Explore Rotary Catalysis of F_1-ATPase by a Markov Model

Yixin Chen

First Supervisor	Prof. Dr. Helmut Grubmüller
Second Supervisor	Prof. Dr. Stefan Klumpp

Abstract

F_1F_o-ATPase is fundamental to energy conversion and utilization for all living beings on the earth. Isolated F_1-ATPase, the soluble sector of F_1F_o-ATPase, catalyzes ATP hydrolysis in aqueous solution. Though having been studied extensively over the past decades, a unified, thorough explanation for the catalytic mechanism of F_1-ATPase is still under controversy.

In this project, we constructed a Markov model incorporating as few essential DOFs as possible to to explore which properties or mechanistic aspects are necessary, sufficient, or dominant for F_1-ATPase function. The DOFs we chose led to the inclusion of 729 Markov states and 17 independent parameters in terms of binding free energies, conformational energies and energy barriers into our model to full specify the dynamics.

We formulated a parameter optimization approach based on Bayesian inference, which enabled us to obtain parameter sets that accurately reproduce experimentally observed catalytic kinetics of F_1-ATPase. In particular, our model exhibits Michaelis-Menten-like dependence of turnover on ATP concentration and near 100% chemo-mechanical coupling efficiency. We further confirmed by numerical calculations and kinetic Monte-Carlo simulations that our model produces effective rotary catalysis by a microscopic mechanism compatible with experimental observations. These results provided us with semi-quantitative understanding of the mechanism of F_1-ATPase in terms of both thermodynamics and kinetics.

As cross-validation of our model, we checked if our model could also reproduce experimentally measured nucleotide binding affinities of F_1-ATPase, which had not been used in parameter optimization to construct the model. The negative results suggest that at least one of the assumptions of our model is wrong or insufficient for describing the observed dynamics of F_1-ATPase. Therefore, we re-examined our model assumptions and proposed ideas for further improving our Markov model.

Contents

1   Introduction

Understanding how F_1F_o-ATPase (F-ATP synthase) works is essential for understanding how life on the earth converts and utilizes energy. While the hydrolysis of ATP (adenosine-5'-triphosphate) into ADP (adenosine-5'-diphosphate) and Pi (free phosphate) could be catalyzed by a large diversity of enzymes to provide energy for various cellular processes, the synthesis of ATP from ADP and Pi, storing energy, is almost exclusively facilitated by members of the F-ATP synthase family. [1]

F-ATPase is ubiquitously found in abundance in bacteria, plant chloroplasts and animal mitochondria, with a highly conserved structure throughout species. As shown in Figure 1a, F-ATPase consists of two major parts, a hydrophobic sector F_O embedded in membranes, and a soluble sector F_1 containing a spherical hexamer, $\alpha_3\beta_3$, and a central stalk made up of a γ-subunit and several other subunits, *e.g.*, δ and ϵ. In the most essential part common to F_1-ATPase found in all species, *i.e.*, the subassembly $\alpha_3\beta_3\gamma$, the three α- and three β-subunits are arranged in alternation around an elongated α-helical structure of the γ-subunit, forming an approximately spherical structure (Figure 1c, left panel). The rest of the γ-subunit protrudes out of the hexameric sphere of $\alpha_3\beta_3$, attaching the F_1 sector to the c-ring of the F_O sector. Every β/α-subunit forms a nucleotide binding site with its adjacent α/β-subunit at the interface (Figure 1b). Those three binding sites located mainly on the three β-subunits exhibit catalytic activity (catalytic site). The other three binding sites located mainly on the α-subunits do not seem to have any direct influence on the catalytic cycle (noncatalytic site). [2, 3]

Under physiological condition, F_O transports protons across the membrane, mining energy out of the proton concentration gradient to drive a rotational motion of the central stalk, which subsequently drives ATP synthesis in the three catalytic sites in $\alpha_3\beta_3$. This process of energy conversion is the final step of oxidative- or photo-phosphorylation. [2, 4–6] Moreover, F_1 has been detached from F_O. [7] Isolated F_1 in aqueous solution exhibits enzymatic activity of catalyzing ATP hydrolysis, which is why it is given the name F_1-ATPase.

Because the mechanism of isolated F_1-ATPase catalyzing ATP hydrolysis is presumed to be exactly the reverse of how it catalyzes ATP synthesis in cooperation with F_O, to understand how F_1-ATPase works would be the first step towards a comprehensive understanding of the whole complex F-ATP synthase.

Much efforts have been made in this direction, leading to the establishment of a paradigm known as rotary catalysis, or binding-change mechanism, originally proposed and developed by Paul Boyer. [10] It depicts F_1-ATPase as a highly efficient molecular machine, coupling conformational changes with nucleotide binding and conversion. Coordinated by the γ-subunit, the three β-subunits adopt different conformations simultaneously, thus differentiated in nucleotide binding affinities. During a catalytic cycle, as the γ-subunit rotates and passes the β-subunits one by one, the three β-subunits go through interconversion among three conformations, namely, "open", "loose" and "tight", during which the nucleotide binding states also change, leading to consecutive hydrolysis of three ATP molecules.

Over the years, numerous experimental studies have accumulated strong evidences corroborating the binding change mechanism. There are three lines of important experimental observations. First, a series of crystal structures of F_1-ATPase obtained by X-ray diffraction (XRD) have revealed that the β-subunits adopt different distinct conformations, where the asymmetric interaction between the γ-subunit and the β-subunits plays an important role (Figure 1c, compare right three panels). [9, 11] Second, differentiated binding affinities of the catalytic sites are measured by fluorescence microscopy studies. [2, 6, 12–18] Third, stepping rotation of γ-subunit in active F_1-ATPase is directly visualized by single-molecule experiments. [19–28]

Experimental observations as outlined above have led to wide acceptance of the fundamental idea of the binding change mechanism. However, they are still insufficient for a full and thorough picture of how F_1-ATPase functions to be formed. Many details of the binding change mechanism remain controversial or unresolved: Is a bisite- or a trisite-mechanism dominant for full-speed ATP hydrolysis under physiological conditions? [2, 29–31] How are β-subunit conformational change, γ-subunit rotation and nucleotide exchange coupled and organized in a catalytic cycle? What stages of the catalytic cycle do the observed crystal structures correspond to? Giving various answers to these questions, new versions of the binding-change mechanism continue to be proposed and published still today. [32–37]
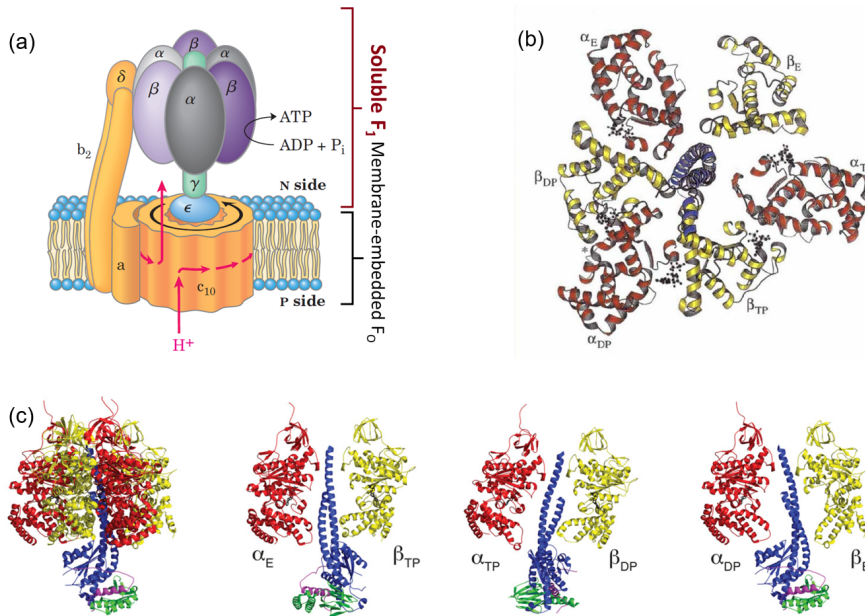
Figure 1: Structure details of F_1-ATPase.

(a) Schematic representation of the structure of the *E. coli* F_1F_o-ATPase. Adapted from Ref. [8].

(b) Bird view of the C-terminal domains of the $\alpha_3\beta_3$ hexameric ring of bovine mitochondria F_1-ATPase by Abrahams *et al.*[9]. The asymmetry of the overall structure and the locations of the nucleotide binding sites are most apparently shown in this view. Adapted from Ref. [9].

(c) Crystal structure of bovine mitochondria F_1-ATPase by Abrahams *et al.*[9] in ribbon representation. The α-, β-, γ-, δ- and ϵ-subunits are red, yellow, blue, green and purple respectively. (Left) The complete F_1-ATPase with three α- and three β-subunits and single copies of γ-, δ-, ϵ-subunits. (Right three) The three different conformations of the catalytic β-subunits present in the left panel.

Theoretical studies have also been done to understand more deeply how F_1-ATPase functions, overcoming some difficulties whereas bearing certain limitations. Molecular dynamics (MD) simulations in which force was applied to rotate the γ-subunit in the synthesis direction provided insights about the molecular mechanism of coupled conformational changes of the γ- and β-subunits. [38–42] However, as these simulations are extremely expensive in computation, the time span covered is limited to a time scale much smaller than that of these conformational changes *in vivo*. Therefore, such simulations may fail to reflect the real transition pathway. [43] Other theoretical studies adopted more coarse-grained approaches and succeeded in explaining specifically one or a few aspects of F_1-ATPase function. [33, 44] Nevertheless, understandings from these studies are hard to be integrated because these models are based on different assumptions, encoding some degrees of freedom (DOF) of F_1-ATPase while omitting the other. They may also introduce phenomenological parameters that are hard to be directly related to known molecular mechanism.

In all, considering the large system size, complicated interactions and dynamics of F_1-ATPase, coarse-graining and identification of essential DOFs are necessary for constructing a model to explain the catalytic mechanism of F_1-ATPase. The many properties and mechanistic aspects of F_1-ATPaserevealed by previous experimental and theoretical studies serve as candidate DOFs that can be integrated into a theoretical model for F_1-ATPase. In this project, we intend to explore which of them are necessary, sufficient, or dominant for F_1-ATPase function.

We constructed a Markov model incorporating as few essential DOFs as possible to study the catalytic mechanism of F_1-ATPase. These DOFs include orientation of the γ-subunit, and conformations and binding states of the β-subunits, the combination of which, considering interaction between the γ- and β-subunits, gives in total 729 Markov states. Between these states, direct transitions including γ-subunit rotation, β-subunit open and closing, substrate binding and unbinding and ATP-ADP conversion at catalytic sites are allowed. To fully specify the rates of these transitions under the rule of microscopic reversibility, we chose 17 independent parameters in terms of binding free energies, conformational free energies and free energy barriers.

We formulated a parameter optimization approach based on Bayesian inference, in which, briefly speaking, the 17 parameters are optimized to maximize the likelihood of our model system showing behaviors in agreement with available

experimental observations. By optimizing the parameters against experimental data of turnover and γ-subunit revolution of F_1-ATPase, we obtained parameter sets that accurately reproduced the data. Further, numerical calculations and kinetic Monte-Carlo simulations using the optimal parameter sets have confirmed that our model produce effective rotary catalysis by a microscopic mechanism compatible with experimental observations. Based on these results, we developed a semi-quantitative description of the catalytic mechanism of F_1-ATPase in terms of both thermodynamics and kinetics, forming a full picture of how our model F_1-ATPase works.

As cross-validation of our model, we checked if our model could also reproduce experimentally measured nucleotide binding affinities of F_1-ATPase, which had not been used in parameter optimization to construct the model. The results of our attempts are negative, suggesting that within the current framework of our model, it is highly likely that there is no parameter set that can reproduce experimental observations of both nucleotide binding properties and catalytic kinetics. This conclusion further suggests that at least one of the assumptions of our model is wrong or insufficient for describing the observed dynamics of F_1-ATPase. Therefore, we re-examined our model assumptions and proposed ideas for further improving our Markov model.

2   Model Assumptions

In this section, we state and explain the basic assumptions of our Markov model by describing the relevant experimental observations on F_1-ATPase that support them.

Assumption 1 There are three different β-subunit conformations (open, half-closed and closed), each associated with a distinct conformational energy.

This assumption is based on the common classification of the multiple β-subunit conformations observed in crystal structures of F_1-ATPase into three groups in papers of X-ray crystallography studies. In the classical crystal structure obtained by Abrahams *et al.* in 1994,[9] $β_E$ revealed a conformation containing an empty, open (more solvent-accessible) catalytic site; $β_{TP}$ and $β_{DP}$ are of very similar conformations in which the catalytic sites are closely packed, although $β_{TP}$ is occupied by an ATP analog, MgAMP-PNP, and $β_{DP}$ is occupied by MgADP. Later, a transition state analogue structure [45] demonstrated the existence of an intermediate conformation between the open ($β_E$) and closed ($β_{TP}$, $β_{DP}$) conformations, $β_{HC}$. The catalytic site of $β_{HC}$ contains a MgADP and a sulfate, probably mimicking a post-hydrolysis ADP·Pi-bound state. Other crystal structures have revealed conformations of β-subunits very similar to these representative open, half-closed and closed conformations.

Assumption 2 Rotation of the γ-subunit is stepwise. Only two substeps of the γ-subunit, 80° and 40°, are considered.

This assumption is based on a series of experimental studies proving that stepping rotation of the γ-subunit is a genuine property of catalytically active F_1-ATPase. The first real-time recording of rotation of the γ-subunit in F_1-ATPase could be attributed to Wolfgang Junge *et al.* [46, 47] who measured polarized absorption relaxation after photobleaching (PARAP) of F_1-ATPase with eosin-labelled γ-subunit and immobilized $α_3β_3$. The decay of polarization indicated rotational motion of the γ-subunit relative

to $\alpha_3\beta_3$. Soon, using single-molecule fluorescent microscopy, scientists observed directly that the γ-subunit of F_1-ATPase rotates in consecutive 120°steps, [4, 5, 48, 49] each of which contains substeps. [3, 9, 34, 50–52] In these experiments, $\alpha_3\beta_3$ was fixed on a glass plate, and a fluorescently labeled actin filament or bead was attached to the γ-subunit. Orientation of the γ-subunit could be indicated by analyzing fluorescence microscopy images obtained during the process of catalyzed ATP hydrolysis. [19, 23, 27] The number and angle of substeps vary subtly in reports from different experiments, [23, 25] and probably in different species, [53] but what has been commonly accepted is the division of one 120° rotation step into two consecutive substeps being approximately 40° and 80°, established as an ATP-binding dwell and a catalytic dwell respectively.

