

## A structure-based sequential mechano-chemical model of hexameric helicases

Wenjun Zheng \*

*Physics Department, University at Buffalo, Buffalo, NY, USA*

*Submitted: 3 Dec. 2008; Accepted: 30 Dec. 2008*

### Abstract

From accumulating studies of hexameric helicases, a mechano-chemical coupling mechanism has emerged which postulates that the nucleoside triphosphate (NTP) hydrolysis is sequentially coordinated together with the binding and translocation of nucleic acid substrate. Various mechano-chemical models have been proposed for different hexameric helicases with no consensus for a common mechanism. To explain both conserved and variable features of the mechano-chemical coupling in various hexameric helicases, we propose a tri-site sequential model for NTP hydrolysis and substrate translocation. It builds upon structurally derived coupling rules that map the chemical state of a nucleotide-binding site to the longitudinal positions of the substrate-binding loops in two adjoined subunits. The coupling rules symbolize the mechano-chemical couplings via both the trans- and cis-residues of a nucleotide-binding site, and facilitate sequentially coordinated hydrolysis involving three consecutive nucleotide-binding sites at one time. Four schemes exist within our model, which differ in hydrolysis direction and substrate-binding mode. By applying our general model to five hexameric helicases (Rho, T7, P4, E1 and LTag), we have predicted various mechano-chemical properties of these NTPase-based motors (including hydrolysis direction, NTP-binding cooperativity, substrate-binding mode, translocation direction and step size, force-generation step). These predictions agree with past experimental data and call for future experimental test.

**Keywords:** sequential hydrolysis; translocation; hexameric helicase; force generation; mechano-chemical coupling.

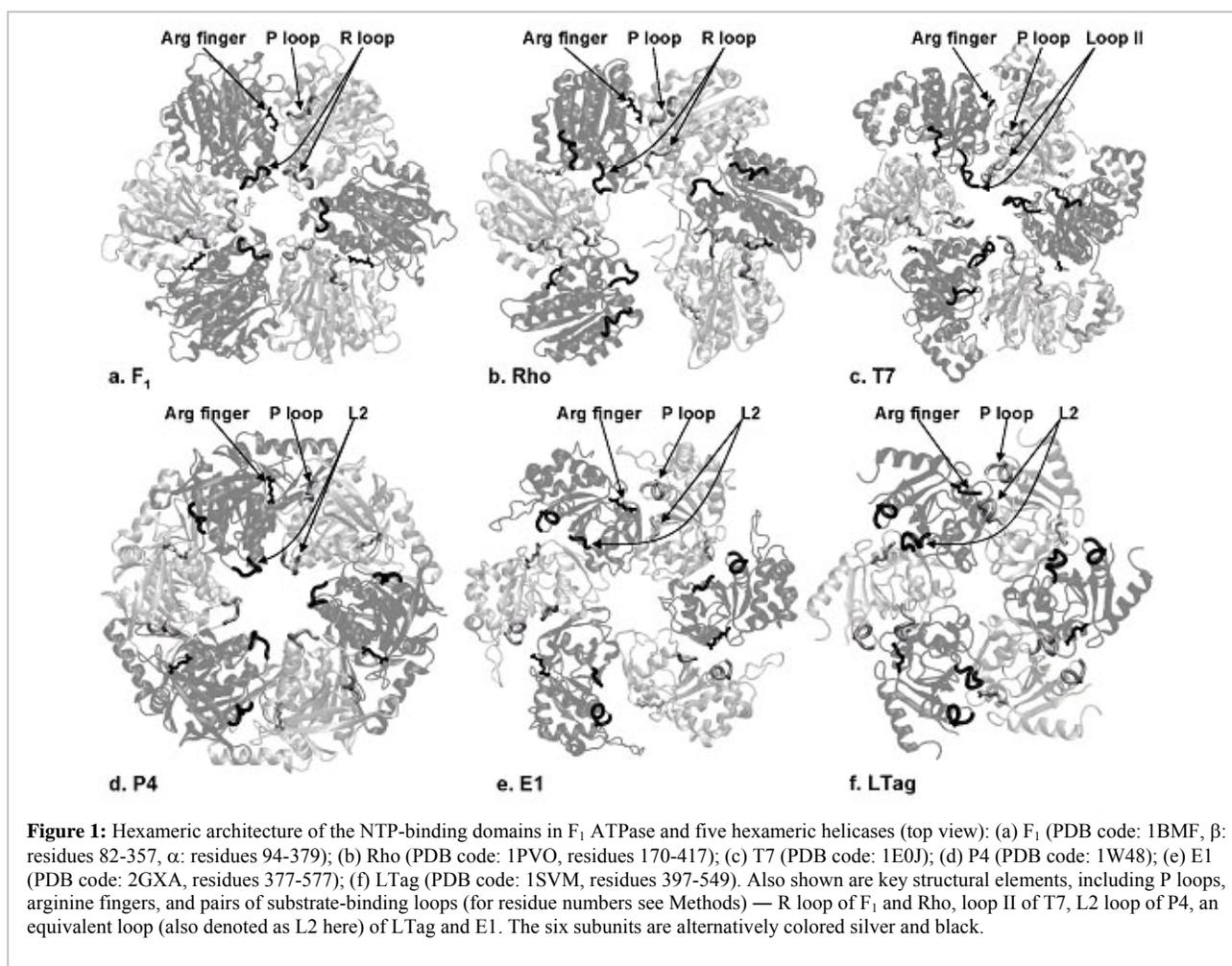
## INTRODUCTION

Helicases are abundant in all living organisms and participate in many essential cellular processes involving nucleic acids (Lohman *et al.*, 1996). The basic activity of a helicase is to drive the unwinding of duplex nucleic acids, which is coupled with the binding and hydrolysis of NTPs. Ring-shaped hexameric helicases are a subset of helicases involved in a diverse range of functions including replication, recombination, packaging of nucleic acids, and transcription regulation in viruses and bacteria. Although some of them do not possess the helicase function of nucleic acids unwinding, they are all molecular motors that translocate unidirectionally along nucleic acid substrate — single-stranded DNA (ssDNA) or RNA (ssRNA) (for review, see Patel *et al.*, 2000; Donmez *et al.*, 2006) (Here we use the term ‘substrate’ exclusively for nucleic acid substrate instead of NTP).

Crystal structures have been solved for a number of hexameric helicases, including *E. Coli* transcription termination factor Rho (Skordalakes *et al.*, 2003; Skordalakes *et al.*, 2006), bacteriophage T7 gene 4 (Sawaya *et al.*, 1999; Singleton *et al.*, 2000; Toth *et al.*, 2003), bacteriophage  $\Phi$ 12 RNA packaging motor P4 (Mancini *et al.*, 2004), Papillomavirus E1 (Enemark *et al.*, 2006), and simian virus 40 large T antigen (LTag) (Li *et al.*, 2003; Gai *et al.*, 2004). Among these hexameric helicases, Rho belongs to the V-F-ATPases family (Gorbalenya *et al.*, 1993), T7 and P4 belong to the DnaB-like family of SF4 helicases (Ilyina *et al.*, 1992), LTag and E1 belong to the SF3 helicases (Gorbalenya *et al.*, 1990; Hickman *et al.*, 2005). According to a more recent classification of helicases (Singleton *et al.*, 2007), Rho belongs to Superfamily 5, T7 belongs to Superfamily 4 (P4 is closely related to Superfamily 4), LTag and E1 belong to Superfamily 3. The crystal structures have revealed a number of structural features relevant to the mechano-chemical coupling mechanism of hexameric helicases (Fig. 1):

### \*Corresponding author:

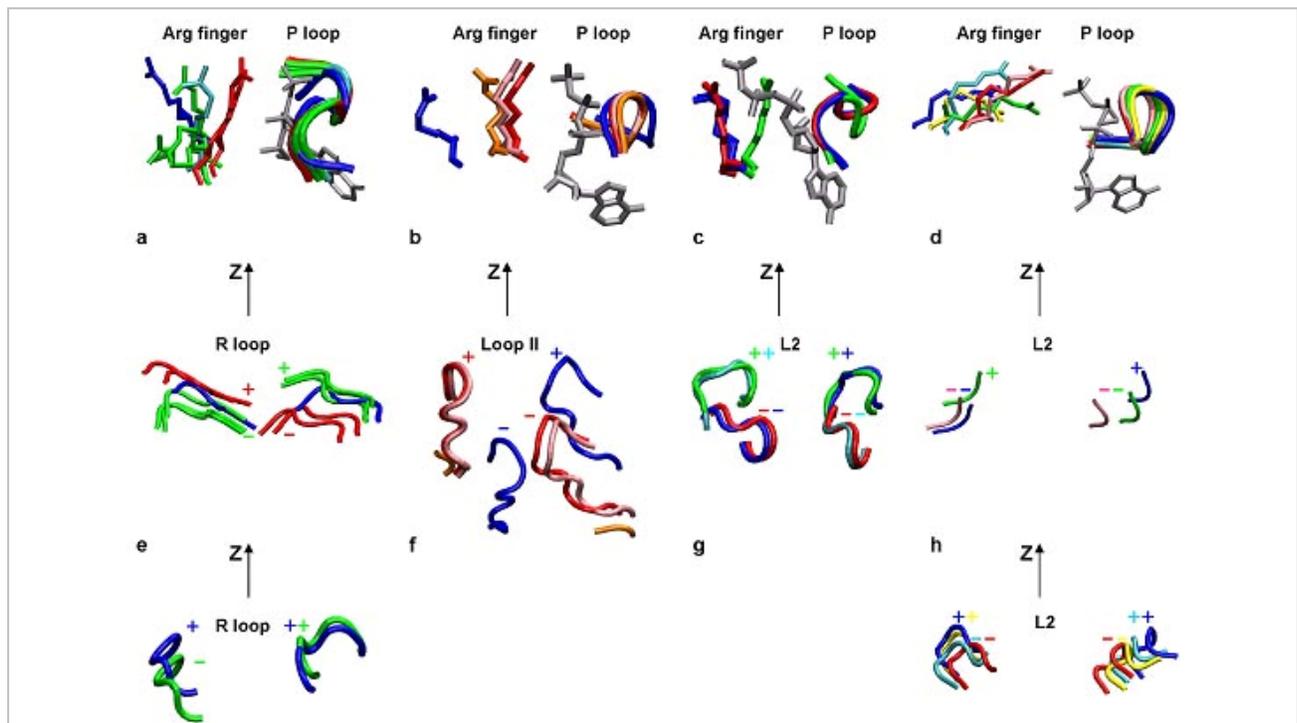
Wenjun Zheng, Ph.D.  
Physics Department, University at Buffalo,  
Buffalo, NY 14260, USA  
Email: wjzheng@buffalo.edu



