3614–3628 *Nucleic Acids Research, 2005, Vol. 33, No. 11*
doi:10.1093/nar/gki674

Evolutionary clues to eukaryotic DNA clamp-loading mechanisms: analysis of the functional constraints imposed on replication factor C AAA+ ATPases

Andrew F. Neuwald*

Cold Spring Harbor Laboratory, 1 Bungtown Road, PO Box 100, Cold Spring Harbor, NY 11724, USA

Received April 25, 2005; Revised and Accepted June 7, 2005

ABSTRACT

Ring-shaped sliding clamps encircle DNA and bind to DNA polymerase, thereby preventing it from falling off during DNA replication. In eukaryotes, sliding clamps are loaded onto DNA by the replication factor C (RFC) complex, which consists of five distinct subunits (A–E), each of which contains an AAA+ module composed of a RecA-like α/β ATPase domain followed by a helical domain. AAA+ ATPases mediate chaperone-like protein remodeling. Despite remarkable progress in our understanding of clamp loaders, it is still unclear how recognition of primed DNA by RFC triggers ATP hydrolysis and how hydrolysis leads to conformational changes that can load the clamp onto DNA. While these questions can, of course, only be resolved experimentally, the design of such experiments is itself non-trivial and requires that one first formulate the right hypotheses based on preliminary observations. The functional constraints imposed on protein sequences during evolution are potential sources of information in this regard, inasmuch as these presumably are due to and thus reflect underlying mechanisms. Here, rigorous statistical procedures are used to measure and compare the constraints imposed on various RFC clamp-loader subunits, each of which performs a related but somewhat different, specialized function. Visualization of these constraints, within the context of the RFC structure, provides clues regarding clamp-loader mechanisms—suggesting, for example, that RFC-A possesses a triggering component for DNA-dependent ATP hydrolysis. It also suggests that, starting with RFC-A, four RFC subunits (A–D) are sequentially activated through a propagated switching mechanism in which a conserved arginine swings away from a position that disrupts the catalytic Walker B region and into contact with DNA thread through the center of the RFC/clamp complex. Strong constraints near regions of interaction between subunits and with the clamp likewise provide clues regarding possible coupling of hydrolysis-driven conformational changes to the clamp's release and loading onto DNA.

INTRODUCTION

Eukaryotic proliferating cell nuclear antigen (PCNA) forms a homotrimeric ring that encircles and slides along DNA and that binds to polymerase during DNA replication [(1–5); reviewed in (6–8)]. PCNA thus prevents DNA polymerase from falling off and thereby facilitates processive replication. Rapid placement of PCNA onto RNA-primed sites by the ATP-dependent clamp-loader complex facilitates efficient DNA synthesis of Okazaki fragments on the lagging strand—a process inherently less efficient than continuous DNA synthesis on the leading strand. PCNA is also involved in DNA repair, DNA modification and chromatin remodeling [reviewed in (8)]. An alternative DNA-repair-specific heterotrimeric sliding clamp (9) associated with checkpoint control (10–12) has also been found.

Eukaryotic replication factor C (RFC) consists of five subunits that form a stable complex with PCNA in the presence of ATP (13–17). This complex binds to RNA-primed DNA and undergoes ATP hydrolysis upon recognition of a 3′-recessed single-stranded/double-stranded junction (the start of an Okazaki fragment), which results in dissociation of RFC and loading of the clamp onto DNA (18). Eukaryotic RFC is evolutionarily related to an archaeal RFC complex composed of one large subunit (RFCL) and four copies of a small subunit (RFCS) (19), the crystal structure of which is known (20). In eukaryotes, the small RFC subunit has diverged into four distinct subunits, each of which presumably has assumed a specialized function. Here, using the convention of Bowman *et al*. (17), the eukaryotic large subunit is denoted RFC-A, while the four RFC small subunits are

*Tel: +1 516 367 6802; Fax: +1 516 367 8461; Email: neuwald@cshl.edu

denoted B–E. Alternative eukaryotic complexes have been identified where an RFC-A-like protein replaces RFC-A. For example, a complex required for sister chromatid cohesion replaces RFC-A with the RFC-A-like protein Ctf18 (21,22). Similarly, replacement of RFC-A by Rad17 (yeast Rad24) creates a Rad–RFC complex that senses and responds to DNA damage (23). The Rad–RFC complex shows specificity for 5′ recessed single-stranded/double-stranded junctions (24), suggesting that specificity for distinct DNA structures is due to the unique large RFC subunit.

The structure of yeast RFC in complex with PCNA and the nucleotide analog ATPγS has been determined (17). Each of the five RFC subunits contains AAA+ modules (25–28), which are characterized by a RecA-like α/β ATPase domain directly followed by a helical domain; bound ATP is sandwiched between these two domains. AAA+ modules possess chaperone-like activity that couples ATP binding and hydrolysis to structural remodeling of a target protein substrate, in this case the PCNA clamp. Within each RFC subunit, the AAA+ modules are followed by another C-terminal helical domain; these C-terminal domains interact with each other to form a circular collar that holds the five subunits together, such that the AAA+ modules hang from the collar-like five fingers from the palm of a left hand. Using this analogy, the thumb corresponds to the large subunit A and the four other fingers, starting from the index finger, correspond to the small subunits B–E, respectively; the tips of the 'thumb' and of the first two 'fingers' (though not the last two) bind PCNA in the crystal structure. The collar is flexibly attached to the AAA+ modules, which are thereby free to undergo potentially dramatic conformational changes upon nucleotide binding, hydrolysis or release without disrupting the complex itself.

AAA+ proteins typically form homo- or heterohexamers with other AAA+ subunits such that an 'arginine finger' (29) from an adjacent subunit interacts with the ATP-binding site to thereby assist in ATP hydrolysis. The RFC complex is unusual inasmuch as its pentameric configuration has the appearance of a hexameric structure that is missing one subunit—a configuration that may facilitate insertion of DNA into the center of the complex and that is postulated to be part of a recognition mechanism for primed DNA (17). Notably, the ATPase domains of the five RFC subunits form a right-handed helix that may wrap around and interact with duplex DNA via certain basic residues conserved in these and in related clamp-loader subunits (17).

Despite remarkable progress, several important questions regarding clamp-loading mechanisms remain. In particular, it is unclear how specific recognition of and binding to primed DNA triggers ATP hydrolysis and how hydrolysis-induced local changes in and around the ATP-binding site are channeled into coordinated global conformational changes leading to release of the clamp onto DNA. These questions can, of course, only be addressed through carefully designed experiments, which in turn require formulation of the right hypotheses. My purpose here is to provide a more informed basis for formulating such hypotheses. This is done by examining the evolutionary evidence left by underlying mechanisms using a statistical approach called contrast hierarchical alignment and interaction network (CHAIN) analysis (30), which identifies and classifies co-conserved patterns within multiply aligned protein sequences and structures. Such patterns presumably correspond to structural features performing critical functions considering that for RFC subunits these patterns have persisted despite 1 to 2 billion years of evolution. The functional constraints imposed on specific residues are inferred statistically from these patterns. This, in turn, provides information regarding underlying mechanisms inasmuch as any hypothetical mechanisms that may be posed need to explain these constraints. In an information theoretical sense, this narrows the search for the true mechanisms (31).

MATERIALS AND METHODS

CHAIN analysis

The analysis here involved the following steps: First, AAA+ protein sequences were detected using iterative search procedures. Second, sequences were very accurately multiply aligned using a combination of motif-based Markov chain Monte Carlo (MCMC) multiple alignment procedures (32) and modified PSI-BLAST searches (33), in which aspects of the MCMC procedures were applied. These procedures only align sequence regions when there is clear statistical support for conservation, which is important for accurately measuring selective constraints. Alignments were manually adjusted based on structural alignments and on close examination of available crystal structures. Third, Bayesian partitioning with pattern selection (BPPS) (30), a statistical procedure based on Gibbs sampling (34), was used to optimally define various categories of selective constraints imposed on RFC subunits. Finally, category-specific constraints were quantified, compared and examined in light of RFC structures and of other empirical data in order to obtain a global perspective and characterization of the selective pressures acting on RFC subunits. These observations suggested the various hypothetical mechanisms discussed below. For further explanations of these steps see Results and Discussion. For a detailed description of CHAIN analysis see (30); for a recent review see (35).

Other computational procedures

Structural alignments were created using the CE program (36), as described previously (37). Protein hydrogen atoms were added to structural coordinates using Reduce (38). Secondary structure assignments were made using the DSSP program (39). Hydrogen bonds and other atomic interactions were