The rotation of the γ-subunit and the chemical composition of the $\alpha_3\beta_3$ hexameric ring combines to give a three-fold symmetry of the $\alpha_3\beta_3\gamma$ complex if the binding states are not considered. There are two sets of equivalent orientations of the γ-subunit, *i.e.*, (80°, 200°, 320°) and (120°, 240°, 360°) as shown in Figure 2a, corresponding to two relative positions of the γ-subunitand the β-subunits, where the γ-β interactions are likely to be different.

Assumption 3 Inter-subunit interaction occurs only between the γ-subunit and one of the three β-subunits. γ-subunit forces the affected β-subunit to adopt either half-closed or open conformation at 80° orientation, and only allows the open conformation of the affected β-subunit at 120° orientation. No interaction within the hexamer $\alpha_3\beta_3$ (between β-subunits) or between the γ-subunit and the other two β-subunits is significant.

This assumption is primarily based on the crystal structures of F_1-ATPase that have already been mentioned when explaining **Assumption 1**. In the 1994 structure, the rigidity and curvature of the α-helical domain of the γ-subunit seems to force the three β-subunits to adopt different conformations. In particular, the γ-subunit pushes against the C-terminal domain of β_E, forcing it to adopt the open conformation, where its nucleotide binding domain hinges outwards (Figure 1b&c, compare right three panels). The other two β-subunits, less perturbed by the γ-subunit, adopt the closed conformations β_{TP} and β_{DP}. [4, 9, 54] In the transition state analogue structure, the orientation of the γ-subunit is shifted by a small angle compared

to the 1994 structure (close to 120° orientation), probably allowing β_E to close partially to adopt the half-closed conformation. [45]

Assumption 4 Each of the three different β-subunit conformations: open, half-closed and closed, as stated in **Assumption 1**, is associated with a distinct ATP binding free energy and a distinct ADP binding free energy.

This assumption is based on the measurements by Trp fluorescence experiments of differentiated nucleotide binding affinities (three for ATP and three for ADP, in the same order for both nucleotides). Senior *et al.* [13] developed a fluorescent probe to measure the binding affinities of the catalytic sites. The approach makes use of an *E. coli* F_1-ATPase mutant (βY331W F_1-ATPase), whose 331th residues in the β-subunits, being tyrosine in the wild-type enzyme, are replaced by tryptophan (Trp). The mutant assumes similar catalytic site structures and catalytic properties as wild-type, while the additional Trp emits substantial fluorescence in the absence of nucleotides. Upon binding of nucleotide to a catalytic site, the fluorescence coming from the Trp in this site is quenched. Measuring the fluorescence quenching of the mutant under a series of nucleotide concentrations covering the whole affinity range provides nucleotide titration curves, *i.e.*, catalytic site occupancy ν (the average number of occupied catalytic sites in one F_1-ATPase molecule) as a function of nucleotide concentration. Binding affinities were estimated by fits of the titration curves, using the model:

$$\nu = \frac{[\text{L}]}{[\text{L}] + K_{d1}} + \frac{[\text{L}]}{[\text{L}] + K_{d2}} + \frac{[\text{L}]}{[\text{L}] + K_{d3}}, \tag{2.1}$$

where [L] is the concentration of nucleotide in bulk solution, and K_{d1}, K_{d2} and K_{d3} are the three different binding affinities of the three catalytic sites. [17] For both ATP and ADP, K_{d1}, K_{d2} and K_{d3} differing by magnitudes were obtained, suggesting the co-existence of a high-affinity site, a medium-affinity site, and a low-affinity site.

There are a few more assumptions of our Markov model which we consider less significant:

Assumption 5 The release of phosphate is fast and causes small energy change, so that phosphate is not explicitly included in our model.

The assumption is based on the low binding affinity of phosphate with the β-subunits in F$_1$-ATPase. [2]

Assumption 6 The catalytic site in a closed β-subunitis catalytically active, where the conversion of ATP and ADP (reversible cleavage and formation of the terminal phosphate bond in ATP) is in perfect equilibrium, *i.e.*, the free energy change is zero and the forward and backward rates are equal.

This assumption is supported by both experimental and theoretical studies. [44, 55, 56]

Assumption 7 The rotation of γ-subunit is depicted as a Brownian ratchet.

3 Method

3.1 A Markov model for F$_1$-ATPase

In brief, we described the dynamics of F$_1$-ATPase by continuous time Markov chain (CTMC), where the protein transitions stochastically among a set of discrete Markov states. Mathematically, the time evolution of the system is described by the master equation:

$$\frac{d\boldsymbol{\rho}}{dt} = -\mathbf{R}\boldsymbol{\rho}, \tag{3.1}$$

where $\boldsymbol{\rho}$ is the vector of the probabilities of the Markov states, and $\mathbf{R}$ is the transition rate matrix. The element of $\mathbf{R}$, R_{mn} ($m \neq n$), is defined as the transition rate from the Markov state n to state m. The diagonal elements of $\mathbf{R}$ are defined:

$$R_{mm} = -\sum_{n \neq m} R_{mn}. \tag{3.2}$$

Consequently, each row of R sums to zero.

The system includes nucleotides explicitly, *i.e.*, the bulk concentrations of ATP and ADP are assumed to be constant ($[\text{ATP}](t) = c_{\text{T}}$, $[\text{ADP}](t) = c_{\text{D}}$), which leads to the linearity of Equation 3.1. We expect this to be a valid approximation for physiological conditions, where the nucleotides are at large excess comparing to F$_1$-ATPase.

3.1.1 Markov states

Summarizing the established works on F$_1$-ATPase, we recognized seven essential degrees of freedom (DOF) to be included in our Markov model, which can be classified into three groups:

I. The orientation of the γ-subunit, ϕ_n. Without losing generality, we adopted the denotations of 80° and 120° which, essentially, only suggests that there are two different relative positions of the γ- and β-subunits.

II. The conformations of the three β-subunits, $C_n^{(1,2,3)}$, each of which could be open (o), half-closed (h) or closed (c), distinguished in the model by their nucleotide binding affinities. Importantly, an extra restraint of the conformation of the β-subunit towards which the γ-subunit orients is included to account for the γ-β interaction. For γ-subunit at the 80° position, the affected β-subunit (containing Catalytic Site 1 as denoted in Figure 2a) can only adopt the half-closed or open conformation, and for γ-subunit at 120° position, it is forced to adopt the open conformation exclusively.

III. The binding states of the three catalytic binding sites, $\mathcal{B}_n^{(1,2,3)}$, each of which could be empty (E), ATP-bound (T) or ADP-bound (D). The probabilities are decided by the conformation of the β-subunit and the bulk concentrations of the nucleotides.

Therefore, every Markov state is denoted by a seven-dimensional vector:

$$\mathbf{s}_n = \left(\phi_n, C_n^{(1)}, C_n^{(2)}, C_n^{(3)}, \mathcal{B}_n^{(1)}, \mathcal{B}_n^{(2)}, \mathcal{B}_n^{(3)} \right). \tag{3.3}$$

The indexing of the β-subunits/catalytic sites is in accordance with Figure 2a, $\phi_n \in \{80°, 120°, 200°, 240°, 320°, 360°\}$, $C_n^{(k)} \in \{\text{o, h, c}\}$, $\mathcal{B}_n^{(k)} \in \{\text{E, T, D}\}$. The possible combinations of the seven DOFs are summarized in Table 1, which account for 729 asymmetric Markov states in total, each having another two symmetric states generated by 120°-rotation of the whole structure. These Markov states could be further grouped by their chemical compositions, *i.e.*, the numbers of bound ATP and ADP, or the orientations of the γ-subunit, giving rise to the chemical space (Figure 2c) and the conformational space (Figure 2b), respectively.

3.1.2 Direct transitions

We attached importance to two principles in deciding which direct transitions to allow among the Markov states. First, microscopic reversibility must be obeyed according to experimental observations that F_1-ATPasecan work reversibly under very likely the same mechanism. [28, 57, 58] Still, we expect to witness the breaking of detailed balance at the steady state of the system in our model if effective catalysis occurs. The non-zero net fluxes persist as ATP is replenished
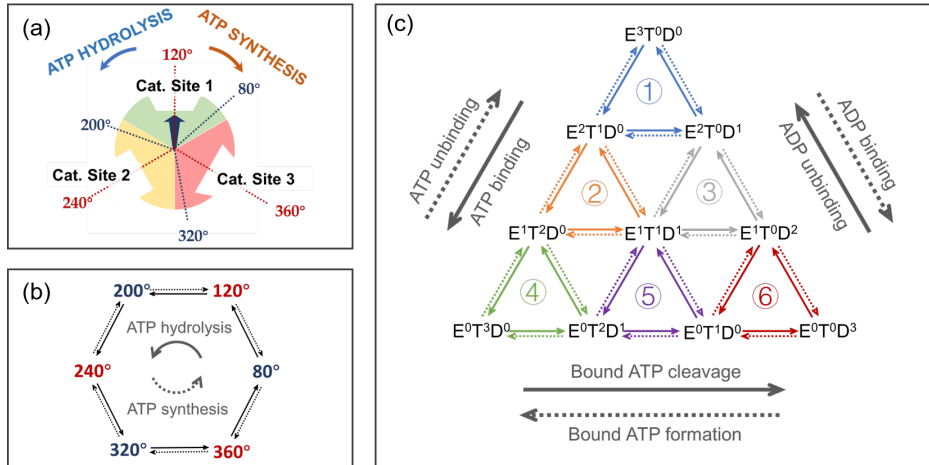
Figure 2: (a) Schematic representation of F$_1$-ATPase in our Markov model. $\alpha_3\beta_3$ is depicted as three $\alpha\beta$ dimers, each of which shown as a 120° sector with a breach representing the catalytic site. The dark blue arrow represents the γ-subunit. The dotted lines denote the possible orientations of the γ-subunit. Although the three $\alpha\beta$ dimers are in principle indistinguishable, we attached artificial labels 1, 2, 3 to their catalytic sites just for clarity of description.

(b) The conformational space of F$_1$-ATPase. The Markov states are grouped by their γ-subunit orientations. The solid and the dotted arrows represent the directions of net ATP hydrolysis and synthesis respectively.

(c) Chemical space of F$_1$-ATPase. The Markov states are grouped by their chemical compositions, *i.e.*, the numbers of bound nucleotides. The denotation E^{x}T^{y}D^z ($x + y + z = 3$, x, y, z are non-negative integers) in the figure represents the group of states containing x empty catalytic site(s), y ATP-bound catalytic site(s) and z ATP-bound catalytic site(s). The 10 groups are interchangeable along the arrows representing nucleotide exchange and reversible ATP cleavage. The 10 groups form six subspaces, each representing a complete catalytic cycle. Counterclockwise flow (solid arrows) and clockwise flow (dotted arrows) along each cycle corresponds to net ATP hydrolysis and synthesis respectively.

Table 1: Possible combinations of the 7 degrees of freedom, each corresponding to one of the 729 asymmetric Markov states.

DOF	Possibility	
γ orientation ϕ_n	80°	120°
cat. site 1 conformation $C_n^{(1)}$	{ open, half-closed }	open
cat. site 2/3 conformation $C_n^{(2,3)}$	{ open, half-closed, closed }	
cat. site 2/3 binding state $\mathcal{B}_n^{(2,3)}$	{ empty, ATP-bound, ADP-bound }	

and ADP is removed continuously to maintain their constant concentrations. Second, the change of the DOFs should be in a stepwise and uncoupled manner. Specifically, four types of direct transitions are allowed in our Markov model:

I. Stepwise rotation of the γ-subunit: 80° $\rightleftharpoons$ 120° $\rightleftharpoons$ 200° $\rightleftharpoons$ 240° $\rightleftharpoons$ 320° $\rightleftharpoons$ 360°($\rightleftharpoons$ 80°). As long as the rotation is not forbidden by configuration of the β-subunit conformations, it is considered as free diffusion, *i.e.*, of a uniform rate in both clockwise and counterclockwise directions.

II. Conformational change of the β-subunits, each can only open or close in a stepwise manner: open $\rightleftharpoons$ half-closed $\rightleftharpoons$ closed.

III. Nucleotide binding or unbinding at one catalytic site.

IV. Reversible cleavage/formation of gamma-phosphate bond in the ATP bound to a closed β-subunit, which has been shown to be the catalytically active site. [44, 55, 56] Because we do not include phosphate in our model explicitly, the reaction is simple taken as ATP $\rightleftharpoons$ ADP in perfect equilibrium, *i.e.*, the free energy change is zero and the forward and backward rates are equal.

By adopting the Markov model, we have already presumed every transition to occur instantaneously. We expect this presumption to be valid because the transitions as listed above, which refer to stochastic attempts as a result of thermal fluctuations rather than steady progress along the reaction coordinate, are sufficiently fast. Consequently, it is highly unlikely that two or more transitions occur simultaneously, which validates our principle that no direct transition should be allowed between two Markov states if more than one DOF is differed.

Table 2: 17 independent parameters defined in our model

Class	Count	Symbol	Comment
Thermodynamics: free energy differences			
β-subunit conformational energies	2	$\Delta G_{\beta,\mathcal{C}}$	$\mathcal{C}$ = o, h, c; $\Delta G_{\beta,h} = 0$
Nucleotide binding free energies	6	$\Delta G^{\theta}_{b,\mathcal{C},\mathcal{B}}$	$\mathcal{C}$ = o, h, c; $\mathcal{B}$ = T, D
Kinetics: free energy barriers			
Nucleotide binding	6	$\Delta G^{\ddagger}_{b,\mathcal{C},\mathcal{B}}$	$\mathcal{C}$ = o, h, c; $\mathcal{B}$ = T, D
β-subunits conformational change	1	$\Delta G^{\ddagger}_{\beta}$	In the direction of energetically favorable transitions
γ-subunit rotation	1	$\Delta G^{\ddagger}_{\gamma}$	Same for counterclockwise and clockwise rotations and 40°/80° substeps
ATP cleavage/formation (only in closed catalytic site)	1	$\Delta G^{\ddagger}_{p}$	ATP and ADP+Pi are assumed to be at equilibrium at a closed catalytic site

3.1.3 Model parameters, Markov state energies and microscopic transition rates

In accordance with the allowed direct transitions, at least 17 independent parameters in terms of free energies and free energy barriers are needed in our model to fully specify the dynamics of F$_1$-ATPase, as listed in Table 2. Taking the energy of a half-closed, empty β-subunit as the zero point, we need: 1) two conformational energies$\Delta G_{\beta,o}$ and $\Delta G_{\beta,c}$ for open and closed empty β-subunits; 2) six binding free energies $\Delta G^{\theta}_{b,o,T}$, $\Delta G^{\theta}_{b,o,D}$, $\Delta G^{\theta}_{b,h,T}$, $\Delta G^{\theta}_{b,h,D}$, $\Delta G^{\theta}_{b,c,T}$, $\Delta G^{\theta}_{b,c,D}$, for ATP / ADP binding with β-subunitassuming the three conformations, and correspondingly, 3) six energy barriers of nucleotide binding; 4) energy barrier of β-subunitconformational change, $\Delta G^{\ddagger}_{\beta}$; 5) energy barrier of γ-subunitrotation, $\Delta G^{\ddagger}_{\gamma}$; 6) energy barrier of reversible ATP cleavage at a closed catalytic site, $\Delta G^{\ddagger}_{p}$.

