Functional state modulation in Pyruvate Phosphate Dikinase (PPDK) connected to long-range structural coupling and multimerization

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Bartholomäus Daniel Ciupka

Aus Ruda Śląska, Polen

Düsseldorf, Juli 2017

Aus dem Institut für Pharmazeutische und Medizinische Chemie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der

Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Holger Gohlke
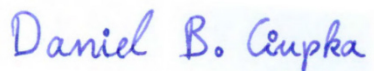
Korreferent: Prof. Dr. Georg Groth

Tag der mündlichen Prüfung: 11.09.2017

Eigenständigkeitserklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf" erstellt worden ist.

Diese Dissertation wurde in der vorgelegten oder einer ähnlichen Form noch bei keiner Institution eingereicht und es wurde bisher keine erfolglosen Promotionsversuche von mir unternommen.

Düsseldorf, den 27.07.2017

(Daniel B. Ciupka)

Part of the following work was already published in cooperation with Alexander Minges, Christian Winkler, Astrid Höppner, Holger Gohlke, and Georg Groth. All experimental work was performed by Alexander Minges, Christian Winkler, Astrid Höppner, and Georg Groth.

Papers

[1] Minges, A.*, Ciupka, D.*, Winkler, C., Höppner, A., Gohlke, H.[#], Groth, G.[#] *Structural intermediates and directionality of the swiveling motion of pyruvate phosphate dikinase.* **Sci. Rep**. **2017**, 7, 45389.

[2] Ciupka, D., Gohlke, H. *On the potential alternate binding change mechanism in a dimeric structure of pyruvate phosphate dikinase*. **Sci. Rep**. **2017**, 7, 8020.

Poster presentation

- Ciupka, D., Minges, A., Groth, G., Gohlke, H. *Potential of mean force calculations predict a stable conformational intermediate in the mechanism of pyruvate phosphate dikinase (PPDK)*. NIC Symposium 2016, Jülich, Germany, February 11 - February 12, **2016.**

- Ciupka, D., Minges, A., Groth, G., Gohlke, H. *Potential of mean force calculations predict a stable conformational intermediate in the mechanism of pyruvate phosphate dikinase (PPDK)*. MCBR4 Scientific Meeting, Heidelberg, Germany, February 23 - February 25, **2015.**

- Ciupka, D., Minges, A., Groth, G., Gohlke, H. *MD simulations and free energy calculations in order to understand the swiveling mechanism of the kinase PPDK*. Workshop on Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, April 11 - April 12, **2014.**

* Both authors contributed equally to the respective work.
[#] These authors share senior authorship.

Für meine Familie, meine Freunde und

alle weiteren, die mich auf meinem Weg

unterstützt, begleitet, und inspiriert haben.

Daß ich erkenne, was die Welt
Im Innersten zusammenhält,

Johann Wolfgang von Goethe,
Faust. Eine Tragödie. Tübingen, 1808.
Erstausgabe digitalisiert durch die Berlin-Brandenburgische Akademie der Wissenschaften [3].

Table of contents

Table of figures

Abbreviations

aa	amino acid
ASU	Asymmetric unit
ATP	Adenosine triphosphate
AMP	Adenosine monophosphate
C. symbiosum	*Clostridium symbiosum*
CD	Central domain
CNA	Constraint Network Analysis
CSS	Complexation significance score
EC	Enzyme Commission number
EM	Electron microscopy
FEM	Feature enhanced maps
F. pringlei	*Flaveria pringlei*
*Fp*PPDK	PPDK from *F. pringlei*
F. trinervia	*Flaveria trinervia*
*Ft*PPDK	PPDK from *F. trinervia*
LD	Linker domain
MD	Molecular dynamics
NBD	Nucleotide-binding domain
NBD1	Nucleotide-binding domain subdomain 1
NBD2	Nucleotide-binding domain subdomain 2
NBD3	Nucleotide-binding domain subdomain 3
NMSim	Normal mode-based geometric simulation
NMR	Nuclear magnetic resonance
PBD	PEP/pyruvate-binding domain
PC	Principal component
PCA	Principal component analysis
PDB	Protein data bank
PEP	Phosphoenolpyruvat
P_i	Orthophosphate
PTS	Phosphotransferase System
PMF	Potential of mean force
PPDK	Pyruvate phosphate dikinase
PP_i	Pyrophosphate
Pyr	Pyruvate
rc	Rigid contact
RMSD	Root-mean-square deviation of atomic positions
SEM	Standard error of the mean
T. brucei	*Trypanosoma brucei*
WHAM	Weighted histogram analysis method
Z. mays	*Zea mays*

Overview of physical variables and constants

G	Free energy
k	Reaction rate coefficient
k_B	Boltzmann constant 1.380648×10^{-23} J K^{-1} (or 3.297623×10^{-27} kcal K^{-1})
$P(x)$	Probability to find the state x
ξ	Reaction coordinate
T	Temperature [in Kelvin]

Abstract

Pyruvate phosphate dikinase (PPDK) is a key enzyme in the energy metabolism that catalyzes the ATP- and Pi-dependent conversion of PEP to pyruvate in C_4 plants, and the reverse ATP forming reaction in bacteria and protista. Several X-ray crystal structures show extreme conformations of the central domain (CD), as well as open and closed conformations of the nucleotide-binding domain (NBD). Accordingly, the CD has been suggested to undergo a swiveling motion to shuttle a phosphoryl group over ~45 Å between two reaction centers, one of the largest single domain movement observed in proteins, so far. However, investigations of the molecular processes that explain and connect the conformational dynamics to functional relevance have remained elusive. The goal of this thesis is therefore to discover the conformational dynamics of PPDK and its connection to PPDK's function.

First, I investigated the correlation of the swiveling to the opening-closing motion in the PPDK *monomer*. For this, I sampled the known conformational space of PPDK using molecular dynamics (MD) simulations. I then analyzed collective and correlated motions, and computed the free energy profiles of the swiveling and the opening-closing motion.

Second, I investigated the intermolecular correlation of the opening-closing motions in the PPDK *dimer* to validate an alternate binding change mechanism, which was proposed based on a crystal structure of the PPDK dimer. For this, I analyzed the structural dynamics via MD simulations, computed the free energy profile of the opening-closing motion of the NBD for different conformations of the other monomer, and investigated the effect of the dimerization on the structural stability via rigidity analysis.

Overall, the results reveal an intramolecular coupling of the swiveling motion of the CD and opening-closing motion of the NBD. Furthermore, the free energy profiles indicate that PPDK might employ a Brownian ratchet mechanism to bias thermal fluctuation to control the directionality of the CD motion. Additionally, I identified an intermediate conformation of the swiveling motion, which was independently resolved by X-ray crystallography. The results support the hypothesis that PPDK dimerization influences the opening-closing motion of the NBDs, and that this influence is mediated via the CDs of both chains. Such an influence would be a prerequisite for an alternate binding change mechanism to occur. To the best of our knowledge, this is the first explanation why PPDK is only active in dimers.

Zusammenfassung

Pyruvate-Phosphat-Dikinase (PPDK) ist ein Schlüsselenzym im Energiestoffwechsel, das die ATP- und Pi-abhängige Umwandung von PEP zu Pyruvat in Pflanzen und die umgekehrte, ATP-bildende Reaktion in Bakterien und Protisten katalysiert. Mehrere Kristallstrukturen zeigen extreme Konformationen der Zentralen Domäne (CD), sowie offene und geschlossene Konformationen einer Nukleotid-Bindungsdomäne (NBD). Folglich wird eine Schwenk-bewegung der CD angenommen um eine Phosphatgruppe über eine Distanz von ~45 Å zwischen zwei entfernten Reaktionszentren zu transportieren, welche die größte Einzeldomäne-bewegung in Proteinen darstellt. Allerdings blieben Analysen der molekularen Prozesse, welche die beobachteten Konformationen in einen funktionellen Zusammenhang bringen könnten, bis jetzt aus. Das Ziel dieser Arbeit ist die Aufzuklärung von PPDK Mechanismus.

Erstens untersuchte ich den Zusammenhang zwischen der Schwenk- und der Öffnungs-/Schließbewegung im PPDK-*Monomer*. Dazu habe ich den konformationellen Raum der PPDK mittels Moleküldynamik (MD)-Simulationen untersucht, gefolgt von Analysen von kollektiven und korrelierten Bewegungen, sowie abschließender Berechnungen der freien Energielandschaften der Schwenk- und Öffnungs-/Schließbewegung.

Zweitens untersuchte ich einen intermolekularen Zusammenhang der Öffnungs-/Schließ-bewegungen im PPDK-*Dimer,* zur Validierung eines alternierenden Bindungswechsel-Mechanismus, der aufgrund einer asymmetrischen Kristallstruktur vorgeschlagen wurde. Dazu habe ich die strukturelle Dynamik mittels MD-Simulationen analysiert, sowie die Energie-landschaft der Öffnungs-/Schließbewegung der NBD für verschieden Konformationen des anderen Monomers errechnet und die Wirkung der Dimerisierung auf die strukturelle Stabilität untersucht.

Insgesamt zeigen die Ergebnisse eine intramolekulare Koppelung der Schwenk-bewegung der CD mit der Öffnungs-/Schließbewegung der NBD. Ferner deuten unsere freien Energielandschaften darauf hin, dass PPDK den Mechanismus einer Brownschen Ratsche anwendet, um thermale Fluktuationen der CD in eine gerichtete Bewegung zu wandeln. Zusätzlich identifizierten wir in unseren MD Simulationen und Energie berechnungen eine Zwischenstruktur der Schwenkbewegung der CD, welche unabhängig davon mittels Röntgen-kristallanalyse aufgelöst wurde. Weiter stützen unsere Ergebnisse die Hypothese, dass die Dimerisierung von PPDK die Öffnungs-/Schließbewegung der NBD beeinflusst und dass diese Einflussnahme durch die CDs beider Monomere vermittelt wird. Ein solcher Einfluss wäre eine Voraussetzung für einen alternierenden Bindungswechsel-Mechanismus. Nach den aktuellen Wissensstand ist das die erste Erklärung, warum nur dimerisiertes PPDK aktiv ist.

1. General introduction

1.1 Importance of proteins – interplay between function, structure, and dynamics

"The key molecular process that makes modern life possible is protein synthesis, since proteins are used in nearly every aspect of living. The synthesis of proteins requires a tightly integrated sequence of reactions, most of which are themselves performed by proteins."

David Goodsell, The Machinery of Life, 2009 [4].

The importance of proteins for all living organisms cannot be underestimated. Proteins are involved in nearly all cellular processes [5]. As enzymes, building blocks, and molecular machines, they are responsible for a remarkably wide range of biological functions. These include biocatalysis [6], transport [6], immune protection [6], growth control, cell differentiation, motion generation [7], and numerous others [6]. As such, detailed knowledge of their respective mode of action at atomic level does not only contribute to the fundamental understanding of most biological functions but also allows us to manipulate them in various applications, such as in medicine, biotechnology, agriculture, and pharmaceutical industry for protein engineering and rational drug design. The key to understand how proteins execute such a variety of functions can be found in their structure and dynamics.