1. Six nucleotide-binding sites are located at the interfaces between adjacent subunits, each of which consists of cis-residues (including Walker A (or P loop) and Walker B motifs, see Walker *et al.*, 1982) from one subunit, and trans-residues (including a conserved arginine finger) from another subunit. The arginine finger (together with other trans-residues, see Gai *et al.*, 2004) is thought to play a key role in enabling cooperative NTP binding and hydrolysis (Scheffzek *et al.*, 1997) involving all six nucleotide-binding sites (Liao *et al.*, 2005; Crampton *et al.*, 2006; Adelman *et al.*, 2006).
2. Positively charged loops (including R loop of Rho, loop II of T7, L2 loop of P4 and an equivalent loop in LTag and E1) are protruded toward the central channel, which are implicated in substrate binding by structural (Skordalakes *et al.*, 2006; Singleton, *et al.*, 2000; Enemark, *et al.*, 2006) and mutational (Crampton *et al.*, 2006; Mori *et al.*, 1989; Miwa *et al.*, 1995; Wei *et al.*, 2001; Notarnicola *et al.*, 1995; Washington *et al.*, 1996; Lisal *et al.*, 2005) studies. Furthermore, structural studies suggested that these loops lever up and down the central axis in response to changes in the nucleotide state (Singleton, *et al.*, 2000; Mancini *et al.*, 2004;

Enemark, *et al.*, 2006; Gai *et al.*, 2004) — possibly driving the substrate translocation in a nucleotide-dependent manner.

A sequential NTP hydrolysis and substrate translocation mechanism has emerged from numerous biochemical, biophysical, genetic and structural studies of hexameric helicases (see Patel *et al.*, 2000; Donmez *et al.*, 2006; Singleton *et al.*, 2007). It postulates that the six nucleotide-binding sites sequentially bind and hydrolyze NTP, which is allosterically coupled to the binding and translocation of substrate. Alternative mechanisms (such as concerted hydrolysis (Gai *et al.*, 2004) or stochastic hydrolysis (Martin *et al.*, 2005)) were also discussed. Various sequential mechano-chemical models have been proposed for Rho (Skordalakes *et al.*, 2006; Adelman *et al.*, 2006; Stitt *et al.*, 1998; Kim *et al.*, 1999), T7 (Singleton, *et al.*, 2000; Liao *et al.*, 2005; Hingorani *et al.*, 1997), P4 (Mancini *et al.*, 2004; Lisal *et al.*, 2005) and E1 (Enemark, *et al.*, 2006). Despite sharing some common features, most of these models were proposed based on experimental studies of a particular helicase. These models differ in their predictions of various mechano-chemical properties of hexameric helicases, including the direction of hydrolysis (counterclockwise or clockwise),



**Figure 2:** Assignment of nucleotide states and positional states of substrate-binding loop pairs in  $F_1$  ATPase and five hexameric helicases by structural comparisons: (a)-(d) show the nucleotide-binding sites (P loops shown as tubes, arginine fingers and bound nucleotides shown as bonds); (e)-(j) show the substrate-binding loop pairs (the central Z axis and +/- assignments are shown). The color schemes are defined as follows:

(a). The nucleotide-binding sites in  $F_1$  (PDB code: 1BMF) and Rho (PDB code: 1PVO): the E, T, D state of  $F_1$  is colored blue, red, green; the D state of Rho is colored cyan; the bound ADP is colored gray. (b). The nucleotide-binding sites in T7 (PDB code: 1E0J, 1CR1): the empty site in 1E0J is assigned to E state (colored blue); the following sites are assigned to T state — the tight/loose ADPNP-bound sites in 1E0J (colored red/pink), the dTTP-bound site in 1CR1 (colored orange); the bound ADPNP is colored gray. (c). The nucleotide-binding sites in P4 (PDB code: 1W4C, 1W46, 1W48): the E, T, D state is colored blue, red, green; the bound ADPCPP is colored gray. (d). The nucleotide-binding sites in LTag (PDB code: 1SVO, 1SVL, 1SVM) and E1 (PDB code: 2GXA): the E, T, D-II, D-I state of LTag is colored blue, red, yellow, cyan; the T, D state of E1 is colored pink, green; the bound ATP is colored gray. (e). The R loop pairs in  $F_1$  (PDB code: 1BMF): the R loop pair adjoined to the nucleotide-binding site at E, T, D state is colored blue, red, green. The T, D state is mapped to (+, -), (-, +). (f). The loop II pairs in T7 (PDB code: 1E0J, 1CR1): the loop II pairs are colored the same as the nucleotide-binding site they are adjoined to. The T, E state is mapped to (+, -), (-, +). Note the loop II of 1CR1 is mostly disordered. (g). The L2 pairs in P4 (PDB code: 1W4C, 1W46, 1W48): the L2 pair is colored green, red if they are adjoined to the nucleotide-binding site at D, T state, and blue or cyan if adjoined to E state. The T, D state is mapped to (-, -), (+, +); the E state is mapped to (+, -) or (-, +). (h). The loop pairs in E1 (PDB code: 2GXA): the loop pair adjoined to the nucleotide-binding site at E, T, D state is colored blue, red, green. The E, T, D state is mapped to (-, +), (-, -), (+, -). (i). The R loop pairs in Rho (PDB code: 2HT1, 1PVO): the R loop pair adjoined to the nucleotide-binding site at E, D state is colored blue, green. The E, D state is mapped to (+, +), (-, +). (j). The loop pairs in LTag (PDB code: 1SVO, 1SVL, 1SVM): the loop pair adjoined to the nucleotide-binding site at E, T, D-II, D-I state is colored blue, red, yellow, cyan. The E, T state is mapped to (+, +), (-, -); the D state is mapped to (+, -) or (-, +).

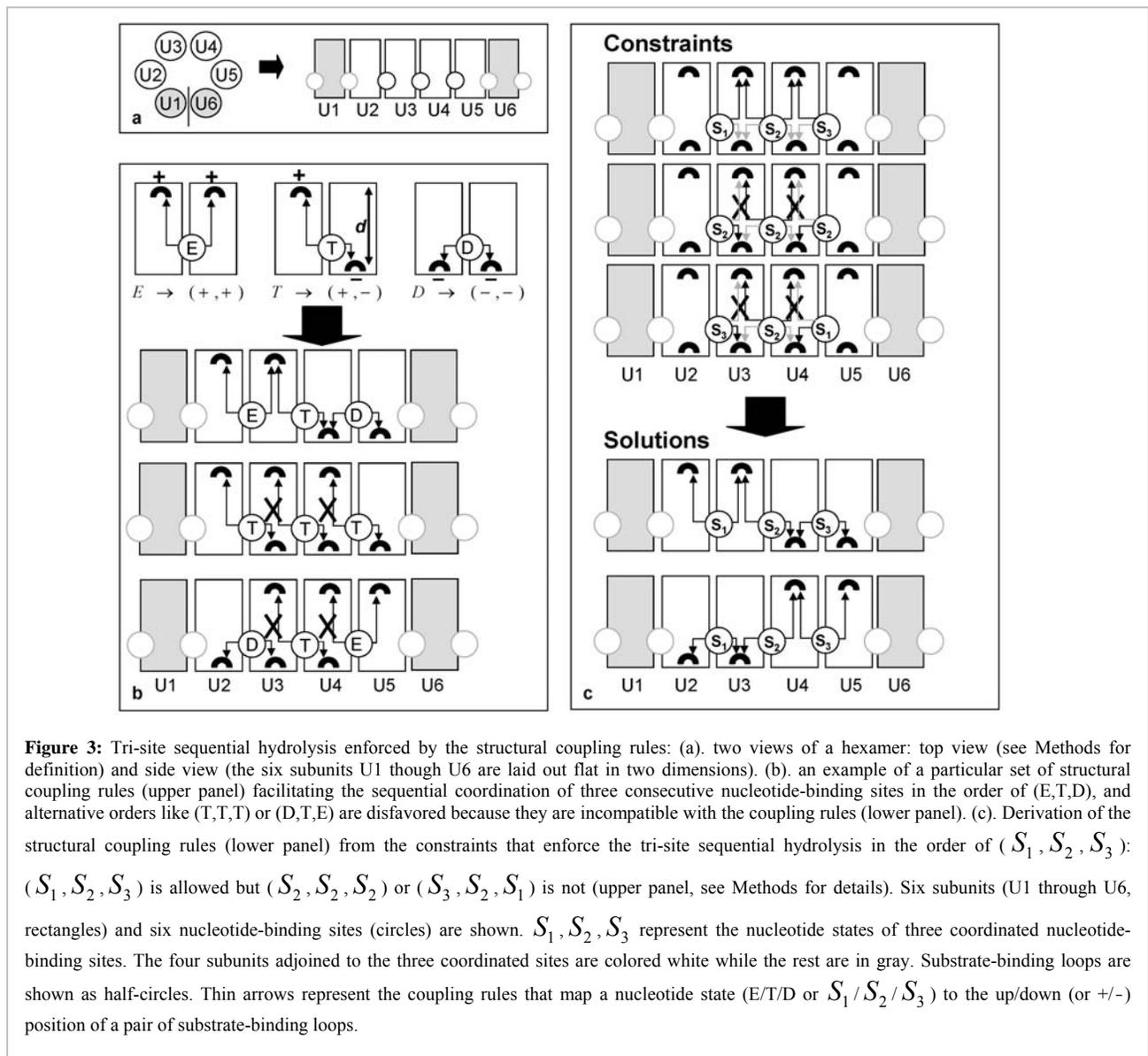
substrate-binding mode (different number of subunits bound to substrate at one time), the direction and step size of translocation, the chemical step coupled with force generation (NTP binding or hydrolysis or product release), etc. The controversy between different models is partly due to the mechanistic diversity in hexameric helicases and the lack of a general model that can accommodate their different mechano-chemical properties.