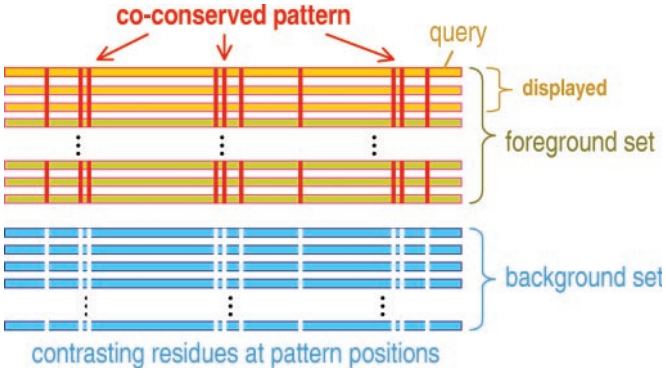
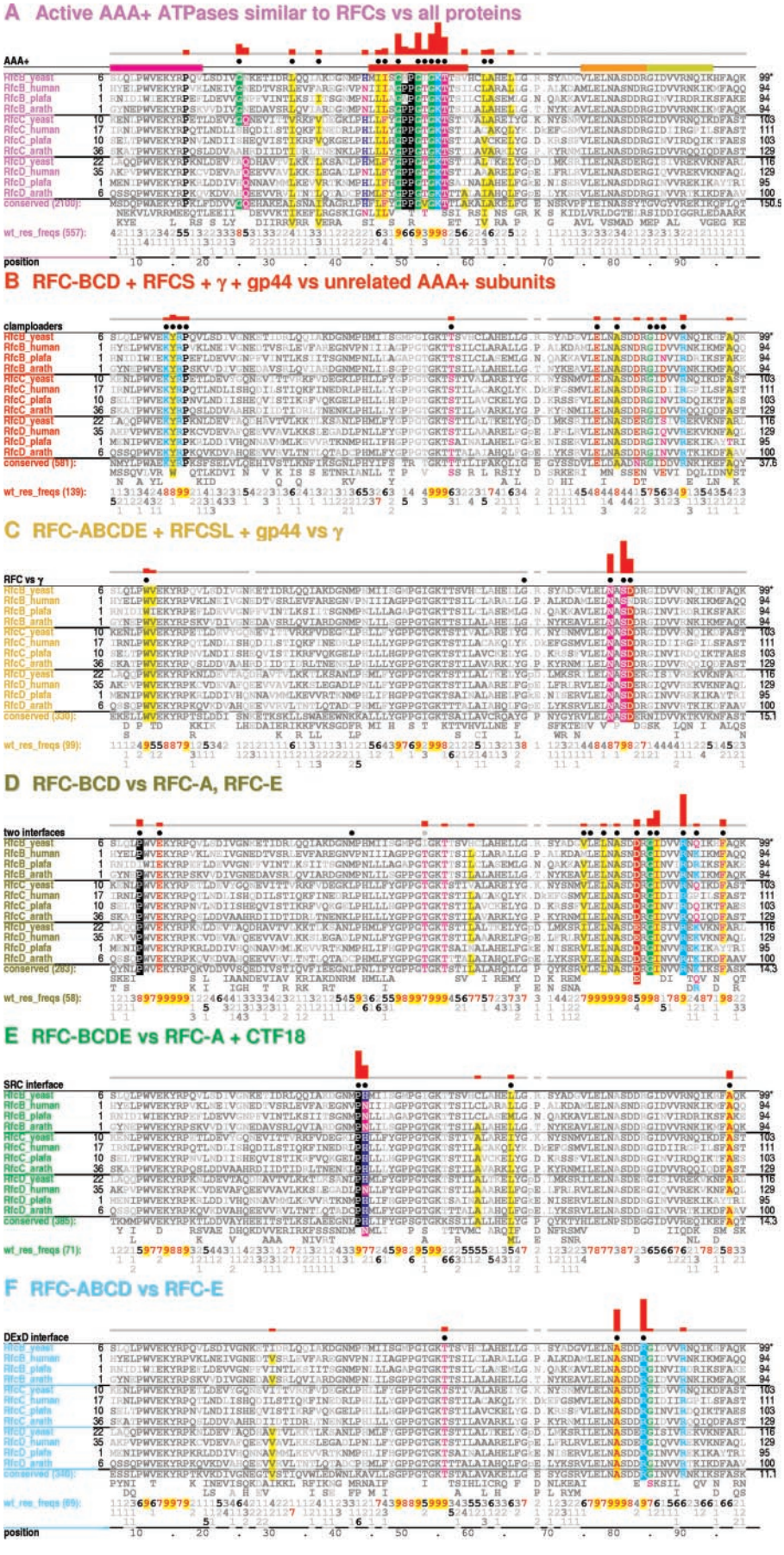
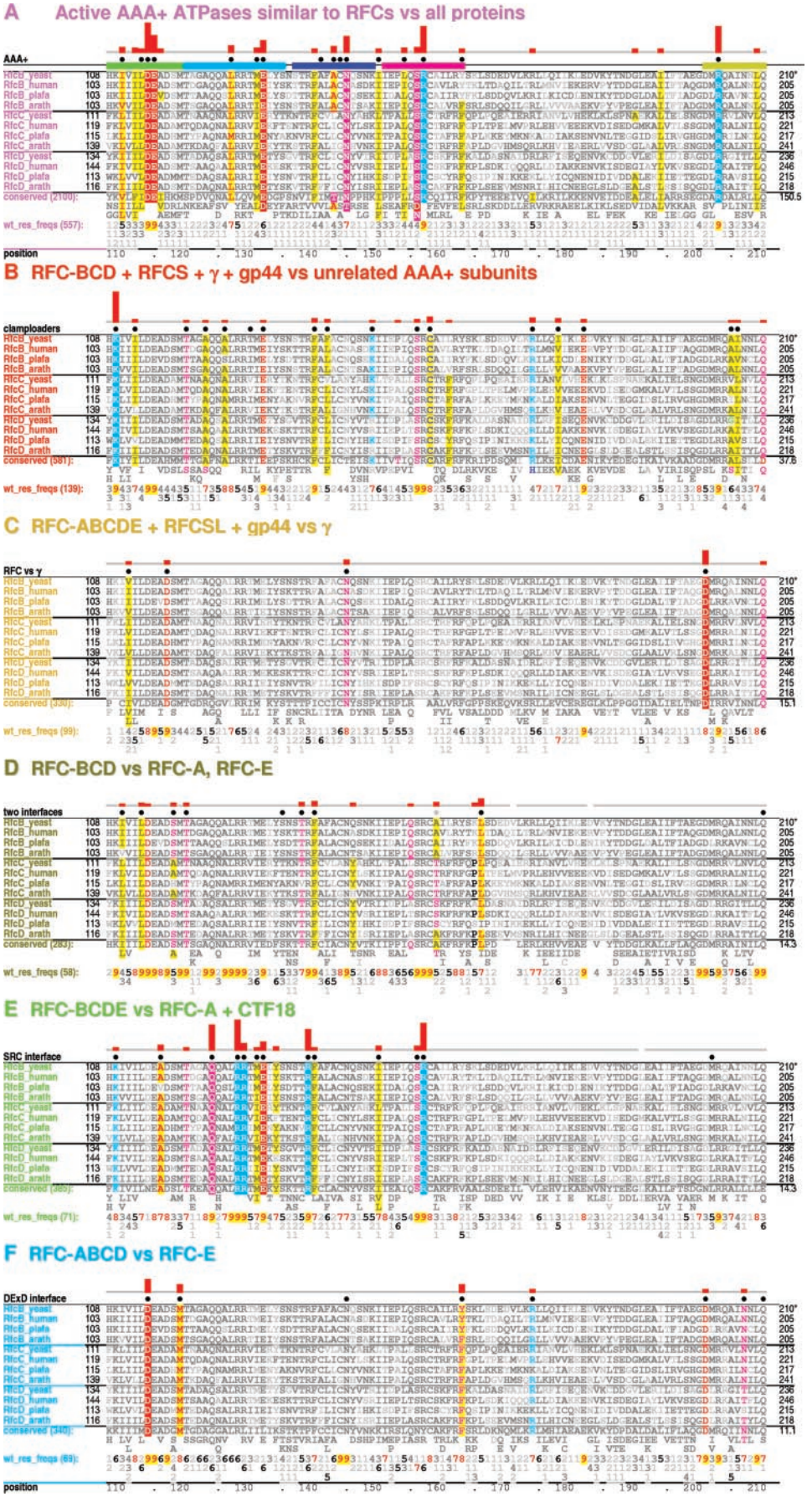


Figure 1. Schematic representation of a contrast hierarchical alignment. Adapted from (35). Copyright 2005, John Wiley and Sons Ltd. Reproduced with permission.


A Active AAA+ ATPases similar to RFCs vs all proteins
B RFC-BCD + RFCS + γ + gp44 vs unrelated AAA+ subunits
C RFC-ABCDE + RFCSL + gp44 vs γ
D RFC-BCD vs RFC-A, RFC-E
E RFC-BCDE vs RFC-A + CTF18
F RFC-ABCD vs RFC-E


A Active AAA+ ATPases similar to RFCs vs all proteins
B RFC-BCD + RFCS + γ + gp44 vs unrelated AAA+ subunits
C RFC-ABCDE + RFCSL + gp44 vs γ
D RFC-BCD vs RFC-A, RFC-E
E RFC-BCDE vs RFC-A + CTF18
F RFC-ABCD vs RFC-E

determined as described previously (30). Structural images were created using Rasmol (40).

Sequences displayed in alignments

NCBI sequence identifiers for the alignments in Figure 2 are RfcB_yeast (RFC4), gi730503 (*Saccharomyces cerevisiae*); RfcB_human (RFC2), gi2507300 (*Homo sapiens*); RfcB_plafa, gi23509093 (*Plasmodium falciparum*); RfcB_arath, gi12323266 (*Arabidopsis thaliana*); RfcC_yeast (RFC3), gi585844 (*S.cerevisiae*); RfcC_human (RFC5), gi728777 (*H.sapiens*); RfcC_plafa, gi23509823 (*P.falciparum*); RfcC_arath, gi21436375 (*A.thaliana*); RfcD_yeast (RFC2), gi730502 (*S.cerevisiae*); RfcD_human (RFC4), gi1703052 (*H.sapiens*); RfcD_plafa, gi16805067 (*P.falciparum*); and RfcD_arath, gi34098857 (*A.thaliana*).

NCBI sequence identifiers for the alignment in Figure 4A are RfcA_yeast (RFC1), gi584899 (*S.cerevisiae*); RfcA_human (RFC1), gi56757608 (*H.sapiens*); RfcA_fly (RFC1), gi12644230 (*Drosophila melanogaster*); RfcA_worm gi3875243 (*Caenorhabditis elegans*); RfcA_plafa gi11999114 (*P.falciparum*); RfcA_amoeba gi56466194 (*Entamoeba histolytica*); and RfcA_arath gi13374860 (*A.thaliana*).

NCBI sequence identifiers for the alignment in Figure 7B are RfcB_yeast (RFC4), gi730503 (*S.cerevisiae*); RfcB_human (RFC2), gi2507300 (*H.sapiens*); RfcB_plafa, gi23509093 (*P.falciparum 3D7*); and RfcB_arath, gi12323266 (*A.thaliana*).

RESULTS

Large and small RFC subunits are evolutionarily related to each other and to bacterial clamp-loader subunits. Very early in eukaryotic evolution, the small RFC subunit diverged into four distinct subunits, while the large subunit diverged into alternative forms. Sequence comparisons of these clamp-loader subunits reveal that each of these is subject both to functional constraints shared by other subunits and to distinct constraints reflecting specialized roles. The analysis here examines these similarities and differences and explores their potential biological significance in light of available crystal structures and biochemical studies.

Measuring changes in selective constraints between RFC functional states

It is important to define functional constraints in a way that mathematically captures what we mean by natural selection 'exerting a pressure or force' on a particular residue in a protein. To do this, we model functional constraints (30) in a manner analogous to the notion of free energy in statistical thermodynamics: just as the degree to which products have shifted away from reactants defines the free energy of a chemical reaction, in CHAIN analysis selective constraints are defined by the degree to which residue positions in a subset of aligned sequences (called the foreground set) have shifted away from the amino acid compositions observed at those positions in the remaining sequences (called the background set). Conceptually, constraints are quantified as follows: the residues conserved in the foreground set at a particular position are represented as distinctly colored balls in an urn with biochemically similar amino acids colored similarly. The selective constraint at that position is then defined by the difficulty of drawing by chance at least as many of the same- or similarly-colored balls from the urn (i.e. the background set) as occur in the foreground set. This measures how 'forcefully' natural selection has pushed the foreground sequences away from the background. Moreover, just as one may determine the difference in free energy (i.e. the ΔG) between distinct thermodynamic states, CHAIN analysis can determine the difference in selective constraints between distinct protein 'functional states': those corresponding to the foreground and background sets. The BPPS procedure optimally defines the foreground and background sets such that sequences in the foreground set strikingly co-conserve a pattern that is strikingly nonconserved in the background sequences (Figure 1). These are displayed as 'contrast hierarchical alignments' (as shown in Figure 2), where the histogram above each alignment provides a direct (essentially logarithmic) measure of selective pressure at each position based on the ball-in-urn model.