With these 17 parameters, we are able to define the free energy of a Markov state *n* as a summation of the conformational energies of the β-subunits and the

binding free energies:

$$G_n(c_{\mathrm{T}}, c_{\mathrm{D}}) = \sum_{k=1,2,3} \left(\Delta G_{\beta, C_n^{(k)}} + \Delta G^{\theta}_{\mathrm{b}, C_n^{(k)}, \mathcal{B}_n^{(k)}} + k_{\mathrm{B}} T \ln c_{\mathcal{B}_n^{(k)}} \right), \qquad (3.4)$$

with the denotations: $C_n^{(k)}$ = o (open), h (half-closed) or c (closed); $\mathcal{B}_n^{(k)}$ = T (ATP-bound), D (ADP-bound) or E (empty); $c_{\mathcal{B}_n^{(k)}}$ is the nucleotide concentration. For an empty catalytic site, the standard binding free energy $\Delta G^{\theta}_{\mathrm{b}, C_n^{(k)}, \mathrm{E}}$ is simply zero, and the concentrations of nucleotides have no influence, which is equivalent with setting $c_{\mathcal{B}_n^{(k)}} = 1$.

Further, the microscopic transition rates could also be defined, assuming a uniform attempt frequency f_{att} for all transitions. The rates of nucleotide binding and unbinding are:

$$\begin{cases} r_{\mathrm{b}, C, \mathcal{B}} = f_{\mathrm{att}} \exp\left(-\dfrac{\Delta G^{\ddagger}_{\mathrm{b}, C, \mathcal{B}}}{k_{\mathrm{B}} T} \right) c_{\mathcal{B}}, & \text{binding} \\[4mm] r_{\mathrm{u}, C, \mathcal{B}} = f_{\mathrm{att}} \exp\left(-\dfrac{\Delta G^{\ddagger}_{\mathrm{b}, C, \mathcal{B}} - \Delta G^{\theta}_{\mathrm{b}, C, \mathcal{B}}}{k_{\mathrm{B}} T} \right), & \text{unbinding} \end{cases} \qquad (3.5)$$

with $C \in \{\mathrm{o}, \mathrm{h}, \mathrm{c}\}$ and $\mathcal{B} \in \{\mathrm{T}, \mathrm{D}\}$.

The rates of γ-subunit rotation ($\lambda = \gamma$) and phosphate bond cleavage/formation in a closed β-subunit ($\lambda = \mathrm{p}$) are:

$$r_{\lambda} = f_{\mathrm{att}} \exp\left(-\frac{\Delta G^{\ddagger}_{\lambda}}{k_{\mathrm{B}} T} \right), \; \lambda \in \{\gamma, \mathrm{p}\}. \qquad (3.6)$$

Defining the rates of β-subunit conformational changes is a little more complicated, because there are many possible transitions connecting different initial and final states. To simplify the situation, we define the rate of an allowed direct transition from state n to state m involving only conformational change of the β-subunits:

$$r_{n \to m} = \begin{cases} k_{\beta}, & (G_m - G_n \le 0) \\[3mm] k_{\beta} \exp\left(-\dfrac{G_m - G_n}{k_{\mathrm{B}} T} \right), & (G_m - G_n > 0) \end{cases} \qquad (3.7)$$

where $k_\beta = f_{att} \exp(-\Delta G_\beta^\ddagger / k_B T)$. This definition ensures the principle of detailed balance while limiting the rates of β-subunit conformational changes in energetically favorable directions to a uniform value k_β.

To obtain physically reasonable results, we set f_{att} to 10^9 s^{-1}, and require all energy barriers to be positive. This means that all the transition rates never exceed 10^9 s^{-1}, which is a reasonable upper limit for the microscopic transitions included in our model.

3.2 Steady state distribution and net fluxes

An analytical solution of the master equation (Equation 3.1) is easily accessible by diagonalization of the transition rate matrix, as shown in Appendix section S1. Two conclusions could be drawn from the solution. First, eventually, the system would reach a steady state distribution $\boldsymbol{\rho}_{st}$ given by a zero solution of the transition rate matrix $\mathbf{R}$:

$$\frac{d\boldsymbol{\rho}_{st}}{dt} = 0. \tag{3.8}$$

Second, the time scale of relaxing to the steady state is decided by the smallest positive eigenvalue ω_1 of the transition rate matrix:

$$T_{relax} \sim 1/\omega_1. \tag{3.9}$$

The non-vanishing net fluxes at the steady state infer the catalytic properties of F$_1$-ATPase. The net flux from Markov state m to state n is

$$j_{m\to n} = \rho_m R_{nm} - \rho_n R_{mn}. \tag{3.10}$$

The summation of all the involved pairs of states gives the fluxes in the chemical space and conformational space as shown in Figure 2b&c, respectively.

With the steady state distribution and stationary fluxes, some important observables can be obtained to be compared with experimental measurements that have been reported. To start with, two quantities that are commonly measured in experimental studies to characterize the catalytic kinetics of F$_1$-ATPase, the turnover number of F$_1$-ATPase, k_{cat}, and the revolution of the γ-subunit, k_{rot}, could be calculated by summing up corresponding net fluxes shown in Figure 2b,c. k_{cat} is the summation of the net fluxes along the edges of the six

subspaces (Figure 2c, direction denoted by the solid arrows), and k_{rot} is equal to the net flux along the six edges in the conformational space (Figure 2b, direction denoted by the solid arrows). The signs of k_{cat} and k_{rot} suggest the directions of the net fluxes. A positive k_{cat} means net hydrolysis of ATP, and a negative one means net synthesis. A positive k_{rot} means net counterclockwise rotation of the γ-subunit (looking from the membrane side of F-ATPase), and a negative one means net clockwise rotation.

With k_{cat} and k_{rot}, a quantity called chemo-mechanical coupling efficiency is defined as

$$\eta = \frac{3k_{rot}}{k_{cat}} \times 100\%. \qquad (3.11)$$

η is commonly used to reflect the efficiency of F_1-ATPase converting chemical energy released by ATP hydrolysis to mechanical energy driving γ-subunit rotation.

Another important quantity is the average occupancy (the number of catalytic sites that are occupied by ATP or ADP), ν, under different nucleotide concentrations [nuc], which could be obtained from the steady state distribution. Numerically obtained function $\nu([nuc])$ by our Markov model is readily to be compared with experimental nucleotide titration curves.

3.3 Kinetic Monte Carlo simulation

To confirm the results by analytical calculations, explore more details of the catalytic mechanism and directly visualize the catalytic cycle of F_1-ATPase, we carried out kinetic Monte Carlo simulations (KMC) based on our Markov model of F_1-ATPase, adopting a direct method of Gillespie algorithm. [59, 60] At every simulation step, a random state is chosen from the accessible states from the current state, and a random holding time τ is sampled from an exponential distribution, $\tau \sim R\exp(-R\tau)$, where R is the total leaving rate from the current state.

From trajectories of simulations, important observables including k_{cat}, k_{rot} and nucleotide titration curves could also be obtained in terms of ensemble averages. Most importantly, some microscopic properties, *e.g.*, substep dwell times and stochasticity of the rotation of γ-subunit, could be much easier understood by analyzing the simulation trajectories directly than by trying to formulate an analytical solution.

3.4 Parameter optimization based on Bayesian inference

The behavior of the F_1-ATPaseas described by our Markov model is dependent on the choice of the parameters. We formulated a parameter optimization procedure based on Bayes theorem to identify the feasible parameter set(s) that best reproduce the experimental measurements of the catalytic properties of F_1-ATPase. [61]

Mathematically, the optimization approach aims at maximizing the conditional probability of a parameter set Ω containing the 17 parameters $\{\omega_i\}$, given the experimental measurements of some observables of interest, Ψ^{exp}. According to Bayes theorem, the conditional probability $P(\Omega \mid \Psi^{\text{exp}})$ is:

$$P(\Omega \mid \Psi^{\text{exp}}) \propto P(\Psi^{\text{exp}} \mid \Omega) \cdot P(\Omega), \tag{3.12}$$

where $P(\Psi^{\text{exp}} \mid \Omega)$ is the conditional probability of observing specific values of the observables Ψ^{exp} for given parameters Ω in our Markov model:

$$P(\Psi^{\text{exp}} \mid \Omega) = \prod_i \left(\frac{1}{\sqrt{2\pi}\sigma_i} e^{\frac{(\psi_i^{\text{exp}} - \psi_i^{\text{mod}}(\Omega))^2}{2\sigma_i^2}} \right), \tag{3.13}$$

where σ_i is the standard deviation of ψ_i^{exp} estimated by experimental errorbar, ψ_i^{mod} is the model prediction of the ith observable with the parameter set Ω. The prior distribution $P(\Omega)$ of a parameter set Ω is

$$P(\Omega) = \prod_j P(\omega_j), \tag{3.14}$$

where $P(\omega_j)$ is the prior distribution of the jth parameter ω_j. Depending on whether there are abundant, reliable experimental measurements of the parameter or not, $P(\omega_j)$ could be chosen to be either a uniform distribution within a reasonable range, or a Gaussian distribution around its experimental value.

We defined the score function, $\mathcal{F}(\Omega)$, as a linear function of the logarithm of $P(\Omega \mid \Psi^{\text{exp}})$. Here we present directly the formula of $\mathcal{F}(\Omega)$ we used in practice:

$$\mathcal{F}(\Omega) = -\sum_i \frac{(\psi_i^{\text{exp}} - \psi_i^{\text{mod}}(\Omega))^2}{\sigma_i^2} - \sum_j \frac{(\omega_j - \mu_j)^2}{\sigma_j^2}. \tag{3.15}$$

The summation over index j accounts for the parameters to be optimized whose priors are assumed to be Gaussian distributions. μ_j and σ_j are the experimental mean value and standard deviation of the parameter ω_j.

Because it is hard to obtain an analytical expression of $\mathcal{F}(\Omega)$, we adopted a derivative-free optimization algorithm in which the parameter set goes through random up-hill walk in the parameter space towards an optimal point maximizing the score. For one iteration of the algorithm, a number (N_{try}) of random guesses $\{\Omega_n^{\text{try}} \mid n = 1, 2, \ldots, N_{\text{try}}\}$ are sampled, each component drawn from a Gaussian distribution ($\omega_i^{\text{opt}}, \sigma_{\text{try}}$), whose mean is its current optimal value ω_i^{opt}. The guess that reaches the highest score (also higher than the current score) is chosen, from which the next iteration starts.

Whereas our Markov model is generally applicable to F_1-ATPasefrom any species in principle, the optimization of parameters is species-specific because experimental measurements of the quantities involved in the score function vary from species to species. Specifically, we investigated a mutant *E. coli* F_1-ATPasealso used in Trp fluorescence experiments, [13, 17] of which Michaelis-Menten kinetics ($V_{\text{max}} = 80$ s^{-1}, $K_{\text{M}} = 0.04$ mM) [13] and nucleotide titration curves are available. [18] This mutant has been shown to behave similarly to the wild-type F_1-ATPase. Although some experiments suggest that F_1-ATPasemay exhibit non-Michaelis-Menten kinetics because of ADP inhibition, the deviation from Michaelis-Menten kinetics is actually small. [20, 23] To be safer, instead of using K_{M} and V_{max} directly as experimental references in the score function, we defined another quantity ζ to evaluate how close our model-predicted k_{cat} is to the experimental value:

$$\zeta \equiv \begin{cases} -10 & (k_{\text{cat}} \leq 0), \\ \log_{10}\left(\dfrac{k_{\text{cat}}}{k_{\text{cat}}^{\text{exp}}}\right) & (k_{\text{cat}} > 0). \end{cases} \tag{3.16}$$

We included ζ of ATP several concentrations covering the range of µM $\sim$in our score function as independent observables, although they may have been inter-dependent. At each ATP concentration, $k_{\text{cat}}^{\text{exp}}$ is calculated using the Michaelis-Menten equation specified by experimental K_{M} and V_{max}. Obviously, $\zeta^{\text{exp}} = 0$. A small technical problem is that ADP concentration is also a necessary input needed by our model, whereas we are not aware of the ADP concentration under which enzyme activity was measured. We tried setting the ADP concentration to

be 10 μM and 1 nM in optimization, and noticed that the influence is rather small. Therefore, we stick to the results using [ADP] = 10 μM in following sections.

We also included the chemo-mechanical coupling efficiencies under the same series of ATP concentrations as independent observables in our score function, assuming that the mutant F_1-ATPase under investigation maintains perfect coupling, *i.e.*, $\eta^{\mathrm{exp}} = 100\%$.

In addition, alternatively, data of nucleotide titration curves measured in Trp fluorescence experiments [18] may also be included in our score function. In such case, we used total catalytic site occupancy (ignore whether it is an ATP or an ADP in place) under a range of ADP concentrations as independent observables, assuming that ATP concentration is sufficiently low and steady state net flux, if not completely vanished, is small enough to be ignored.

4 Results and Discussion

## 4.1 Optimized parameters reproduce rotary catalysis

As described in section 3.4, we optimized the 17 parameters to maximize the conditional probability of reproducing the experimental data of ATP-concentration-dependent turnover, $k_{cat}([ATP])$, and an ideal chemo-mechanical coupling efficiency η of 100% , assuming uniform priors for all the 17 parameters. Starting from different initial values of the parameters, we carried out 1876 independent optimization runs, all of which reached convergence fast. The final scores approached zero (Figure 3, upper left panel), the ideal upper limit indicated by Equation 3.15, suggesting that the optimal parameter sets obtained almost perfectly reproduce the experimental turnovers and high chemo-mechanical coupling efficiency.

Interestingly, these independent runs of optimization don't seem necessarily to converge to a unique solution, which in itself is not surprising because the number of free parameters, which is 17, exceeds the number of experimental values used as input. The distributions of the optimized parameters are shown in Figure 3, and the mean values and standard deviations are listed in Table 3. Notably, the distributions of the binding free energies are generally unimodal and show smaller variances. In contrast, the distributions of the energy barriers are over larger ranges, some of which even tend to be bimodal. Clearly, the catalytic kinetics of F_1-ATPasehas different sensitivities towards the parameters.

4.1.1 Michaelis-Menten-like catalytic kinetics

Because we only used four points of $k_{cat}([ATP])$ calculated by Michaelis-Menten equation in optimization, we further confirmed that the optimal parameter sets can really reproduce the whole curve of $k_{cat}([ATP])$ by calculating the steady state catalytic kinetics over ATP concentrations from 1 nM to 1 M for 1 nM, 1 μM, 10 μM and 1 mM ADP. As an example, the results for one of the obtained optimal parameter sets are shown in Figure 4.