The objective of this work is three fold:

1. Shed some light on the structural basis of mechano-chemical coupling and cooperative NTP hydrolysis in hexameric helicases using abstracted structural coupling rules deduced from structural data.

2. Build a general model with multiple schemes to account for both conserved and variable features of the mechano-chemical coupling in five hexameric helicases.
3. Make predictions for various mechano-chemical properties of five hexameric helicases for experimental validations.

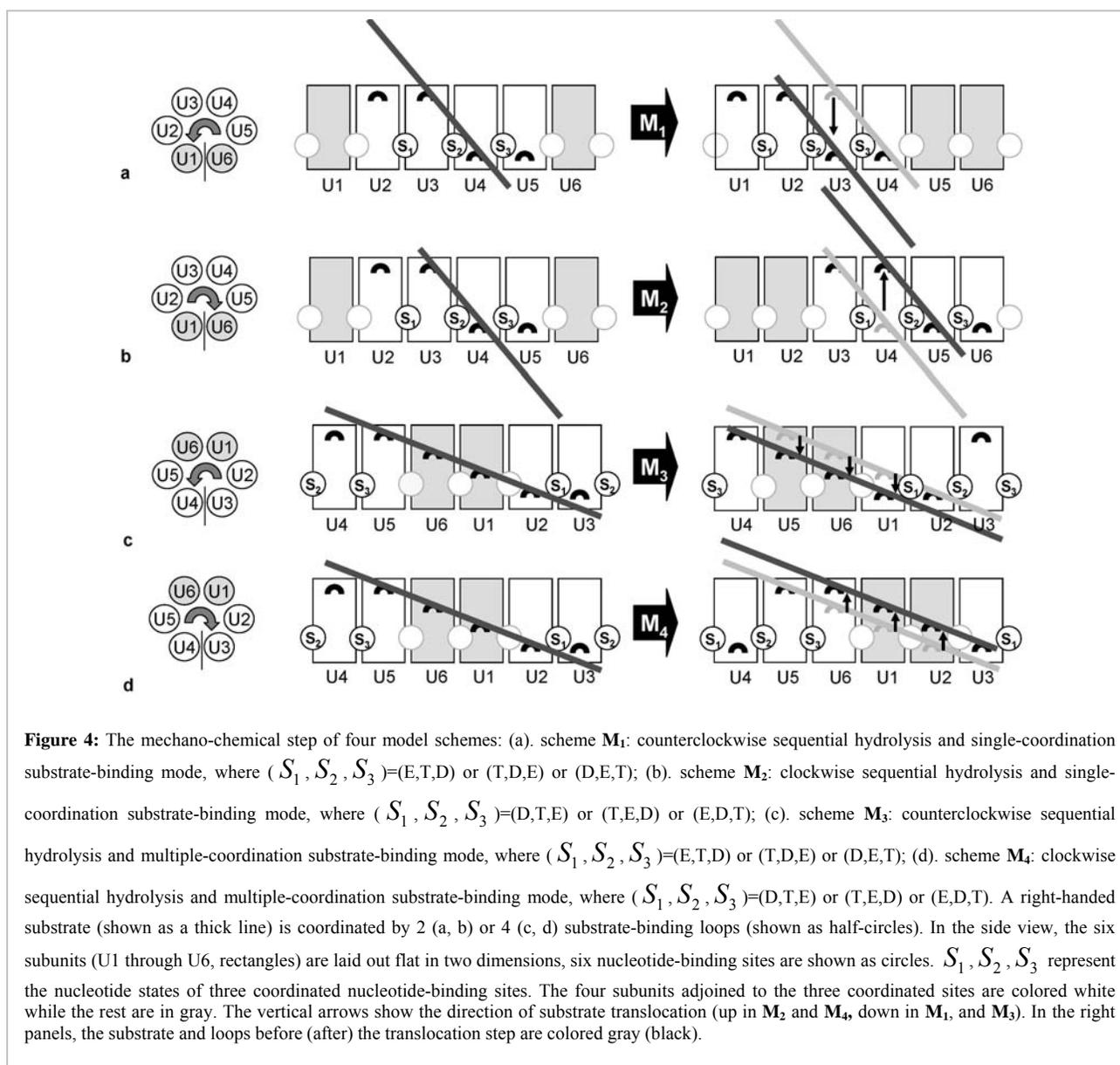
To this end, we will propose a tri-site sequential model of NTP hydrolysis and substrate translocation in hexameric helicases. Our model is motivated by the tri-site sequential hydrolysis model (Weber *et al.*, 2000) of  $F_1$  ATPase — a hexameric rotary motor. In  $F_1$ , a cooperative change in nucleotide-binding affinity (Boyer 1993) was proposed to drive the three catalytically active nucleotide-binding sites to cycle through three distinct chemical states (empty, NTP-



bound and NDP-Pi bound) (Weber *et al.*, 2000). A similar mechanism in hexameric helicases has long been sought for based on the striking structural similarity and the conservation of NTPase motifs between  $F_1$  and hexameric helicases — especially Rho (a sequence homologue of  $F_1$ , see Mori *et al.*, 1989) and T7 (Washington *et al.*, 1996; Hingorani *et al.*, 1997). A major difference from the tri-site model of  $F_1$  is that our model assumes all six nucleotide-binding sites are catalytically active although only three of them are sequentially coordinated at one time (see Subsection III).

Our model is based on a set of structural coupling rules between the chemical state of a nucleotide-binding site and the longitudinal positions of two substrate-binding loops in the subunits adjoined to this site. These rules are deduced from a comprehensive analysis of multiple crystal structures of  $F_1$  ATPase and five hexameric

helicases (Rho, T7, P4, E1 and LTag). These rules allow adjacent nucleotide-binding sites to be coupled together to facilitate a sequential coordination of three consecutive sites at one time. To further couple the NTPase activities to substrate translocation, we consider two alternative substrate-binding modes involving different number of subunits. Four distinct schemes ( $M_1 \sim M_4$ ) exist in our model, which differ in both hydrolysis direction and substrate-binding mode (Fig. 4). We then apply our model to Rho, T7, P4, E1 and LTag, and make specific predictions for a variety of mechano-chemical properties (hydrolysis direction, NTP-binding cooperativity, substrate-binding mode, translocation direction and step size, force-generation step). Some of these predictions agree with past experiments whereas the others call for future experimental verifications.



## MATERIALS AND METHODS

### Determination of directions of sequential hydrolysis and substrate translocation

To allow a comparison between different hexameric helicases, we determine the directions of hydrolysis and translocation using the following procedures:

1. We adopt the top view of a hexamer such that the arginine finger (P loop) is located on the left (right) side of an inter-subunit interface (Fig. 1). From this top view, a sequential hydrolysis can proceed in either counterclockwise or clockwise direction (Fig. 4).

2. The direction of Z axis of the central channel is set to be opposite to the viewing direction of the top view (the Z axis points out of the paper in Fig. 1). The substrate-binding loops move up or down along the Z axis in response to changes in nucleotide state. As a result, the substrate is translocated along either +Z or -Z direction.

### Assignment of nucleotide state by structural alignment

We assign the three nucleotide states (E: empty, T: NTP-bound, D: NDP-Pi-bound or NDP-bound, where Pi stands for inorganic phosphate) to the conformations of nucleotide-binding sites from crystal structures of hexameric helicases and  $F_1$  ATPase as follows: first, we overlay these conformations (together with the bound nucleotide) by structurally aligning the alpha carbon atoms of the P loop residues (residues 159-164 of the

$\beta$  subunits in  $F_1$ , 172-177 of the  $\alpha$  subunits in  $F_1$ , 181-186 of Rho, 315-320 of T7, 133-138 of P4, 436-441 of E1, 429-434 of LTag) and the arginine finger (R356 of the  $\beta$  subunits in  $F_1$ , R373 of the  $\alpha$  subunits in  $F_1$ , R366 of Rho, R522 of T7, R279 of P4, R538 of E1, R540 of LTag); second, we inspect the position and orientation of the arginine sidechain relative to the bound nucleotide to assign these conformations to T/D/E state (see Subsection II of Results).

We have made assignments for the conformations of nucleotide-binding sites in the following structures:

**$F_1$  and Rho:** In a structure of  $F_1$  (PDB code: 1BMF, Fig. 1a), among the three catalytically active nucleotide-binding sites, the empty one is assigned to the E state, and the other two (named DP and TP in Abrahams *et al.*, 1994) are both assigned to the T state because the arginine sidechains in these two sites are structurally very similar, and both are in close proximity to the bound nucleotide (Fig. 2a). The three catalytically inactive sites are assigned to the D state, where the arginine sidechain conformations are intermediate between the E state and the T state (two different intermediate arginine conformations are found in these three sites, Fig. 2a). Next, we overlay the nucleotide-binding site of an open-ring structure of Rho bound with ATP analogs (PDB code: 1PVO, Fig. 1b) to those of  $F_1$ , and assign it to the D state (Fig. 2a). In another closed-ring structure of Rho (PDB code: 2HT1), its nucleotide-binding site is empty with the arginine finger disordered, so we assign it to the E state.