Categories of functional constraints. Various categories of functional constraints imposed on RFC subunits were examined (Figure 2). One category corresponds to RFC subunits that are catalytically assisted by the SRC-motif arginine finger donated by an adjacent subunit (i.e. subunits A–D, but not E). Although RFC-E binds nucleotide, it may lack ATPase activity because it poorly conserves the active site Walker A and B residues (17). The functional constraints associated with this category are shown in Figure 2F and are color-coded cyan in this and other figures. As expected, the residues conserved in this category generally occur in regions facing the adjacent subunit's SRC-motif interface. A second category corresponds to subunits (B–E, but not A) that donate an SRC-arginine finger (Figure 2, E; color coded green). The residues conserved in this category generally occur in regions facing an adjacent

Figure 2. Contrast hierarchical alignments showing category-specific functional constraints on subunits RFC-B, RFC-C and RFC-D. For each alignment, the foreground and background sets (as explained in the text) are indicated (i.e. as 'foreground' versus 'background'). The foreground set includes both the sequences shown in the alignment and those whose conserved properties are merely denoted by the consensus pattern and corresponding residue frequencies ('res_freqs') below the alignment. Residue frequencies are indicated in integer tenths where, for example, a '5' indicates that the corresponding residue directly above it occurs in 50–60% of the (weighted) sequences. Residue positions subject to significant functional constraints are highlighted in color (with biochemically similar residues colored similarly). The histogram above each alignment quantifies the selective pressure imposed on these positions (using what is essentially logarithmic scaling). For RFC subunits B–D, sequences representing metazoans (human), fungi (baker's yeast), plants (thale cress) and protozoans (malaria parasite) are shown. (See Material and Methods for sequence ids and organism scientific names.) (**A**) Constraints acting on active AAA+ ATPases similar to RFC subunits. More specifically, the foreground set consists of active ATPases possessing canonical Walker A, Walker B, and sensor 1 and 2 motifs as well as a conserved acidic residue in helix 5 and a conserved arginine finger in helix 6. Structural regions are indicated above this alignment by the bars, which are colored as the corresponding regions in Figures 4–7 (see Figure 4D for a global view). (**B**) Constraints that eukaryotic and archaeal RFC subunits share with bacterial and bacteriophage clamp loader subunits. (**C**) Constraints generally imposed on RFC subunits but not on bacterial γ subunits. (**D**) Constraints characteristic of RFC subunits that interface on both sides with other subunits. (**E**) Constraints characteristic of RFC subunits that interact with the ATP-binding region of an adjacent subunit. (**F**) Constraints characteristic of RFC subunits whose ATP-binding sites interact with an adjacent subunit.

unit's ATP-binding site. A third category corresponds to subunits (B–D, but not A and E) that possess both types of interfaces with adjacent subunits (Figure 2, D; color coded yellow). The residues in this category may function to transduce conformational signals between interfaces. A fourth category corresponds to constraints typically shared by eukaryotic and archaeal RFC subunits and by the analogous phage protein gp44, but that are absent from bacterial γ clamp-loader subunits (Figure 2, C; color coded gold). This category defines those features that most distinguish RFCs from bacterial clamp-loader ATPases. Incidentally, because the bacterial γ complex lacks many RFC constraints and, conversely, is subject to constraints that RFCs lack (data not shown), there presumably are clear-cut mechanistic differences between them. A fifth category corresponds to constraints imposed both on small RFC subunits (B–D) and on bacterial γ clamp-loader subunits, but that are absent from non-clamp-loader AAA+ proteins (Figure 2, B; color coded red). This category identifies residues that may play roles fundamental to clamp-loading. RFC subunits also share certain features with many other active AAA+ ATPases (Figure 2, A; color coded magenta). These features include the Walker A (GKT) and Walker B (DExx) motifs (41) also found in non-AAA+ P-loop NTPases, the sensor 1 motif (140–145B in Figure 2, A) and an arginine finger (R203B in Figure 2, A). The Walker A region binds ATP, while the Walker B region binds Mg^{++} and is required for ATP hydrolysis.

Each of the eukaryotic RFC subunits is also subject to unique constraints. For example, CTF18 is an alternative form of RFC-A that may perform clamp-loading functions associated with sister chromatid cohesion. CTF18 is widely conserved across diverse phyla and is also very similar in sequence to RFC-A. A comparison of RFC-A with CTF18 thus identifies functional constraints acting on RFC-A but not on CTF18 (Figure 4A; color coded blue). Similarly, functional constraints acting on RFC-B but absent from other RFC subunits are shown in Figure 7B (color coded blue). Finally, certain conserved residues fall into categories not specifically examined here (such as, for example, those shared by RFC-B and RFC-D but not by other small subunits); these are placed into an intermediate category (color coded light gray).

DISCUSSION

The selective constraints characterized in Figures 2, 4A and 7B are empirically based inasmuch as they were inferred from sequence data using statistical procedures designed to optimally define each constraint category while ignoring sequence similarities due to chance. Although these constraints are of interest in their own right, more importantly they provide clues to associated underlying mechanisms. Precisely what this tells us must be explored in light of other experimental data—structural data in particular. Nevertheless, because crystal structure models are static while the RFC complex is dynamic, these constraints can fully make sense only in light of all biologically relevant RFC structural conformations—and right now only a single crystal structure of the RFC complex is available. Nevertheless, by carefully examining those constraints most characteristic of each category (and thus presumably highly critical for function) certain features, which presumably reflect underlying mechanisms, clearly begin to immerge.

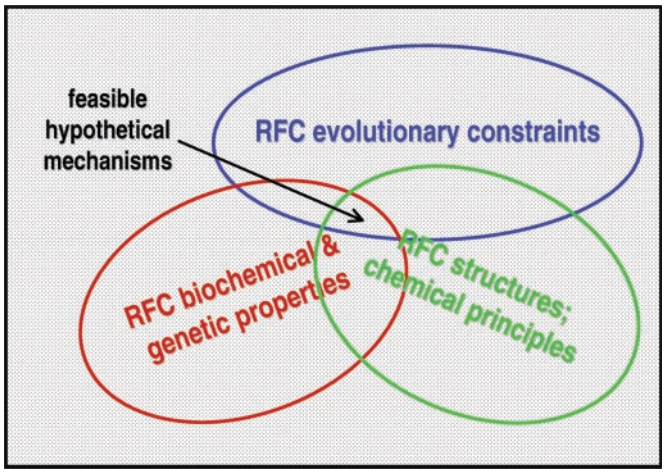


Figure 3. Venn diagram showing how the characterization of RFC evolutionary constraints, in conjunction with biochemical, genetic and structural data, helps identify feasible hypothetical mechanisms. The box represents the space of all possible RFC mechanisms with each point representing a specific mechanism. Each circle represents the set of hypothetical mechanisms consistent with the data source indicated. The intersection of all three circles corresponds to mechanisms consistent with all three sources of data. See Discussion in text. Adapted from (35). Copyright 2005, John Wiley and Sons Ltd. Reproduced with permission.

Examination of evolutionary constraints in this way is similar to the examination of other non-hypothesis driven data (such as, for example, that of a protein structure) inasmuch as it provides information relevant to our understanding of protein function, despite the fact that it fails to test (much less confirm) specific hypotheses. While such observations lack sufficient focus to narrow the possibilities down to a single mechanism, they typically strongly or weakly favor certain competing mechanisms, disfavor others and totally exclude others still. This is because, whatever the true mechanism may be, it needs to be consistent with observations regarding the selective constraints imposed upon the protein, as well as with the protein's structural and other properties (see Venn diagram in Figure 3). In theory, a rigorous Bayesian statistical formulation of this approach would assign posterior probabilities to each among many competing mechanisms based on the available empirical data. While CHAIN analysis is being further developed with this idealistic goal in mind (35), such a formulation does not yet exist. Nevertheless, even in the absence of a mathematically rigorous formulation, a probabilistic assembly of feasible mechanisms conceptually emerges from the detailed analysis described here. This can help guide further experimentation, leading in turn to further pruning of hypothetical mechanisms, so that over time our understanding of RFC-mediated clamp loading may become clear.

Possible RFC-A trigger mechanism for initiating ATP hydrolysis

Subunits within the eukaryotic RFC complex are arranged in the order A–E with A corresponding to the large subunit and B–E to the small subunits (Figure 4E). RFC-A is proposed to recognize primed DNA as a signal for DNA-dependent hydrolysis (24). If so then ATP hydrolysis is likely to be initiated by

