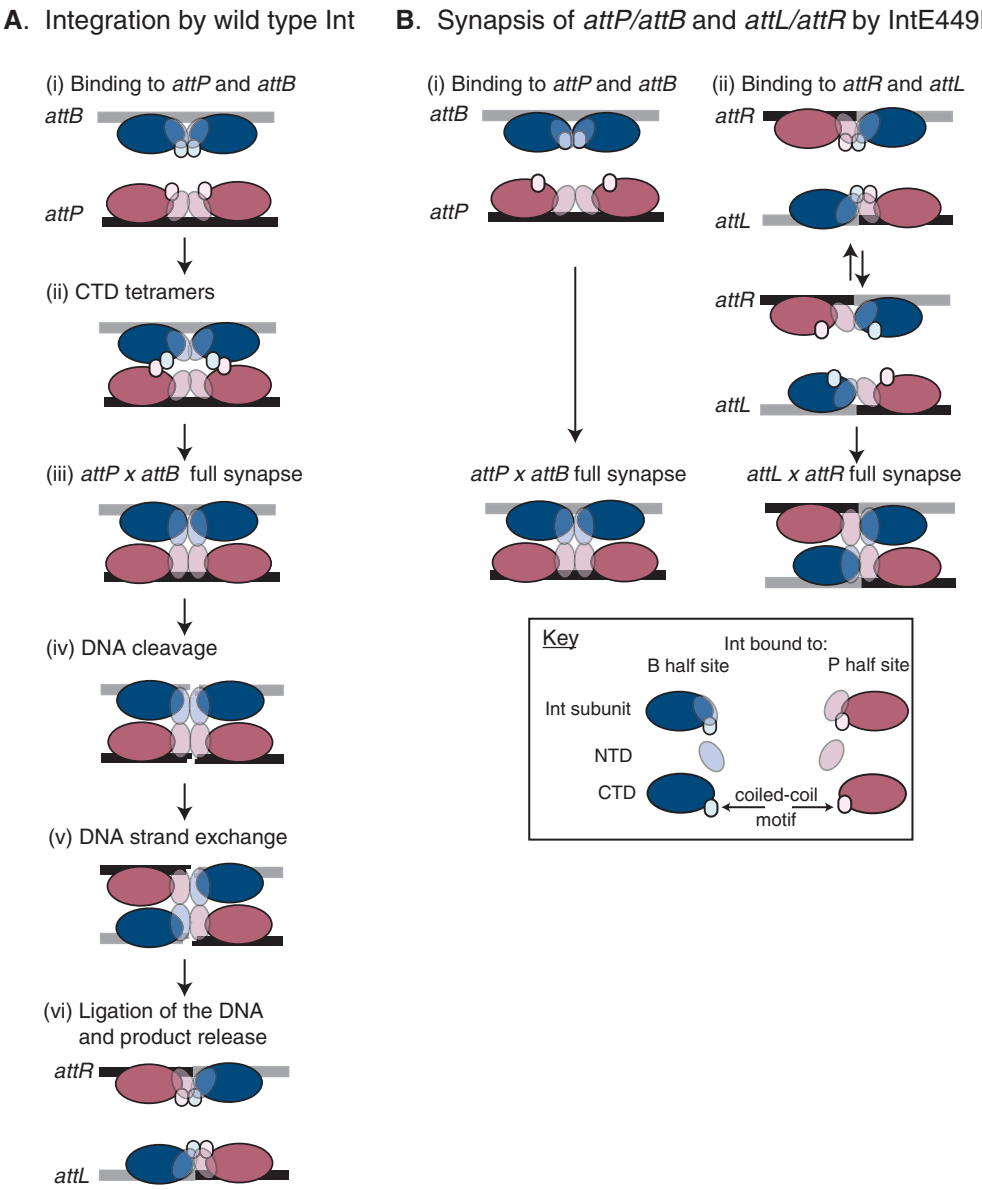

Figure 6. Model for synapsis and recombination by ϕC31 Int. The substrates *attP* and *attB* are shown as grey and black lines. Int subunits bound to a half site derived from *attB* and *attP* are shown in blue and red, respectively, and are meant to suggest subtle differences in conformation. Int subunits bound to *attP* are also shown spaced wider apart than Int subunits bound to *attB*. Three functional motifs are indicated within each Int subunit as described in the Key. (**A**) (i) Coiled-coil interactions occur between Int subunits bound to *attB* but not when bound to *attP*. (ii) The free coiled coil motifs in *attP* could interact with the CTD of Int subunits bound to *attB* to form a CTD synapse containing a tetramer of CTD domains. (iii) We propose that the coiled-coil motif blocks the formation of the NTD synaptic interface and this inhibition must be removed for the formation of a productive synapse. (iv) and (v) Formation of the productive synapse then triggers the DNA cleavage and strand exchange activities of the NTD which occur by a mechanism that resembles that of the resolvase/invertases (11). (vi) After religation of the DNA we propose that there are further conformation changes that establish the coiled-coil interaction once more between adjacently bound Int subunits, now bound to *attL* and *attR*. The role, if any, of the coiled-coil motif in the full synapse or during catalysis and strand exchange is not known. (**B**) A model for synapsis by IntE449K. (i) The coiled-coil motif is shown in an aberrant position where its ability to block oligomerization through the NTDs is reduced. The net result is formation of the full synapse without forming the CTD synapse. (ii) In an *attL* × *attR* reaction, we propose that weak CTD:CTD interactions between adjacently bound subunits results in a loss of inhibition on the oligomerization activity of the NTDs, which can lead to formation of a productive synapse. This model shows how the asymmetric binding by Int to *attL* and *attR* could lead to the observed preference for complementary subunit interactions at synapsis (as shown here) rather than non-complementary interactions (see text for more details).