The functional richness of proteins very well match the complexity of their structure at a global level [6]. At the molecular level, proteins consist of one or several polypeptide chain(s) of 20 commonly occurring amino acids, complemented in some cases by cofactors, *e.g.* metal ions or special organic molecules. The amino acid sequence, termed primary structure (Figure 1A), defines the protein and encodes the folding into a characteristic three-dimensional shape. Repeating structural elements, stabilized by hydrogen bonds between the backbone on an local level, form the secondary structure of a protein with α-helices and β-sheets as main types (Figure 1B). The interactions of the residues of a single peptide chain on the global level lead to a specific spatial arrangement, the tertiary structure (Figure 1C). Interactions between several peptide chains lead to the formation of multi-subunit complexes, referred to as quaternary structure (Figure 1D).
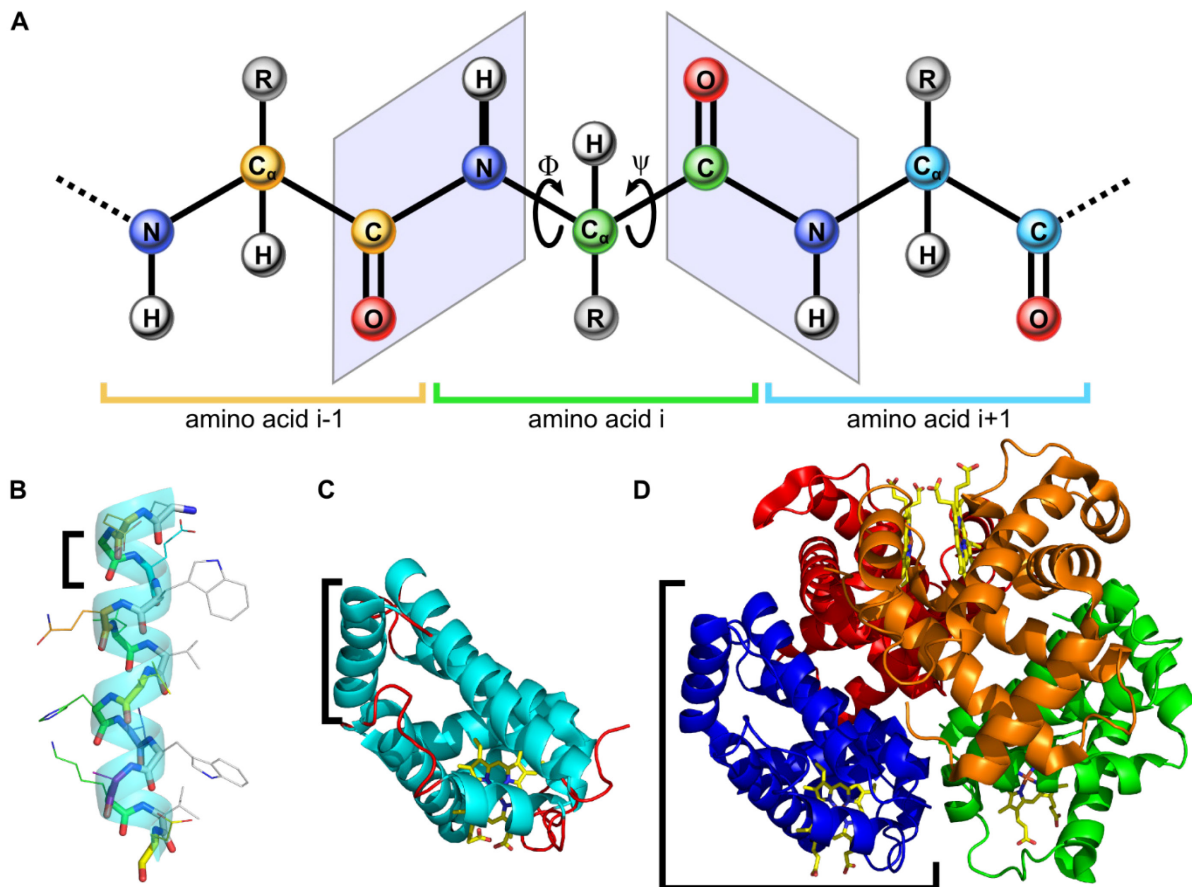
Figure 1: The structure of proteins. (A) The linear sequence of amino acids linked by peptide bonds (primary structure) with the 20 commonly occurring side chains (marked with an "R") attached to the polypeptide backbone. The peptide bond is planar (light blue box), and rotation take place about the Φ and Ψ angle. **(B)** An α-helix is shown as an example of secondary structure, with a bracket spanning one amino acid. **(C)** Structure of myoglobin (PDB ID: 1MBN, which was the first resolved protein structure [8]) is shown as an example of tertiary structure with a bracket spanning one helix, and **(D)** the hemoglobin structure (PDB ID: 2HHB [9]) as an example for quaternary structure with a bracket spanning one protomer.

The interactions responsible for the folding of the protein into its native conformation are primary hydrophobic interactions, due to the tendency of hydrophobic residues to aggregate in water to minimize their exposed surface area, but also salt bridges, hydrogen bonds, pi stacking interactions, and cysteine disulfide bonds. The structure of the protein represents therefor an optimal balance [6] of the various interactions between the amino acids themselves and the solvents.

As such, the structure of proteins is not static. Substantial structural fluctuations occur in proteins [10] and this internal motion, which often carries out the particular function, is considered to be the essential link between structure and function [11, 12].

1.2 Dynamics of the protein structure – key to understand function

"...if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms."

Richard Feynman, Lecture of Physics, 1965 [13].

Protein dynamics, defined as *any time-dependent change in atomic coordinates* [14], is a fundamental part of all proteins and needs to be considered for a complete description. As a result of thermal energy acting on atoms, protein motion takes place on a broad time scale of more than nine orders of magnitude, ranging from femtoseconds (10^{-15} sec) to well beyond seconds [15], and involves a wide range of conformational changes, from the molecular vibration of single bonds to large rearrangements of protein domains over several nanometers (Figure 2). These intrinsic motions are crucial for all proteins [14, 16]. Already when a protein is produced as a linear polypeptide chain by the ribosome, it starts to fold into the three-dimensional structure encoded in the sequence. Even after the folding, internal motions of proteins are constantly present at physiological conditions, and many proteins perform their function by cycling between multiple conformational states [17] (Figure 3).

fs (10^{-15} s)	ps (10^{-12} s)	ns (10^{-9} s)	µs (10^{-6} s)	ms (10^{-3} s)	s (10^{0} s)
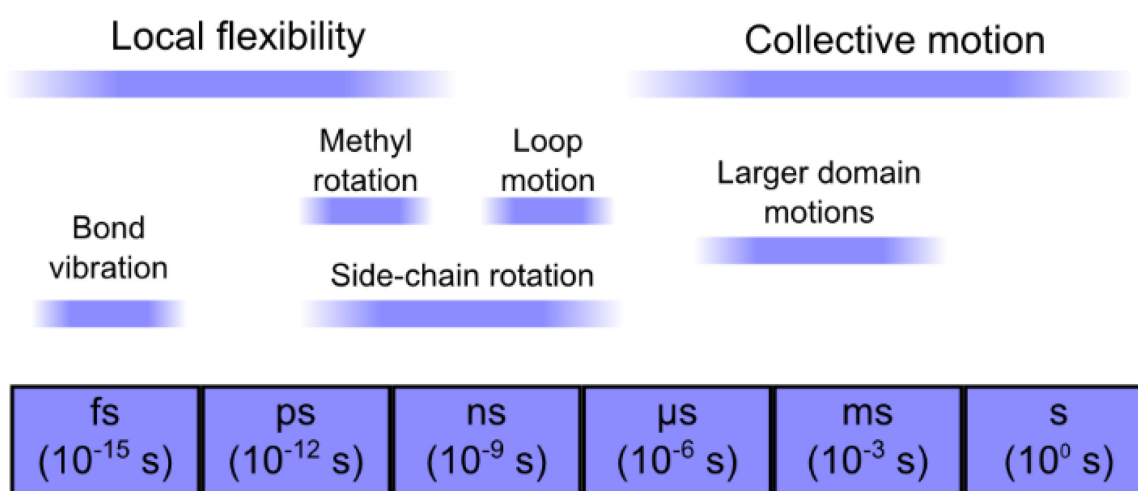
Figure 2: Timescale of dynamic processes in proteins, showing different types of motions occurring in proteins (top) at the associated typical time-scales (bottom). Adapted from [14].
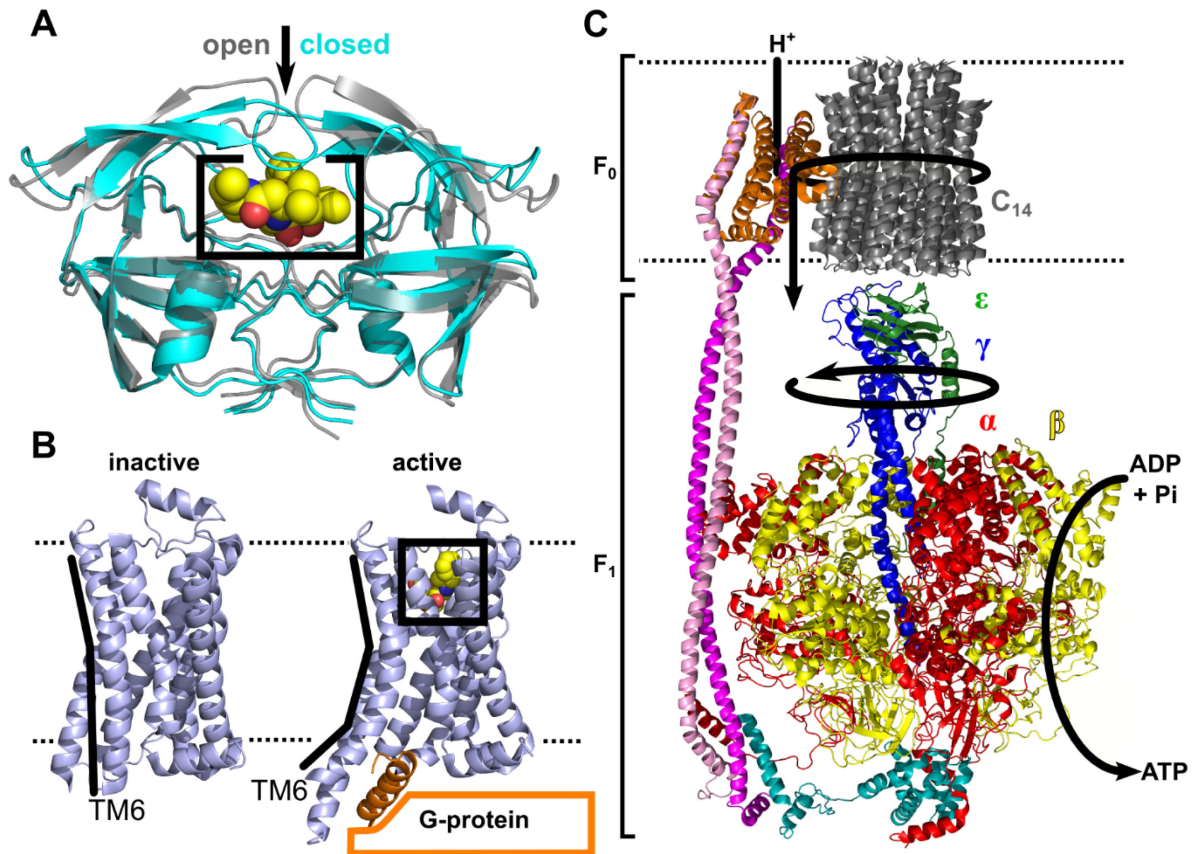
Figure 3: Examples of function-related protein motions, showing **(A)** the HIV-2 protease in open (gray, PDB ID: 1HSI [18]) and closed (cyan, PDB ID: 1JLD [19]) conformation, with the closing related to binding of a ligand (yellow), **(B)** the G-protein coupled receptor in inactive (PDB ID: 3D4S [20]) and active (PDB ID: 3SN6 [21]) conformations, where upon ligand binding (black box) transmembrane helix 6 (TM6) bends (illustrated by a black line), allowing the binding of the G-protein (orange) at the other side of the membrane, **(C)** the structure of ATP synthase (PDB ID: 5T4O [22]) obtained by combined cryo-EM, X-ray crystallography and NMR-spectroscopy [22]. A proton gradient drives the rotation of the C-ring and the associated $\gamma\epsilon$ subunit, changing the conformation of the catalytic sites (formed by the α and β subunit) in the $\alpha_3\beta_3$-ring between "open", "loose binding", and "tight binding" states according to the binding change mechanism, and, hence, providing the energy for the synthesis of ATP.