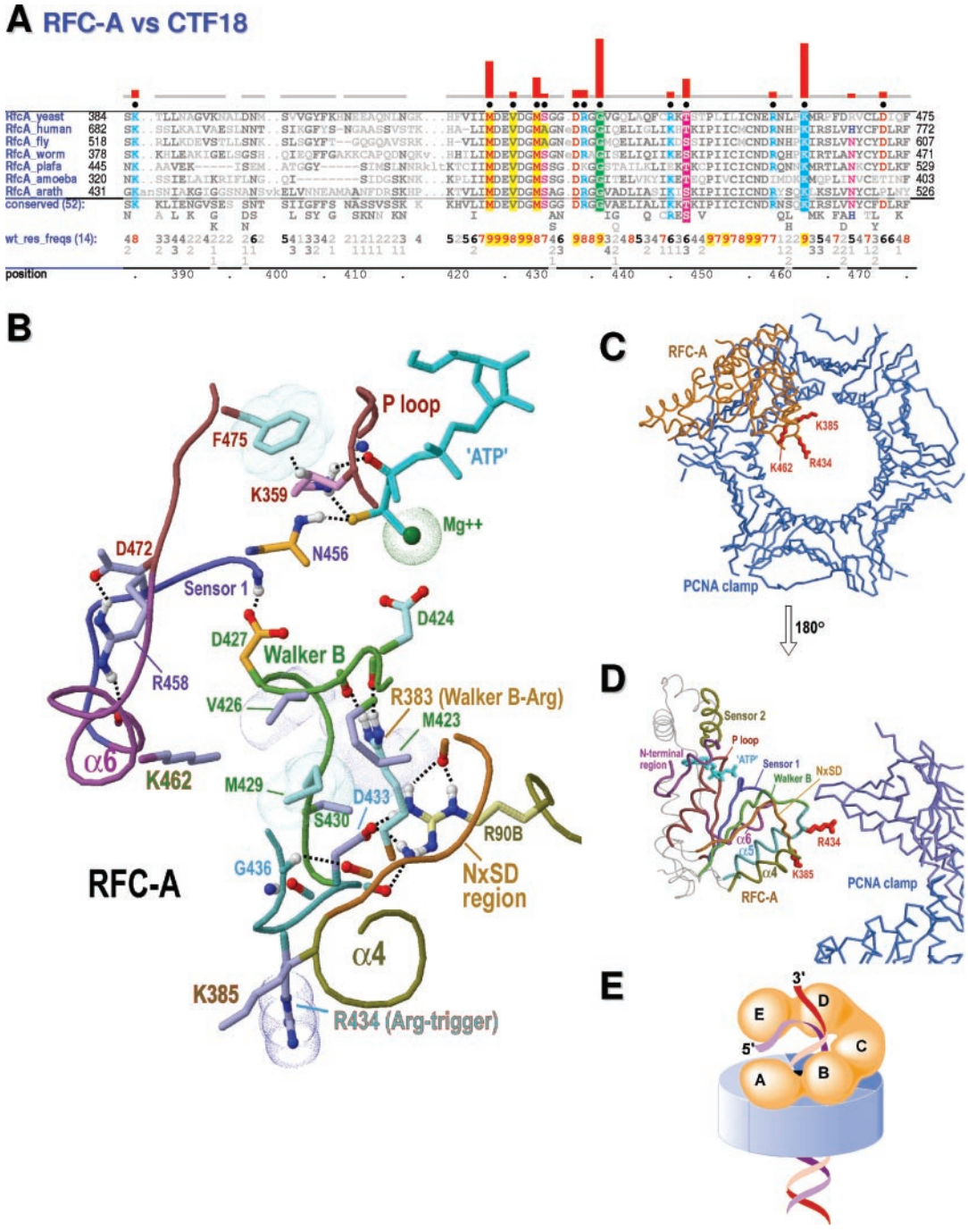


Figure 4. The RFC-A putative triggering component and its context within the RFC/PCNA complex. See text for details. (**A**) Contrast hierarchical alignment showing RFC-A-specific functional constraints absent from (RFC-A-like) CTF18 subunits. See legend to Figure 2 for descriptions. (**B**) Structural view of the putative triggering component suggested by the RFC-A-specific constraints in A (as explained in the text). The ATP analogue ATP-γS is indicated as 'ATP'. Color scheme: main chain of conserved structural elements (colored by regions as indicated by the bars above the alignment in Figure 2, A); ATP-γS, cyan; magnesium, ion, green; oxygen, nitrogen, sulfur and hydrogen atoms establishing hydrogen bonds (red, blue, yellow and white, respectively); side chains are colored by functional categories (colored as the first column of the corresponding alignments in Figures 2 and 4A). Hydrogen bonds are depicted as dotted lines. Electrostatic and van der Waals interactions are depicted by dot clouds. Dotted lines into clouds depict CH–π or NH–π interactions. (**C**) Location of possible trigger residues (K385A, R434A and K462A) relative to the hole in the center of the PCNA clamp. (**D**) View of RFC-A from below the clamp. For clarity, the clamp subunit contacting RFC-A is not shown. Key regions of the AAA+ module discussed in this analysis are indicated (color coded according to the bars above the alignment in Figure 2, A). (**E**) Locations of RFC subunits A–E in the proposed RFC/PCNA/DNA complex [after (17)].

RFC-A, which then sequentially triggers hydrolysis by the small subunits. Thus, under this model, RFC-A is likely to adopt a conformation that is primed both for catalysis and for propagating conformational changes onto the adjacent (RFC-B) subunit. Consistent with this notion, the authors of the RFC/PCNA/ATP crystal structure (17) identify RFC-A's conformation as 'close to being optimal for catalysis'. They also note that the A:B interface 'locks down the relative

orientation of domain 2 of RFC-A with respect to the nucleotide' such that RFC-B's SRC serine (S156B) interacts with RFC-A's sensor 2 arginine (R515A), which, in turn, is stabilized for optimal coordination with the β and γ phosphates of ATP by a glutamate (E132B) in helix 5 of RFC-B. (This glutamate corresponds to an acidic residue often conserved in AAA+ ATPases.) Furthermore, among all the subunits, RFC-A binds most directly to the PCNA clamp, suggesting that its release from the clamp may play a key role in clamp loading.

The putative arginine trigger. The constraints imposed on RFC-A but absent from the RFC-A-like CTF18 subunit (Figure 4A) presumably correspond to a functionally critical RFC-A component, the structural features of which (Figure 4B) suggest a triggering mechanism for DNA-dependent ATP hydrolysis. The putative trigger corresponds to a conserved arginine of the RFC-A component (Figure 4, R434A) (or, in a few organisms, to another basic residue at this position) that strikingly protrudes into the PCNA clamp's center hole through which DNA is thread (Figure 4C). Along with this arginine, two co-conserved lysines (K385A and K462A) also protrude into the center of the clamp, though to a lesser extent (Figure 4C), and thus may function as additional triggers. Coupling between the arginine trigger and ATPase activation is suggested by the structural location (17) of an RFC-A-conserved aspartate (D433A) that immediately precedes this arginine and that hydrogen bonds to the backbone of a second arginine that is highly characteristic of RFC subunits possessing canonical Walker A and B motifs (i.e. subunits A–D) (R84B in Figure 2, F). This second arginine (R383A in Figures 4B and 5A), which I term the Walker B-arginine, is in the loop preceding helix 4 (termed the NxSD loop; Figure 4D) and is linked via its main chain to one of the proposed additional triggers (K385A in Figure 4B and D).

The Walker B-arginine. In all four RFC subunits possessing canonical Walker A and B motifs, the Walker B-arginine (R383A, R84B, R88C and R101D in Figure 5A–D) hydrogen bonds to the main chain oxygens of the two adjacent Walker B acidic residues (D114B and E115B in Figure 2, A) that coordinate with the ATP-associated Mg^{++} ion and that play critical roles in catalysis. In other AAA+ and in related P-loop ATPases, the main chain oxygen between the two Walker B acidic residues normally hydrogen bonds to the adjacent β-strand. Its interaction with arginine thus appears to disrupt the standard Walker B geometry in the RFC/ATP/PCNA complex, which seems likely to inhibit catalytic activity. However, this arginine is dramatically repositioned in the structure of the archaeal RFCS subunit co-crystallized with ADP (20) and instead is on the surface of the protein (R88 in Figure 5E) where, within the full RFC complex, it presumably would protrude into the center cavity and thus could interact with DNA. Indeed, this arginine was proposed to interact with DNA within a hypothetical complex of RFC, PCNA and DNA (17). Furthermore, in the archaeal RFCS structure, the main chain oxygen between the Walker B residues now hydrogen bonds to the adjacent β-strand in the manner usually observed for P-loop ATPases. Taken together, these observations suggest that, upon association of the RFC/PCNA/ATP complex with DNA, the Walker B-arginine may undergo a conformational switch bringing it into contact with DNA, while simultaneously allowing the Walker B acidic residues to assume catalytically active configurations. This would, of course, directly couple sensing of DNA to ATPase activation. The conservation of this arginine in four adjacent RFC subunits along with other functional constraints imposed on these regions (see below) suggests the possibility of a cascading series of Walker B-arginine switches, each induced by a preceding switch.

RFC-A residues associated with the putative trigger. Other RFC-A-conserved residues (Figure 4A) provide clues as to how the Walker B-arginine in RFC-A (R383A) may undergo a conformational switch upon association with DNA. First, the conserved aspartate (D433A) that immediately precedes the putative arginine trigger (R434A) hydrogen bonds to the backbone of the Walker B-arginine (R383A) (Figure 4B). Thus movement of the arginine trigger could propagate conformational changes through its backbone and the side chain of this aspartate to the Walker B-arginine. The two conserved glycine residues (G435A and G436A) immediately following the arginine trigger may provide the necessary backbone flexibility for this event. Similarly, two methionines (M423A and M429A) co-conserved with this RFC-A component (Figure 4A) may facilitate a Walker B-arginine conformational switch, because these are located on either side of the Walker B (DExD) motif and also pack up against the Walker B-arginine on either side—an arrangement facilitating conformational rearrangements in both regions. (Methionine is highly adaptable and often occurs in buried regions that undergo conformational changes. Methionine is the only unbranched hydrophobic residue, and thus is the most flexible, and it contains a side chain sulfur atom that provides some hydrogen-bonding capability.) Two positions conserving small residues—one next to one of the methionines (S430A in Figure 4A) and an alanine whose backbone typically forms a β-strand with the Walker B backbone (A379A corresponding to A80B in Figure 2F, which is strongly co-conserved with the Walker B-arginine)—may likewise facilitate conformational rearrangements.

It should, however, be noted that since both the RFC-A (16) and the CTF18 (22) forms of the RFC complex load PCNA onto DNA and exhibit DNA-dependent ATPase activity, under the proposed model ATP hydrolysis by the CTF18 complex would have to be triggered by another mechanism.

Structural context of the Walker B-arginine

RFC subunits are distinguished by an NxSD motif. An NxSD motif most distinguishes RFC subunits from corresponding bacterial clamp-loader γ subunits (Figure 2, C), which lack the Walker B-arginine and thus presumably any associated mechanisms. The NxSD motif is located two positions before the Walker B-arginine and corresponds both to the β-strand that hydrogen bonds with the Walker B region and to the loop following this strand. In the structure of the archaeal small subunit (RFCS) co-crystallized with ADP (20) (Figure 5E), a Walker B main chain oxygen hydrogen bonds to the main chain –NH directly following the NxSD–asparagine, but in the eukaryotic RFC/PCNA/ATP complex this oxygen hydrogen bonds to the Walker B-arginine instead. Furthermore, in the eukaryotic complex, the NxSD–asparagine (N378A,