For relatively high ADP concentration, k_{cat} may be drastically decreased , *i.e.*,
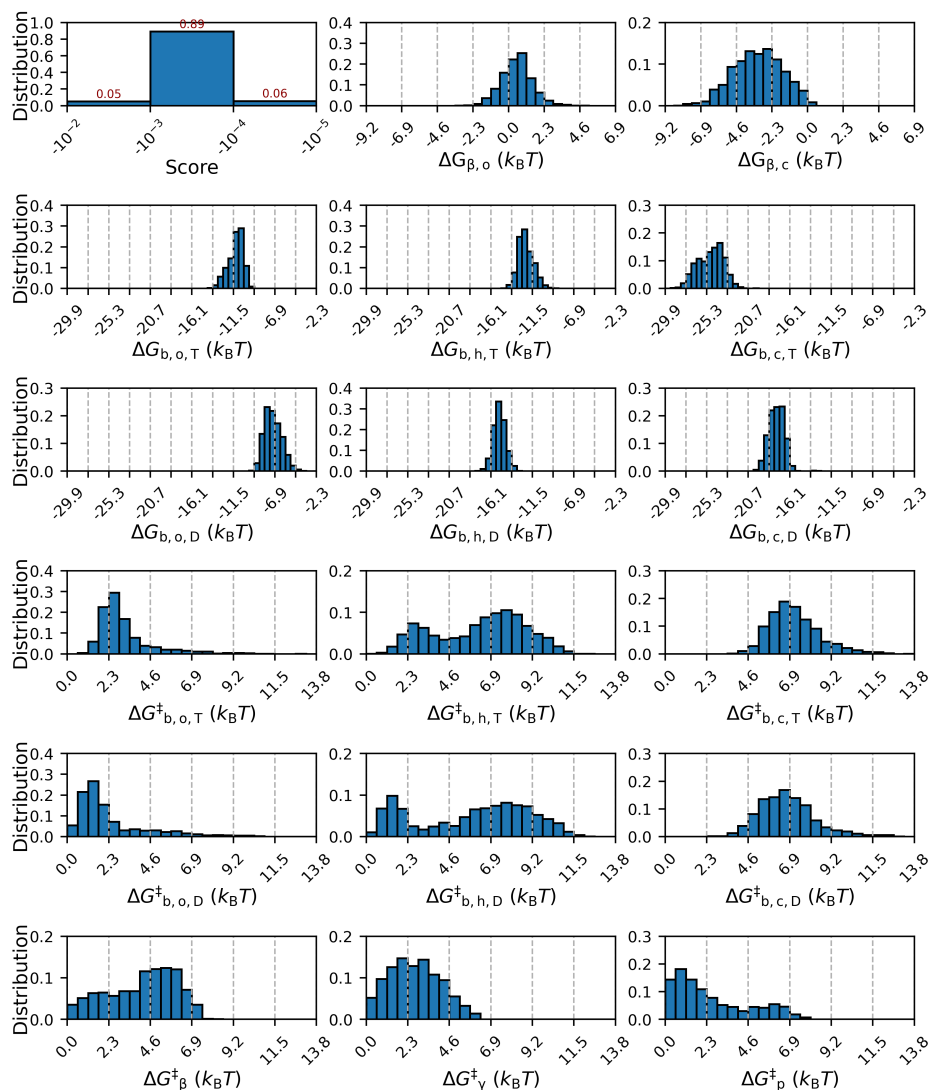
Figure 3: Distributions of the final scores (upper left panel) and optimal parameter values (all the other 17 panels) obtained by the 1876 runs of optimization. For panels showing the distributions of optimal parameter values, the width of grid (grey dotted lines) is 2.303 $k_B T$, where 2.303 ≈ ln(10). When a free energy or energy barrier changes by 2.303 $k_B T$, the corresponding rate constant differs by one magnitude.

Table 3: Mean values and standard deviations of the 1876 optimal parameter sets

Parameter	mean (k_BT)	standard deviation (k_BT)
$\Delta G_{\beta,o}$	0.5	1.0
$\Delta G_{\beta,c}$	-3.2	1.6
$\Delta G_{b,o,T}^{\theta}$	-11.2	0.8
$\Delta G_{b,o,D}^{\theta}$	-7.1	0.9
$\Delta G_{b,h,T}^{\theta}$	-12.2	0.8
$\Delta G_{b,h,D}^{\theta}$	-15.2	0.7
$\Delta G_{b,c,T}^{\theta}$	-24.9	1.4
$\Delta G_{b,c,D}^{\theta}$	-17.6	0.9
$\Delta G_{\beta}^{\ddagger}$	4.2	1.9
$\Delta G_{\gamma}^{\ddagger}$	2.7	1.4
$\Delta G_{p}^{\ddagger}$	2.6	2.1
$\Delta G_{b,o,T}^{\ddagger}$	3.1	1.5
$\Delta G_{b,o,D}^{\ddagger}$	2.4	2.0
$\Delta G_{b,h,T}^{\ddagger}$	6.3	2.5
$\Delta G_{b,h,D}^{\ddagger}$	5.8	3.1
$\Delta G_{b,c,T}^{\ddagger}$	7.2	1.4
$\Delta G_{b,c,D}^{\ddagger}$	6.8	1.6
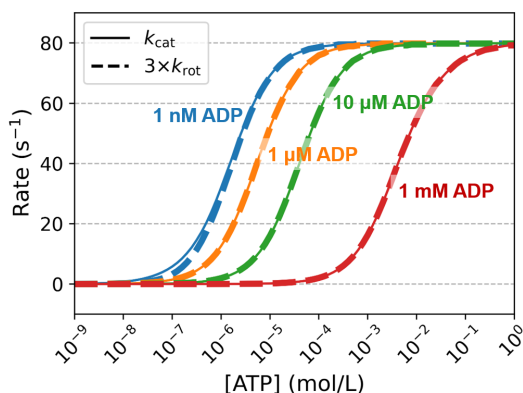
Figure 4: Steady state catalytic kinetics predicted by one optimal parameter set. Solid lines are for k_{cat}, and dotted lines are for $3 \times k_{rot}$. Four pairs of curves of k_{cat} and $3 \times k_{rot}$ are shown, with ADP concentrations being 1 nM (blue), 1 μM (orange), 10 μM (green) and 1 mM (red) respectively. For the latter three pairs ([ADP] = 1 μM, 10 μM and 1 mM), k_{cat} and $3 \times k_{rot}$ almost completely overlap with each other.

$k_{cat} \approx 0$, or even reversed, *i.e.*, $k_{cat} < 0$, especially for the left part of the curve for 1 mM ADP. Nevertheless, within the region of [ATP] > [ADP], the curves of k_{cat} and k_{rot} as functions of ATP concentration can both be perfectly fit to Michaelis-Menten equations, giving K_M and V_{max} as listed in Table 4. Especially, K_M and V_{max} of k_{cat} for 10 μM ADP are very close to what we input for optimization (0.05 mM, 80^{-1} s), [13] confirming that the curve of k_{cat} is well reproduced by the parameter set. In addition, K_M of k_{rot} for 10 μM ADP is equal to that of k_{cat}, and V_{max} of k_{rot} is one third of that of k_{cat}, giving $\eta \sim 100\%$ in the investigated range of ATP concentration.

4.1.2 Tri-site catalysis dominates

Further analysis on the steady state turnover offers a potential explanation resolving the dispute about whether a bi-site mechanism or a tri-site mechanism is fundamental for F_1-ATPase at full speed of catalysis.

There are different definitions for these terms, namely, "uni-site", "bi-site" and "tri-site", in the literature. For clarity, here we use these terms in the meaning of the largest number of nucleotide-occupied catalytic sites observed simultane-

Table 4: Curve fitting results of $k_{cat}([ATP])$ and $k_{rot}([ATP])$ by Michaelis-Menten equation $k = V_{max}[ATP]/(K_M + [ATP])$. All the fits obtained $R^2 > 0.99$.

| [ADP] (mol/L)	k_{cat}		k_{rot}	
	K_M (mol/L)	V_{max} (s^{-1})	K_M (mol/L)	V_{max} (s^{-1})
10^{-9}	6.68×10^{-7}	50.9	2.21×10^{-6}	20.1
10^{-6}	5.35×10^{-6}	79.9	5.48×10^{-6}	26.6
10^{-5}	3.98×10^{-5}	79.8	3.98×10^{-5}	26.6
10^{-3}	3.81×10^{-3}	77.4	3.81×10^{-3}	25.8

ously in a catalytic cycle. The chemical space as shown in Figure 2c depicts all the possible pathways of a catalytic cycle included in our Markov model, in which subspace 1 corresponds to a uni-site mechanism, subspaces 2 and 3 correspond to bi-site mechanisms and subspaces 4, 5 and 6 correspond to tri-site mechanisms. In principle, our Markov model does not forbid any of these mechanisms, and they may occur simultaneously. Nevertheless, the steady state net fluxes in some of the subspaces may be much larger than others, indicating that one or more specific mechanisms contribute more to the total net fluxes of ATP hydrolysis, *i.e.*, k_{cat}.

We calculated the ratio of the steady state net fluxes in chemical space to k_{cat}, for ATP concentration from 1 nM to 1 M, in the presence of 1 pM, 1 nM and 1 μM ADP, and plotted the results in Figure 5. For most of the nucleotide concentrations investigated, fluxes in subspaces 1 ("uni-site"), 2 and 3 ("bi-site") are negligible in comparison to those in subspaces 4, 5 and 6 ("tri-site"), *i.e.*, tri-site catalysis is definitely dominant. Only when both ADP and ATP concentrations are sufficiently low can uni-site and bi-site catalysis become dominant, suggesting that multi-site catalysis involves occupation by both ATP and ADP.

In all, our model suggests that under physiological conditions, a tri-site mechanism for catalysis is dominant, whereas the contribution of bi-site and/or uni-site catalysis increases for sufficiently low nucleotide concentrations (unphysiological).

4.1.3 Three catalytic sites work in parallel

We have shown that our optimal parameters predict a ratio of k_{cat} to k_{rot} very close to 3, agreeing with the highly coupled rotation with catalysis observed
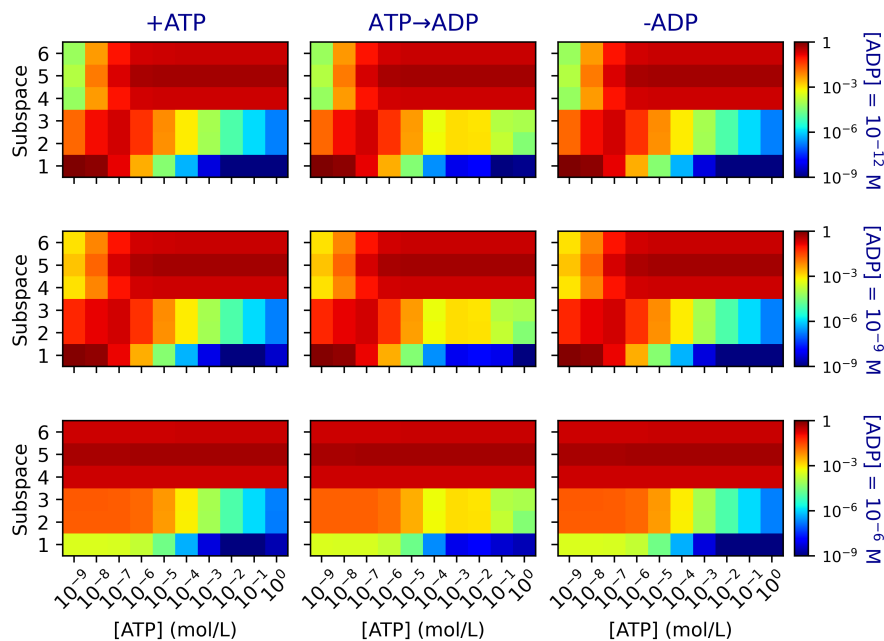
Figure 5: Ratio of the steady state net fluxes in chemical space (Figure 2c) to k_{cat}, under ATP concentration from 10^{-9}-1 M, and ADP concentrations 1 fM, 1 nM and 1 µM (from top to bottom). Color coding reflects the ratio (dimensionless). Warmer color indicates larger contribution of the particular flux to the total rate of catalysis (k_{cat}). Index of subspace is in accordance with Figure 2c, where at most 1 catalytic site is occupied in subspace 1 (uni-site mechanism), at most 2 catalytic sites are occupied simultaneously in subspaces 2 and 3 (bi-site mechanism) and at most 3 catalytic sites are occupied simultaneously in subspaces 4, 5 and 6 (tri-site mechanism).
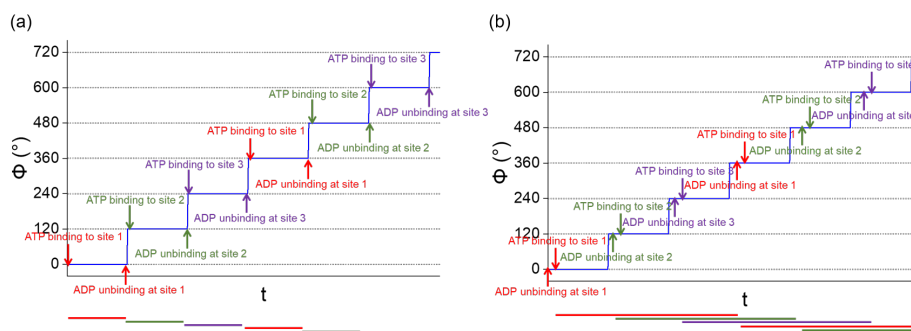
Figure 6: Two possible schemes by which catalysis and rotation are coupled. x-coordinate represents time (t), and y-coordinate represents the cumulative angle of γ-subunitrotation (Φ). Stepped curves (blue) represent trajectories of γ-subunit rotation. The arrows denote the initial (substrate ATP binding to the catalytic site) and final points (product ADP released from the same catalytic site) of a complete event of ATP hydrolysis at one catalytic site. The solid lines at the bottom of each panel represent the time span of complete events of ATP hydrolysis.
(a) The three catalytic sites work in sequence, so that during every 120° rotation of the γ-subunit, one ATP hydrolysis event is completed at one catalytic site.
(b) The three catalytic sites work in parallel, so that only after one revolution of the γ-subunit, one ATP hydrolysis event is completed at one catalytic site.

in experiments. Still, there are multiple probabilities how this factor 3 can be achieved microscopically. Defining a complete event of ATP hydrolysis in a catalytic site as initialized by substrate ATP binding to this site and finalized by product ADP being release in the very same site, two possible schemes coupling catalysis and rotation are shown in Figure 6: (a) the three catalytic sites work in sequence, so that during every 120° rotation of the γ-subunit, one ATP hydrolysis event is completed at one of the three catalytic sites; (b) the three catalytic sites work in parallel, so that only after one revolution of the γ-subunit, one ATP hydrolysis event is completed at one catalytic site. Only scheme b is supported by experimental observations. [28, 58]

To check if our model agrees with scheme b, we want to directly visualize the catalytic cycle. Therefore, instead of doing numerical calculations, we ran 10 KMC simulations assuming physiological nucleotide concentrations ([ATP] = 1 mM, [ADP] = 0.25 mM), each of time length 10 s. In total, 4095 events of ATP hydrolysis completed, in contrast to only 1 complete event of ATP synthesis. The
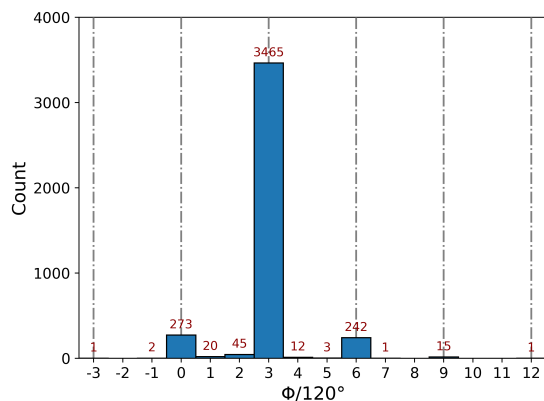
Figure 7: Distribution of the number of 120° steps γ-subunit took ($\Phi/120°$, where Φ is the cumulative angle of γ-subunit rotation) during one complete event of ATP hydrolysis seen in KMC simulations.

average rates, $k_{cat} = (40.9 \pm 2.1)$ s^{-1} and $k_{rot} = (13.7 \pm 0.7)$ s^{-1}, agree well with the values calculated numerically ($k_{cat} = 40.3$ s^{-1}, $k_{rot} = 13.5$ s^{-1}). The average population of states also agrees well with the analytical steady state distribution, confirming that the simulations have well converged.