Collective domain motions represent spectacular examples of the flexibility of proteins and important mechanisms for a variety of protein functions, including: catalysis, regulation of activity, transport of metabolites, formation of protein assemblies, and cellular motion [23]. Domains are often closed around bindings sites, therefore being part of an induced fit mechanism [24], excluding water from the active site, trapping substrates, assisting the positioning of catalysis groups around the substrate, and preventing the escape of reaction intermediates [23].

The close-packing of proteins, with buried polar groups of the main chain and side chain, strongly restrains possible conformational changes [23]. A concept to explain large-scale domain motions by the repertoire of low-energy conformational changes that are available to proteins has been introduced by Gerstein *et. al* (1994) [23]. In this concept, domain motion is derived from two principles: hinge motions and shear motions (Figure 4). Hinge motion takes place at regions not constrained by tertiary packing interactions by the low-energy rotation of one or few backbone torsion angles [23]. The hinge region typically connects two domains, which can be considered to move essentially as rigid bodies with all deformation located at the linking hinge region [25]. Shear motion takes place at closed packed segments and are divided into two types: perpendicular and parallel to the interface [23]. Parallel shear motions involve the interdigitating of side chains [23]. Large shifts require the switching between different interdigitating configurations [23]. In perpendicular shear motions the interface exists only in one state [23]. As such, perpendicular shear motions can be a result of hinge motions outside of the interface region [23] and involve the mostly energetic unfavorable breaking of the interface interactions. It is important to realize that hinge and shear motions are ideal paradigms for describing large domain motions. A real domain motion will often have a combination of both motions, *i.e.* an hinge in one part of the protein and shearing interfaces elsewhere [23].
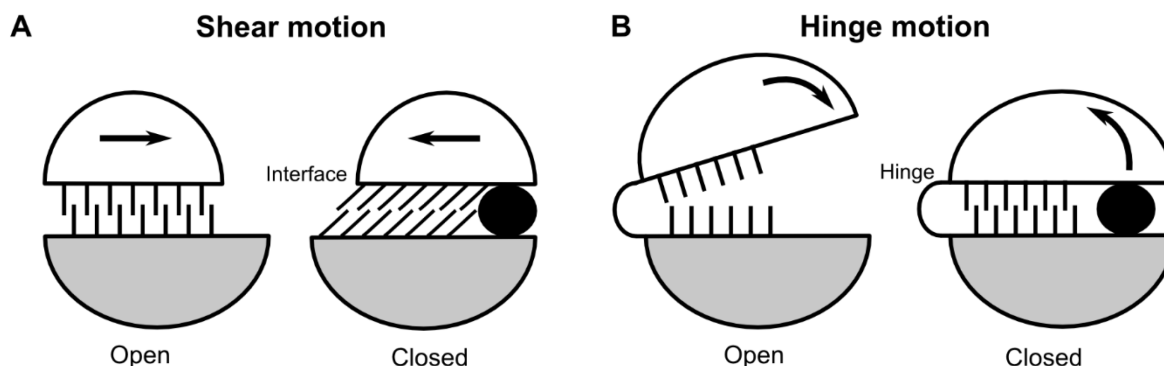
Figure 4: Shear and hinge mechanisms of protein domain motion. Schematic representation of **(A)** shear and **(B)** hinge mechanism of domain motion showing an open and a closed (ligand-bound) state. Domain motions are executed as **(A)** shear motions at closely packed segments involving the interdigitating of side chains or as **(B)** hinge motions at a region not constrained by tertiary packing interactions (hinge) by the rotation of one (or a few) backbone torsions. Hinge motions from outside of the region of the interface can produce a shear motion perpendicular to the plane of the interface (so that the interface exists only in one state). Based on Gerstein *et al.* (1994) [23], adapted from [11, 26].

1.3 Free energy landscape – representation of the structural dynamics

"For molecules, moving in a straight line would seem to be as difficult

as walking in a hurricane is for us."

R. Dean Astumian and Peter Hänggi, Brownian motors, 2002 [27].

To understand the extent of the thermal fluctuations on protein conformations, we have to be aware that proteins operate in a regime where thermal, chemical and mechanical energies are of comparable magnitude (in most cases within a range of 1-1.5 of each other) [28]. In more detail, the thermal energy at room temperature (k_BT with $T = 300$ K) is $9 \cdot 10^{-25}$ kcal ($4 \cdot 10^{-21}$ J). This is just one order of magnitude lower than the hydrolysis of ATP to ADP at physiological condition (the chemical energy mostly associated with biological processes), which is equal to $1,3 \cdot 10^{-23}$ kcal [29] ($5 \cdot 10^{-20}$ J) or, in another way, ~10 k_BT) [27]. This situation is even more extreme when considering that a typical molecular motor consumes 100-1000 ATP molecules per second, equal to $5 \cdot 10^{-16}$ to $5 \cdot 10^{-17}$ W, while with a thermal relaxation time on the order of 10^{-13} s, the power of thermal fluctuation is about 10^{-8} W [27].

Thus, the conformational dynamics of proteins resulting from thermal fluctuation is the reason why the (common) representation of the protein as a single static structure does not fully capture its stochastic nature. The structure of the protein is better represented as a large ensemble of conformations, fluctuating around an average structure [30]. Such an average of a statistical representative set of conformations might be useful for comparing and evaluating the microscopic interactions with the macroscopic behavior observed in experiments [30].

For characterizing the complex nature of protein dynamics, one of the most important concepts is the free energy landscape [15]. A free energy landscape describes the free energy of the protein as a function of the conformations [31]. Since proteins inherently have 3N configurational degrees of freedom, the energy landscape is highly multidimensional [14, 32]. Consequently, the protein can execute numerous motions, which are not all coupled to functions [15, 31]. Therefore, a one-dimensional (or more-dimensional) cross section of the energy landscape along (a) reaction coordinate(s) is often used to study functionally important motions (see Figure 5). Such reduced description of the energy landscape can be provided by a potential of mean force (PMF) calculation [17] (see 1.5.3 Potential of mean force computation).
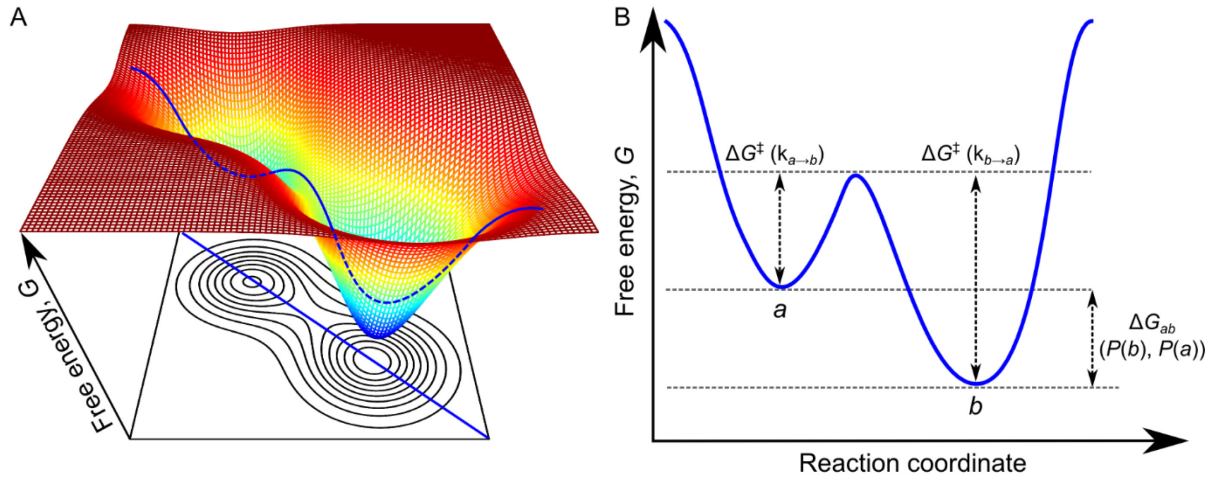
Figure 5: Simplified free energy landscape. (A) Two-dimensional surface (top) and contour map (bottom) of a hypothetical free energy landscape. **(B)** One-dimensional cross section of the higher dimensional free energy landscape (shown in (A)), provides information of the minima (conformational states (a and b)), thermodynamics (relative probabilities of conformational states calculated based on the energy difference (ΔG_{ab}) with the Boltzmann distribution (eq. 2)), and kinetics (transition rate between the states calculated based on the barrier high ($\Delta G^{\ddagger}$) with the Arrhenius equation (eq. 3)).

In a sense, a free energy landscape (see Figure 5) resembles a mountain landscape with the local minima representing highly occupied, low-energy conformations that are separated by energy barriers. A transition is represented as a trajectory in the free energy landscape, beginning from a local minimum (state a), proceeding through an energetic bottleneck, corresponding to a saddle point, and ending at another local energetic minimum (state b) [17]. The energy given in kcal mol^{-1} (or J mol^{-1}), is typically related to the thermal energy available at room temperature (k_BT at 300 K). The probability to find the state x ($P(x)$) is proportional to the energy of the state ($G(x)$) and the thermal energy (k_BT) according to the relationship of the Boltzmann factor (eq. 1)

$$P(x) \propto \exp\left\{-\frac{G(x)}{k_BT}\right\} \qquad \text{(eq. 1)}$$

Moreover, the difference between the energetic state of two conformations ($P(b)/P(a)$) provides precise information on the relative probability of the conformational states (thermodynamic) by the Boltzmann distribution (eq. 2), which relates free energy to a probability:

$$\frac{P(b)}{P(a)} = \exp\left\{-\frac{\Delta G_{ab}}{k_BT}\right\} \qquad \text{(eq. 2)}$$

The height of the energy barriers, which separate the states, allows to approximate the transition rate between the states (kinetic) by the Arrhenius equation (eq. 3):

$$\text{k} = A \exp\left\{-\frac{\Delta G^{\ddagger}}{k_B T}\right\}$$

(eq. 3)

With k being the rate coefficient, A being the frequency factor, and $\Delta G^{\ddagger}$ being the barrier height.

Protein motions can be described as transitions between the conformational states of proteins [31] and can take two forms: fluctuations and relaxations [14, 15]. A change in the system, *e.g.* ligand binding, protein mutation, or changing of external conditions, influences the energy landscape and might shift the system into a non-equilibrium substate (Figure 6A). From a non-equilibrium substate, a protein *relaxes* towards equilibrium by jumping from substate to substate [15]. In equilibrium, a protein *fluctuates* from substate to substate [15].