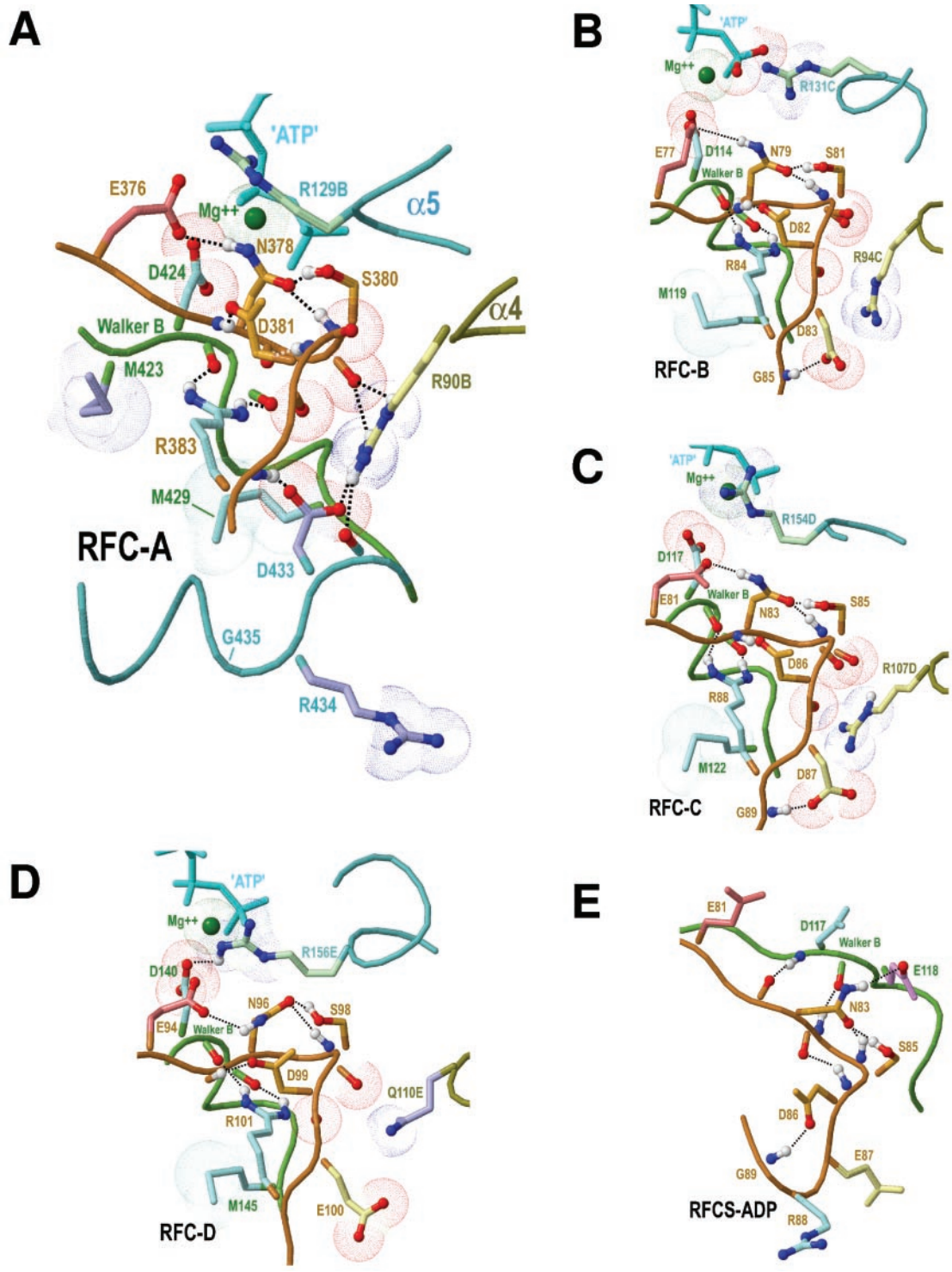


Figure 5. Structural view of the Walker B-arginine, the NxSD region (in orange) and two interacting arginines from the adjacent subunit. See legend to Figure 4 for descriptions and color schemes. (**A**) RFC-A subunit. The Walker B-arginine is R383A. The putative arginine trigger (R434A) is also shown. (**B**) RFC-B subunit. The Walker B-arginine is R84B. (**C**) RFC-C subunit. The Walker B-arginine is R88C. (**D**) RFC-D subunit. The Walker B-arginine is R101D. The adjacent RFC-E subunit conserves a glutamine at the position corresponding to the α4-arginine. (**E**) The archaeal RFCS subunit co-crystallized with ADP. The Walker B-arginine (R88) is rearranged relative to the crystal structure of the RFC/PCNA complex.

N79B, N83C and N96D) establishes characteristic interactions with the NxSD–serine and NxSD–aspartate, as well as with other nearby residues (Figure 5A–D), while in the archaeal RFCS-ADP structure some of these interactions are rearranged (Figure 5E). Taken together, these observations suggest that these rearrangements may accompany a conformational switch of the Walker B-arginine and help reposition the catalytically critical Walker B region.

The NxSD–asparagine also establishes considerable surface contact with the second of two sequence adjacent conserved arginines (R129B in Figure 2, E) in helix 5 of a neighboring RFC subunit (Figure 5A–D). In fact, in all subunits interfacing with a neighboring helix 5 region in this way (i.e. subunits A–D), the NxSD–asparagine makes greater surface contact with this α5-arginine (R129B, R131C, R155D and R156E in Figure 5) than does any other residue. The NxSD–serine (S380A, S81B, S84C and S98D) also contacts this α5-arginine and, in fact, both this serine and the arginine establish one of the highest surface area contacts with the facing subunit. Together, both of the sequence adjacent α5-arginines also interact with the Walker B motif and with phosphate oxygens of bound ATP (17). Thus, these residues and interactions could couple conformational changes associated with the Walker B-arginine switch and with ATP hydrolysis to the adjacent subunit. (The possible roles of these α5-arginines and of other residues co-conserved with them are discussed further below.)

The α4-arginine. An arginine in helix 4 (R90B, R94C and R107D) also establishes considerable surface contact with an adjacent subunit (Figure 5A–C). This arginine primarily interacts with residues of the loop associated with the adjacent NxSD motif, with which it forms electrostatic interactions and hydrogen bonds involving main chain and side chain oxygens (Figure 5A–C). (The nature of the interaction between RFC-A and RFC-B is somewhat different; however, inasmuch as it also involves the putative RFC-A trigger component.) Structural rearrangements involving the Walker B-arginine and the NxSD region would presumably perturb this adjacent α4-arginine as well. Indeed, the α4-arginine may itself play a role in this putative switch inasmuch as it is predicted to interact with DNA (17). Moreover, the α4-arginine is a highly characteristic feature of those RFC subunits interfacing with two adjacent subunits (see R90B in Figure 2, D), suggesting that it—along with similarly conserved residues such as an acidic residue (from an adjacent subunit) with which it electrostatically interacts (D83B, D87C and E100D in Figure 6B–D)—may play a role in propagating conformational signals between subunit interfaces. Indeed, helix 4 both directly interacts with the NxSD loop of an adjacent subunit (primarily via this arginine) and is connected via

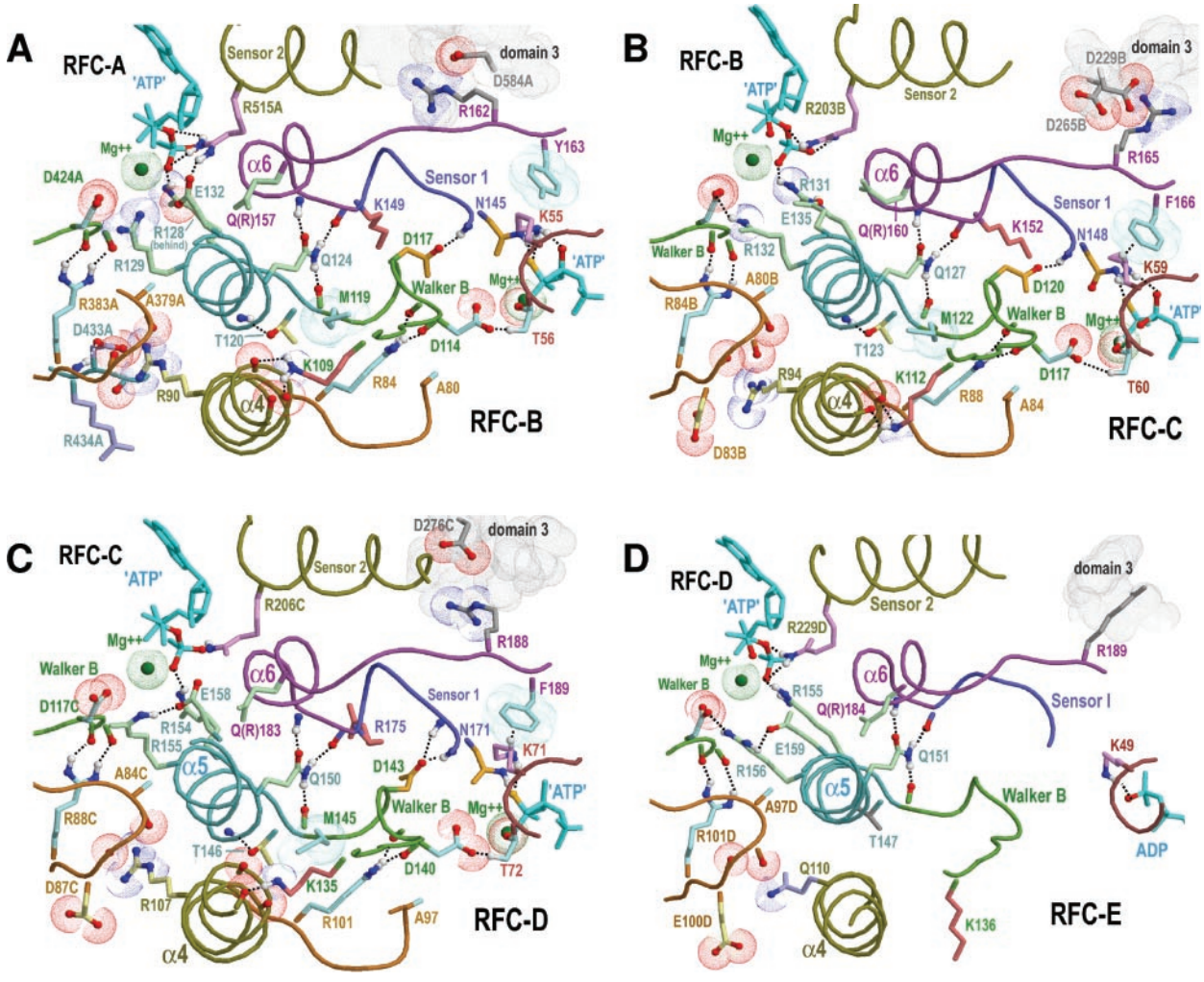


Figure 6. Structural view centered on the α5-helix and the α5-glutamine within small RFC subunits. See legend to Figure 4 for descriptions and color schemes. Note that the SRC-arginine fingers of these subunits (R157B, R160C, R183D and R184E) are mutated to glutamine in this crystal structure. See text for details. (**A**) RFC-B and ATP-binding region of adjacent RFC-A. (**B**) RFC-C subunit and ATP-binding region of adjacent RFC-B. (**C**) RFC-D subunit and ATP-binding region of adjacent RFC-C. (**D**) RFC-E subunit and ATP-binding region of adjacent RFC-D.

its main chain to the Walker B-arginine (and to the NxSD region) of its own subunit. Thus, this arginine may help mediate propagation of the proposed arginine switch from one subunit to the next. Such events would likely also be coupled to clamp loading, given that helix 4 also binds to the PCNA clamp (17,42) (see below).

Because the NxSD motif is also conserved (albeit weakly so) in RFC-E, these residues (at least in some organisms) apparently also perform a function independent of the Walker B-arginine, which RFC-E lacks. It is possible that this involves helix 4 conformational changes associated with binding and/or release of the clamp.