**T7:** In a ‘dimer of trimers’ structure of T7 (PDB code: 1E0J, Fig. 1c), among the three distinct nucleotide-binding sites, the empty one is assigned to the E state, and the other two (bound with ATP analogs, assigned to ATP-bound state and ADP-Pi-bound state in Singleton *et al.*, 2000) are both assigned to the T state, because the arginine sidechain is structurally very close between them (Fig. 2b). Then we overlay the nucleotide-binding site of another dTTP-bound T7 structure (PDB code: 1CR1), and assign it to the T state (Fig. 2b). The D state is unassigned for T7.

**P4:** In three P4 structures — apo (PDB code: 1W4C), ATP-analog-bound (PDB code: 1W48, Fig. 1d) and Mg-ADP-bound (PDB code: 1W46), their nucleotide-binding sites are assigned to the E, T and D state (Fig. 2c), which is in agreement with Mancini *et al.*, 2004.

**E1:** In an E1 structure (PDB code: 2GXA, Fig. 1e), the three types of nucleotide-binding sites (form I, II and III as defined in Enemark *et al.*, 2006) are assigned to the T, D and E state (Fig. 2d), which is in agreement with Enemark *et al.*, 2006.

**LTag:** In three LTag structures — apo (PDB code: 1SVO), ATP-bound (PDB code: 1SVM, Fig. 1f) and ADP-bound (PDB code: 1SVL), their nucleotide-binding sites are assigned to the E, T and D state (Fig. 2d), which is in agreement with Gai *et al.*, 2004. Note there are two forms of D state (D-I and D-II, see Gai *et al.*, 2004) with different arginine sidechain conformations (Fig. 2d).

### Assignment of positional states to substrate-binding loop pairs by structural alignment

To assign a positional state ( $\pm, \pm$ ) (+: up, -: down) to a pair of substrate-binding loops from adjacent subunits, we structurally align pairs of adjacent subunits and then determine the up/down (+/-) positions of the substrate-binding loops and the offset value  $d$  (difference in Z coordinate between the two loops). Here the substrate-binding loop is chosen to be the R loop of  $F_1$  and Rho (residues 314-318 of the  $\beta$  subunits in  $F_1$ , 331-335 of the  $\alpha$  subunits in  $F_1$ , 321-325 of Rho), the loop II of T7 (residues 466-474), the L2 loop of P4 (residues 238-242) and a structurally equivalent loop of E1 and LTag (residues 479-481 of E1, 480-485 of LTag). The longitudinal position of this loop is found to be correlated with the other substrate-binding loops, so the results do not depend on the choice of a particular loop. Combined with the assignment of nucleotide states, we can obtain a set of coupling rules that map a nucleotide state to the positional state of a pair of substrate-binding loops (Table 1 [Supplementary data]). Note that in the structurally deduced coupling rules, the offset value  $d$  may vary between structures (represented by  $d, d', d''$  etc). For simplicity, we assume such variations do not lead to additional states beyond the four discrete states: (+,-), (-,+), (-,-) and (+,+).

We have made assignments for the conformations of substrate-binding loop pairs in the following structures:

**$F_1$  and Rho:** In a structure of  $F_1$  (PDB code: 1BMF, Fig. 1a), the two R loop pairs adjoined to the two nucleotide-binding sites at the T state are both assigned to (+,-) (offset  $d' \sim 4\text{-}5\text{\AA}$ , Fig. 2e); the three R loop pairs adjoined to the three nucleotide-binding sites at the D state are all assigned to (-,+) (offset  $d'' \sim 2\text{-}4\text{\AA}$ , Fig. 2e). Based on the alignment between two Rho structures (PDB code: 1PVO and 2HT1) we assign 1PVO to (-,+) and 2HT1 to (+,+) (offset  $d \sim 8.5\text{\AA}$ , Fig. 2i).

**T7:** In a ‘dimer of trimers’ structure of T7 (PDB code: 1E0J, Fig. 1c), the loop II pair adjoined to the nucleotide-binding site at the E state is assigned to (-,+) (offset  $d \sim 11\text{\AA}$ , Fig. 2f); the two loop II pairs

adjoined to the two nucleotide-binding sites at the T state are both assigned to  $(+,-)$  (offset  $d' \sim 6\text{\AA}$ , Fig. 2f). Another dTTP-bound T7 structure (PDB code: 1CR1) is also assigned to  $(+,-)$  (Fig. 2f).

**P4:** In two P4 structures — ATP-analog-bound (PDB code: 1W48, Fig. 1d) and Mg-ADP-bound (PDB code: 1W46), it is straightforward to assign  $(-,-)$  and  $(+,+)$  to their L2 loop pairs (offset  $d \sim 6\text{\AA}$ , Fig. 2g) respectively. In a third apo structure (PDB code: 1W4C), three different positional states exist:  $(+,-)$ ,  $(-,+)$  and  $(-,-)$  (Fig. 2g).  $(-,-)$  is removed from the coupling rule for the E state because it collides with that of the T state.

**E1:** In an E1 structure (PDB code: 2GXA, Fig. 1e), the T, D and E state are mapped to  $(-,-)$ ,  $(+,-)$  (offset  $d' \sim 3\text{\AA}$ ) and  $(-,+)$  (offset  $d \sim 11\text{\AA}$ , Fig. 2h).

**LTag:** In two LTag structures — apo (PDB code: 1SVO) and ATP-bound (PDB code: 1SVM, Fig. 1f), the E and T state are mapped to  $(+,+)$  and  $(-,-)$  (offset  $d \sim 17\text{\AA}$ , Fig. 2j). In a third ADP-bound structure (PDB code: 1SVL), three different positional states exist:  $(+,-)$ ,  $(-,+)$  and  $(+,+)$  (offset  $d' \sim 6\text{\AA}$ , Fig. 2j).  $(+,+)$  is removed from the coupling rule for the D state because it collides with that of the E state.

Note that in LTag and E1, we use the displacements of the DNA-binding  $\beta$  hairpins (residues 501-514 of E1, 507-520 of LTag) to estimate the offset values.

### Determination of structural coupling rules to ensure sequential coordination of tri-site hydrolysis

To ensure a sequential coordination of three consecutive nucleotide-binding sites in the order of three distinct nucleotide states ( $S_1, S_2, S_3$ ) — a permutation of (E,T,D), we impose the following constraint: the  $S_2$  state at a nucleotide-binding site is only compatible with  $S_1$  (not  $S_2$  or  $S_3$ ) state at its left neighboring site, and  $S_3$  (not  $S_1$  or  $S_2$ ) state at its right neighboring site (Fig. 3c). If we represent the unknown structural coupling rules

as  $S_1 \rightarrow (Z_{\alpha_1}, Z_{\beta_1}), S_2 \rightarrow (Z_{\alpha_2}, Z_{\beta_2}), S_3 \rightarrow (Z_{\alpha_3}, Z_{\beta_3})$ ,

where  $Z_{\alpha_i}, Z_{\beta_i}, Z_{\alpha_2}, Z_{\beta_2}, Z_{\alpha_3}, Z_{\beta_3} \in \{+,-\}$  represent the up/down positions of substrate-binding loops, the above constraints can be symbolically formulated as

$$Z_{\beta_1} = Z_{\alpha_2}, Z_{\beta_2} = Z_{\alpha_3}, Z_{\alpha_2} \neq Z_{\beta_2}, Z_{\beta_3} \neq Z_{\alpha_2}, Z_{\beta_2} \neq Z_{\alpha_1}.$$

For example, the compatibility of  $S_2$  with  $S_1$  at the left neighboring site gives  $Z_{\beta_1} = Z_{\alpha_2}$ , and the incompatibility of  $S_2$  with  $S_2$  at the left neighboring site gives  $Z_{\alpha_2} \neq Z_{\beta_2}$ . Only two sets of coupling rules satisfy these constraints (Fig. 3c):

$$S_1 \rightarrow (+,+), S_2 \rightarrow (+,-), S_3 \rightarrow (-,-),$$

and

$$S_1 \rightarrow (-,-), S_2 \rightarrow (-,+), S_3 \rightarrow (+,+).$$

No other mappings satisfy the above constraints. For example,  $S_1 \rightarrow (+,+), S_2 \rightarrow (+,-), S_3 \rightarrow (-,+)$  would allow both  $S_1$  and  $S_3$  at the left neighboring site of  $S_2$  which violates the requirement of tri-site coordination by  $S_2$  state.

These two solutions combined with two possible hydrolysis direction (counterclockwise or clockwise) result in four schemes of sequential models (Fig. 4a-4d).

## RESULTS

We will divide the Results section into the following parts: In Subsection I, we will summarize the experimental evidence for the cis- and trans-coupling between a nucleotide-binding site and the substrate-binding loops in its two adjoined subunits. In Subsection II, we will symbolize the cis- and trans-coupling with a set of structural coupling rules that map the nucleotide state of a nucleotide-binding site to the longitudinal positions of the two substrate-binding loops in its adjoined subunits. These coupling rules will be deduced by a comprehensive analysis of crystal structures of F<sub>1</sub>, Rho, T7, P4, E1 and LTag. In Subsection III and IV, based on the structural coupling rules, we will introduce a sequential mechano-chemical model — including four distinct model schemes which differ in the direction of sequential hydrolysis and the substrate-binding mode. In Subsection V, we will apply our general model to five hexameric helicases (Rho, T7, P4, E1 and LTag) in light of existing experimental data and make predictions for various mechano-chemical properties of these hexameric helicases.