A nomenclature introduced by Frauenfelder *et al*. [31], based on his work on myoglobin, used the term conformational states to distinguish between the numbers of macrostates, *e.g.* with and without a bound ligand, and introduced a hierarchical classification of the conformational substates (tier-0 to tier-2 dynamics) based on the height of the separating energy barriers (Figure 6B). Tier-0 dynamics are separated by energy barriers of several $k_B T$, corresponding to time-scales of microseconds and slower at physiological conditions [14]. Typically, these are large-amplitude collective motions such as domain motions. Owing to the relatively long substate lifetimes, these individual substates can either be observed directly or be trapped experimentally [14]. Because they are few in numbers and different enough to be described in detail they are termed taxonomic substates [15]. Taxonomic substates should also not be imaged as single conformations. The protein still fluctuates within the wall of the tier-0 substate around an average structure on faster time-scales, exploring a large ensemble of closely related conformations [14, 15]. On tier-1 dynamics, collective fluctuation of small groups of atoms (such as loops) takes place on nanosecond timescale within the wall of the tier-0 substate [14]. The number of substates from this point on is so large that their properties can only be described by distributions and therefore are termed statistical substates. Each taxonomic substate contains a large number of statistical substates. On tier-2 dynamics, local atomic fluctuations such as side chain rotation take place on a picosecond timescale [14]. Even higher tiers such as bond vibrations on the femtoseconds time-scales exist [14].
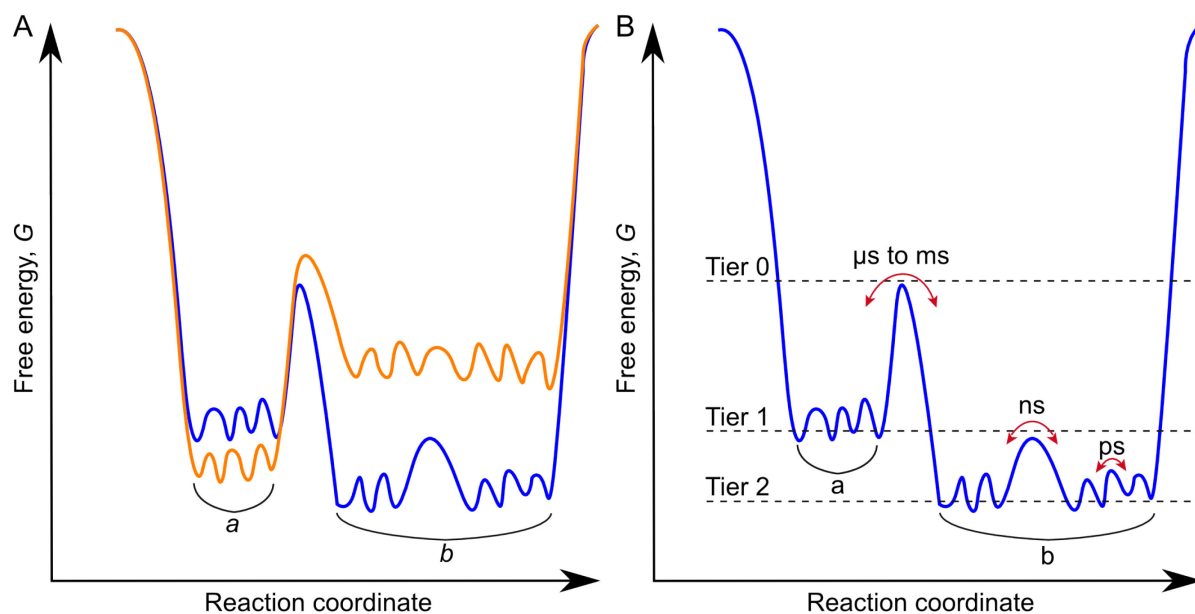
Figure 6: Hierarchy of conformational (sub)states, (A) The state of the system (*e.g.* ligand binding, protein mutation, or changing of external conditions) influences the energy landscape and might shift the equilibrium between the conformations. **(B)** Conformational substates are separated into slow tier-0 dynamics (involving energy barriers of several k_BT and timescales slower than µs) and fast tier-1 and tier-2 dynamics (involving energy barriers of a few k_BT and timescales in the ns and ps range, respectively). Based on the work of Frauenfelder *et al.* [33], adapted from [14].

To what extent the hierarchy proposed by Frauenfelder *et al.* (Figure 6) is generally applicable for proteins other than myoglobin is still a matter of discussion [14, 15], but it is important to keep in mind that even the lowest energy (sub)states are in fact large ensembles of (nearly) isoenergetic conformations [31, 32].

1.4 Proteins as molecular machines – driving forces of directed motions

"The most basic entity of a living cell is a molecule and, consequently, the question we have to answer is: How can a single molecule create directed motion in the presence of strong thermal fluctuations? Or more precisely, how can a molecule, immersed in thermal noise, but provided with some kind of low entropy energy (either in the form of chemical or electrostatic energy) perform directed work?

How does chemistry turn into 'clockwork'?*"*

Peter M. Hoffmann, How molecular motors extract order from chaos, 2016 [28].

Proteins that transduce energy, *e.g.* ATP hydrolysis or a electro-chemical gradient at a membrane, into mechanical work (directed motion) as part of its working cycle are described as molecular machines [7, 28, 34]. At this position we have to mention that also other (and stricter) definitions exist in the scientific literature [28, 35, 36], which partially demands that the input has to be chemical energy and/or that the *final* output has to be mechanical work [28]. Furthermore, the terms molecular machine and molecular motor are sometimes used interchangeably [16] and sometimes molecular motors are presented as a special subtype of molecular machines [28]. Typical examples of molecular machines include linear motors like myosin, kinesin, and dynein, and rotary motors like the ATP synthase (Figure 3C) and flagella. While different input energy sources for protein and artificial molecular machines are available [37], we will stick in the following to a chemical reaction as energy source for simplicity, as ATP hydrolysis is the most common energy source in biological systems.

Our understanding of how directional motion in proteins takes place has been shaped by the work of Huxley in 1957 [38], proposing a mechanism where chemical energy was not used to move muscle filaments relative to each other but to break the connection between them. The motion was supposed to result from random thermal fluctuation and the function of chemical energy was supposed to bias or rectify the random thermal fluctuation into directed motion. Directed motion resulting from undirected thermal fluctuation has been ruled out for systems at equilibrium by the second law of thermodynamics [39], demonstrated by the thought experiment known as 'Maxwell's demon' [40] and later by mechanical models from Smoluchowskis [41] and from Feynmans [13] (Figure 7).
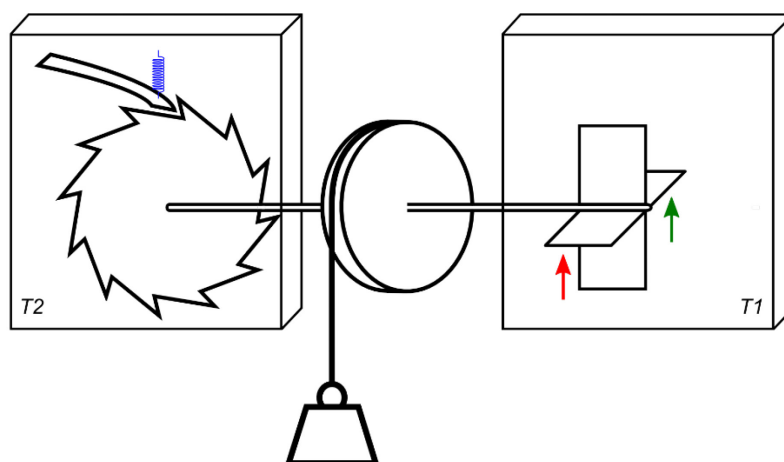
Figure 7: Feynman molecular ratchet model. Ratchet and pawl with spring **(left)** preventing anti-clockwise motion and rectifying random thermal fluctuation, therefore driving the paddle wheel **(right)** and producing work and unidirectional motion **(middle)** for systems where the random thermal force in chamber *T1* are large enough to overcome the spring-force in chamber *T2*. Therefore, the system will not work if the temperature is $T1 = T2$. Directional motion is only possible in a non-equilibrium system (here $T1 > T2$). Adapted from [13].

Structural features alone therefore cannot bias Brownian motion [13, 41, 42]. A requirement to extract directed motion from thermal fluctuation is, therefore, that the system is in out-of-equilibrium condition [28]. In Feynmans molecular ratchet model, this is achieved by a thermal gradient [13]. In proteins, a thermal gradient necessary to drive significant motion is not realistic [28, 42, 43].

The idea of the Brownian ratchet reemerged in the early 1990s by several groups [44, 45], which presented models for extracting directed motion from thermal fluctuations in isothermal conditions [46]. In these models, the non-equilibrium condition comes from chemical reaction(s) and binding events. In chemical systems at equilibrium, forward and backward reaction rates are equal [28]. Thus, in chemically driven molecular machines the evolution of the system toward the equilibrium of the chemical cycle moves the mechanical cycle [28].

There are two views of how molecular machines couple a chemical cycle with a mechanical cycle, termed power stroke [47] and Brownian ratchet [48], with the main differences lying in the generation of the force for directed motion [35]. In a pure power stroke mechanism [35, 47], the chemical cycle is directly responsible for performing mechanical work. The chemical energy is therefore converted into potential energy of a "tight spring" conformation and then released in the directed motion [35, 48]. In a pure Brownian ratchet mechanism, thermal fluctuations drive the motion and the (chemical) cycle is used to rectify the random thermal fluctuations into directed motion. The motions there are resulting only from thermal fluctuation and the (chemical) energy is used to block the backward fluctuation [46] and to capture 'favorable' conformations [49].

The molecular ratchet in a one-dimensional free energy profile shows a sawtooth-like character (Figure 8) with two variants: Fluctuating ratchets (also termed fluctuating potential) model and fluctuating force (also termed tilting ratchet) model. The fluctuation is assumed to be a result of weak and strong, or attached and detached, binding states in both fluctuation ratchets and fluctuating force models [28]. In the fluctuation ratchet model (Figure 8A) the free energy profile alternates between a sawtooth-like and a flat profile [28, 50]. In the fluctuating force model (Figure 8B), the sawtooth-like potential is tilted up and down during the fluctuations, and as such, this later type of ratchet share characteristics with a powerstroke driven molecular motor [28].
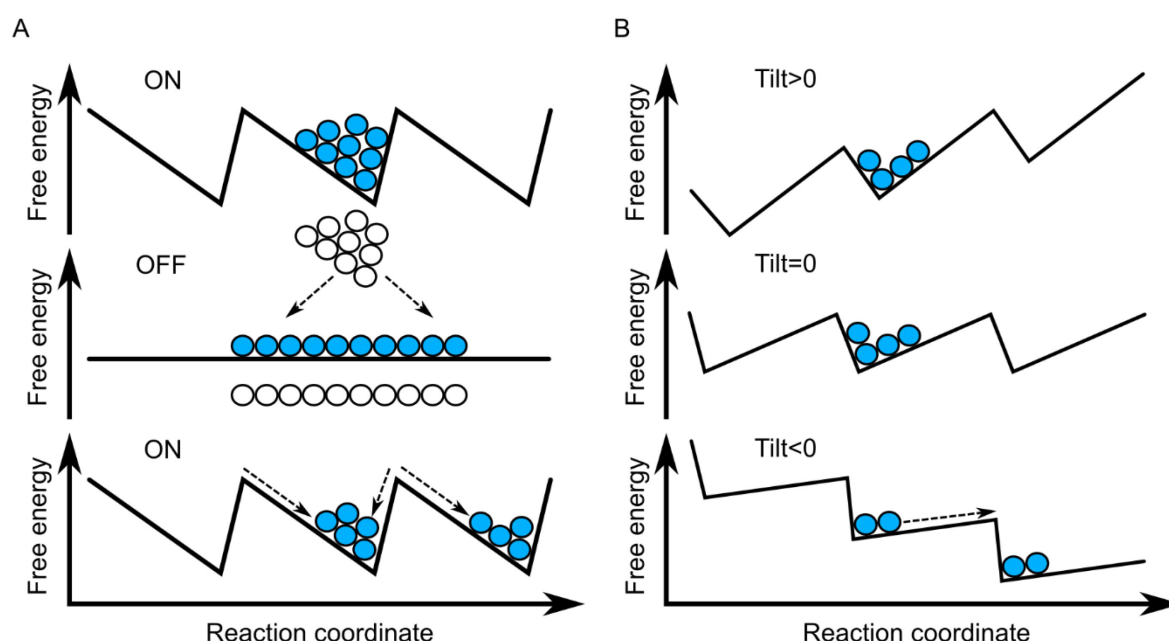
Figure 8: Sawtooth-like energy profiles of fluctuation potential and fluctuating force ratchet accomplishing net transport from left to right. **(A)** Fluctuating ratchet alternates between sawtooth-like potential (ON state) and flat potential (OFF state), with the ON state **(top)** trapping particles in potential wells, the OFF state **(middle)** resulting in free diffusion of the particles, and the ON state **(bottom)** transporting parts of the particle to the next potential well and part back to their original position. **(B)** Tilting ratchet alternates the tilt of the ratchet potential between positive and negative directions, where the shape of the potential enables right-ward motion and limits left-ward motion, adapted from [51].