An aromatic residue contacting the P-loop lysine

An aromatic residue (either a phenylalanine or tyrosine) (F475A, Y163B, F166C and F189D) that, similar to the Walker B-arginine, is conserved in RFC subunits with typical Walker A and B motifs (Y163B in Figure 2, F) typically forms a CH–π interaction with the Walker A (or P-loop) lysine (K359A, K55B, K59C and K71D) (Figure 4B and Figure 6A–C). Indeed, in all of these subunits the P-loop lysine makes its greatest surface area contact with this aromatic residue. Since the NH_3 group of the P-loop lysine coordinates with both the β- and γ-phosphate of ATP, hydrolysis of ATP may transmit conformational changes to this aromatic residue. Alternatively, this aromatic residue may influence positioning of this lysine (and possibly its chemical properties) and thereby positioning of the β- and γ-phosphates and hydrolysis of the bond between them. Repositioning of this aromatic residue (in subunits B–D) could, in turn, be influenced by conformational changes in a neighboring subunit via a conserved arginine (R162B, R165C and R188D; Figure 6A–C). This arginine, which is sequence adjacent to this aromatic residue, establishes one of the highest surface area contacts with domain 3 of the neighboring RFC subunit. Notably, this region of contact often includes a highly conserved aspartate (D584A, D265B and D276C; Figure 6A–C) that can interact electrostatically with the arginine. (This arginine is well conserved in subunits B–E, but weakly so in RFC-A, and thus was placed into an intermediate category.) In some subunits, the contacts between the arginine, the aromatic residue and the P-loop lysine appear to be reinforced by other conserved aromatic residues (e.g. Y52C, F63D, Y64D, F164C and F187D).

Conserved N-terminal and helical domain arginines coordinating with ATP

The RFC complex presumably converts the energy of ATP hydrolysis into coordinated conformational changes associated with clamp loading. What functional constraints might correspond to these mechanisms? A reasonable place to start is to examine the functional constraints surrounding bound ATP and the adjacent regions into which the force generated by hydrolysis is likely to be directed. In AAA+ ATPases nucleotide binds at the interface between the α/β domain, the helical domain and the N-terminal region of the AAA+ module. Studies of various nucleotide states of P97/VCP ATPase (43,44), which contains two AAA+ modules, indicate that upon nucleotide hydrolysis the AAA+ α/β domain is held constant (mainly undergoing rotational and translational changes relative to adjacent α/β domains), while the AAA+ helical domain (and the sensor 2 region in particular) undergoes large conformational changes. In addition, an N-terminal region of the second P97/VCP AAA+ module (residues 461–480, which correspond to residues 292–311 of RFC-A) appear to undergo striking conformational changes upon ATP hydrolysis based on the order–disorder transitions seen between distinct nucleotide bound states (44). Since only the second P97/VCP AAA+ module is believed to be fully catalytically active, these results suggest that ATP hydrolysis by an active AAA+ module induces major conformational changes primarily in its sensor 2 and N-terminal regions.

Channeling energy through arginine levers. When the α- and β-phosphates are separated from the γ-phosphate upon hydrolysis, these presumably are driven apart by a strong electrostatic repulsive force. In the RFC/PCNA complex, these phosphate groups coordinate with the side chains of multiple conserved arginines (e.g. R16B, R128B, R129B, R157B and R203B in Figure 2), whose rigid planar guanidinium groups possess multiple hydrogen bond donor atoms. (Lysine can also coordinate with phosphate groups, but its side chain seems too flexible to act as a lever.) An arginine's side chain thus may act like a crankshaft to couple the explosive force of hydrolysis into translational motion in a specific direction. Such a role in ATP hydrolysis has been proposed for R189 of the β subunit of F_1-ATPase, for example, based on quantum and molecular mechanical simulations (45).

One such arginine occurs in the RFC N-terminal region within a KYRP motif that also is conserved within bacterial γ subunits (R16B in Figure 2, B), indicating that it may play a fundamental role in clamp loading. This arginine (R32D) appears to be conformationally constrained by the adjacent proline (P33D) and by a nearby acidic residue (E29D) that can hydrogen bond to this arginine (as occurs in the RFC-D structure) (data not shown). Thus, although an acidic residue cannot directly coordinate with a phosphate group of ATP, it may leverage conformational changes indirectly through coordination in this way with an arginine side chain (or, alternatively, with an ATP-associated Mg^{2+} ion). The sensor 2 arginine (R515A, R203B, R206C and R229D), which is conserved in many AAA+ ATPases (Figure 2, A), also coordinates with ATP phosphates in this way and with an AAA+ conserved acidic residue in helix 5 of the adjacent subunit (see R515A and E132B in Figure 7A). Thus, these two arginines seem well positioned to channel the energy of ATP hydrolysis into conformational changes associated with the N-terminal and sensor 2 regions in which they occur. These arginines may also help position ATP and/or ATP-associated residues for catalysis.

Residues co-conserved with the α5- and α6-arginines coordinating with ATP

Similarly, [and as previously noted (17)], one of the sequence adjacent arginines in helix 5 (R128B in Figure 2, E) (Figure 6) as well as the SRC-arginine finger in helix 6 (R157B in Figure 2, E) both coordinate with ATP in the neighboring subunit. Thus, these may also serve as levers for directing hydrolysis-driven conformational changes into the adjacent subunit. However, these are only two among about a half-dozen residues that are subject to strong functional constraints within this category (Figure 2, E) and that thus likely play important roles in this regard.

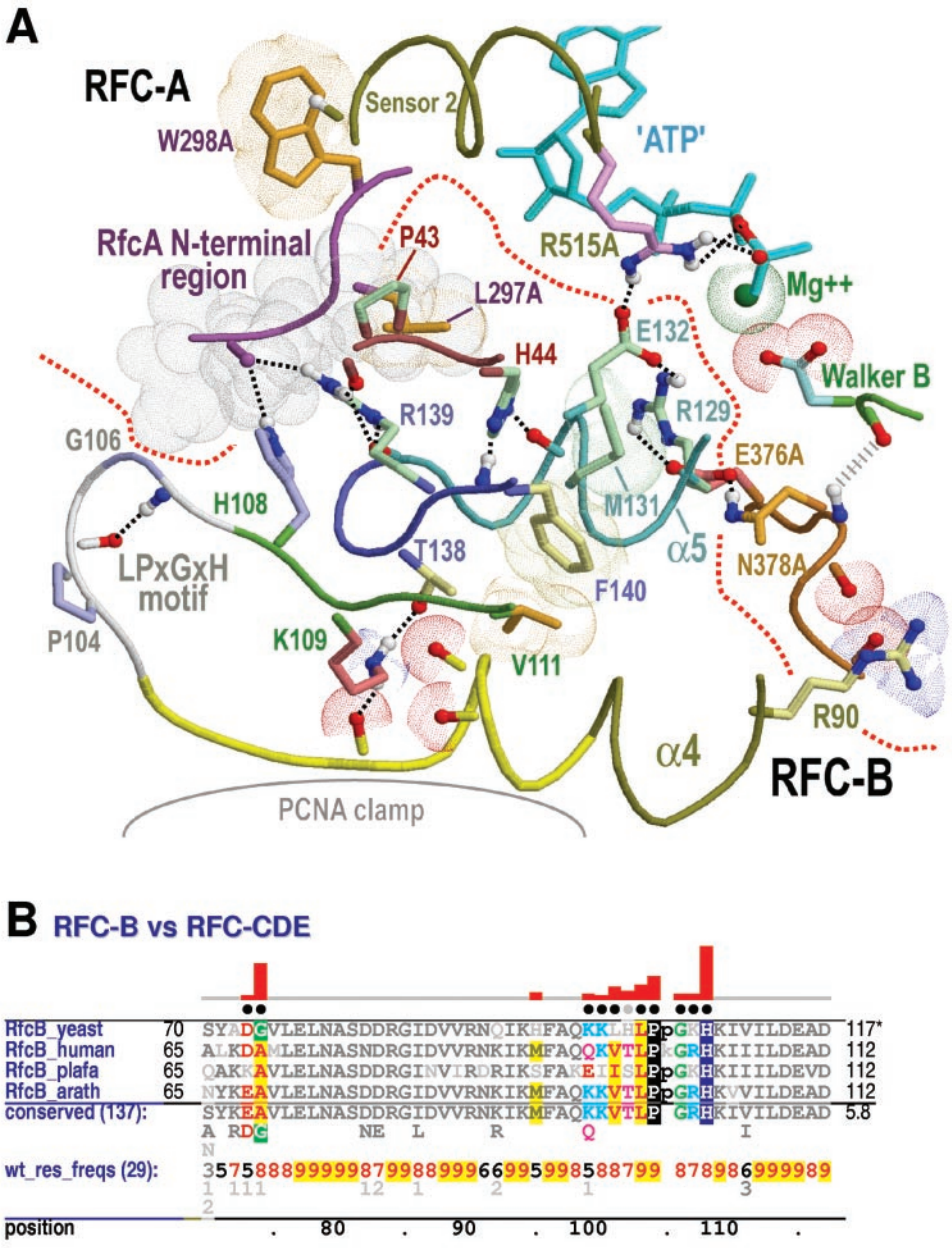


Figure 7. Functional constraints in regions of contact between RFC-B and RFC-A and between RFC-B and the PCNA clamp. (**A**) Structural view of these regions. The backbone trace of the region in RFC-B that interacts with the PCNA clamp is show in bright yellow. The gray dot cloud corresponds to a region in RFC-A making substantial surface contact with RFC-B. See text for details and the legend to Figure 4 for other descriptions and color schemes. (**B**) Contrast hierarchical alignment showing RFC-B-specific constraints. A highly characteristic feature of RFC-B is the LPxGxH motif following the region of contact with the PCNA clamp. The [DE]-[GA] motif (residues 73–74) interacts with the β-strand between K109B and V111B, but, for clarity, this is not shown in A. See legend to Figure 2 for further descriptions.

The α5-glutamine. For example, a glutamine within helix 5 (Q124B, Q127C, Q150D and Q151E in Figures 6A–D, respectively) is also highly characteristic of this functional category (Figure 2, E) and establishes three well-conserved side chain hydrogen bonds: (i) a hydrogen bond to a backbone nitrogen near the base of helix 6, which contains the SRC-arginine finger; (ii) a hydrogen bond to the backbone oxygen of a conserved basic residue (K149B, K152C and R175D) predicted to interact with DNA (17) and also conserved within the bacterial clamp-loader γ subunit (K161-γ) (17,46); and (iii) a hydrogen bond to the backbone oxygen located between a conserved methionine following the Walker B (DExD) motif and a conserved threonine (T120B, T123C and T146D) that functions as a helix 5 N-cap (47,48). This methionine (M119B in Figure 2, F) is conserved only in subunits containing a Walker B-arginine (R84B in Figure 2, F) and, in fact, packs against it (Figures 5 and 6). It also corresponds to one of the two methionines conserved in RFC-A (but not CTF18) (Figure 4) that is proposed to facilitate a Walker B-arginine conformational switch (see above). The

α5-glutamine thus may propagate conformational changes involving helix 5 (and possibly helix 6) to regions associated with the Walker B-arginine and with sensing of DNA.