### I. Two communication paths from nucleotide-binding site to substrate-binding loops: cis- vs. trans-coupling

The foundation of our mechano-chemical model is the postulation that the chemical state of a nucleotide-binding site is strongly coupled to the longitudinal positions of the substrate-binding loops in the two adjoined subunits. Because the nucleotide-binding sites of a hexameric helicase are located at inter-subunit interfaces, such coupling may be mediated by two communication paths:

1. **cis-coupling**: between the cis-residues of a nucleotide-binding site (such as P loop and Walker B motif) and the substrate-binding loops in the subunit adjoined to the cis-residues (referred as  $\beta$  subunit, following the subunit naming of  $F_1$ );
2. **trans-coupling**: between the trans-residues of a nucleotide-binding site (such as arginine finger) and the substrate-binding loops in the subunit adjoined to the trans-residues (referred as  $\alpha$  subunit, following the subunit naming of  $F_1$ ).

The cis-coupling has been explored in several structural studies of Rho, T7 and P4, where key residues and interactions involved in this coupling have been identified. In a Rho structure, the substrate-binding R loop is in close proximity with P loop (Skordalakes *et al.*, 2006). In a T7 structure, H465 of loop II (a substrate-binding loop equivalent to the R loop of Rho and  $F_1$ ) forms a hydrogen bond with the  $\gamma$ -phosphate of dTTP (Sawaya *et al.*, 1999). In P4, a similar  $\gamma$ -phosphate sensing residue (N234) is also found in loop L2 (a substrate-binding loop equivalent to loop II of T7) (Mancini *et al.*, 2004). In SF3 helicases (LTag, E1), a loop equivalent to the R loop of Rho lies adjacent to the Walker B motif, allowing its direct coupling to ATPase activity (Fig. 4 of Hickman *et al.*, 2005).

Evidence for the trans-coupling has also been found. In many NTPase proteins, NTP hydrolysis requires the alignment of the arginine finger to the  $\gamma$  phosphate of NTP, which is coupled to a relative rotation between the NTP-binding domains of the two subunits adjoined to the nucleotide-binding site. This nucleotide-dependent domain rotation can in turn move the substrate-binding loops longitudinally, as observed in T7 (Singleton *et al.*, 2000), E1 (Enemark *et al.*, 2006) and LTag (Gai *et al.*, 2004). Besides coupling through inter-subunit rotations, the trans-coupling may also be transmitted via intra-subunit conformational changes. For example, in P4, a conformational coupling was found between loop L2 and the arginine finger (R279) via a hydrogen bond between R279 and R251 (Mancini *et al.*, 2004), and another trans-residue S252 was found to serve as a  $\gamma$ -phosphate sensor that controls the position of loop L2 (Kainov *et al.*, 2008).

Therefore, both cis- and trans- couplings are likely present in hexameric helicases to mediate nucleotide-dependent movements of substrate-binding loops, although the coupling details may vary between different hexameric helicases (see Subsection II).

## II. Structural coupling rules map nucleotide state to positions of two substrate-binding loops

The cis- and trans- couplings allow a change of chemical state at a nucleotide-binding site to alter the positions of two adjoined substrate-binding loops (one in the  $\alpha$  subunit and the other in the  $\beta$  subunit). To quantify these allosteric couplings, we will introduce a set of structural coupling rules that map the nucleotide state to the positional state of a pair of substrate-binding loops in two adjacent subunits. These coupling rules, which vary between different hexameric helicases, will be deduced from a comprehensive structural analysis of multiple hexameric helicases at different nucleotide states (see Methods). These coupling rules will form the structural basis of our sequential mechano-chemical model of hexameric helicases (see Subsection III).

To simplify the description of the structural coupling rules, we make the following assumptions:

1. **States of a nucleotide-binding site**: Only three nucleotide states (E: empty, T: NTP-bound, D: NDP-Pi-bound or NDP-bound, where Pi stands for inorganic phosphate) are considered here. Therefore the NTPase cycle consists of three chemical steps: NTP binding ( $E \Rightarrow T$ ), hydrolysis ( $T \Rightarrow D$ ) and product release ( $D \Rightarrow E$ ). The assignment of a conformation of a nucleotide-binding site to T/D/E state is based on the position of the arginine finger relative to the bound nucleotide — if the arginine's sidechain is positioned in close proximity to the  $\gamma$ -phosphate of the bound NTP (or NTP analog), it is assigned to the T state; if the arginine's sidechain is positioned far away from the bound nucleotide, it is assigned to the E state; the D state is assigned if an intermediate position of the arginine finger is seen. Our assignments are mostly consistent with those used in the structural studies of  $F_1$ , T7, P4, E1 and LTag (Abrahams *et al.*, 1994; Singleton *et al.*, 2000; Mancini *et al.*, 2004; Enemark *et al.*, 2006; Gai *et al.*, 2004) with a few exceptions (see Methods).
2. **Positional states of two substrate-binding loops**: A substrate-binding loop (chosen to be R loop of  $F_1$ /Rho, loop II of T7, loop L2 of P4, and an equivalent loop of E1/LTag, see Methods for residue numbers) is assumed to adopt two discrete positions that are energetically favorable — up (+) or down (-) position along the central Z axis (see Methods for the definition of Z axis). The up and down positions are separated by an offset  $d$  along the Z axis ( $d$  determines the step size of translocation, see Subsection IV). Hence there are only four possible states for the longitudinal positions of a pair of substrate-binding loops: (+, -), (-, +), (+, +) and (-, -). We note that deviations from the four discrete states (with

an intermediate offset value between 0 and  $d$ ) are energetically unfavorable but may be induced and stabilized by substrate binding (see Subsection IV).

Given the above assumptions, we have performed a comprehensive structural analysis to determine the structural coupling rules for Rho, T7, P4, E1 and LTag (for details see Methods, for results see Table 1 & Fig. 2). This analysis is made possible by the availability of crystal structures of F<sub>1</sub>, T7, P4, E1 and LTag with more than one nucleotide state (Abrahams *et al.*, 1994; Singleton *et al.*, 2000; Mancini *et al.*, 2004; Enemark *et al.*, 2006; Gai *et al.*, 2004). These coupling rules are symbolized by a mapping from three nucleotide states (E/T/D) to four positional states ( $\pm, \pm$ ) such as  $E \rightarrow (+,+)$ ,  $T \rightarrow (+,-)$ ,  $D \rightarrow (-,-)$ . Because of the close homology between Rho and F<sub>1</sub>, we assume that the coupling rule for the T state in F<sub>1</sub> also applies to Rho. Among the three nucleotide states,  $T \rightarrow (+,-)$  in F<sub>1</sub>/Rho and T7,  $T \rightarrow (-,-)$  in P4, E1 and LTag. For the D and E state, the mapping is in some cases not unique (see the D state of LTag and the E state of P4 in Table 1) or undefined (see the D state of T7 in Table 1). Such ambiguity may be due to the uncertainty in crystallographically capturing nucleotide states using nucleotide analogues.

### III. Construction of a tri-site sequential model

The structural coupling rules deduced above underlie the mechano-chemical coupling between NTPase events and movements of substrate-binding loop in the same subunit, which was emphasized in various existing models (Singleton *et al.*, 2000; Mancini *et al.*, 2004; Enemark *et al.*, 2006; Gai *et al.*, 2004). In our new model, we argue that these coupling rules not only facilitate intra-subunit mechano-chemical coupling, but also indirectly couple two adjacent nucleotide-binding sites via their trans- and cis-coupling to the same substrate-binding loop, whose up/down position is constrained by the nucleotide states of both nucleotide-binding sites. This constraint allows a nucleotide-binding site to be kinetically coordinated with its two neighboring sites and thus facilitates cooperative NTP binding and hydrolysis among the three consecutive sites. For example, if the structural coupling rules are  $E \rightarrow (+,+)$ ,  $T \rightarrow (+,-)$ ,  $D \rightarrow (-,-)$ , then it can be inferred that the T state at a nucleotide-binding site favors the E state at its left neighboring site and the D state at its right neighboring site (Fig. 3b). Therefore, the binding of NTP in a nucleotide-binding site facilitates product release ( $D \Rightarrow E$ ) and hydrolysis ( $T \Rightarrow D$ ) in its two neighboring sites, which give rise to a sequentially coordinated NTP hydrolysis akin to the tri-site sequential model of F<sub>1</sub> (Weber *et al.*, 2000).

The presence or absence of negative cooperativity in NTP binding can also be deduced from the structural coupling rules. For example, if the T state is mapped to  $(+,-)$ , then the incompatibility of two adjacent T states (Fig. 3b) leads to negative cooperativity in NTP binding — only three NTPs can be bound with high affinity in a hexameric helicase as observed in Rho (Stütt *et al.*, 1990) and T7 (Hingorani *et al.*, 1996). Otherwise, if the T state is mapped to  $(+,+)$  or  $(-,-)$ , two adjacent T states would be compatible with each other, allowing six NTPs to be bound simultaneously, in agreement with the structural and biochemical studies on LTag (Gai *et al.*, 2004), E1 (Enemark *et al.*, 2006) and P4 (Mancini *et al.*, 2004; Lísál *et al.*, 2005).

To generalize the idea that certain structural coupling rules may enforce tri-site sequential hydrolysis (Fig. 3b), we will determine what structural coupling rules are required to facilitate tri-site sequential hydrolysis. Assuming there are six catalytically active nucleotide-binding sites in hexameric helicases (Liao *et al.*, 2005; Crampton *et al.*, 2006; Adelman *et al.*, 2006), we partition them into three coordinated and three uncoordinated sites as follows:

1. **Coordinated nucleotide-binding sites:** three consecutive nucleotide-binding sites (located at the interfaces of subunit pairs U2-U3, U3-U4, U4-U5, where U1 through U6 represent six subunits, Fig. 3c) are sequentially coordinated such that their nucleotide states are maintained in the order of  $(S_1, S_2, S_3)$  which is a fixed permutation of (E, T, D);
2. **Uncoordinated nucleotide-binding sites:** the remaining three nucleotide-binding sites (located at the interfaces of U1-U2, U5-U6, U6-U1, Fig. 3c) are not involved in sequential coordination, and their nucleotide states are assumed to be intermediate between  $S_1$  and  $S_3$ .