Although it is instructive to regard power strokes and Brownian ratchets as separate mechanisms, the boundaries between thermal and power stroke driven mechanisms are rather flexible [28, 35, 52] and most proteins employ a combination of both strategies [49]. A rough but useful distinction between power strokes and Brownian ratchets mechanisms can be given by the height of the separating barrier [49]: Thermal fluctuation trapped by rectifying nanometer-size or larger diffusional displacements ('big ratchets') with barrier heights much larger than $k_B T$ indicates for the predominance of a Brownian ratchets mechanism, while a sequence of small angstrom-sized steps ('small ratchets'), with barrier heights comparable to or smaller than $k_B T$ [49], put the mechanism closer to a power strokes [49].

1.5 Methods to analyze the dynamics of proteins

"I suppose it is tempting,

if the only tool you have is a hammer, to treat everything as if it were a nail."

Abraham H. Maslow, The Psychology of Science, 1966 [53].

The dimension and time scale involved in domain motion makes this topic difficult to study both experimentally and computationally [17]. As several experimental and computational methods are available, which have their field of applications and limitations, the search for understanding the mode of action of a protein often requires the combination, independent validation, and interpretation of the available information from several methods. A detailed description of all methods to analyze proteins dynamics would go beyond the scope of this thesis. Herein, I provide a short overview of the methods of X-ray crystallography whose key information serves as foundation of my work, and of the computational methods, which are crucial to understand my thesis. For a broader view on the methods to analyze the structure, dynamics, and function relationship, I would refer the reader to the excellent reviews from Adcock and McCammon [12], Henzler-Wildman and Kern [14], and Gran *et al.* [54].

1.5.1 Protein structure determination with X-ray crystallography

Experimentally determined three-dimensional structures of proteins represent an important tool to get insights into the function and construction of proteins, and serve as starting points for further computational methods. Methods to obtain structure in atomic resolution include X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM). X-ray crystallography is the main source of protein structures as ~89% of all resolved structures in the PDB [55] and all complete PPDK structures were obtained by this method.

A large part of our knowledge regarding the mechanism of domain motion has come from comparing two (or more) different conformations derived from X-ray crystallography [23]. For this technique, conditions have to be found in which a purified sample of protein crystallizes [56]. This is the rate limiting step and mostly determines whether or not the structural determination with X-ray crystallography will be possible [56]. Obtained protein crystals are exposed to an X-ray beam, resulting in diffraction patterns, which are then processed to calculate a three-dimensional electron density map [56]. Different approaches are used to

determine the phases [56]. At the end, the three-dimensional structure of the protein is modeled into the electron density and refined [56]. Since X-ray crystallography provides mainly static information, it is often complemented by techniques that provide information on the dynamics such as Förster resonance energy transfer (FRET), NMR or MD simulations.

1.5.2 Molecular dynamics simulations

MD simulations are a widely used computational technique and often the method of choice to obtain fine details concerning the motion of atoms as a function of time. Since the first MD simulation of proteins in 1977 [10], this technique developed into a robust and trusted tool, which now is widely and routinely applied to gain insights into the dynamics of proteins [12]. MD simulations have the unbeatable edge in that it can describe protein dynamics on an atomic level: the precise position of each atom at any instant in time for a single protein molecule can be followed, along with the corresponding energies [14, 57].

In MD simulations, a potential energy function, called a force field, is used to derive the forces acting on each atom of a given structure, followed by numerical solution of Newton's equation of motion, to derive the time evolution of the system [6, 12]. However, despite the massive increase in computational performance over the last decades and the fast improvement of MD simulation techniques, the amounts of computational resources is still a factor that limits simulations of biomolecules, both in time and system size [57, 58]. To study biologically relevant processes beyond the restriction of conventional MD simulation, various sophisticated techniques have been devised.

1.5.3 Potential of mean force computation using the umbrella sampling method

As many biological processes are controlled by alterations in rate and relative populations rather than by a simple 'on-off' switch [14], knowledge of the energetics increases the understanding of the involved processes. A free energy profile of a process of interest can be obtained by the potential of mean force (PMF, [59]) method, which describes the free energy F as a function of a reaction coordinate ξ according to (eq. 4) [60].

$$F(\xi) = -k_B T \ln[P(\xi)] + C \qquad \text{(eq. 4)}$$

Where k_B is the Boltzmann factor, T the temperature, $P(\xi)$ the probability of the system for the respective reaction coordinate values, and C a constant used for normalization. The reaction coordinate may be one- or more-dimensional and is usually defined on geometrical grounds such as an angle, or a distance [61, 62].

While the free energy could in principle be computed directly from the probability distribution derived from one sufficiently long MD simulation, this approach is impractical due the lack of conventional MD simulations to efficiently overcome and sample high energy barriers of several k_BT, which in most cases prevents accurate sampling of the whole reaction coordinate within the available computational time [63].

The umbrella sampling method [64] is commonly used as an efficient approach to sample conformations along a reaction coordinate by introducing an additional energy term $w(\xi)$ to the force field [62, 64]. In the "conventional" variant, this additional energy term $w(\xi)$ has the form of harmonic restraint potentials (Figure 9A), tethering the conformations with a force constant K to their respective reaction coordinate value ξ_0 (eq. 5):

$$w(\xi) = \frac{1}{2}K(\xi-\xi_0)^2 \qquad \text{(eq. 5)}$$

This confines the simulation to a small interval around the reaction coordinate value ξ_0, helping to sample this region efficiently (Figure 9B) [61]. Performing a series of individual MD simulations, restraining each of these windows to different ξ_0 values, results in overlapping distributions, which sample the whole range of the reaction coordinate (Figure 9B) [62].

The results of the umbrella sampling have then to be unbiased and combined. For this, the Weighted Histogram Analysis Method (WHAM) [65] has been shown to be the most reliable approach [61], in comparison with the weighted distribution function (W-DF) method [66], and the weighted potential of mean force (W-PMF) method [67, 68], providing comparable results as the more computationally expensive free energy perturbations (FEP) [69] approach [61]. In WHAM, the biasing potential introduced in (eq. 5) is subsequently removed to obtain unbiased distributions for each window (eq. 6) [64, 70]:

$$P(\xi)_i^{unbiased} = P(\xi)_i^{biased} \, exp\left(-\frac{w_i(\xi)}{k_BT}\right) \left\langle exp\left(\frac{w_i(\xi)}{k_BT}\right)\right\rangle \qquad \text{(eq. 6)}$$

where $<\,>$ denotes the ensemble average.

The unbiased potential $P(\xi)_i^{unbiased}$ of the individual windows are recombined by WHAM to extract the global distribution $P(\xi)$ (Figure 9C) by calculating a weighted average of the distributions of the individual windows (eq. 7) [62].

$$P(\xi)^{unbiased} = \sum_{i=1}^{windows} P(\xi)_i^{unbiased} \, p_i(\xi) \qquad \text{(eq. 7)}$$

With the weight $p_i(\xi)$ fulfilling the condition $\sum p_i(\xi) = 1$ and calculated via (eq. 8).

$$p_i(\xi) = N_i \exp\left(-\frac{w_i(\xi) + F_i}{k_b T}\right) \qquad \text{(eq. 8)}$$

With the free energy constant F_i being calculated by (eq. 9).

$$exp\left(-\frac{F_i}{k_b T}\right) = \int P(\xi)^{unbiased} \, exp\left(-\frac{w_i(\xi)}{k_b T}\right) d\xi \qquad \text{(eq. 9)}$$

Because the global (unbiased) distribution $P(\xi)^{unbiased}$ enters (eq. 9) and the free energy constant F_i enters (eq. 7) via (eq. 8), these equations have to be solved iteratively until convergence is reached [62].
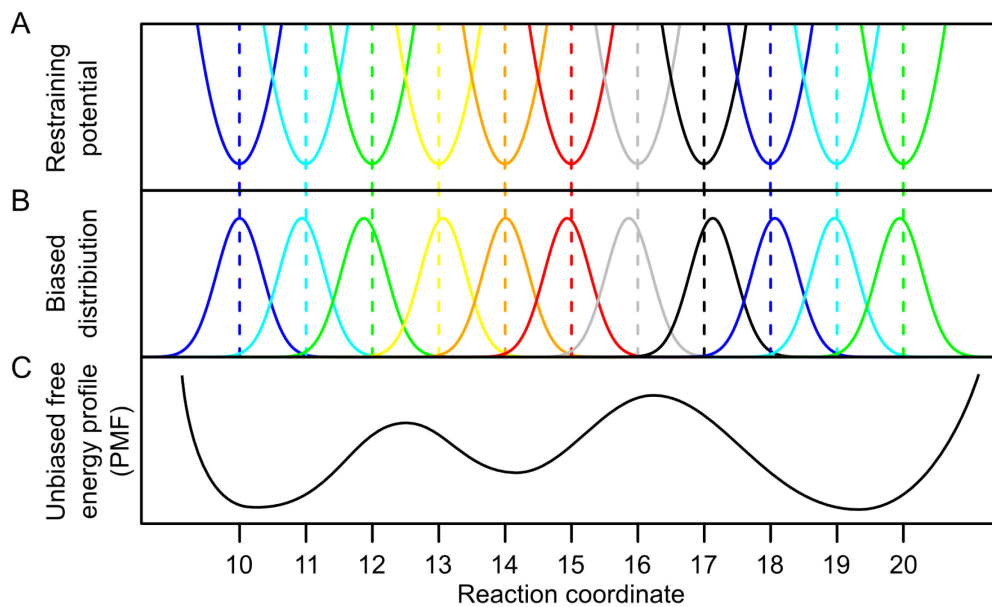
Figure 9: Umbrella sampling method of potential of mean force computation. (A) Harmonic umbrella sampling potentials are used to restrain MD simulations within small intervals of the reaction coordinate (windows, each color represents a window). **(B)** Performing restrained MD simulations in series results in overlapping distributions sampling over the whole range of the reaction coordinate (with each color representing one MD simulation). **(C)** The free energy profile (PMF) can then be obtained by unbiasing and recombining the overlapping distributions using WHAM.

An alternative approach termed adaptive bias umbrella sampling [71] covers the whole range of the reaction coordinate in one simulation by iteratively adapting the bias w' to fit the underlying energy landscape $w'(\xi) = -F(\xi)$, which flattens the effective energy surface of the restrained MD simulations and allows uniform sampling along the reaction coordinate. The choice between "conventional" umbrella and adaptive biases sampling represents a computational tradeoff as the simulations of a large number of windows in case of "conventional" umbrella sampling can be done in parallel and, therefore, be more CPU-time efficient but, on the contrary, the CPU time needed for equilibration increases with the number of windows [62].