Both the α5-arginine that coordinates with ATP and the α5-glutamine are highly conserved within NtrC and related AAA+ proteins ($R253^{NtrC}$ and $Q248^{NtrC}$, respectively in pdb entry 1ny5) (49). NtrC activates transcription from a distant enhancer DNA sequence by remodeling the closed complex between promoter DNA and RNA polymerase to an open complex (50). The side chain of the α5-glutamine in NtrC forms hydrogen bonds to the corresponding backbone atoms as in RFCs. Also conserved (albeit weakly so) is a basic residue ($K291^{NtrC}$) that is positioned to interact with DNA in the same way as the corresponding residue (K149B) in RFC-B. Thus, these α5 residues are conserved and may play similar functional roles outside of the RFC family.

Interaction of the N-terminal region of RFC-A with RFC-B. Several other strong constraints in the RFC-BCDE functional category (Figure 2, E) correspond to regions of contact between RFC-A and RFC-B (Figure 7A). In RFC-A, the region of contact involves the N-terminal end of the AAA+ module (residues 295–298) and corresponds to one of the two regions of P97/VCP (residues 461–480) proposed to undergo dramatic conformational changes upon ATP hydrolysis (43,44) (see above). Indeed, this region contains the two residues of the RFC-A AAA+ module exhibiting the largest contact surface areas with RFC-B, namely D295A and L297A. The residue position corresponding to D295A (Q8B in Figure 2) is nonconserved while the position corresponding to L297A conserves a proline in subunits B–D (P10B in Figure 2, D) and either a proline or (more often) a leucine in subunits A and E. Incidentally, the residue following this proline/leucine position in the sequence is a highly conserved tryptophan (position 11 in Figure 2, C) that packs against the sensor 2 helix (W298A in Figure 7A) and thus provides a connection between these two AAA+ regions implicated in hydrolysis-driven conformational changes. The 295–297 region of RFC-A packs around several residues highly conserved in subunit B: (i) a histidine (H108B) and an arginine (R139B), which form π orbital (aromatic–aromatic) interactions with each other, and (ii) the residues of a P-[HN] motif (P43B and H44B in Figure 2, E). The histidine (H108B) is a strikingly characteristic residue of RFC-B (Figure 7B). (In other subunits another aromatic residue often occurs at the histidine position and similarly interacts with the corresponding arginine.) The other residues (P43B, H44B and R139B) are all co-conserved with the α4- and α5-arginines that coordinate with an adjacent ATP (Figure 2, E).

Interpreting constraints in light of RFC-A and RFC-B interactions. The crystal structure interactions between RFC-A and RFC-B were closely examined because these appear to be mechanistically more meaningful than those between other subunits (17) (see above). This revealed that interactions between co-conserved residues within RFC-B interconnect sites of contact with RFC-A to the binding site for the PCNA clamp (Figure 7A). The arginine (R139B) that packs up against the N-terminal end of the RFC-A AAA+ module (residues 295–297 of RFC-A) also hydrogen bonds (i) to the backbone of this region, (ii) to the loop attached to the C-terminal end of helix 5 of RFC-B and (iii) to a backbone oxygen just before the P[HN] motif (Figure 7A). The histidine of this motif (H44B in Figure 7A), in turn, hydrogen bonds to the backbone nitrogen of R139B. It also hydrogen bonds to the backbone oxygen located between the acidic residue in helix 5 (E132B) (which interacts with the sensor 2 arginine of RFC-A) and a methionine (M131B), the side chain of which packs against this histidine. This methionine also packs against a conserved phenylalanine (F140B) directly following the arginine (R139B) in the sequence. A conserved threonine (T138B) directly preceding this arginine hydrogen bonds to a lysine (K109B) that is by far the residue most strikingly conserved in both RFCs and γ (Figure 2, B)—suggesting that it plays a fundamental role in clamp loading. This lysine, which is located in the β-strand preceding the Walker B (DExD) motif, electrostatically interacts with and/or hydrogen bonds to backbone oxygens of the C-terminal end of helix 4, which is near the center of a region that binds directly to the clamp (Figure 7A). This lysine is sequence adjacent to the RFC-B histidine (H108B) that—together with the conserved arginine with which it forms an aromatic–aromatic interaction (R132B)—packs against the most striking region of contact between the RFC-A and RFC-B AAA+ modules. Taken together, the functionally constrained residues in this region thus appear to couple RFC-B's clamp interacting region to the RFC-A-AAA+ module's N-terminal and sensor 2 regions, which (based on studies of P97/VCP AAA+ ATPase) may be associated with dramatic conformational changes upon ATP hydrolysis. Similar observations apply to the other small RFC subunits, as these also conserve this lysine and most of the associated residues and interactions.

Possible experiments

Certain aspects of proposed RFC mechanisms may be explored using simple mutagenesis experiments. For example, one or more of proposed basic trigger residues of RFC-A (K385A, R434A and K462A) could be mutated to alanines to explore their DNA sensing roles. Similarly, the role of the conserved lysine associated with the center of the PCNA binding site (K109B in Figure 2, B) could be explored by mutating it to serine or glycine—two residues that commonly occur at this position in other AAA+ ATPases and that thus are unlikely to be structurally disruptive. Mutation of the Walker B-arginine (R84B) to another naturally occurring amino acid is likely to be structurally disruptive and thus not particularly easy to interpret. However, mutation of the associated methionine (M119B) to a less flexible leucine, for example, might be expected to at least hamper the proposed Walker B-arginine conformational switch, which might be detected through kinetic analysis of DNA-dependent ATP hydrolysis in the mutant versus the wild-type complexes.

More direct exploration of the role of the Walker B-arginine may be possible through unnatural amino acid mutagenesis [reviewed in (51,52)]. Though technically difficult, several research groups have been and are engineering novel synthetase and cognate tRNA systems that (utilizing a stop codon) replace a native amino acid with an unnatural one in a site-specific manner. This approach has been used to generate over 100 unnatural amino acid substitutions (52) and has been applied within bacteria, yeast and mammalian

cells (51). Thus, for example, the Walker B-arginine might be replaced by L-citrulline, a naturally occurring arginine analog in which a guanidinium $–C–NH_2$ group is replaced by a –C=O group. This would abolish one of the hydrogen bonds that the native arginine establishes with the Walker B main chain and thus could probe the functional importance of this bond. This mutation presumably would not significantly disrupt the protein's overall structure given the spatial similarity of these amino acids. An arginine was replaced by L-citrulline in a small zinc finger protein using an expressed protein ligation method (53). Doing so for the much larger RFC subunits, however, requires an artificially evolved L-citrulline-recognizing synthetase and corresponding cognate tRNA, which is available (52). Mutagenesis with other arginine analogs could introduce additional similarly subtle perturbations into the Walker B-arginine structure. Such analogs include L-canavanine, which replaces the side chain carbon directly bound to the guanidinium group with oxygen; L-homoarginine, which adds a carbon to the arginine side chain; N^G-monomethyl-L-arginine, which possesses a $–CH_3$ group attached to a guanidinium $–NH_2$ group; and the L-arginine isostere 2(S)-amino-6-boronohexanoic acid. Mutagenesis using these arginine analogs would be useful for exploring the roles of other conserved arginines, such as the α4-arginine and the proposed arginine levers. The incorporation of isotopic derivatives (54) could facilitate solution NMR spectroscopy (55) of the RFC complex.

The proposed sequentially propagated conformational switches involving Walker B-arginines could be tested through kinetic studies similar to those used to establish ordered ATP hydrolysis within the bacterial γ clamp-loader complex (56). In such an analysis, each Walker B-arginine containing subunit (A–D) would be mutated, one subunit at a time, and the effect of that mutation on the kinetics of ATP hydrolysis in the mutant versus the wild-type complex could be compared. Predicted conformational changes, both in wild-type and mutant structures, could be tested using hydrogen/deuterium exchange mass spectrometry (57–59). This approach has been used, for example, to explore conformational changes associated with ERK2 protein kinase upon activation loop phosphorylation and substrate binding (60,61).

CONCLUSION

Understanding RFC clamp-loader mechanisms in atomic detail is a daunting task that will require a long series of carefully chosen hypotheses and experiments to fully sort out. Nevertheless, such hypotheses are not formulated in a conceptual vacuum, but rather are based on information or 'clues' obtained from previous observations. This analysis provides such clues by identifying and categorizing the evolutionary constraints imposed on RFCs—presumably by those very mechanisms we seek to understand. Because natural selection imposes these constraints on the genomic sequences of living organisms within their native environments, such observations lack the artifactual biases sometimes associated with *in vitro* experimental systems or with *in vivo* cell cultures and may reveal functionally critical features that have been overlooked due to the inherent limitations of current experimental methods. For example, although the roles associated with the Walker A and B motifs in ATP binding and hydrolysis (41) have been appreciated for a long time, the roles of the various RFC residues examined here are presumably just as important (given that they are just as highly conserved across major taxa), yet they have not been studied thus far, presumably due to a lack of functional clues. This analysis provides such clues and suggests feasible hypotheses. Such hypotheses include, for example, the involvement of certain residues and interactions in the coupling of DNA sensing to ATP hydrolysis and in coupling of conformational changes within the AAA+ N-terminal and sensor 2 regions to clamp loading and signal propagation to adjacent subunits. The possibilities described here are not intended to be exhaustive, as further examination of RFC evolutionary constraints will suggest other aspects of underlying mechanisms. On an unrelated note, this study also underscores the potential importance of arginines as triggers, levers or fingers in structural mechanisms.

ACKNOWLEDGEMENTS

The author thanks Tristan Fiedler for critical reading of the manuscript. This work was supported by NIH (NLM) grant LM06747. Funding to pay the Open Access publication charges for this article was provided by NIH (NLM) grant LM06747.

Conflict of interest statement. None declared.