The partition of six nucleotide-binding sites into coordinated and uncoordinated sites is only temporary. As the tri-site sequential hydrolysis proceeds along the ring, the set of coordinated sites rotates in counterclockwise or clockwise direction by 60° per step --- a previously uncoordinated site joins it while a previously coordinated site leaves (Fig. 4). Therefore, each nucleotide-binding site is under sequential coordination during 50% of the NTPase cycle time. The two hydrolysis directions are discussed in details as follows:

**Counterclockwise sequential hydrolysis:** after a counterclockwise rotation by 60°, the three new coordinated sites are now located at the interfaces of U1-U2, U2-U3, and U3-U4 (Fig. 4a and 4c). We postulate that the cooperative changes of

nucleotide state in the three new coordinated sites are induced by the chemical step  $S_1 \Rightarrow S_2$  at the middle site (at the U2-U3 interface), which favors  $S_1$  state at its left neighboring site (at the U1-U2 interface) and  $S_3$  state at its right neighboring site (at the U3-U4 interface). Thus we impose the constraint that the  $S_2$  state at a nucleotide-binding site is only compatible with  $S_1$  state at its left neighboring site, and  $S_3$  state at its right neighboring site (Fig. 3c). Only two sets of coupling rules satisfy these constraints (Methods):

$$S_1 \rightarrow (+,+), S_2 \rightarrow (+,-), S_3 \rightarrow (-,-) \quad \text{or} \\ S_1 \rightarrow (-,-), S_2 \rightarrow (-,+), S_3 \rightarrow (+,+).$$

They correspond to scheme  $M_1$  (Fig. 4a) and  $M_3$  (Fig. 4c) of our sequential model, respectively. Under the condition that the chemical step  $S_1 \Rightarrow S_2$  proceeds in the hydrolysis direction ( $E \Rightarrow T \Rightarrow D \Rightarrow E$ ), there are three possible ways of assigning E/T/D to  $S_1/S_2/S_3$ :  $(S_1, S_2, S_3) = (E, T, D)$  or  $(T, D, E)$  or  $(D, E, T)$ .

**Clockwise sequential hydrolysis:** after a clockwise rotation by  $60^\circ$ , the three new coordinated sites are now located at the interfaces of U3-U4, U4-U5, and U5-U6 (Fig. 4b and 4d). We postulate that the cooperative changes of nucleotide state in the three new coordinated sites are induced by the chemical step  $S_3 \Rightarrow S_2$  at the middle site (at the U4-U5 interface), which favors  $S_1$  state at its left neighboring site (at the U3-U4 interface) and  $S_3$  state at its right neighboring site (at the U5-U6 interface). Then we impose the same constraint as in the counterclockwise case, which are satisfied by two sets of coupling rules (see Methods):

$$S_1 \rightarrow (+,+), S_2 \rightarrow (+,-), S_3 \rightarrow (-,-) \quad \text{or} \\ S_1 \rightarrow (-,-), S_2 \rightarrow (-,+), S_3 \rightarrow (+,+).$$

They correspond to scheme  $M_2$  (Fig. 4b) and  $M_4$  (Fig. 4d) of our sequential model, respectively. Under the condition that the chemical step  $S_3 \Rightarrow S_2$  proceeds in the hydrolysis direction ( $E \Rightarrow T \Rightarrow D \Rightarrow E$ ), there are three possible ways of assigning E/T/D to  $S_1/S_2/S_3$ :  $(S_1, S_2, S_3) = (E, D, T)$  or  $(D, T, E)$  or  $(T, E, D)$ .

#### IV. Substrate binding and translocation

Next we discuss how to couple the tri-site sequential hydrolysis with substrate binding and translocation. Here we will consider the following two alternative

substrate-binding modes in light of structural data, and then determine the translocation direction and step size.

1. **Single-coordination substrate-binding mode:** the substrate is bound to two adjacent subunits (U3 and U4 in Fig. 4a and 4b) adjoined to the nucleotide-binding site at  $S_2$  state. This substrate-binding mode is likely present in T7, where an ssDNA substrate was found to only interact with 1~2 subunits at one time (Kim *et al.*, 2002). Because of the right-handedness of DNA/RNA substrate,  $S_2 \rightarrow (+,-)$  is required so that the two substrate-binding loops in subunits U3 and U4 can simultaneously bind the substrate. So this binding mode is only possible in scheme  $M_1$  and  $M_2$  (Fig. 4a and 4b). In scheme  $M_1$  ( $M_2$ ), the substrate is translocated when the substrate-binding loop in subunit U3 (U4) moves down (up) by  $d$ , meanwhile the substrate is released by subunit U4 (U3) and transferred to the next substrate-binding subunit U2 (U5) (Fig. 4a and 4b). Therefore, in scheme  $M_1$  ( $M_2$ ), the translocation is toward  $-Z$  ( $+Z$ ) direction with step size of  $d$ . In scheme  $M_1$ , because the ‘force-generating’ subunit U3 is adjoined to two nucleotide-binding sites at  $S_1$  and  $S_2$  state, the force-generation step is coupled with two chemical steps ( $S_2 \Rightarrow S_3$  and/or  $S_1 \Rightarrow S_2$ ) (Fig. 4a). Similarly, the force-generation step is coupled with  $S_3 \Rightarrow S_2$  and/or  $S_2 \Rightarrow S_1$  in scheme  $M_2$  (Fig. 4b). Note that the above translocation mechanism requires that the substrate binds to subunit U3 (U4) with high affinity and subunit U4 (U3) with low affinity in scheme  $M_1$  ( $M_2$ ).
2. **Multiple-coordination substrate-binding mode:** the substrate is bound to four subunits (U1, U2, U5, and U6 in Fig. 4c and 4d) adjoined to three uncoordinated nucleotide-binding sites in a spiral form (Fig. 4c and 4d). This spiral binding mode is evident from a number of structural studies. In a T7 structure, the DNA-binding loops in three consecutive subunits are positioned in a spiral pattern complementary to a right-handed ssDNA (Singleton *et al.*, 2000). In an E1 structure, an ssDNA is engaged by 5-6 subunits simultaneously (Enemark *et al.*, 2006). Because of the right-handedness of DNA/RNA substrate,  $S_2 \rightarrow (-,+)$  is required to ensure that the four substrate-binding loops in subunits U1, U2, U5, and U6 can simultaneously bind the substrate. So this binding mode is only possible in scheme  $M_3$  and  $M_4$  (Fig. 4c and 4d). In scheme  $M_3$  ( $M_4$ ), the substrate is translocated when the substrate-binding loops in subunits U5, U6, and U1 (U6, U1,

and U2) move down (up) by  $\sim d/3$ , meanwhile the substrate is released by subunit U2 (U5) and bound to the next substrate-binding subunit U4 (U3) (Fig. 4c and 4d). Therefore, in scheme  $\mathbf{M}_3$  ( $\mathbf{M}_4$ ), the translocation is toward  $-Z$  ( $+Z$ ) direction with step size  $\sim d/3$ , if four subunits are engaged in the spiral binding. More generally, if we allow the spiral binding to involve more (5-6) subunits as observed in an E1 structure (Enemark *et al.*, 2006), or fewer (3) subunits as proposed for P4 (Mancini *et al.*, 2004) and T7 (Singleton *et al.*, 2000), the step size will fall in the range  $0.2d-0.5d$ . In scheme  $\mathbf{M}_3$ , because one of the force-generating subunits (U1) is adjoined to a nucleotide-binding site undergoing a chemical step toward  $S_1$  state, the force-generation step is coupled with the chemical step  $S_3 \Rightarrow S_1$  (Fig. 4c). Similarly, the force-generation step is coupled with  $S_1 \Rightarrow S_3$  in scheme  $\mathbf{M}_4$  (Fig. 4d).

The spiral binding requires three pairs of substrate-binding loops to adopt positions that deviate from the four discrete states —  $(+, -)$ ,  $(-, +)$ ,  $(+, +)$  and  $(-, -)$  (see loop pair positions in subunit pairs U5-U6, U6-U1, and U1-U2 in Fig. 4c and 4d). These energetically unfavorable states can be induced and stabilized by substrate binding.

## V. Application to five hexameric helicases

Finally we will apply our tri-site sequential model to five hexameric helicases (Rho, T7, P4, E1, LTag) for which sufficient experimental data are available to fit them to one of the four model schemes (Fig. 4). We will fit the structurally deduced coupling rules for these helicases (Table 1) to the coupling rules for the four model schemes (Table 2 [Supplementary data] and Fig. 4), and identify the best-fit scheme to predict the mechano-chemical properties of these NTPase-based motors (including hydrolysis direction, NTP-binding cooperativity, substrate-binding mode, translocation direction, step size, and force-generation step). Some of our predictions will be compared with past experimental data and alternative models.

In case when some coupling rules can not be simultaneously fitted (for example,  $T \rightarrow (+, -)$  and  $D \rightarrow (-, +)$  cannot be both fitted to any scheme, Table 2) we allow partial fitting using a subset of the structurally deduced coupling rules. This may be justified considering the uncertainty in assigning nucleotide states based on crystal structures (i.e. some coupling rules may be inaccurate).

The translocation direction ( $+Z$  or  $-Z$ ) has been determined for T7 (Egelman *et al.*, 1995), P4 (Mancini

*et al.*, 2004) and Rho (Richardson *et al.*, 2002) from experimental data. To further limit the number of schemes that can fit, we will impose an additional constraint of translocation direction in case of T7 and P4.