1.5.4 Structural stability investigation with the Constraint Network Analysis

Constraint Network Analysis (CNA) is a graph theory-based rigidity analysis that links biomolecular structure, (thermo-)stability, and function [72, 73] [2]. In the CNA approach, a protein is represented as a constraint network with bodies, representing atoms, connected by sets of bars, constraints, representing covalent and noncovalent interactions [74] [2]. A rigid cluster decomposition of the constraint network then identifies rigid parts that are connected by flexible links [2]. By gradually removing noncovalent constraints from an initial network representation of a biomolecule, a succession of network states is generated that yields a 'constraint dilution trajectory' [75, 76] [2]. Analyzing such a trajectory via rigidity analysis reveals a hierarchy of rigidity that reflects the modular structure of biomolecules in terms of secondary, tertiary, and supertertiary structure [2].

2. Background

2.1 The pyruvate phosphate dikinase (PPDK)

The enzyme pyruvate phosphate dikinase (PPDK, EC 2.7.9.1) catalyzes the ATP and orthophosphate (P_i) dependent reverse conversion of pyruvate to phosphoenolpyruvate (PEP), pyrophosphate (PP_i), and AMP according to the following reaction (1):

$$Pyruvate\ +\ ATP\ +\ P_i\ \rightleftharpoons\ PEP\ +\ AMP\ +\ PP_i \tag{1}$$
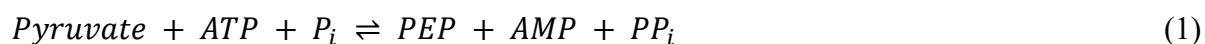
PPDK can be found in C_4-plants, where it catalyzes the regeneration of the primary CO_2 acceptor PEP during photosynthesis, and in low amount in C_3-plants, where the physiological role is largely unknown. The enzyme is also found in bacteria, archaea, and protists, where it catalyze the reverse ATP-forming reaction during gluconeogenesis [77]. The biologically active form of PPDK in bacteria is a dimer, and PPDK also appears as tetramer in several plants [1, 78-80]. The essential role of PPDK in the cellular energy metabolism of C_4-plants and human pathogenic protists combined with its absence in mammals make this enzyme an interesting target for the development of C_4 plant-specific herbicides, *e.g.*, against the weeds *Echinocloa crus-galli*, *Eleusine indiva* and *Elymus repens*, and anti-parasitic drugs, *e.g.*, against the protists *Entamoeba histolytica* or *Trypanosoma cruzi*. Because pyruvate and ATP are prominent substrates in other parts of the metabolism, substrate-analogs of pyruvate and ATP used as competitive inhibitors of PPDK will likely be non-selective and could proof fatal to hosts. This drawback would not occur if herbicides or anti-parasitic drugs allosterically inhibited this enzyme. To achieve this goal, understanding PPDK's molecular mechanisms is in this context of high interest.

The overall reaction (1) involves two partial reactions (2-3) in plants [81] and protists [82], with a phosphoryl transfer via a catalytic histidine residue.

$$Enzyme\ +\ ATP\ +\ P_i \rightleftharpoons\ Enzym-His-P\ +\ AMP\ +\ PP_i \tag{2}$$
$$Enzyme-His-P\ +\ Pyruvate \rightleftharpoons Enzyme\ +\ PEP \tag{3}$$
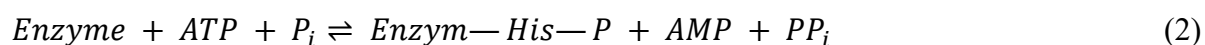
This temporal separation of reactions is mirrored by a spatial separation of the reaction centers: The structure of PPDK consists of three different domains, an N-terminal nucleotide-binding domain (NBD), which carries out partial reaction (2), a C-terminal PEP/pyruvate-binding domain (PBD), which carries out partial reaction (3), and a central domain (CD), which contains the catalytic histidine that acts as phosphoryl acceptor and donor (see Figure 10).
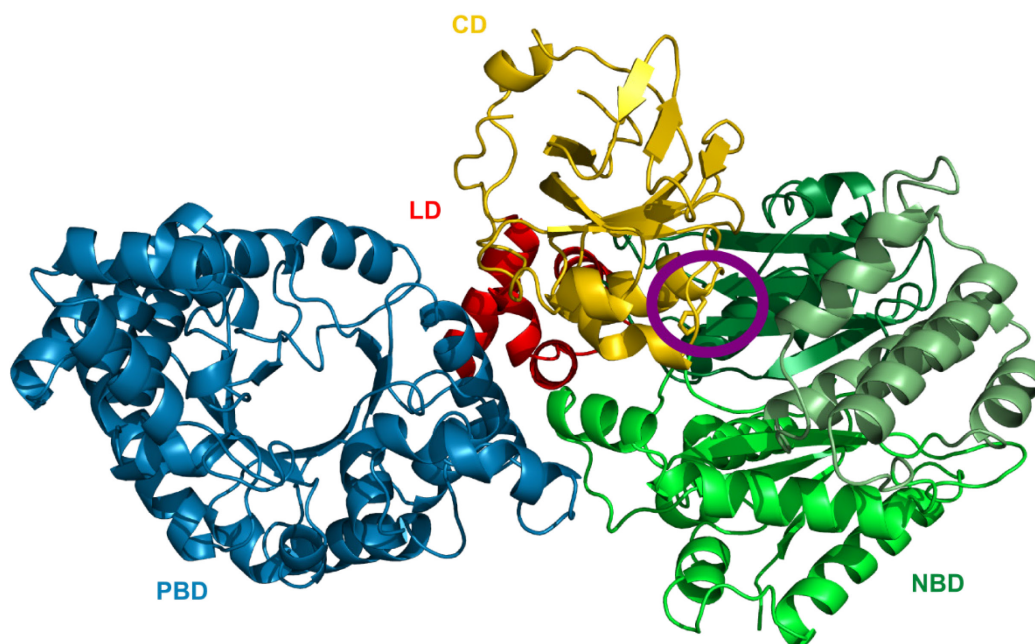
Figure 10: The structure of PPDK Cartoon representation of the first resolved PPDK structure (PDB ID: 1DIK [83]). The central domain (CD, yellow) contains the catalytic histidine (circled in deep purple). At the nucleotide-binding domain (NBD, with the three subdomains depicted by three different greens) and the PEP/pyruvate-binding domain (PBD, colored blue), where partial reactions (2) and (3) take place. The linker domains (LD, colored red) connects the CD, PBD, and NBD.

By the start of my thesis, ten crystal structures of PPDK have been resolved, for PPDKs of the bacterium *Clostridium symbiosu*m, the protozoan *Trypanosoma brucei*, and the plants *Zea mays* (see Figure 11).
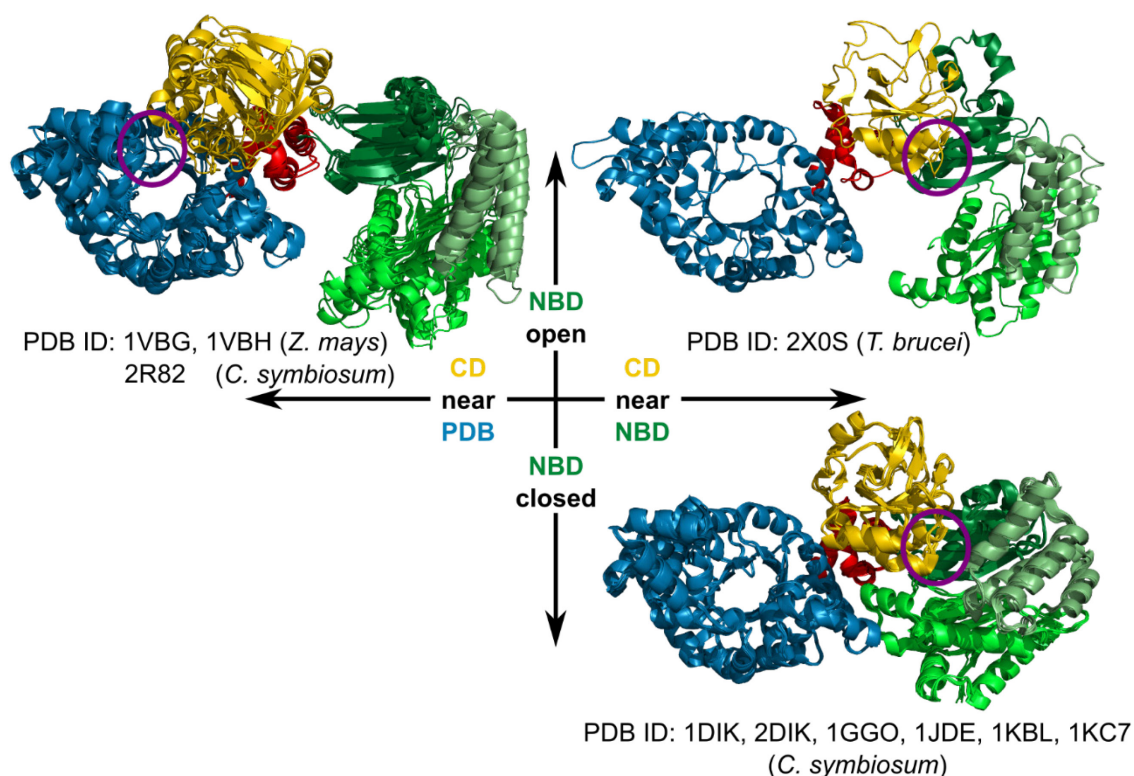
Figure 11: Overlay of the crystal structures of PPDK available at the beginning of the thesis, showing the CD in ether near the PBD (left), or near the NBD (right), and the NBD in an open (top) or closed (bottom) conformation. The position of the catalytic histidine is indicated by a purple circle. Adapted from [1, 2].

The overlay of the PPDK structures (see Figure 11) suggests two principal modes of PPDK movement: First, the CD is either located near the PBD, or near the NBD. Accordingly, the CD needs to undergo a swiveling motion to shuttle the phosphoryl group between the two active sites, followed by a backward motion to initiate the next cycle of phosphoryl transfer. This swiveling motion involves a rotation of ~110° of the CD, which transports the catalytic histidine across a distance of ~45 Å. This transporting motion is one of the largest found in enzymes. Second, the NBD, which consists of three subdomains forming the ATP-grasp motif, shows an opening-closing motion, which is assumed to be associated with ATP binding.

Research of almost 50 years shaped our understanding of PPDKs mode of action and led to several models over time, which aim to explain all observations (Figure 13). Insights into the enzyme mechanism, domain organization, and conformational preferences have led to the current "swiveling domain model" [83], which underlines the importance of the central domain movement (Figure 13).