The picture of rotary catalysis illustrated by the simulations agrees with scheme b, *i.e.*, the three catalytic sites work in parallel, and it in average takes three 120° steps for one ATP to be hydrolyzed in a catalytic site. Figure 7 shows the histogram of the number of 120° steps γ-subunit took ($\Phi/120°$, where Φ is the cumulative angle of γ-subunit rotation) during one complete event of ATP hydrolysis. The majority of complete ATP hydrolysis events takes three 120° steps.

Still, rotary catalysis depicted by our model is stochastic in nature. As can be seen in Figure 7, a small portion of ATP hydrolysis events completed not exactly after one revolution of γ-subunit. Some events took less than one revolution of γ-subunit, indicating that even at full-speed catalysis the γ-subunit went through slippery steps (futile rotation without completing ATP hydrolysis). The opposite case, *i.e.*, ATP hydrolysis event completed after less than one revolution of the γ-subunit, also occurred at almost the same frequency, leading to the near 100% chemo-mechanical coupling efficiency.

4.2 Discussion: how does our model F$_1$-ATPase work?

As mentioned in section 4.1, the optimal parameter sets we obtained are probably distributed over multiple regions. To understand why different optimal parameter sets lead to similar catalytic kinetics, we further checked if, despite the variances, these parameter sets reflect a common behavior of the catalytic mechanism of our model F$_1$-ATPase. To be more specific, we raised two questions: Which ones of the model parameters are the limiting factors of k_{cat}, and which ones determine the direction and rate of γ-subunitrotation? We will propose and discuss answers to these questions using a network scheme of dominant catalytic pathways (Figure 8) and corresponding energy diagram (Figure 9).

4.2.1 Dominant catalytic pathways

First of all, we identified the common dominant catalytic pathways predicted by these optimal parameter sets by analyzing the distribution of steady state net fluxes. By "dominant catalytic pathways", we mean those pathways maintaining the largest steady state net fluxes.

Figure 8 summarizes all the microscopic transitions and Markov states that are dominant in a catalytic cycle predicted by the optimal parameter sets we obtained. The scheme depicts one 120°-step of the γ-subunit, during which catalytic sites 1 and 2 undergo a series of changes, while catalytic site 3 remains closed and occupied. Every dotted box includes a group of Markov states that have the same configuration of β-subunit conformations and γ-subunit orientation, but differ in their binding states. The catalytic site containing a question mark can be empty (E), ADP-bound (D) or ATP-bound (T). Any nucleotide bound to a closed catalytic site is considered to be in fast conversion between ATP and ADP (T/D), as suggested by the high rate of reversible phosphate bond cleavage, $k_{\text{p}} \sim 10^8 s^{-1}$. Every arrow denotes one or more microscopic transitions (grey for β-subunit conformational change and red for γ-subunit rotation), each of which connects two states of the same binding state configuration from the two groups.

The initial state, state 1-D , and the final state, state 1', are equivalent (can be superimposed after 120° rotation of the whole structure). During the transition from state 1-D to state 1', one ADP produced in previous 120° step(s) is released from catalytic site 1, and one ATP from bulk solution binds to catalytic site 1. γ-subunit rotates to the next orientation, 200°, allowing the β-subunits to adjust

their conformations and get ready for the next 120° step initiated at state 1'. These processes combined result in the net hydrolysis of one ATP molecule.

Multiple transition pathways from state 1-D to state 1' can be seen in the network. For nucleotide exchange, there are alternative pathways where nucleotide can bind and unbind in either an open β-subunit or a half-closed β-subunit, *i.e.*, within group 1, 2 or 3. Fluxes of nucleotide exchange in a closed β-subunit are negligible, thus are not shown in Figure 8. Anyway, before getting ready for the 80° substep (from state 6 to 7), ATP must have already occupied catalytic site 1. Then, states 6 and 7 must always be visited because the 80° substep of the γ-subunit is only allowed by such β-subunitconfiguration. These three states 1-D (1'), 6, 7 are like "checkpoints" along the transition pathways.

To understand why the most dominant states participating in catalytic cycles are as shown in Figure 8, note that all the optimal parameter sets we obtained predict that a nucleotide-bound β-subunit is generally more stable than an empty β-subunit (under standard condition), and that the stability of nucleotide-bound β-subunit increases from open to half-closed to closed. This phenomenon leads to the order of energies of states 1-D (1'), 6 and 7: state 1-D (1'), whose β-subunit configuration is $\beta_{HC}(ADP)\beta_C(ATP \rightleftharpoons ADP)\beta_C(ATP \rightleftharpoons ADP)$, has the lowest free energy among all Markov states. States 6 and 7 of the same β-subunit configuration $\beta_O(ATP)\beta_{HC}(ADP)\beta_C(ATP \rightleftharpoons ADP)$, on the one hand, are of the highest free energy among the states participating in the shown network; on the other hand, they are also of the lowest free energy among the states that allow 80° substep of the γ-subunit. Our numerical calculations show that although the steady state distribution deviates from Boltzmann distribution, populations of the Markov states still maintain roughly the dependence on energy, *i.e.*, states of lower energies are more populated. In fact, all the optimal parameter sets we obtained produce steady state distributions in which the lowest-energy state, state 1-D (1'), takes up more than 80% of the population, for a large range of ATP concentrations. Therefore, the states included in Figure 8 are the most energetically favorable states for fulfilling a catalytic cycle, among which the net fluxes are also the largest.

Taking the energy of state 1-D as the baseline, we plot the energy diagrams along one possible transition pathway in Figure 9, assuming physiological nucleotide concentrations ([ATP] = 1 mM, [ADP] = 0.25 mM). Whereas state 1-D and state 1' are equivalent in structure, in Figure 9, there is an energy drop from state 1-D to state 1', which is due to the fact that one ATP has been hydrolyzed
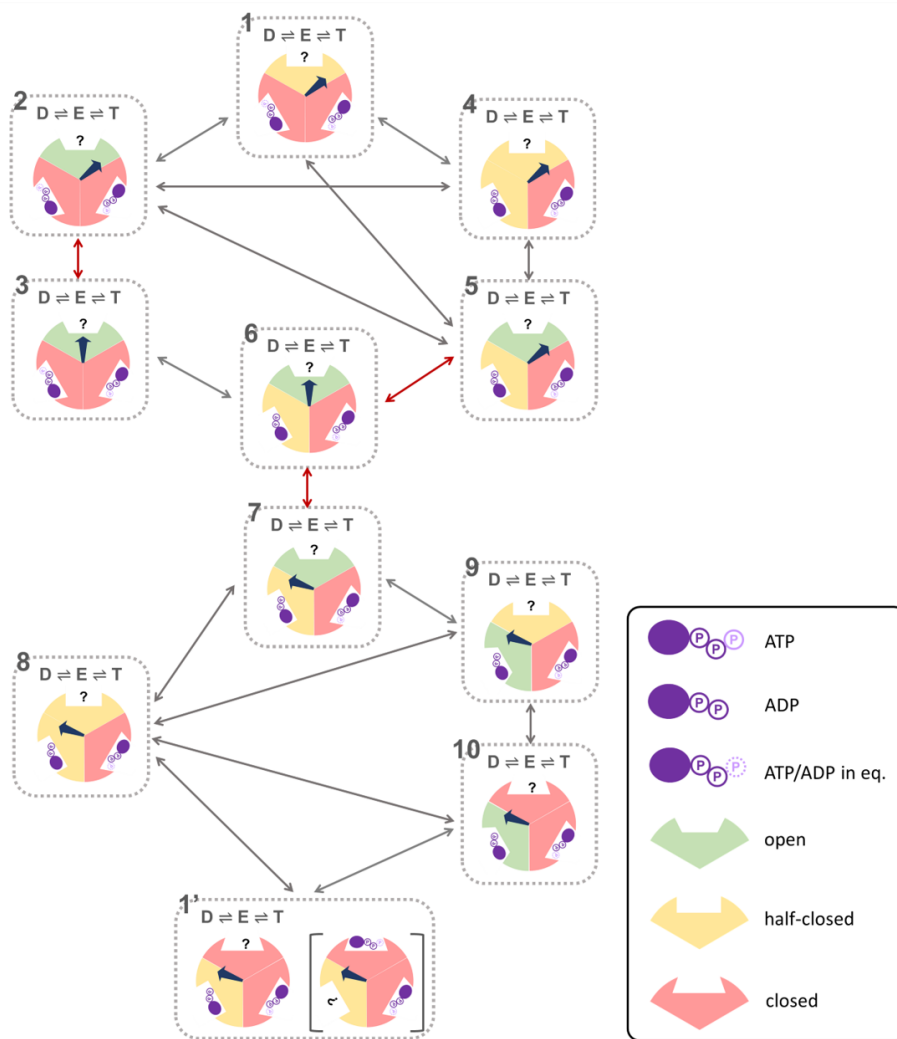
Figure 8: Dominant catalytic pathways predicted by the optimal parameter sets in common. One 120°-step of the γ-subunit is depicted. Every dotted box includes a group of Markov states that have the same configuration of β-subunit conformations and γ-subunit orientation, but differ in their binding states. The catalytic site containing a question mark can be empty (E), ADP-bound (D) or ATP-bound (T). Any nucleotide bound to a closed catalytic site is considered to be in fast conversion between ATP and ADP, denoted by "ATP/ADP in eq." as shown in legend. Every arrow denotes one or more microscopic transitions (grey for β-subunit conformational change and red for γ-subunit rotation), each of which connects two states of the same binding state configuration from the two groups.
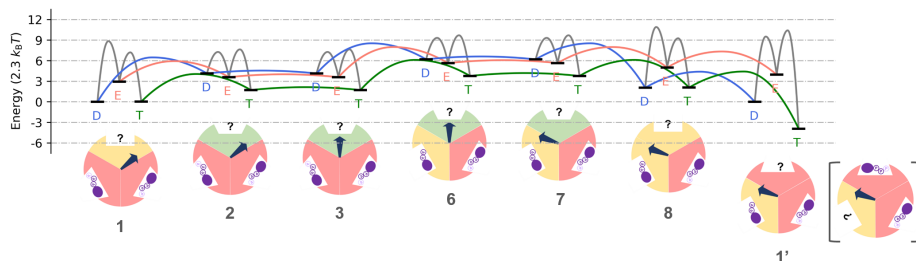
Figure 9: Energy diagram corresponding to the dominant catalytic pathways shown in Figure 8. Representations are in accordance with Figure 8.

and released in the system. Therefore, the energy difference between state 1-D and state 1' under standard condition ([ATP] = [ADP] = 1 M), which is simply the difference between binding free energies of ATP and ADP with a closed β-subunit, $\Delta G^{\theta}_{b,c,T} - \Delta G^{\theta}_{b,c,D}$, is equal to the standard hydrolysis energy of ATP in solution:

$$\Delta G^{\theta}_{hyd} = \Delta G^{\theta}_{b,c,T} - \Delta G^{\theta}_{b,c,D} \tag{4.1}$$

Whereas we didn't implement any restraint or constraint on the binding free energies in the optimization procedure, the optimal parameter sets produce $\Delta G^{\theta}_{b,c,T} - \Delta G^{\theta}_{b,c,D} \approx (4.4 \pm 1.0)$ kcal/mol, which is close to the experimental value of ΔG^{θ}_{hyd} (-7.3 kcal/mol). Because we have assumed that the nucleotides are in large excess compared to F_1-ATPase, nucleotide concentrations can be regarded as constants. We define a 120° step such that this energy drop is compensated before the next 120° step initiates, by replacing the ADP produced in the previous 120° step with an ATP, so that the system remains the same at the beginning of every 120° step, fulfilling a thermodynamic cycle.

4.2.2 Limiting factors of k_{cat}

More quantitatively, based on the energy diagram as shown in Figure 9, we derived an approximate analytical solution for k_{cat} in terms of the model parameters.

The core idea of the derivation is: During one 120° step of the γ-subunit, there are several rate limiting steps, corresponding to the several energy barriers shown in Figure 9. The overall transition rate between the initial and final states,

k_{app}, is limited by the height of the highest barrier relative to the initial state, $\mathcal{A}$, such that $k_{app} = f_{att}\exp(-\mathcal{A}/(k_B T))$, where $f_{att}=10^9$ s^{-1} is the attempt frequency we have chosen when defining the microscopic transition rates. Consequently, if there are multiple alternative transition pathways between two states, the rate and flux along the pathway of the lowest barrier height $\mathcal{A}$ will be the largest. For details of the derivation, see Appendix section S3. Below, we directly present the major conclusions.

First, in principle, as our model does not forbid net ATP synthesis, k_{cat} may become negative under appropriate conditions, indicating net ATP synthesis. Only when ATP hydrolysis is energetically favorable, *i.e.*, if

$$\Delta G_{hyd}(c_T, c_D) = \Delta G^{\theta}_{b,c,T} - \Delta G^{\theta}_{b,c,D} + k_B T\ln(c_T/c_D) < 0 \qquad (4.2)$$

holds, net ATP hydrolysis will occur, giving $k_{cat} > 0$.