readily forms a productive synapse with an expanded repertoire of attachments sites. The simplest explanation of these properties is that the interactions between the NTDs of Int to form the productive synaptic interface are normally inhibited by the coiled-coil motif and IntE449K is defective in this inhibitory activity. We propose that in the IntE449K subunits the coiled-coil motifs are misplaced, which severely weakens the CTD tetramer interactions and greatly reduces inhibition of the NTD tetramer interface [Figure 6B(i)]. Evidence for the misplaced coiled-coil motifs in IntE449K comes from the aberrant binding activity of $hCTD^{E449K}$ on *attP* in which we observed

higher order complexes, possibly binding of trimers and tetramers of hCTD to substrates.

Int is an irreversible recombinase in the absence of any accessory factors. No recombination or topoisomerase activity can be detected when Int is provided with *attL* and *attR* in biochemical assays (13,21). We propose that, although Int binds with similar affinities to *attL* and *attR* as to *attP* and *attB*, the conformations of Int bound to *attL* and *attR* are strongly inhibitory of formation of both the CTD and the NTD tetramer interfaces. Binding of hCTD to *attL* and *attR* was highly cooperative suggesting strong CTD–CTD interactions between adjacently bound subunits. The loss of cooperativity by $hCTD^{L460P}$ binding to *attL* and *attR* again implicates the coiled-coil as mediating the interaction between adjacently bound CTDs (Figure 4E and G). Thus when Int is bound to *attL* and *attR* the coiled-coil motifs in the four Int subunits are all sequestered [Figure 6A(vi)]. However $hCTD^{E449K}$ bound to *attL* and *attR* yielded significantly more complex I in binding assays suggesting that the CTD interactions between adjacently bound CTD domains could be weakened compared the hCTD (Figure 4E and F). Thus in a small fraction of bound complexes there could be sufficient relaxation of the normal inhibition of the NTD tetramer interface to lead to formation of a productive synapse, which in turn leads to $attL \times attR$ recombination [Figure 6B(ii)].

The absence of the CTD synapse in $hCTD^{E449K}$ suggests that IntE449K forms tetramers entirely through NTD interactions. Previously we described the dimer–dimer interactions between the IntE449K subunits that form the synaptic interfaces in integration and excision as either complementary or non-complementary (13). Complementary interactions occur between IntE449K subunits bound to a P-type arm and a B type arm whereas non-complementary interactions occur between subunits bound to both B-type and both P-type arms. $attP \times attB$ recombination always involves complementary IntE449K (or wild-type Int) subunit interactions regardless of the orientations in which the *att* sites collide during synapsis [in Figure 6A(i)–(iii) flipping either the *attB*-Int or *attP*-Int complex by 180° about a vertical axis does not change the nature of the subunit interactions on tetramer formation]. Dimer–dimer interactions between IntE449K bound to *attL* and *attR* can be complementary or non-complementary depending on how the DNA bound IntE449K subunits collide (in Figure 6 IntE449K subunit interactions on tetramer formation are complementary, but by flipping the IntE449K–*attL* complex by 180° about a vertical axis switches the interactions to a non-complementary format). We showed that complementary interactions are greatly preferred by IntE449K, and are therefore an inherent property of Int subunit interactions. If IntE449K only synapses through the NTDs, then it is at the NTD synaptic interface that the complementary interactions are preferred. Moreover it is at this step that Int discriminates against a tetramer that contains mixed complementary and non-complementary or all non-complementary interactions ($attL \times attP$, $attP \times attP$, $attP \times attR$, $attB \times attB$, $attB \times attR$ and $attB \times attL$). The model illustrates the asymmetric binding of the Int subunits to *attL* and *attR* containing each one B-type and one P-type half site and we propose that this asymmetry could lead to the preference for complementary rather than non-complementary interactions during synapsis.

In summary we have shown here the important role of the putative coiled-coil motif in ϕC31 Int activity. Located in the large CTD of Int, the coiled-coil motif mediates protein–protein interactions, dependent on substrate binding. It is proposed that the coiled-coil motif also has a strong inhibitory activity on the ability of the NTDs to form the productive synapse. Finally we note that our model also predicts a possible target, i.e. the coiled-coil motif, where an accessory protein could intervene to switch the directionality of recombination.