However, a final conclusion of PPDK's mode of action remains elusive as not all observations can be explained with the current model, including the interaction of the opening-closing motion of the NBD motion with the swiveling motion of the CD and the requirement of PPDK to form a dimer (Figure 12) to be active.
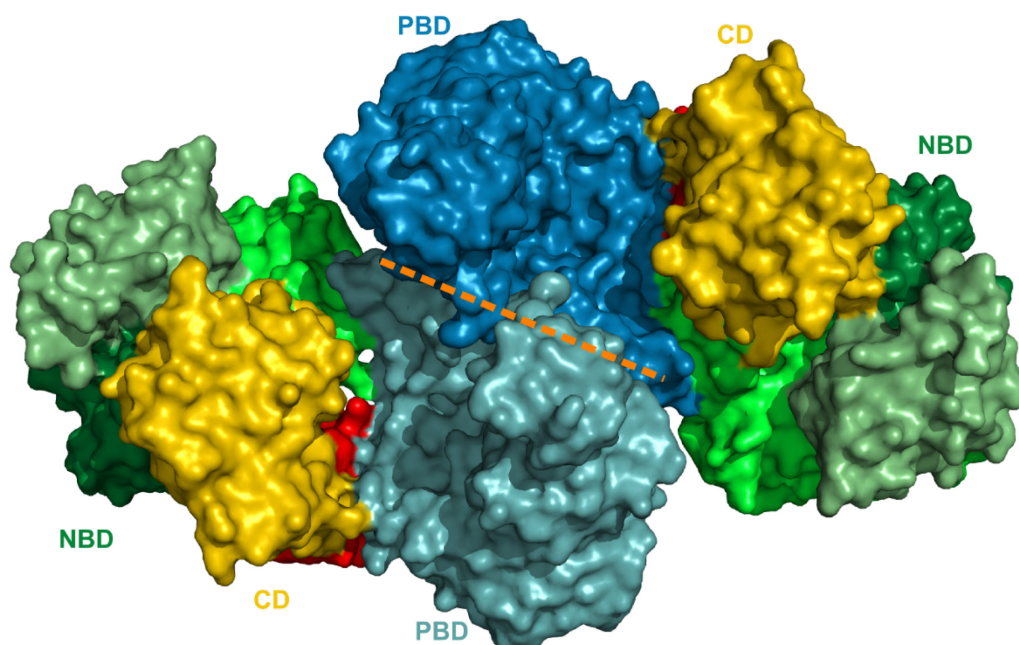
Figure 12: Structure of the PPDK dimer, proposed based on crystal packing of PDB ID 1DIK [83], with the dimer interface formed by the PBDs (indicated by the orange dashed line). Figure adapted from [83].
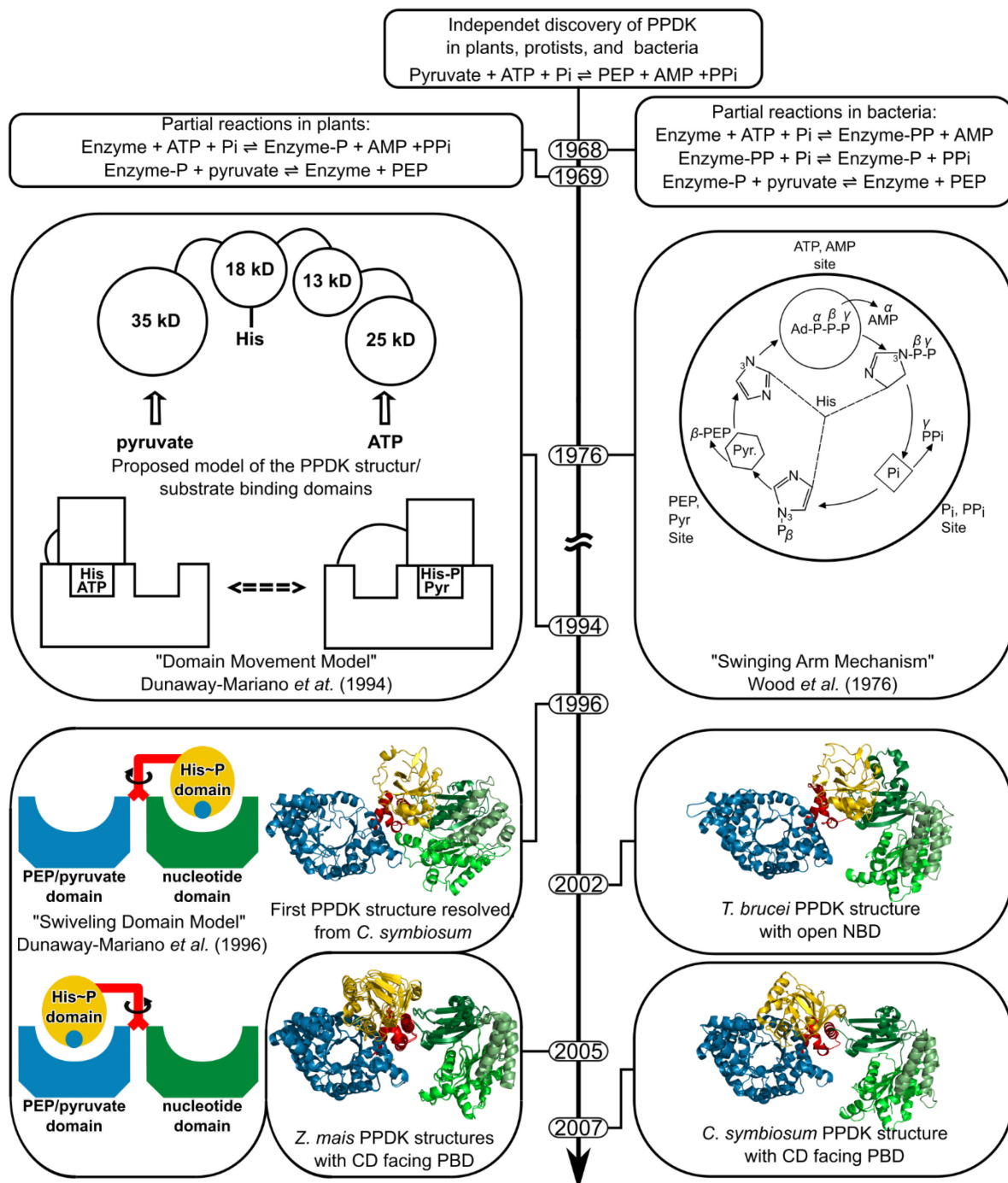
Figure 13: Historical timeline of understanding PPDK mechanism. (1968) PPDK was discovered independently by three groups with respect to its ability to catalysis the ATP and P_i-dependent conversion of pyruvate to phosphoenolpyruvate (PEP) in plants, protists, and bacteria by Hatch and Slack [84], Reeves [85], and Evans and Wood [86], respectively. **(1968)** Further studies revealed the existence of three partial reactions in bacteria [86] and **(1969)** two partial reactions in plants [81]. **(1976)** Insights from reaction kinetics led to the first PPDK model of the "Swinging Arm Mechanism" involving a histidine residue that shuttles the phosphoryl group between different reaction sites [87]. **(1994)** The proteolysis of PPDK into catalytically active fragments allowed to form a model of the PPDK structure together with the "Domain Movement Model" of the mode of action [88]. **(1996)** The first three-dimensional PPDK structure was resolved with X-ray crystallography from *C. symbiosum* and led to an improved model, which characterizes the domain movement as swiveling motion and predicted a PPDK conformation with the CD near the PBD [83]. **(2002)** A X-ray structure resolved from *T. brucei* provided a conformation of the PPDK with the CD facing the NBD and the NBD in an open conformation [89]. **(2005)** A PPDK structure from *Z. mays* [90] was also resolved in the conformation predicted from the "Swinging Domain Model". **(2007)** A PPDK structure in the predicted conformation were resolved from the same organism as the first PPDK structure (*C. symbiosum*), by including repulsive mutations between CD and NBD [91]. Figures adapted from [83, 87, 88].

2.2 The PPDK structures from *Flaveria trinervia* and *Flaveria pringlei*

In parallel to this work, the crystal structure of PPDK from *Flaveria trinervia* and *Flaveria pringlei* have been resolved by Minges, Winkler, Höppner, and Groth, providing insight into the ATP binding mode, exposing for the first time the dimeric form of PPDK in the asymmetric unit and resolving a structural intermediate of the swiveling motion. Their results were published in our paper Minges and Ciupka *et al.* [1], and are reproduced verbatim in subchapter 2.2:
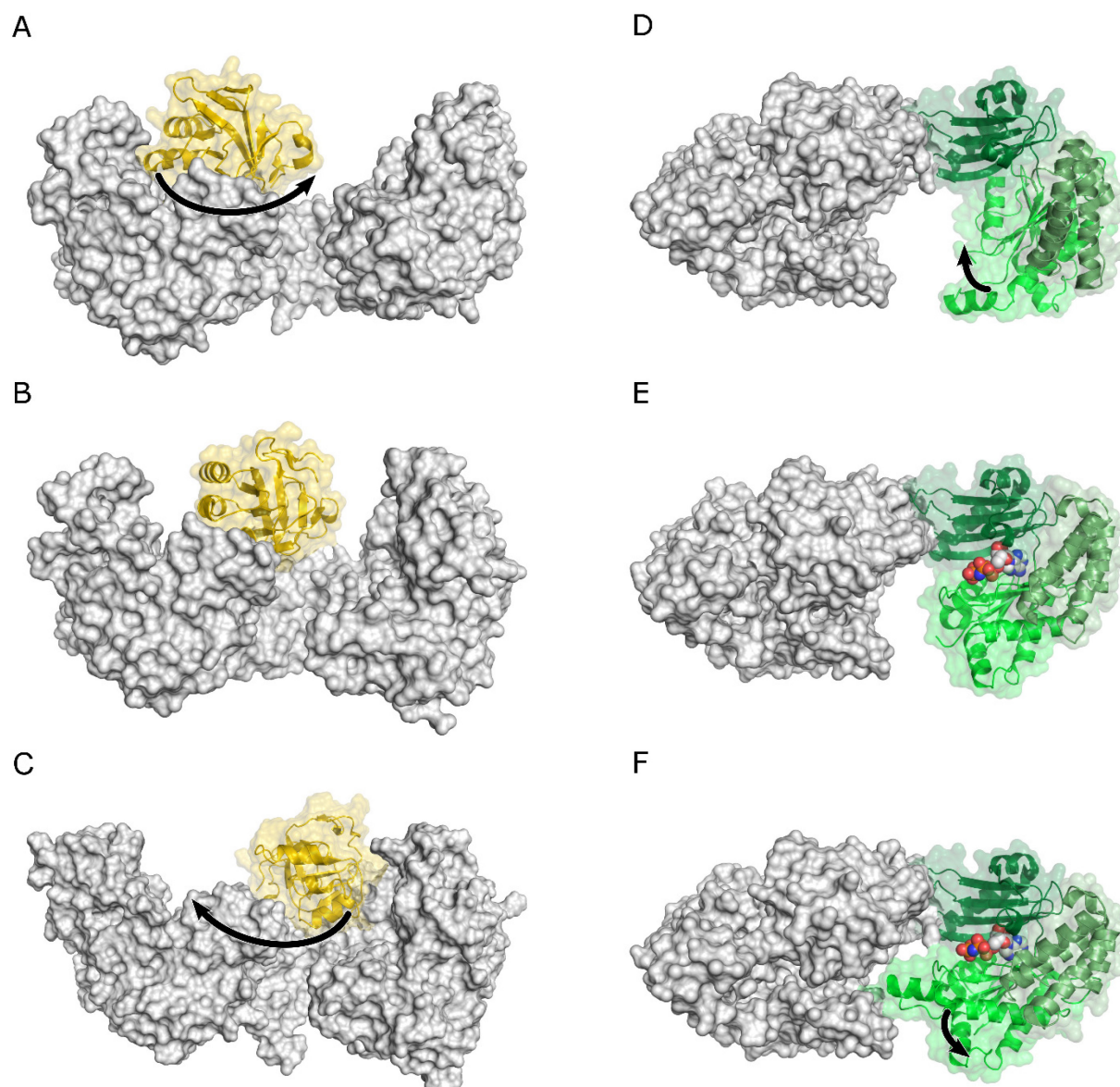
Figure 14: Movement of the PPDK domains. (A–C) Movement of the CD (yellow). **(A)** *Ft*PPDK in PBD-facing conformation (PDB ID 5JVL/C). **(B)** *Fp*PPDK (PDB ID 5JVN) with CD in an intermediate position. **(C)** *Tb*PPDK (PDB ID 2X0S) [89] with CD in NBD-facing conformation. **(D-F)** Movement of the NBD (the three subdomains are depicted by three different greens). **(D)** *Ft*PPDK in nucleotide-unbound state (PDB ID 5JVJ/A). **(E)** *Ft*PPDK in semi-closed, nucleotide-bound state (PDB ID 5JVL/C). **(F)** *Ft*PPDK in fully closed, nucleotide-bound state (PDB ID 5JVL/A). Taken from [1].