Then, in the range of net ATP hydrolysis, we have an approximate analytical solution for k_{cat}:

$$k_{cat} \approx f_{att}\,\exp\left(-\frac{\max\{\mathcal{A}_{u,D}, \mathcal{A}_{b,T}, \mathcal{A}_{ccw}\}}{k_B T}\right), \qquad (4.3)$$

where the barrier heights $\mathcal{A}_{u,D}$, $\mathcal{A}_{b,T}$, $\mathcal{A}_{ccw}$ are related with ADP unbinding, ATP binding and conformational changes enabling counterclockwise rotation of the γ-subunit respectively. Especially,

$$\mathcal{A}_{ccw} = \Delta G_{\beta,o} - \Delta G_{\beta,c} - \Delta G^{\theta}_{b,c,D} + \Delta G^{\theta}_{b,h,D} + \Delta G^{\theta}_{b,o,T} - \Delta G^{\theta}_{b,h,T} + \max\{\Delta G^{\ddagger}_{\beta}, \Delta G^{\ddagger}_{\gamma}\}. \qquad (4.4)$$

The approximate solution of k_{cat} shown by Equation 4.3 roughly agrees with the numerical calculation of k_{cat}. To illustrate that, we define an overall energy barrier for k_{cat}:

$$\mathcal{A}_{cat} = -\ln\left(\frac{k_{cat}}{f_{att}}\right), \qquad (4.5)$$

which can be directly compared with the barrier heights appearing in Equation 4.3, $\mathcal{A}_{u,D}$, $\mathcal{A}_{b,T}$ and $\mathcal{A}_{ccw}$. Two examples of comparison are shown in Figure 10. For every column in Figure 10, a parameter was varied (x-coordinate) while the others were kept constant. k_{cat} was numerically calculated from steady state solution of the master equation to obtain $\mathcal{A}_{cat}$, and the other barrier heights

were calculated directly by Equations S3.1-S3.9 listed in Appendix sections S3. As demonstrated in the lower panels of Figure 10, $\mathcal{A}_{cat}$ correlates well with the highest energy barrier.

Figure 10 also explains why the distributions of the optimal parameters shown in Figure 3 are different: some are more concentrated, whereas others are more diffused. The upper left panel in Figure 10 shows that the changes of k_{cat} and k_{rot} are mild over a large range of $\Delta G_{\beta}^{\ddagger}$, the energy barrier of β-subunits conformational change, because barrier heights dependent on $\Delta G_{\beta}^{\ddagger}$ has not become the rate limiting step until $\Delta G_{\beta}^{\ddagger}$ is large enough, as shown in the left lower panel. In contrary, k_{cat} and k_{rot} are much more sensitive to $\Delta G_{b,h,D}^{\theta}$, the binding free energy of ADP with half-closed β-subunit, as shown in the right column. $\Delta G_{\beta}^{\ddagger}$ and $\Delta G_{b,h,D}^{\theta}$ are typical representatives of the parameters more diffused and those more concentrated in Figure 3.

4.2.3 Direction of γ-subunit rotation

Further, we understood how the direction of γ-subunit rotation is controlled by the parameters. By our definition, $k_{rot} < 0$ means counterclockwise rotation and $k_{rot} > 0$ means clockwise rotation.

Giving the condition that inequality 4.2 holds, *i.e.*, net ATP hydrolysis occurs, the direction of γ-subunitrotation is kinetics-controlled, because the final state reached after one 120° clockwise rotation is equivalent to that reached after one 120° counterclockwise rotation. Therefore, a counterclockwise 120° rotation and a clockwise one are actually alternative pathways. We inferred the most-likely pathways of a clockwise 120° rotation as shown in Figure 11.

In comparison with a step of counterclockwise 120° rotation already shown in Figure 8, the highest barrier height related with conformational changes enabling the clockwise 120° rotation is decided by the energy of state 12-T (13-T):

$$\mathcal{A}_{cw} = \Delta G_{\beta,o} - \Delta G_{\beta,c} - \Delta G_{b,o,D}^{\theta} + \max\{\Delta G_{\beta}^{\ddagger}, \Delta G_{\gamma}^{\ddagger}\}. \tag{4.6}$$

If $\mathcal{A}_{ccw}$ is lower than $\mathcal{A}_{cw}$, net flux of counterclockwise rotation will be larger than that of clockwise rotation. Therefore, to ensure counterclockwise rotation, the binding free energies should maintain the inequality:

$$\Delta G_{b,o,T}^{\theta} + \Delta G_{b,h,D}^{\theta} - \Delta G_{b,o,D}^{\theta} - \Delta G_{b,h,T}^{\theta} < 0. \tag{4.7}$$
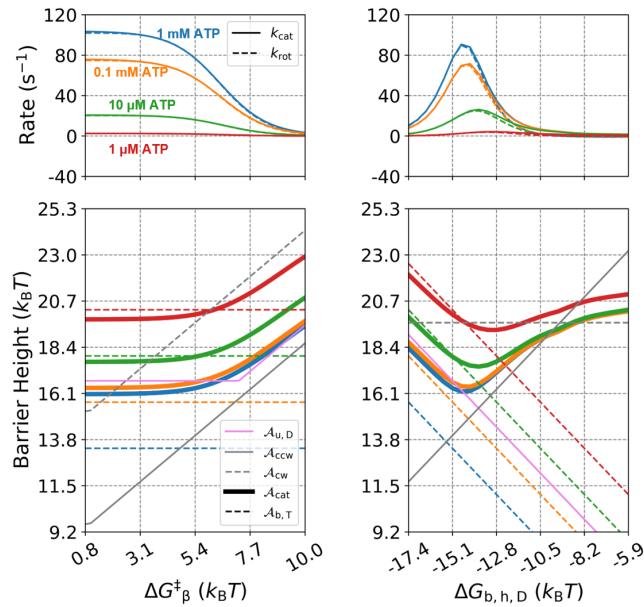
Figure 10: Comparison of the approximate analytical solution of k_{cat} shown by Equation 4.3 with numerical calculation. (Upper panels) k_{cat} (solid lines) and k_{rot} (dotted lines) as functions of the parameters. Four pairs of curves of k_{cat} and k_{rot} are shown, for ATP concentrations being 1 mM (blue), 0.1 mM (orange), 10 μM (green) and 1 μM (red). (Lower panels) Barrier heights $\mathcal{A}_{u,D}$, (violet solid lines), $\mathcal{A}_{ccw}$ (defined by Equation 4.4, grey solid lines), $\mathcal{A}_{cw}$ (grey dotted lines), $\mathcal{A}_{cat}$, (color representation in accordance with the upper panels), $\mathcal{A}_{b,T}$, (colored dotted lines, color representation in accordance with the upper panels), as functions of the parameters. $\mathcal{A}_{cat}$ correlates well with the highest energy barrier among $\mathcal{A}_{b,T}$ and $\mathcal{A}_{u,D}$.
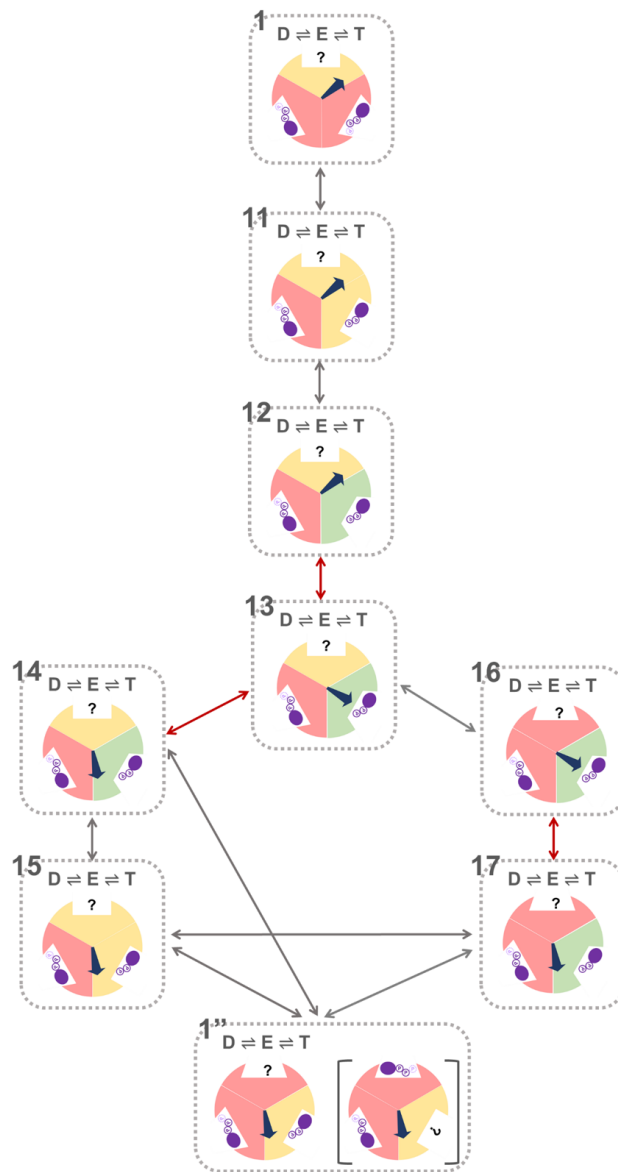
Figure 11: The most-likely pathways of a clockwise 120° γ-subunit rotation we inferred. Representations are in accordance with Figure 8.
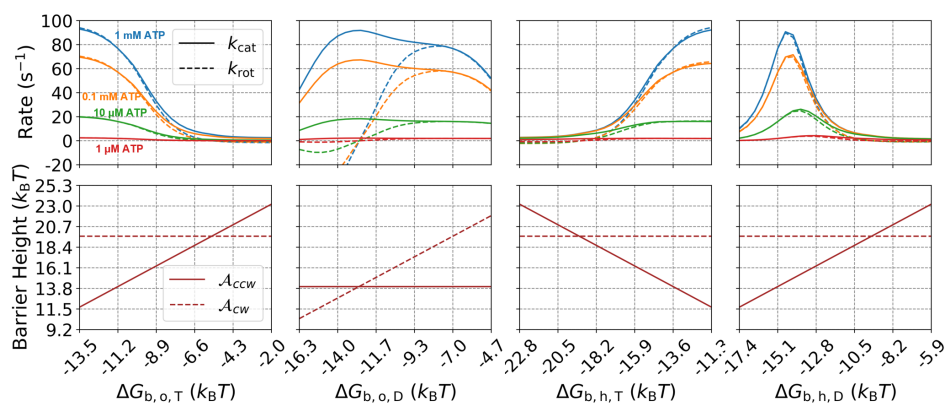
Figure 12: (Upper panels) k_{cat} (solid lines) and k_{rot} (dotted lines) as functions of the binding free energies. Four pairs of curves of k_{cat} and k_{rot} are shown, for ATP concentrations being 1 mM (blue), 0.1 mM (orange), 10 μM (green) and 1 μM (red). (Lower panels) Barrier heights of clockwise rotation ($\mathcal{A}_{cw}$, dotted lines) and counterclockwise rotation ($\mathcal{A}_{cw}$, solid lines) as functions of the binding free energies. Comparing the upper and lower panels in every column, the zero points of k_{rot} coincide with the cross points of $\mathcal{A}_{ccw}$ and $\mathcal{A}_{cw}$, as suggested by the inequality 4.7.

Figure 12 shows k_{cat} and k_{rot} for varying the four binding free energies $\Delta G^{\theta}_{b,o,T}$, $\Delta G^{\theta}_{b,o,D}$, $\Delta G^{\theta}_{b,h,T}$, $\Delta G^{\theta}_{b,h,D}$, which play a role in $\mathcal{A}_{ccw} - \mathcal{A}_{cw}$. Comparing the upper and lower panels in every column, the zero points of k_{rot} coincide with the cross points of $\mathcal{A}_{ccw}$ and $\mathcal{A}_{cw}$, validating our proposal of the inequality 4.7 as a criterion for the direction of γ-subunit rotation.

4.2.4 A plausible explanation for coupling between k_{cat} and k_{rot}

Finally, we can give a plausible explanation for the nice coupling of k_{cat} and k_{rot} predicted by the optimal parameter sets as shown in section 4.1.1. In addition to fulfilling the inequality 4.7 which ensures counterclockwise rotation of the γ-subunit, the conformational energies and binding free energies of the β-subunits obtained by optimization ensure the following tendencies:

First, because the energy of a closed β-subunit is much lower than that of a half-closed β-subunit, a closed, nucleotide-bound β-subunit can hardly open spontaneously unless the γ-subunit rotates to the orientation as to forbid it from closing. Also, because of the relatively low binding rate constants and high

binding affinities, nucleotide exchange occurs hardly in a closed β-subunit, but dominantly in half-closed and/or open β-subunits. Consequently, to complete an event of ATP hydrolysis at one catalytic site, the β-subunitcontaining the site need to adopt the half-closed or open conformation twice, at which points the γ-subunit should be either at the 80° or 120° orientation interacting with the β-subunit. Therefore, it takes an integer (or around integer) number of γ-subunit revolution to complete an event of ATP in one catalytic site, which agrees with observations in KMC simulation as stated in section 4.1.3.

Second, the energy required for a closed, ATP-bound β-subunit to open partially is much higher than that for a closed, ADP-bound β-subunit, so that a closed β-subunitis more likely to open and let the γ-subunit pass only after a newly bound ATP has been converted to ADP.

Third, when the γ-subunit is in place for a catalytic site to release ADP, if the product ADP is not released from the catalytic site, or after ADP being released, the catalytic site remains empty, the γ-subunit can rotate 120° without favorable energy change (comparing state 1-D with states 1'-D and 1'-E in Figure 9), so that the 120° rotation is reversible, contributing nothing to the total net flux of γ-subunit rotation, *i.e.*, k_{rot}. In this way, transition pathways where net fluxes contribute to k_{rot} also contribute proportionally to k_{cat}, giving rise to near perfect coupling of ATP hydrolysis and γ-subunit rotation.

4.3 Cross-validation by nucleotide binding affinities

So far, we have shown that our model is able to reproduce the experimental data of catalytic kinetics which were used in parameter optimization to construct our model. Here, as a cross-validation of our model, we want to test if our model can also predict the differentiated nucleotide binding affinities of the catalytic sites in F_1-ATPase measured in Trp fluorescence experiment, [18] which have not been used in construction of our model.

The first thinking is: are the low, middle and high binding affinities reported in the experimental papers directly comparable to the binding affinities of open, half-closed, and closed β-subunits defined in our model, which are also in the order from low to high, as listed in Table 5?

One reason for questioning the simple correspondence is that, in fact, that high, middle and low binding affinities reported by the Trp fluorescence studies are not directly attributed to certain β-subunitconformations identified in structure

Table 5: Apparent nucleotide binding affinities measured in Trp fluorescence experiment[18] and microscopic nucleotide binding affinities predicted by our Markov model by parameter optimization against catalytic kinetics. Our model-predicted binding affinities are listed from high to low (which coincides with the sequence of binding affinities of closed, half-closed and open β-subunits for both ATP and ADP) to be compared with the experimentally measured binding affinities.