### Rho:

The following two schemes are identified:

$\mathbf{M}_1$  (Fig. 4a, where  $S_1 = E$ ,  $S_2 = T$ ,  $S_3 = D$ ): in this scheme, our model predicts the following mechano-chemical properties: counterclockwise sequential hydrolysis (similar to  $F_1$ ), negative cooperativity in NTP binding, single-coordination substrate-binding mode, translocation in  $-Z$  direction with step size  $\sim 4-5 \text{ \AA} \sim 1-2$  bases per NTP hydrolyzed, and the force-generation step coupled to NTP binding ( $E \Rightarrow T$ ) and/or hydrolysis ( $T \Rightarrow D$ ). The predictions regarding NTP binding cooperativity (Patel *et al.*, 2000) and translocation direction (Richardson *et al.*, 2002) agree with previous experiments. The predicted step size also roughly agrees with the measured kinetic step size of Rho (0.5-1 base per NTP hydrolyzed (Walstrom *et al.*, 1997).

$\mathbf{M}_3$  (Fig. 4c, where  $S_1 = T$ ,  $S_2 = D$ ,  $S_3 = E$ ): this scheme predicts counterclockwise sequential hydrolysis, absence of cooperativity in NTP binding, multiple-coordination substrate-binding mode, translocation in  $-Z$  direction, and the force-generation step coupled to NTP binding ( $E \Rightarrow T$ ). Some of these predictions overlap with  $\mathbf{M}_1$ . However, the prediction of no cooperativity in NTP binding contradicts experimental results (Patel *et al.*, 2000).

Therefore, we favor scheme  $\mathbf{M}_1$  for describing Rho's sequential NTP hydrolysis and substrate translocation. This model scheme largely agrees with and complements another sequential model of Rho (Adelman *et al.*, 2006), which proposed that the substrate translocation is driven by the weak to tight nucleotide-binding transition, and is coupled to the NTPase-coordinated changes in substrate-binding affinity of adjacent subunits. This model focused on the NTPase kinetics while our model focuses on the structural couplings underlying the cooperative NTP hydrolysis and substrate translocation.

### T7:

The following two schemes are identified:

$\mathbf{M}_1$  (Fig. 4a, where  $S_1 = E$ ,  $S_2 = T$ ,  $S_3 = D$ ): this scheme predicts counterclockwise sequential hydrolysis, negative cooperativity in NTP binding, single-coordination substrate-binding mode, translocation in  $-Z$  direction with step size  $\sim 6 \text{ \AA} \sim 2$  bases per NTP hydrolyzed, and the force-generation step coupled to

NTP binding ( $E \Rightarrow T$ ) and/or hydrolysis ( $T \Rightarrow D$ ). The prediction regarding NTP binding cooperativity is supported experimentally (Patel *et al.*, 2000). The predicted step size agrees with the measured kinetic step size of T7 (2-3 bases per NTP hydrolyzed (Kim *et al.*, 2002)).

**M<sub>3</sub>** (Fig. 4c, where  $S_1 = D, S_2 = E, S_3 = T$ ): similar to Rho, this scheme predicts no cooperativity in NTP binding which disagrees with the experimental observations of negatively cooperative NTP binding in T7 (see Patel *et al.*, 2000).

Therefore, similar to Rho, we favor scheme **M<sub>1</sub>** for describing T7's sequential NTP hydrolysis and substrate translocation. This model scheme essentially agrees with a previously proposed sequential model (Liao *et al.*, 2005). Our model has complemented this model by providing structural coupling details for the cooperative steps underlying the sequential DNA bind-release cycles in T7 (Liao *et al.*, 2005). Interestingly, recent kinetic studies have revealed significant similarity in NTPase kinetics between T7 (Liao *et al.*, 2005) and Rho (Adelman *et al.*, 2006), which is in line with our finding that the same scheme of a tri-site sequential model applies to both of them. The prediction of possible coupling of hydrolysis step ( $T \Rightarrow D$ ) to force generation agrees with the recent kinetic finding that the force-producing step during DNA unwinding by T7 is associated with dTTP hydrolysis or Pi release (Donmez *et al.*, 2008).

#### **P4:**

The following scheme is identified:

**M<sub>4</sub>** (Fig. 4d, where  $S_1 = T, S_2 = E, S_3 = D$ ): in this scheme, our model predicts clockwise sequential hydrolysis, lack of NTP binding cooperativity, multiple-coordination substrate-binding mode, translocation in +Z direction with step size  $\sim 3\text{Å} \sim 1$  base per NTP hydrolyzed (assuming three subunits are engaged with substrate at one time (Kainov *et al.*, 2008)), and the force-generation step coupled to NTP hydrolysis ( $T \Rightarrow D$ ). Among these predictions, the prediction of no NTP-binding cooperativity agrees with a kinetic study of P4 (Lísal *et al.*, 2005). A different kinetic step size of two bases per NTP hydrolyzed was deduced from the observed NTP consumption in a DNA packaging motor gp16 of  $\Phi 29$  (Guo *et al.*, 1987). A direct measurement of mechanical step size of P4 is needed to resolve this difference. The prediction that force generation in P4 is coupled with NTP hydrolysis ( $T \Rightarrow D$ ) agrees with the observation that substrate binding to P4 has no effect on the kinetics of nucleotide binding or release, and only affects NTP hydrolysis (Lísal *et al.*, 2005).

Our model scheme (**M<sub>4</sub>**) is similar to another sequential model of P4 (Mancini *et al.*, 2004) — both predict clockwise hydrolysis (Note: Mancini *et al.*, 2004 adopted an opposite top view so their 'counterclockwise hydrolysis' is the same as our 'clockwise hydrolysis'), spiral binding of substrate (involving at least three subunits), translocation in +Z direction, and force generation coupled to NTP hydrolysis. However, unlike our three-state (E, T, D) model, their model has only two nucleotide states (T and D). If we assume rapid binding of NTP and slow NTP hydrolysis in scheme **M<sub>4</sub>**, then our three-state model is reduced to the two-state model of Mancini *et al.*, 2004.

Despite belonging to the DnaB-like family of SF4 helicases, P4 and T7 differ significantly in their mechano-chemical mechanisms as described by two different model schemes — they differ in hydrolysis direction, NTP binding cooperativity, substrate-binding mode, translocation direction, and force-generation step. Despite such extensive differences, they can both be described with the framework of our general sequential model.

#### **E1:**

The following two schemes are identified:

**M<sub>2</sub>** (Fig. 4b, where  $S_1 = E, S_2 = D, S_3 = T$ ): it predicts clockwise sequential hydrolysis, no NTP binding cooperativity, single-coordination substrate-binding mode, translocation in +Z direction with step size  $\sim 3\text{Å} \sim 1$  base per NTP hydrolyzed, and the force-generation step coupled to NTP hydrolysis ( $T \Rightarrow D$ ) and/or product release ( $D \Rightarrow E$ ).

**M<sub>4</sub>** (Fig. 4d, where  $S_1 = T, S_2 = E, S_3 = D$ ): it predicts clockwise sequential hydrolysis, no NTP binding cooperativity, multiple-coordination substrate-binding mode, translocation in +Z direction with step size  $\sim 2-6\text{Å} \sim 1-2$  bases per NTP hydrolyzed, and the force-generation step coupled to NTP hydrolysis ( $T \Rightarrow D$ ).

The two schemes share several common features but differ in substrate-binding mode. We favor scheme **M<sub>4</sub>** for describing E1's sequential NTP hydrolysis and substrate translocation because it agrees with the structural observation of spiral binding of substrate (Enemark *et al.*, 2006).

The predictions of scheme **M<sub>4</sub>** mostly agree with a coordinated escort model (Enemark *et al.*, 2006) which postulates that all six subunits simultaneously escort the substrate translocation via the nucleotide-dependent movements of substrate-binding hairpins. Scheme **M<sub>4</sub>** has provided a different structural mechanism for the 'coordinated escort' involving the sequential

coordination of three rather than all six nucleotide-binding sites.

### LTag:

The following two schemes are identified:

**M<sub>2</sub>** (Fig. 4b, where  $S_1 = E, S_2 = D, S_3 = T$ ): it predicts clockwise sequential hydrolysis, no NTP binding cooperativity, single-coordination substrate-binding mode, translocation in +Z direction with step size  $\sim 2$ -6 bases per NTP hydrolyzed, and force-generation step coupled to NTP hydrolysis ( $T \Rightarrow D$ ) and/or product release ( $D \Rightarrow E$ ).

**M<sub>3</sub>** (Fig. 4c, where  $S_1 = T, S_2 = D, S_3 = E$ ): it predicts counterclockwise sequential hydrolysis, no NTP binding cooperativity, multiple-coordination substrate-binding mode, translocation in -Z direction with step size  $\sim 1$ -9Å  $\sim 1$ -3 bases per NTP hydrolyzed, and force-generation step coupled to NTP binding ( $E \Rightarrow T$ ).

The two schemes differ in many mechano-chemical properties. Future measurements of these properties (particularly translocation direction and step size) will help to distinguish between the two schemes of sequential hydrolysis and an alternative concerted hydrolysis model (Gai *et al.*, 2004).

## DISCUSSION

The basic assumption of our model is the sequential coordination of NTP hydrolysis among six nucleotide-binding sites. Although sequential models are consistent with many experimental studies of hexameric helicases, we can by no means rule out alternative mechanisms (such as concerted hydrolysis (Gai *et al.*, 2004) or stochastic hydrolysis (Martin *et al.*, 2005)). For example, based on the crystal structures of LTag bound with six NTPs and NDPs, a concerted model was proposed (Gai *et al.*, 2004), which postulates the NTP binding and hydrolysis in six nucleotide-binding sites proceed in an 'all-or-none' fashion. We have applied our sequential model to LTag and have identified two schemes that are consistent with the structural coupling rules deduced from LTag structures. Future experiments are needed to resolve these alternative models.

Compared with previously proposed sequential models, our model shares several common features (such as sequential hydrolysis involving all six active sites, substrate transferred between adjacent subunits, spiral binding of substrate, etc). However, our model is novel in the following ways:

1. It is based on the structural coupling rules deduced from a comprehensive analysis of crystal structures of five hexameric helicases and F<sub>1</sub> ATPase —

these coupling rules facilitate both nucleotide-dependent movements of substrate-binding loops and sequential coordination between adjacent nucleotide-binding sites.