REFERENCES

1. Huang,C.C., Hearst,J.E. and Alberts,B.M. (1981) Two types of replication proteins increase the rate at which T4 DNA polymerase traverses the helical regions in a single-stranded DNA template. *J. Biol. Chem.*, **256**, 4087–4094.
2. Prelich,G., Kostura,M., Marshak,D.R., Mathews,M.B. and Stillman,B. (1987) The cell-cycle regulated proliferating cell nuclear antigen is required for SV40 DNA replication *in vitro*. *Nature*, **326**, 471–475.
3. Kong,X.P., Onrust,R., O'Donnell,M. and Kuriyan,J. (1992) Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell*, **69**, 425–437.
4. Krishna,T.S., Kong,X.P., Gary,S., Burgers,P.M. and Kuriyan,J. (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell*, **79**, 1233–1243.
5. Stukenberg,P.T., Studwell-Vaughan,P.S. and O'Donnell,M. (1991) Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *J. Biol. Chem.*, **266**, 11328–11334.
6. Kelman,Z. and O'Donnell,M. (1995) DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu. Rev. Biochem.*, **64**, 171–200.
7. Waga,S. and Stillman,B. (1998) The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.*, **67**, 721–751.
8. Majka,J. and Burgers,P.M. (2004) The PCNA-RFC families of DNA clamps and clamp loaders. *Prog. Nucleic Acid Res. Mol. Biol.*, **78**, 227–260.
9. Majka,J. and Burgers,P.M. (2003) Yeast Rad17/Mec3/Ddc1: a sliding clamp for the DNA damage checkpoint. *Proc. Natl Acad. Sci. USA*, **100**, 2249–2254.
10. Longhese,M.P., Foiani,M., Muzi-Falconi,M., Lucchini,G. and Plevani,P. (1998) DNA damage checkpoint in budding yeast. *EMBO J.*, **17**, 5525–5528.
11. Melo,J. and Toczyski,D. (2002) A unified view of the DNA-damage checkpoint. *Curr. Opin. Cell Biol.*, **14**, 237–245.
12. Huberman,J.A. (1999) DNA damage and replication checkpoints in the fission yeast, *Schizosaccharomyces pombe*. *Prog. Nucleic Acid Res. Mol. Biol.*, **62**, 369–395.
13. Tsurimoto,T. and Stillman,B. (1989) Purification of a cellular replication factor, RF-C, that is required for coordinated synthesis of leading and lagging strands during simian virus 40 DNA replication *in vitro*. *Mol. Cell. Biol.*, **9**, 609–619.

14. Tsurimoto,T. and Stillman,B. (1990) Functions of replication factor C and proliferating-cell nuclear antigen: functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4. *Proc. Natl Acad. Sci. USA*, **87**, 1023–1027.
15. Lee,S.H., Kwong,A.D., Pan,Z.Q. and Hurwitz,J. (1991) Studies on the activator 1 protein complex, an accessory factor for proliferating cell nuclear antigen-dependent DNA polymerase delta. *J. Biol. Chem.*, **266**, 594–602.
16. Gomes,X.V. and Burgers,P.M. (2001) ATP utilization by yeast replication factor C. I. ATP-mediated interaction with DNA and with proliferating cell nuclear antigen. *J. Biol. Chem.*, **276**, 34768–34775.
17. Bowman,G.D., O'Donnell,M. and Kuriyan,J. (2004) Structural analysis of a eukaryotic sliding DNA clamp–clamp loader complex. *Nature*, **429**, 724–730.
18. Tsurimoto,T. and Stillman,B. (1991) Replication factors required for SV40 DNA replication *in vitro*. I. DNA structure-specific recognition of a primer-template junction by eukaryotic DNA polymerases and their accessory proteins. *J. Biol. Chem.*, **266**, 1950–1960.
19. Miyata,T., Oyama,T., Mayanagi,K., Ishino,S., Ishino,Y. and Morikawa,K. (2004) The clamp-loading complex for processive DNA replication. *Nature Struct. Mol. Biol.*, **11**, 632–636.
20. Oyama,T., Ishino,Y., Cann,I.K., Ishino,S. and Morikawa,K. (2001) Atomic structure of the clamp loader small subunit from *Pyrococcus furiosus*. *Mol. Cell*, **8**, 455–463.
21. Naiki,T., Kondo,T., Nakada,D., Matsumoto,K. and Sugimoto,K. (2001) Chl12 (Ctf18) forms a novel replication factor C-related complex and functions redundantly with Rad24 in the DNA replication checkpoint pathway. *Mol. Cell. Biol.*, **21**, 5838–5845.
22. Bermudez,V.P., Maniwa,Y., Tappin,I., Ozato,K., Yokomori,K. and Hurwitz,J. (2003) The alternative Ctf18-Dcc1-Ctf8-replication factor C complex required for sister chromatid cohesion loads proliferating cell nuclear antigen onto DNA. *Proc. Natl Acad. Sci. USA*, **100**, 10237–10242.
23. Parrilla-Castellar,E.R., Arlander,S.J. and Karnitz,L. (2004) Dial 9-1-1 for DNA damage: the Rad9-Hus1-Rad1 (9-1-1) clamp complex. *DNA Repair* (*Amst.*), **3**, 1009–1014.
24. Ellison,V. and Stillman,B. (2003) Biochemical characterization of DNA damage checkpoint complexes: clamp loader and clamp complexes with specificity for 5′ recessed DNA. *PLoS Biol.*, **1**, E33.
25. Guenther,B., Onrust,R., Sali,A., O'Donnell,M. and Kuriyan,J. (1997) Crystal structure of the delta' subunit of the clamp-loader complex of *E.coli* DNA polymerase III. *Cell*, **91**, 335–345.
26. Neuwald,A.F., Aravind,L., Spouge,J.L. and Koonin,E.V. (1999) AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.*, **9**, 27–43.
27. Ogura,T. and Wilkinson,A.J. (2001) AAA+ superfamily ATPases: common structure—diverse function. *Genes Cells*, **6**, 575–597.
28. Lupas,A.N. and Martin,J. (2002) AAA proteins. *Curr. Opin. Struct. Biol.*, **12**, 746–753.
29. Ahmadian,M.R., Stege,P., Scheffzek,K. and Wittinghofer,A. (1997) Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. *Nature Struct. Biol.*, **4**, 686–689.
30. Neuwald,A.F., Kannan,N., Poleksic,A., Hata,N. and Liu,J.S. (2003) Ran's C-terminal, basic patch and nucleotide exchange mechanisms in light of a canonical structure for Rab, Rho, Ras and Ran GTPases. *Genome Res.*, **13**, 673–692.
31. Wilbur,W.J. and Neuwald,A.F. (2000) A theory of information with special application to search problems. *Comput. Chem.*, **24**, 33–42.
32. Neuwald,A.F. and Liu,J.S. (2004) Gapped alignment of protein sequence motifs through Monte Carlo optimization of a hidden Markov model. *BMC Bioinformatics*, **5**, 157.
33. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
34. Liu,J.S. (2001) *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag, New York.
35. Neuwald,A.F. and Liu,J.S. (2005) Measuring evolutionary constraints as protein properties reflecting underlying mechanisms. In: Jorde,L.B., Little,P.R.R., Dunn,M.J. and Subramaniam.S. *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*. John Wiley & Sons, Hoboken, NJ.
36. Shindyalov,I.N. and Bourne,P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.*, **11**, 739–747.
37. Neuwald,A.F. (2003) Evolutionary clues to DNA polymerase III beta clamp structural mechanisms. *Nucleic Acids Res.*, **31**, 4503–4516.
38. Word,J.M., Lovell,S.C., LaBean,T.H., Taylor,H.C., Zalis,M.E., Presley,B.K., Richardson,J.S. and Richardson,D.C. (1999) Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.*, **285**, 1711–1733.
39. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
40. Sayle,R.A. and Milner-White,E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374.
41. Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.*, **1**, 945–951.
42. Jeruzalmi,D., Yurieva,O., Zhao,Y., Young,M., Stewart,J., Hingorani,M., O'Donnell,M. and Kuriyan,J. (2001) Mechanism of processivity clamp opening by the delta subunit wrench of the clamp loader complex of *E. coli* DNA polymerase III. *Cell*, **106**, 417–428.
43. DeLaBarre,B. and Brunger,A.T. (2003) Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains. *Nature Struct. Biol.*, **10**, 856–863.
44. Delabarre,B. and Brunger,A.T. (2005) Nucleotide dependent motion and mechanism of action of p97/VCP. *J. Mol. Biol.*, **347**, 437–452.
45. Dittrich,M., Hayashi,S. and Schulten,K. (2003) On the mechanism of ATP hydrolysis in F1-ATPase. *Biophys. J.*, **85**, 2253–2266.
46. Jeruzalmi,D., O'Donnell,M. and Kuriyan,J. (2001) Crystal structure of the processivity clamp loader gamma (γ) complex of *E.coli* DNA polymerase III. *Cell*, **106**, 429–441.
47. Richardson,J.S. and Richardson,D.C. (1988) Amino acid preferences for specific locations at the ends of alpha helices. *Science*, **240**, 1648–1652.
48. Presta,L.G. and Rose,G.D. (1988) Helix signals in proteins. *Science*, **240**, 1632–1641.
49. Lee,S.Y., De La Torre,A., Yan,D., Kustu,S., Nixon,B.T. and Wemmer,D.E. (2003) Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains. *Genes Dev.*, **17**, 2552–2563.
50. Wyman,C., Rombel,I., North,A.K., Bustamante,C. and Kustu,S. (1997) Unusual oligomerization required for activity of NtrC, a bacterial enhancer-binding protein. *Science*, **275**, 1658–1661.
51. Cropp,T.A. and Schultz,P.G. (2004) An expanding genetic code. *Trends Genet.*, **20**, 625–630.
52. England,P.M. (2004) Unnatural amino acid mutagenesis: a precise tool for probing protein structure and function. *Biochemistry*, **43**, 11623–11629.
53. Jantz,D. and Berg,J.M. (2003) Expanding the DNA-recognition repertoire for zinc finger proteins beyond 20 amino acids. *J. Am. Chem. Soc.*, **125**, 4960–4961.
54. Deiters,A., Geierstanger,B.H. and Schultz,P.G. (2005) Site-specific *in vivo* labeling of proteins for NMR studies. *Chembiochem.*, **6**, 55–58.
55. Fernandez,C. and Wider,G. (2003) TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr. Opin. Struct. Biol.*, **13**, 570–580.
56. Johnson,A. and O'Donnell,M. (2003) Ordered ATP hydrolysis in the gamma complex clamp loader AAA+ machine. *J. Biol. Chem.*, **278**, 14406–14413.
57. Eyles,S.J. and Kaltashov,I.A. (2004) Methods to study protein dynamics and folding by mass spectrometry. *Methods*, **34**, 88–99.
58. Busenlehner,L.S. and Armstrong,R.N. (2005) Insights into enzyme structure and dynamics elucidated by amide H/D exchange mass spectrometry. *Arch. Biochem. Biophys.*, **433**, 34–46.
59. Hoofnagle,A.N., Resing,K.A. and Ahn,N.G. (2003) Protein analysis by hydrogen exchange mass spectrometry. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 1–25.
60. Hoofnagle,A.N., Resing,K.A., Goldsmith,E.J. and Ahn,N.G. (2001) Changes in protein conformational mobility upon activation of extracellular regulated protein kinase-2 as detected by hydrogen exchange. *Proc. Natl Acad. Sci. USA*, **98**, 956–961.
61. Lee,T., Hoofnagle,A.N., Kabuyama,Y., Stroud,J., Min,X., Goldsmith,E.J., Chen,L., Resing,K.A. and Ahn,N.G. (2004) Docking motif interactions in MAP kinases revealed by hydrogen exchange mass spectrometry. *Mol. Cell*, **14**, 43–55.