	K_d with ATP		K_d with ADP	
	Experimental	Model prediction	Experimental	Model prediction
High	16 nM	0.15 nM	41 nM	23 nM
Middle	1.5 μM	5.0 μM	6.0 μM	0.25 μM
Low	29 μM	14 μM	42 μM	825 μM

determination studies. In all likelihood, definition of the binding affinities in our model (microscopic binding affinities) could be different from what the Trp fluorescence studies reported (apparent binding affinities).

4.3.1 Re-interpretation of the experimental measurements of binding affinities

To check if the hypothesis mentioned above, that the microscopic binding affinities defined in our model may be different from the apparent binding affinities measured in Trp fluorescence experiments, is valid, we did a synthetic experiment on our model F_1-ATPase. We calculated numerically an ADP titration curve using our model, with a set of binding affinities for open, half-closed and closed β-subunits as input: $K_{d,o} = 0.87$ mM, $K_{d,h} = 28$ μM, $K_{d,c} = 28$ nM. Then, we tried fitting the model-predicted titration curve by Equation 2.1 just as what the experimental papers did, and obtained $K_{d1} = 15$ nM, $K_{d2} = 15$ nM, $K_{d3} = 41$ μM.

In comparison with the binding affinities that we originally input into the model, K_{d1} and K_{d2} are close to $K_{d,c}$, and K_{d3} is close to $K_{d,h}$, whereas $K_{d,o}$ is somehow not captured by fitting of the titration curve. In all, the synthetic experiment failed to reflect faithfully the microscopic binding affinities we had intended to measure.

What causes the discrepancy? A plausible explanation is that, as we have indicated in section 4.2.1, in our Markov model, $\beta_h\beta_c\beta_c$ is the most stable and most populous configuration of F_1-ATPase, and β_o occurs much rarer. Therefore,

instead of showing three distinct binding affinities, only $K_{d,c}$ and $K_{d,h}$, the high and middle binding affinities, are sensed by analyzing the titration curve.

More mathematically speaking, the general form of 2.1 to describe titration curves does apply to our model, because a summation of three fraction terms is intrinsic for a protein containing three independent binding sites. Nevertheless, the three fitting parameters K_{d1}, K_{d2} and K_{d3}, instead of being simply the microscopic binding affinities we defined for open, half-closed and closed β-subunits, are apparent binding affinities influenced by both the microscopic binding free energies and the β-subunit conformational energies. Detailed derivation of exact expressions of the binding affinities in terms of the relevant microscopic parameters is included in Appendix section S2.

In conclusion, the result of our synthetic experiment has shown that there is no simple correspondence of the experimentally measured binding affinities to the microscopic binding affinities defined in our model for open, half-closed and closed β-subunits.

4.3.2 Can our model reproduce experimental titration curves?

We have proved in section 4.3.1 that there is no simple correspondence of the experimentally measured binding affinities to the microscopic binding affinities defined in our model for open, half-closed and closed β-subunits. Moreover, the optimal parameter sets we obtained in the attempt to reproduce catalytic kinetics as discussed in section 4.1 predict nucleotide titration curves obviously different from the experimentally measured ones. Figure 13 shows one example (using one of the optimal parameter sets) of model-predicted ADP titration curve, in comparison with the experimentally measured one. Significant deviation between the two curves is seen.

Still, can we choose the microscopic binding affinities appropriately, such as to reproduce the experimental titration curves in addition to other experimental data, especially catalytic kinetics?

To answer this question, we first carried out another round of optimization, maximizing the conditional probability of parameters, given experimental data of both the catalytic kinetics (*i.e.*, Michaelis-Menten-like ATP-concentration-dependent k_{cat} and $\eta \sim 100\%$) and the ADP titration curve. Unexpectedly, the optimization failed to find any parameter set that could well reproduce both parts of experimental observations.
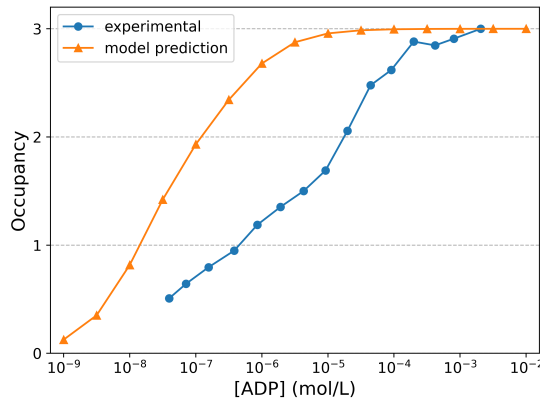
Figure 13: Our model-predicted (using one of the optimal parameter sets) ADP titration curve (orange) in comparison with the experimentally measured one (blue, data from Ref. [18]).

As a control group for comparison, we carried out one more round of optimization, maximizing the conditional probability of parameters given only the experimental ADP titration curve (without evaluation on reproducing k_{cat} and η). The optimal parameter sets found, being very different from those obtained previously by optimization considering only k_{cat} and η, produced ADP titration curve very close to the experimental observation; nevertheless, they gave small k_{cat} and k_{rot} showing almost no change for ATP concentration in the range from μM to mM, indicating the absence of any effective rotary catalysis.

Putting all these results together, our tentative conclusion is that our current Markov model is incapable of reproducing both strongly-coupled rotary catalysis and nucleotide binding behaviors observed in Trp fluorescence experiments. This conclusion suggests that at least one of the assumptions of our current model is wrong, or insufficient for explaining the known dynamics of F_1-ATPase. Therefore, in the following section, we discuss the underlying problems of our current assumptions and propose possible directions for further refinement of our model.

4.3.3 Possible directions for refinement

In this section, we re-consider our model assumptions that may have caused the failure of our model to reproduce both rotary catalysis and the nucleotide titration curves measured in Trp fluorescence experiment. [18] We kindly remind our readers that in chapter 2, we have given a detailed description of the assumptions of our model, as well as relevant experimental observations that support these assumptions. Therefore, in the following text, we refer to these assumptions directly without repeating them in detail.

Is it enough to have only three β-subunit conformations, each associated with one conformational energy, one ATP binding free energy and one ADP binding free energy, regardless of the orientation of the γ-subunit? (Questioning on Assumptions 1 and 4)

Some studies on F_1-ATPase suggest the possibility of a fourth β-subunit conformation or even more. For example, although β_{TP} and β_{DP} seen in crystal structures are of very similar conformations commonly classified into the closed conformation, they could actually have different binding affinities that make β_{TP} prefer ATP-binding and β_{DP} prefer ADP-binding. If this is the case, at least an additional conformation c′ should be included in our model. The conformational energy of c′ could be set to be equal to that of the current closed conformation, but the binding free energies of c′ should be included into our model as two additional independent parameters.

There may also be additional β-subunit conformation(s) or altered binding affinities of the conformations already included that appear only at certain orientation(s) of the γ-subunit. A recent study used tryptophan fluorescence to measure the binding affinities of F_1-ATPase locked in the catalytic dwell (γ-subunit at 80° orientation) and ATP waiting dwell (γ-subunit at 120° orientation) by disulfide crosslink. [32] The ATP waiting dwell structure shows much reduced binding affinities in comparison with the catalytic dwell structure, especially for the low affinity site, which almost completely loses the ability to bind nucleotide at up to 5 mM concentration, potentially suggesting that the conformation of the low affinity site has changed dramatically when γ-subunit rotates from 80° position to 120° position. If so, an extra low-affinity β-subunit conformation o′ should be included in addition to the open conformation, or the binding affinities of the open conformation should be allowed to change when the γ-subunit is at 120° orientation.

It should also be noted that the existence of extra β-subunitconformations of distinct binding affinities is not in controversy with previous Trp fluorescence experiments reporting only three binding affinities, [17, 18] because the ATP waiting dwell is likely to be transient, and the related β-subunit conformation(s) may be of populations so low that it/they cannot be sensed out of the titration curves, just like what we saw in our synthetic experiment on our model F_1-ATPase – the low binding affinity $K_{d,o}$ is omitted because of the low population of open β-subunit.

Is it enough to consider only the inter-subunit interaction between the γ-subunit and one of the β-subunits? (Questioning on Assumption 3)

Considering only the inter-subunit interaction between the γ-subunit and one of the β-subunits may be invalid, especially if we do not include more β-subunit conformations. Without assuming the existence of more β-subunit conformations, there is one possible scheme in which the apparent binding affinities obtained in Trp fluorescence experiments correspond directly to microscopic binding affinities of certain β-subunit conformations: for some reason (most likely to be appropriate inter-subunit interactions), the configuration of F_1-ATPase where the three β-subunits adopt a low-affinity conformation, a middle-affinity conformation, and a high-affinity conformation respectively ($\beta_L\beta_M\beta_H$, L, M and H are for low-, middle- and high-affinities respectively) is actually the most stable state and takes up most of the population.

Therefore, if the configuration where the three β-subunits are open, half-closed and closed ($\beta_o\beta_h\beta_c$) in our model could be extra stabilized to become the dominant state by including appropriate, more complicated inter-subunit interactions, we may be able to reproduce the experimental data of both catalytic kinetics and nucleotide binding properties.

Potential choices of additional inter-subunit interactions are: including interactions between the γ-subunit and all the three β-subunits either by introducing more parameters of interaction energies, or by forbidding certain conformation(s) of the β-subunits being affected by the γ-subunit; assuming that interactions within the $\alpha_3\beta_3$ hexamer contributes to extra stabilization of the configuration $\beta_o\beta_h\beta_c$. The latter choice, *i.e.*, including interactions within the $\alpha_3\beta_3$ hexamer, receives extra support from an experimental study on axeless F_1-ATPase(*i.e.*, the subcomplex $\alpha_3\beta_3$). [62] This study observed symmetry breaking and catalytic activity of the $\alpha_3\beta_3$ when ATP was present. The three β-subunits, adopting a

$\beta_o\beta_c\beta_c$ configuration, cyclically propagated conformational states in the counterclockwise direction, similar to rotation of the γ-subunit in intact F_1-ATPase. Such results suggest that the asymmetric structure and rotary catalysis of F_1-ATPase may also originate from, at least partially, intrinsic interactions within the $\alpha_3\beta_3$ sphere.

5   Conclusion

Previous experimental and theoretical studies on F_1-ATPase have suggested that coarse-graining and identification of essential DOFs are necessary for constructing a model to explain the catalytic mechanism of F_1-ATPase. In this project, we explored which properties and mechanistic aspects of F_1-ATPase are necessary, sufficient, or dominant for F_1-ATPase function.

We constructed a Markov model for F_1-ATPase incorporating as few as seven DOFs, including orientation of the γ-subunit, and conformations and binding states of the β-subunits, the combination of which gives in total 729 Markov states. 17 independent parameters in terms of binding free energies, conformational free energies and free energy barriers were introduced to fully specify the dynamics of the system.

We formulated a parameter optimization approach based on Bayesian inference, which enables straightforward inclusion of quantitative experimental observations on F_1-ATPase, *e.g.*, catalytic kinetics and nucleotide binding affinities. By optimizing the parameters against experimental data of turnover and γ-subunit revolution of F_1-ATPase, we obtained parameter sets that accurately reproduced the data. In particular, our model exhibits Michaelis-Menten-like dependence of turnover on ATP concentration and near 100% chemo-mechanical coupling efficiency.

We further confirmed by numerical calculations and kinetic Monte-Carlo simulations that our model produce effective rotary catalysis by a microscopic mechanism compatible with experimental observations. Based on these results, we developed a semi-quantitative description of the catalytic mechanism in terms of both thermodynamics and kinetics, forming a full picture of how our model F_1-ATPase works.

As cross-validation of our model, we checked if our model could also reproduce experimentally measured nucleotide binding affinities of F_1-ATPase, which had not been used in previous parameter optimization to construct the model. The negative results of our attempts suggested that within the current framework of our model, it is highly likely that there is no parameter set that can reproduce

experimental observations of both nucleotide binding properties and catalytic kinetics. This conclusion further suggests that at least one of the assumptions of our model is wrong or insufficient for describing the observed dynamics of F_1-ATPase. Therefore, we re-examined our model assumptions and proposed ideas for further refinement.

In conclusion, our model has shown that including as few as three distinct conformations for each β-subunit and only 17 model parameters suffice for producing efficient rotary catalysis of F_1-ATPase, realized by a Brownian ratchet mechanism. However, to reproduce experimentally observed nucleotide binding properties highly likely requires the inclusion of more β-subunit conformations of potentially different binding affinities, and/or better description of inter-subunit interactions.

Bibliography

(1) Lipmann, F. **Metabolic generation and utilization of phosphate bond energy**, *Advances in Enzymology and Related Areas of Molecular Biology* **1941**, *1*, 99–162 (see page 1).

(2) Weber, J.; Senior, A. E. **Catalytic mechanism of F1-ATPase**, *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1997**, *1319*:1, 19–58 (see pages 1, 2, 10).

(3) Malyan, A. N. **Noncatalytic nucleotide binding sites: Properties and mechanism of involvement in ATP synthase activity regulation**, *Biochemistry (Moscow)* **2013**, *78*:13, 1512–1523 (see pages 1, 8).

(4) Walker, J. E. **The ATP synthase: the understood, the uncertain and the unknown**, *Biochemical Society Transactions* **2013**, *41*:1, 1–16 (see pages 1, 8).

(5) Junge, W.; Nelson, N. **ATP synthase**, *Annual Review of Biochemistry* **2015**, *84*, 631–657 (see pages 1, 8).

(6) Nadanaciva, S.; Weber, J.; Senior, A. E. **New probes of the F1-ATPase catalytic transition state reveal that two of the three catalytic sites can assume a transition state conformation simultaneously**, *Biochemistry* **2000**, *39*:31, 9583–9590 (see pages 1, 2).

(7) Penefsky, H. S.; Pullman, M. E.; Datta, A.; Racker, E. **Partial resolution of the enzymes catalyzing oxidative phosphorylation II. Participation of a soluble adenosine triphosphatase in oxidative phosphorylation**, *Journal of Biological Chemistry* **1960**, *235*:11, 3330–3336 (see page 1).

(8) Nelson, D. L.; Lehninger, A. L.; Cox, M. M., **Lehninger principles of biochemistry**; Macmillan: 2008 (see page 3).

(9) Abrahams, J. P.; Leslie, A. G.; Lutter, R.; Walker, J. E. **Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria**, *Nature* **1994**, *370*:6491, 621–628 (see pages 2, 3, 7, 8).

(10) Boyer, P. D. **The binding change mechanism for ATP synthase—some probabilities and possibilities**, *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1993**, *1140*:3, 215–250 (see page 2).