2. Its four distinct schemes allow mechanistic variations in hydrolysis direction and substrate-binding mode to be accommodated in a unified modeling framework. So it can be potentially applied to various hexameric helicases which differ in mechano-chemical coupling details. Indeed, we have obtained promising results by applying it to five hexameric helicases (Rho, T7, P4, E1 and LTag).
3. It makes specific predictions for a variety of mechano-chemical properties (hydrolysis direction, NTP-binding cooperativity, substrate-binding mode, translocation direction and step size, force-generation step). Some of these predictions are already in agreement with past experimental results whereas the others call for future experimental verifications.

The deduction of structural coupling rules is of general significance. The key assumption of protein crystallography is that a crystal structure faithfully captures a snapshot of an active multi-subunit protein complex. However, concerns about perturbation effects of crystallization need to be addressed. In an effort to weaken our dependence on the above assumption, we choose to deduce from crystal structures those local coupling rules involving subunits in contact. Intuitively, those local couplings are likely to be stronger than nonlocal couplings (between subunits not in contact) and are thus more robust to perturbation effects of crystallization. Therefore, this study does not require the crystal structures to be at active functional states.

In this work, we have focused on the signal transmission from NTP hydrolysis to the mechanical movements of substrate-binding loops. In the reverse direction, substrate binding can also stimulate NTPase activities as shown experimentally for Rho (Engel *et al.*, 1984; Kim *et al.*, 2001), T7 (Crampton *et al.*, 2006) and P4 (Lisal *et al.*, 2005). In the context of our sequential model, the coupling rules in reverse direction (from positional state of substrate-binding loops to nucleotide state), if combined with the cooperative interactions between substrate-binding loops and nucleic acid substrate, will allow us to model the kinetics of substrate-binding stimulated NTPase.

In future, our model will be refined in light of the new structural data that capture different nucleotide states and substrate-bound conformations, which will remove the inaccuracy and ambiguity in the coupling rules extracted from existing structures. This model will be adapted to study other multi-subunit NTPase machines such as AAA+ proteins (Erzberger *et al.*, 2006) like ClpX unfoldase (Hersch *et al.*, 2005).

## References

- Abrahams JP, Leslie AG, *et al.* (1994) Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370**: 621-628.
- Adelman JL, Jeong YJ, *et al.* (2006) Mechanochemistry of transcription termination factor Rho. *Mol. Cell* **22**: 611-621.
- Boyer PD (1993) The binding change mechanism for ATP synthase--some probabilities and possibilities. *Biochim. Biophys. Acta.* **1140**: 215-250.
- Crampton DJ, Mukherjee S, *et al.* (2006) DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. *Mol. Cell* **21**: 165-174.
- Donmez I and Patel SS (2006) Mechanisms of a ring shaped helicase. *Nucleic Acids Res.* **34**: 4216-4224.
- Donmez I and Patel SS (2008) Coupling of DNA unwinding to nucleotide hydrolysis in a ring-shaped helicase. *EMBO J.* **27**: 1718-26.
- Egelman EH, Yu X, *et al.* (1995) Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases. *Proc. Natl. Acad. Sci. U S A.* **92**: 3869-3873.
- Enemark EJ and Joshua-Tor L (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**: 270-275.
- Engel D and Richardson JP (1984) Conformational alterations of transcription termination protein rho induced by ATP and by RNA. *Nucleic Acids Res.* **12**: 7389-7400.
- Erzberger JP and Berger JM (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu Rev Biophys Biomol Struct.* **35**: 93-114.
- Gai D, Zhao R, *et al.* (2004) Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* **119**: 47-60.
- Gorbalenya AE, Koonin EV, *et al.* (1990) A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Lett.* **262**: 145-148.
- Gorbalenya AE and Koonin EV (1993) Helicases, amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.* **3**: 419-429.
- Guo P, Peterson C, *et al.* (1987) Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage  $\phi$ 29. *J. Mol. Biol.* **197**: 229-236.
- Hersch GL, Burton RE, *et al.* (2005) Asymmetric interactions of ATP with the AAA+ ClpX6 unfoldase: allosteric control of a protein machine. *Cell.* **121**: 1017-27.
- Hickman AB and Dyda F (2005) Binding and unwinding, SF3 viral helicases. *Curr. Opin. Struct. Biol.* **15**: 77-85.
- Hingorani MM and Patel SS (1996) Cooperative interactions of nucleotide ligands are linked to oligomerization and DNA binding in bacteriophage T7 gene 4 helicases. *Biochemistry.* **35**: 2218-28.
- Hingorani MM, Washington MT, *et al.* (1997) The dTTPase mechanism of T7 DNA helicase resembles the binding change mechanism of the F1-ATPase. *Proc. Natl. Acad. Sci. U S A.* **94**: 5012-5017.
- Ilyina TV, Gorbalenya AE, *et al.* (1992) Organization and evolution of bacterial and bacteriophage primase-helicase systems. *J. Mol. Evol.* **34**: 351-357.
- Kainov DE, Mancini EJ, *et al.* (2008) Structural basis of mechano-chemical coupling in a hexameric molecular motor. *J. Biol. Chem.* **283**: 3607-17.
- Kim DE and Patel SS (1999) The mechanism of ATP hydrolysis at the noncatalytic sites of the transcription termination factor Rho. *J. Biol. Chem.* **274**: 32667-32671.
- Kim DE and Patel SS (2001) The kinetic pathway of RNA binding to the Escherichia coli transcription termination factor Rho. *J. Biol. Chem.* **276**: 13902-10.
- Kim DE, Narayan M, *et al.* (2002) T7 DNA helicase, a molecular motor that processively and unidirectionally translocates along single-stranded DNA. *J. Mol. Biol.* **321**: 807-819.
- Li D, Zhao R, *et al.* (2003) Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature* **423**: 512-518.
- Liao JC, Jeong YJ, *et al.* (2005) Mechanochemistry of t7 DNA helicase. *J. Mol. Biol.* **350**: 452-475.
- Lísal J, Lam TT, *et al.* (2005) Functional visualization of viral molecular motor by hydrogen-deuterium exchange reveals transient states. *Nat. Struct. Mol. Biol.* **12**: 460-466.
- Lísal J and Tuma R (2005) Cooperative mechanism of RNA packaging motor. *J. Biol. Chem.* **280**: 23157-23164.
- Lohman TM and Bjornson KP (1996) Mechanisms of helicase-catalyzed DNA unwinding. *Annu. Rev. Biochem.* **65**: 169-214.
- Mancini EJ, Kainov DE, *et al.* (2004) Atomic snapshots of an RNA packaging motor reveal conformational changes linking ATP hydrolysis to RNA translocation. *Cell* **118**: 743-755. Patel SS and Picha KM (2000) Structure and function of hexameric helicases. *Annu. Rev. Biochem.* **69**: 651-697.
- Martin A, Baker TA, *et al.* (2005) Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature.* **437**: 1115-1120.
- Miwa Y, Horiguchi TK, *et al.* (1995) Structural and functional dissections of transcription termination factor rho by random mutagenesis. *J. Mol. Biol.* **254**: 815-837.
- Mori H, Imai M, *et al.* (1989) Mutant rho factors with increased transcription termination activities. II. Identification and functional dissection of amino acid changes. *J. Mol. Biol.* **210**: 39-49.
- Notarnicola SM, Park K, *et al.* (1995) A domain of the gene 4 helicase/primase of bacteriophage T7 required for the formation of an active hexamer. *J. Biol. Chem.* **270**: 20215-20224.
- Richardson JP (2002) Rho-dependent termination and ATPases in transcript termination. *Biochim. Biophys. Acta.* **1577**: 251-260.
- Sawaya MR, Guo S, *et al.* (1999) Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. *Cell* **99**: 167-177.
- Scheffzek K, Ahmadian MR, *et al.* (1997) The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science.* **277**: 333-338.
- Singleton MR, Sawaya MR, *et al.* (2000) Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* **101**: 589-600.

- Singleton MR, Dillingham MS, *et al.* (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem.* **76**: 23-50.
- Skordalakes E and Berger JM (2003) Structure of the Rho transcription terminator, mechanism of mRNA recognition and helicase loading. *Cell* **114**: 135-146.
- Skordalakes E and Berger JM (2006) Structural insights into RNA-dependent ring closure and ATPase activation by the Rho termination factor. *Cell* **127**: 553-564.
- Stitt BL and Xu Y (1998) Sequential hydrolysis of ATP molecules bound in interacting catalytic sites of Escherichia coli transcription termination protein Rho. *J. Biol. Chem.* **273**: 26477-26486.
- Toth EA, Li Y, *et al.* (2003) The crystal structure of the bifunctional primase-helicase of bacteriophage T7. *Mol. Cell* **12**: 1113-1123.
- Walker JE, Saraste M, *et al.* (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.
- Walstrom KM, Dozono JM, *et al.* (1997) Kinetics of the RNA-DNA helicase activity of Escherichia coli transcription termination factor rho. 2. Processivity, ATP consumption, and RNA binding. *Biochemistry* **36**: 7993-8004.
- Washington MT, Rosenberg AH, *et al.* (1996) Biochemical analysis of mutant T7 primase/helicase proteins defective in DNA binding, nucleotide hydrolysis, and the coupling of hydrolysis with DNA unwinding. *J. Biol. Chem.* **271**: 26825-26834.
- Weber J and Senior AE (2000) ATP synthase: what we know about ATP hydrolysis and what we do not know about ATP synthesis. *Biochim Biophys Acta.* **1458**: 300-309.
- Wei RR and Richardson JP, *et al.* (2001) Mutational changes of conserved residues in the Q-loop region of transcription factor Rho greatly reduce secondary site RNA-binding. *J. Mol. Biol.* **314**: 1007-1015.