ACKNOWLEDGEMENTS

We thank Dr Wael Houssen for introducing the G485E mutation into Integrase.

FUNDING

Biotechnology and Biological Science Research Council of the UK [grant reference: BB/D007836/1]; University of Aberdeen studentship (to P.A.R.). Funding for open access charge: Biotechnology and Biological Science Research Council.

Conflict of interest statement. None declared.

REFERENCES

1. Groth,A.C. and Calos,M.P. (2004) Phage integrases: biology and applications. *J. Mol. Biol.*, **335**, 667–678.
2. Bibb,L.A., Hancox,M.I. and Hatfull,G.F. (2005) Integration and excision by the large serine recombinase φRv1 integrase. *Mol. Microbiol.*, **55**, 1896–1910.
3. Kim,A.I., Ghosh,P., Aaron,M.A., Bibb,L.A., Jain,S. and Hatfull,G.F. (2003) Mycobacteriophage Bxb1 integrates into the *Mycobacterium smegmatis groEL1* gene. *Mol. Microbiol.*, **50**, 463–473.
4. Thorpe,H.M. and Smith,M.C.M. (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl Acad. Sci. USA*, **95**, 5505–5510.
5. Ghosh,P., Wasil,L.R. and Hatfull,G.F. (2006) Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biol.*, **4**, e186.
6. Ghosh,P., Pannunzio,N.R. and Hatfull,G.F. (2005) Synapsis in phage Bxb1 integration: selection mechanism for the correct pair of recombination sites. *J. Mol. Biol.*, **349**, 331–348.
7. Thorpe,H.M., Wilson,S.E. and Smith,M.C.M. (2000) Control of directionality in the site-specific recombination system of the *Streptomyces* phage ϕC31. *Mol. Microbiol.*, **38**, 232–241.
8. Smith,M.C.A., Till,R., Brady,K., Soultanas,P., Thorpe,H. and Smith,M.C.M. (2004) Synapsis and DNA cleavage in ϕC31 integrase-mediated site-specific recombination. *Nucleic Acids Res.*, **32**, 2607–2617.
9. Gupta,M., Till,R. and Smith,M.C.M. (2007) Sequences in *attB* that affect the ability of ϕC31 integrase to synapse and to activate DNA cleavage. *Nucleic Acids Res.*, **35**, 3407–3419.
10. Smith,M.C.M. and Thorpe,H.M. (2002) Diversity in the serine recombinases. *Mol. Microbiol.*, **44**, 299–307.
11. Grindley,N.D.F., Whiteson,K.L. and Rice,P.A. (2006) Mechanisms of site-specific recombination. *Ann. Rev. Biochem.*, **75**, 567–605.

12. Yuan,P., Gupta,K. and Van Duyne,G.D. (2008) Tetrameric structure of a serine integrase catalytic domain. *Structure*, **16**, 1275–1286.
13. Rowley,P.A., Smith,M.C.M., Younger,E. and Smith,M.C.A. (2008) A motif in the C-terminal domain of ϕC31 integrase controls the directionality of recombination. *Nucleic Acids Res.*, **36**, 3879–3891.
14. Rowley,P.A., Smith,M.C.A., Younger,E. and Smith,M.C.M. (2008) Role of the N-terminal domain of ϕC31 integrase in *attB-attP* synapsis. *J. Bacteriol.*, **190**, 6918–6921.
15. Liu,H. and Naismith,J.H. (2009) A simple and efficient expression and purification system using two newly constructed vectors. *Protein Express. Purif.*, **63**, 102–111.
16. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A laboratory Manual,* 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
17. Lupas,A. (1997) Predicting coiled-coil regions in proteins. *Curr. Opin. Struct. Biol.*, **7**, 388–393.
18. Yang,W. and Steitz,T.A. (1995) Crystal-structure of the site-specific recombinase gamma-delta resolvase complexed with a 34 bp cleavage site. *Cell*, **82**, 193–207.
19. Mason,J.M. and Arndt,K.M. (2004) Coiled coil domains: stability, specificity, and biological implications. *Chembiochem.*, **5**, 170–176.
20. Groth,A.C., Olivares,E.C., Thyagarajan,B. and Calos,M.P. (2000) A phage integrase directs efficient site-specific integration in human cells. *Proc. Natl Acad. Sci. USA*, **97**, 5995–6000.
21. Smith,M.C.A., Till,R. and Smith,M.C.M. (2004) Switching the polarity of a bacteriophage integration system. *Mol. Microbiol.*, **51**, 1719–1728.
22. Lupas,A., Van Dyke,M. and Stock,J. (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.
23. Cole,C., Barber,J.D. and Barton,G.J. (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.*, **36**, W197–W201.