(11) Menz, R. I.; Leslie, A. G.; Walker, J. E. **The structure and nucleotide occupancy of bovine mitochondrial F1-ATPase are not influenced by crystallisation at high concentrations of nucleotide**, *FEBS Letters* **2001**, *494*:1-2, 11–14 (see page 2).

(12)   Senior, A. E.; Nadanaciva, S.; Weber, J. **The molecular mechanism of ATP synthesis by F1F0-ATP synthase**, *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **2002**, *1553*:3, 188–211 (see page 2).

(13)   Weber, J.; Wilke-Mounts, S.; Lee, R.; Grell, E.; Senior, A. **Specific placement of tryptophan in the catalytic sites of Escherichia coli F1-ATPase provides a direct probe of nucleotide binding: maximal ATP hydrolysis occurs with three sites occupied**, *Journal of Biological Chemistry* **1993**, *268*:27, 20126–20133 (see pages 2, 9, 20, 26).

(14)   Weber, J.; Wilke-Mounts, S.; Senior, A. E. **Cooperativity and stoichiometry of substrate binding to the catalytic sites of Escherichia coli F1-ATPase. Effects of magnesium, inhibitors, and mutation**, *Journal of Biological Chemistry* **1994**, *269*:32, 20462–20467 (see page 2).

(15)   Weber, J.; Bowman, C.; Senior, A. E. **Specific tryptophan substitution in catalytic sites of Escherichia coli F1-ATPase allows differentiation between bound substrate ATP and product ADP in steady-state catalysis**, *Journal of Biological Chemistry* **1996**, *271*:31, 18711–18718 (see page 2).

(16)   Löbau, S.; Weber, J.; Senior, A. E. **Nucleotide occupancy of F1-ATPase catalytic sites under crystallization conditions**, *FEBS Letters* **1997**, *404*:1, 15–18 (see page 2).

(17)   Weber, J.; Senior, A. E. **Fluorescent probes applied to catalytic cooperativity in ATP synthase**, *Methods in Enzymology* **2004**, *380*, 132–152 (see pages 2, 9, 20, 45).

(18)   Mao, H. Z.; Gray, W. D.; Weber, J. **Does F1-ATPase have a catalytic site that preferentially binds MgADP?** *FEBS Letters* **2006**, *580*:17, 4131–4135 (see pages 2, 20, 21, 40, 41, 43–45).

(19)   Noji, H.; Yasuda, R.; Yoshida, M.; Kinosita, K. **Direct observation of the rotation of F1-ATPase**, *Nature* **1997**, *386*:6622, 299–302 (see pages 2, 8).

(20)   Yasuda, R.; Noji, H.; Kinosita Jr, K.; Yoshida, M. **F1-ATPase is a highly efficient molecular motor that rotates with discrete 120 steps**, *Cell* **1998**, *93*:7, 1117–1124 (see pages 2, 20).

(21)   Adachi, K.; Yasuda, R.; Noji, H.; Itoh, H.; Harada, Y.; Yoshida, M.; Kinosita, K. **Stepping rotation of F1-ATPase visualized through angle-resolved single-fluorophore imaging**, *Proceedings of the National Academy of Sciences* **2000**, *97*:13, 7243–7247 (see page 2).

(22)   Hirono-Hara, Y.; Noji, H.; Nishiura, M.; Muneyuki, E.; Hara, K. Y.; Yasuda, R.; Kinosita, K.; Yoshida, M. **Pause and rotation of F1-ATPase during catalysis**, *Proceedings of the National Academy of Sciences* **2001**, *98*:24, 13649–13654 (see page 2).

(23) Yasuda, R.; Noji, H.; Yoshida, M.; Kinosita, K.; Itoh, H. **Resolution of distinct rotational substeps by submillisecond kinetic analysis of F1-ATPase**, *Nature* **2001**, *410*:6831, 898–904 (see pages 2, 8, 20).

(24) Masaike, T.; Koyama-Horibe, F.; Oiwa, K.; Yoshida, M.; Nishizaka, T. **Cooperative three-step motions in catalytic subunits of F1-ATPase correlate with 80 and 40 substep rotations**, *Nature Structural & Molecular biology* **2008**, *15*:12, 1326 (see page 2).

(25) Yasuda, R.; Masaike, T.; Adachi, K.; Noji, H.; Itoh, H.; Kinosita, K. **The ATP-waiting conformation of rotating F1-ATPase revealed by single-pair fluorescence resonance energy transfer**, *Proceedings of the National Academy of Sciences* **2003**, *100*:16, 9314–9318 (see pages 2, 8).

(26) Shimabukuro, K.; Yasuda, R.; Muneyuki, E.; Hara, K. Y.; Kinosita, K.; Yoshida, M. **Catalysis and rotation of F1 motor: cleavage of ATP at the catalytic site occurs in 1 ms before 40 substep rotation**, *Proceedings of the National Academy of Sciences* **2003**, *100*:25, 14731–14736 (see page 2).

(27) Nishizaka, T.; Oiwa, K.; Noji, H.; Kimura, S.; Muneyuki, E.; Yoshida, M.; Kinosita, K. **Chemomechanical coupling in F1-ATPase revealed by simultaneous observation of nucleotide kinetics and rotation**, *Nature Structural & Molecular biology* **2004**, *11*:2, 142–148 (see pages 2, 8).

(28) Adachi, K.; Oiwa, K.; Nishizaka, T.; Furuike, S.; Noji, H.; Itoh, H.; Yoshida, M.; Kinosita Jr, K. **Coupling of rotation and catalysis in F1-ATPase revealed by single-molecule imaging and manipulation**, *Cell* **2007**, *130*:2, 309–321 (see pages 2, 12, 29).

(29) Milgrom, M. Y.; Murataliev, B. M.; Boyer, D. P. **Bi-site activation occurs with the native and nucleotide-depleted mitochondrial F1-ATPase**, *Biochemical Journal* **1998**, *330*:2, 1037–1043 (see page 2).

(30) Sakaki, N.; Shimo-Kon, R.; Adachi, K.; Itoh, H.; Furuike, S.; Muneyuki, E.; Yoshida, M.; Kinosita Jr, K. **One rotary mechanism for F1-ATPase over ATP concentrations from millimolar down to nanomolar**, *Biophysical Journal* **2005**, *88*:3, 2047–2056 (see page 2).

(31) Kinosita Jr, K.; Adachi, K.; Itoh, H. **Rotation of F1-ATPase: how an ATP-driven molecular machine may work**, *Annual Review of Biophysics and Biomolecular Structure* **2004**, *33*, 245–268 (see page 2).

(32) Li, Y.; Valdez, N. A.; Mnatsakanyan, N.; Weber, J. **The nucleotide binding affinities of two critical conformations of Escherichia coli ATP synthase**, *Archives of Biochemistry and Biophysics* **2021**, *707*, 108899 (see pages 2, 44).

(33) Sumi, T.; Klumpp, S. **Is F1-ATPase a rotary motor with nearly 100% efficiency? Quantitative analysis of chemomechanical coupling and mechanical slip**, *Nano Letters* **2019**, *19*:5, 3370–3378 (see pages 2, 4).

(34) Kühlbrandt, W. **Structure and mechanisms of F-type ATP synthases**, *Annual Review of Biochemistry* **2019**, *88*, 515–549 (see pages 2, 8).

(35) Nam, K.; Karplus, M. **Insights into the origin of the high energy-conversion efficiency of F1-ATPase**, *Proceedings of the National Academy of Sciences* **2019**, *116*:32, 15924–15929 (see page 2).

(36) Mukherjee, S.; Warshel, A. **The FOF 1 ATP synthase: from atomistic three-dimensional structure to the rotary-chemical function**, *Photosynthesis Research* **2017**, *134*:1, 1–15 (see page 2).

(37) Sugawa, M.; Okazaki, K.-i.; Kobayashi, M.; Matsui, T.; Hummer, G.; Masaike, T.; Nishizaka, T. **F1-ATPase conformational cycle from simultaneous single-molecule FRET and rotation measurements**, *Proceedings of the National Academy of Sciences* **2016**, *113*:21, E2916–E2924 (see page 2).

(38) Böckmann, R. A.; Grubmüller, H. **Nanoseconds molecular dynamics simulation of primary mechanical energy transfer steps in F1-ATP synthase**, *Nature Structural Biology* **2002**, *9*:3, 198–202 (see page 4).

(39) Böckmann, R. A.; Grubmüller, H. **Conformational dynamics of the F1-ATPase 03b2-subunit: A molecular dynamics study**, *Biophysical Journal* **2003**, *85*:3, 1482–1491 (see page 4).

(40) Czub, J.; Grubmüller, H. **Rotation triggers nucleotide-independent conformational transition of the empty 03b2 subunit of F1-ATPase**, *Journal of the American Chemical Society* **2014**, *136*:19, 6960–6968 (see page 4).

(41) Czub, J.; Wieczór, M.; Prokopowicz, B.; Grubmüller, H. **Mechanochemical energy transduction during the main rotary step in the synthesis cycle of F1-ATPase**, *Journal of the American Chemical Society* **2017**, *139*:11, 4025–4034 (see page 4).

(42) Karplus, M.; Gao, Y. Q. **Biomolecular motors: the F1-ATPase paradigm**, *Current Opinion in Structural Biology* **2004**, *14*:2, 250–259 (see page 4).

(43) Mukherjee, S.; Warshel, A. **Electrostatic origin of the mechanochemical rotary mechanism and the catalytic dwell of F1-ATPase**, *Proceedings of the National Academy of Sciences* **2011**, *108*:51, 20550–20555 (see page 4).

(44) Baylis Scanlon, J. A.; Al-Shawi, M. K.; Le, N. P.; Nakamoto, R. K. **Determination of the partial reactions of rotational catalysis in F1-ATPase**, *Biochemistry* **2007**, *46*:30, 8785–8797 (see pages 4, 10, 14).

(45) Menz, R. I.; Walker, J. E.; Leslie, A. G. **Structure of bovine mitochondrial F1-ATPase with nucleotide bound to all three catalytic sites: implications for the mechanism of rotary catalysis**, *Cell* **2001**, *106*:3, 331–341 (see pages 7, 9).

(46) Sabbert, D.; Engelbrecht, S.; Junge, W. **Intersubunit rotation in active F-ATPase**, *Nature* **1996**, *381*:6583, 623–625 (see page 7).

(47) Junge, W.; Sabber, D.; Engelbrecht, S. **ATP-synthesis. Rotatory catalysis by F-ATPase: Real-time recording of intersubunit rotation**, *Berichte der Bunsengesellschaft für physikalische Chemie* **1996**, *100*:12, 2014–2019 (see page 7).

(48) Von Ballmoos, C.; Wiedenmann, A.; Dimroth, P. **Essentials for ATP synthesis by F1F0 ATP synthases**, *Annual Review of Biochemistry* **2009**, *78*, 649–672 (see page 8).

(49) Watanabe, R.; Noji, H. **Chemomechanical coupling mechanism of F1-ATPase: catalysis and torque generation**, *FEBS Letters* **2013**, *587*:8, 1030–1035 (see page 8).

(50) Gogol, E. P.; Johnston, E.; Aggeler, R.; Capaldi, R. A. **Ligand-dependent structural variations in Escherichia coli F1 ATPase revealed by cryoelectron microscopy**, *Proceedings of the National Academy of Sciences* **1990**, *87*:24, 9585–9589 (see page 8).

(51) Grüber, G.; Capaldi, R. A. **The Trapping of Different Conformations of the Escherichia coli F1 ATPase by Disulfide Bond Formation: EFFECT ON NUCLEOTIDE BINDING AFFINITIES OF THE CATALYTIC SITES**, *Journal of Biological Chemistry* **1996**, *271*:51, 32623–32628 (see page 8).

(52) Hausrath, A. C.; Grüber, G.; Matthews, B. W.; Capaldi, R. A. **Structural features of the γ subunit of the Escherichia coli F1 ATPase revealed by a 4.4-Å resolution map obtained by x-ray crystallography**, *Proceedings of the National Academy of Sciences* **1999**, *96*:24, 13697–13702 (see page 8).

(53) Suzuki, T.; Tanaka, K.; Wakabayashi, C.; Saita, E.-i.; Yoshida, M. **Chemomechanical coupling of human mitochondrial F1-ATPase motor**, *Nature Chemical Biology* **2014**, *10*:11, 930–936 (see page 8).

(54) Walker, J. E. **ATP synthesis by rotary catalysis (Nobel Lecture)**, *Angewandte Chemie International Edition* **1998**, *37*:17, 2308–2319 (see page 8).

(55) Nakamoto, R. K.; Ketchum, C. J.; Al-Shawi, M. K. **Rotational coupling in the F0F1 ATP synthase**, *Annual review of biophysics and biomolecular structure* **1999**, *28*:1, 205–234 (see pages 10, 14).

(56) Boyer, P. D.; Cross, R. L.; Momsen, W. **A new concept for energy coupling in oxidative phosphorylation based on a molecular explanation of the oxygen exchange reactions**, *Proceedings of the National Academy of Sciences* **1973**, *70*:10, 2837–2839 (see pages 10, 14).

(57) Itoh, H.; Takahashi, A.; Adachi, K.; Noji, H.; Yasuda, R.; Yoshida, M.; Kinosita, K. **Mechanically driven ATP synthesis by F1-ATPase**, *Nature* **2004**, *427*:6973, 465–468 (see page 12).

(58) Adachi, K.; Oiwa, K.; Yoshida, M.; Nishizaka, T.; Kinosita, K. **Controlled rotation of the F1-ATPase reveals differential and continuous binding changes for ATP synthesis**, *Nature Communications* **2012**, *3*:1, 1–12 (see pages 12, 29).

(59) Gillespie, D. T. **A general method for numerically simulating the stochastic time evolution of coupled chemical reactions**, *Journal of Computational Physics* **1976**, *22*:4, 403–434 (see page 18).

(60) Gillespie, D. T. **Stochastic simulation of chemical kinetics**, *Annual Review of Physical Chemistry* **2007**, *58*, 35–55 (see page 18).

(61) Bock, L. V.; Caliskan, N.; Korniy, N.; Peske, F.; Rodnina, M. V.; Grubmüller, H. **Thermodynamic control of -1 programmed ribosomal frameshifting**, *Nature Communications* **2019**, *10*:1, 1–11 (see page 19).

(62) Uchihashi, T.; Iino, R.; Ando, T.; Noji, H. **High-speed atomic force microscopy reveals rotary catalysis of rotorless F1-ATPase**, *Science* **2011**, *333*:6043, 755–758 (see page 45).
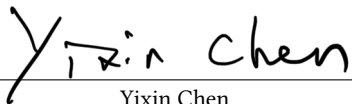
Acknowledgments

Declaration of Authorship

I hereby declare that this thesis is my own unaided work. All direct or indirect sources used are acknowledged as references.

Göttingen, October 15, 2021

Yixin Chen
Yixin Chen