Overall structure of *Flaveria* PPDKs.

The structure of the C4-isoform of PPDK from the flowering plant *Flaveria trinervia* was determined by molecular replacement at 2.9 Å resolution using the maize structure (PDB ID 1VBH) [90] as a template [1]. The structure (PDB ID 5JVJ) includes two monomers in the asymmetric unit (ASU) forming a dimer that corresponds to the previously described biological assembly of bacterial and maize PPDK [83, 90] with an overall well-defined electron density for the entire monomer A and for the PBD of monomer B [1]. Parts of the NBD of monomer B revealed only poorly defined electron density, and direct tracing of monomer B in these regions was hampered [1]. Yet, both monomers show electron density in the PBD for the co-crystallized substrate PEP [1]. Besides, the NBD of monomer B exhibits additional density in both the $mF_\mathrm{o} - DF_c$ difference map and the feature enhanced maps (FEM) probably reflecting a bound adenine nucleotide [1]. The overall shape of this additional density is consistent with structural requirements and binding mode of adenine nucleotides in other nucleotide-binding proteins with the ATP-grasp fold [92, 93] [1]. In addition, this density complies with those observed when PPDK was crystallized in the presence of the nucleotide analog 2'-Br-dAppNHp (see PDB ID 5JVL and Figure 15C) [1]. However, since the molecular identity of the bound compound was not fully resolved at the present resolution, no compound was placed in this density in the deposited structure [1]. Large parts of monomer B were successfully built using monomer A as a template by iterative manual model building and refinement [1]. Yet, no conclusive electron density was found for residues 18-22, 47-65, 83-87, 101-106, 120-124, 163-166, 192-198 and 216-236 [1]. An overall root-mean-square deviation (RMSD) of 4.8 Å was calculated from a structural alignment of the individual monomers of the PPDK dimer in 5JVJ, indicating a substantial difference in their conformation [1]. The main difference is found in the NBD of both monomers with the A monomer reflecting an open conformation and the B monomer reflecting a closed conformation of this domain [1]. Overall, the orientation of NBD1 (aa 1-111 and 197-243) and NBD2 (aa 112-196) relative to NBD3 (aa 244-340) is greatly changed in the two monomer conformations (Figure 14D-F) [1]. Superimposition of the NBD subdomains highlights that NBD1 is reoriented by a large motion of about 40° around a hinge region consisting of two short peptide linkers formed by residues 112-115 and 195-200 towards NBD3 [1]. At the same time, NBD2 is displaced by about 40° to accommodate for the new position of NBD1 [1]. The hinge motion of NBD1 and NBD2 results in the closing of the large cleft formed in the open configuration of the PPDK monomer (Figure 14D-F) [1]. This cleft is no longer accessible in the closed form of the PPDK monomer [1]. The presence of two structurally distinct conformations of the PPDK monomer (NBD open versus closed) within a

single crystal structure may suggest that these structural asymmetries reflect functional asymmetries in substrate binding and/or catalytic turnover in the individual subunits of the PPDK dimer [1]. A similar structural asymmetry in the monomer arrangement (open versus closed) was observed in the crystal structure of bacterial ATP-dependent DNA helicases and related to functional asymmetry in the mechanism of ATP hydrolysis in solution in each of the two subunits forming the functional dimer [94, 95] [1]. When 2'-Br-dAppNHp was used for co-crystallization (PDB ID 5JVL), the crystal contains four monomers in the ASU arranged in terms of two dimers [1]. The dimers correspond to the subunit arrangement in 5JVJ [1]. Electron density maps clearly indicate main chain connectivity and most of the side chain orientations for subunits A, C and D, whereas the majority of the NBD in chain B is not traceable [1]. The overall structures of both *Ft*PPDK crystal forms (Figure 14A,D,E,F) are highly similar to PPDKs from *C. symbiosum*, *T. brucei* and *Z. mays* [83, 89, 90] consisting of an N-terminal NBD (aa 1-340), a CD (aa 380-515) of β-strands with associated α-helices that contain the catalytic H456 in helix 20 involved in the phosphoryl group transfer between NBD and PBD, and a C-terminal TIM barrel containing the PBD (aa 535-874) [1].
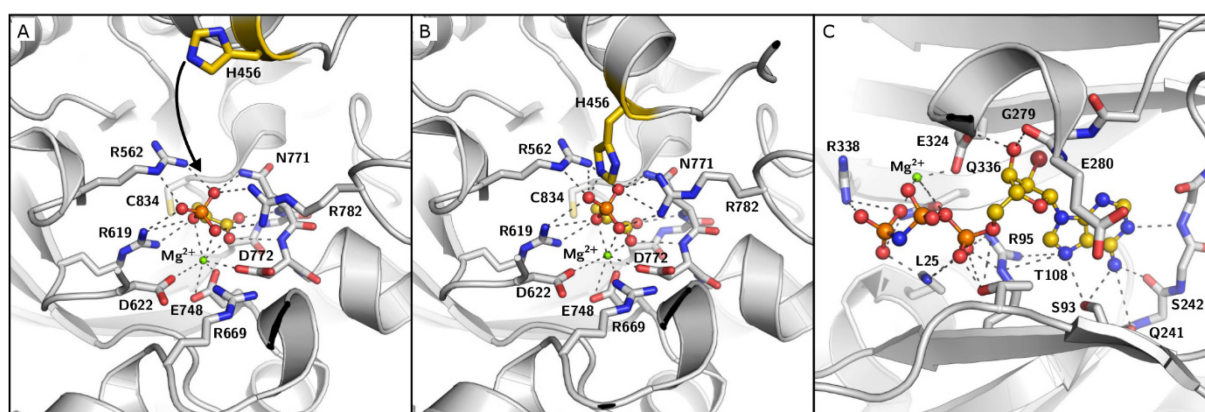
Figure 15: Substrate binding sites of *Ft*PPDK. **(A)** Semi-closed state of the PEP binding site (PDB 5JVL/A) with the catalytic H456 (yellow) pointing away from PEP. **(B)** Closed state of the PEP binding site (PDB 5JVL/C) revealing tight interactions between PEP and surrounding residues, including the catalytic H456 (yellow). **(C)** Closed state of the nucleotide-binding site of 5JVL/D occupied with 2'-Br-dAppNHp. Interacting residues are highlighted. Taken from [1].

For monomers C and D of 5JVL, the catalytic H456 is located in close position and appropriate orientation to the bound PEP substrate to mediate phosphoryl group transfer from the NBD to the PBD (Figure 15A,B) [1]. Such a close contact has never been observed in previously reported PPDK structures, but has been resolved in Enzyme I of the E. coli Phosphoenolpyruvate:Sugar Phosphotransferase System (PTS), a bacterial carbohydrate import system bearing a homologous PEP binding domain [96] [1]. For structural comparison of the catalytic sites in PTS Enzyme I and the PPDK a set of reference atoms was selected [1]. While the catalytic histidine in the PTS Enzyme I structure is in the phosphorylated state the

corresponding histidine in the active site of PPDK structure 5JVL is in the non-phosphorylated form [1]. On the contrary, PPDK substrate PEP carries a phosphoryl group, whereas PTS Enzyme I substrate inhibitor oxalate does not [1]. Hence, atoms Nϵ2 of the catalytic histidine oxygen atoms O2 in oxalate and PEP in 5JVL were used for distance measurements and structural comparison of both enzymes at their PEP binding site. In the PTS Enzyme I structure, the distance between the phosphorylated Nϵ2 atom of the catalytic H189 and the oxygen atom O2 of the oxalate substrate is 4.9 Å [1]. The distance between Nϵ2 of the catalytic H456 and O2 of the PEP substrate in chain C and D of the 2'-Br-dAppNHp-*Ft*PPDK structure is similar (4.5 Å) (Figure 15B), which emphasizes the close structural relationship of the newly described extreme conformation observed in 5JVL chains C and D and the PTS Enzyme I structure [1]. This close spatial arrangement of the catalytic histidine and the phosphoryl group substrate in the PEP binding site enables efficient phosphoryl group transfer from PEP to H456 or H456 to pyruvate in PPDK as for PTS Enzyme I [1]. The distance between the nucleophilic nitrogen atom Nϵ2 of H456 in the 2'-Br-dAppNHp-FtPPDK structure and the attacked phosphorus of PEP is around 3 Å (Figure 15B) [1]. This relatively short distance is indicative for an associative in-line displacement of the phosphoric monoester [97, 98] [1]. Therefore, the structural conformations resolved in the 2'-Br-dAppNHp-*Ft*PPDK structure visualize and verify for the first time the proposed phosphoryl group transfer mechanism in the PBD and the involvement of the catalytic histidine in the CD in this process [1]. Considering the overall CD, only small conformational changes between the plant structures from maize and *Flaveria* presented here and the non-plant structures from *Clostridium* and *Trypanosoma* are evident indicating that this domain primarily rotates as a rigid body [1]. However, in contrast to the crystal structures of the non-plant PPDKs where the CD is close to the NBD, the CD of *Ft*PPDK rests alongside the PBD (Figure 14A) as observed for PPDK from maize [90] and the *Clostridium* triple mutant R219E/E271R/S262D (PDB ID 2R82) [91] [1]. The different conformations of the NBD observed in the *Flaveria* structures presented in this work cover the whole range of conformational intermediates observed with other members of the ATP-grasp family such as synapsin [99] or biotin carboxylase [100] and reflect individual snapshots of the closing motion of this domain [1]. These structural intermediates of PPDK have been previously proposed based on structures of homologous nucleotide-binding enzymes [101] but have not been experimentally verified until now [1].

Structural intermediate of the swiveling motion.

In contrast to the C_4 plant crystal structures from *F. trinervia* and maize, the structure of the C_3-isoform of PPDK from *F. pringlei* (Figure 14B) contains only one monomer in the ASU (PDB ID 5JVN) [1]. Remarkably, in this structure the CD is positioned in an intermediate state between the NBD and PBD, thereby enlarging the solvent accessibility of the PEP binding pocket [1]. The CD position is not enforced by crystal contacts and therefore represents a genuine structural intermediate of the proposed swiveling mechanism [1]. The entire domain is rather mobile resulting in higher B-factors in this region [1]. Consequently, the electron density of several side chains in this region is not well defined [1]. Concerning the substrate and cofactor binding site, density is visible for both ligands, PEP and 2'-Br-dAppNHp [1]. Additional density in the conformational intermediate was observed in close proximity of the phosphoryl group of the PEP substrate [1]. This additional density might reflect an alternative position of the PEP substrate in the PBD [1]. However, the available resolution and data quality prevent modeling of the PEP substrate in those positions in the active site [1]. Compared to the near-PBD positioned conformation of the CD in the C_4-isoform (PDB ID 5JVL/C and 5JVL/D) the entire CD in the *Fp*PPDK structure is rotated by 45° towards the NBD around a rotation axis formed by the linker region (Figure 16) [1]. When compared to the PPDK structure from *Clostridium symbiosum* (*Cs*PPDK, PDB ID 1KBL) [83], the CD is rotated in the opposite direction and displaced by 52° towards the PBD [1]. Focusing on the catalytic H456 in the CD, this residue is displaced by 17 Å (C_α-distance) when the *F. pringlei* conformational intermediate (5JVN) and the PBD-facing *Ft*PPDK structures (*e.g.* 5JVL/C) are aligned, and by 24 Å in the opposite direction when the NBD-facing conformation observed in the *Cs*PPDK is compared to the *Fp*PPDK structure, respectively [1]. Hence, compared to the two extreme conformations of the CD next to the NBD or PBD previously resolved, the CD in the *F. pringlei* conformational intermediate shows about 50% of the rotational and translational movement of the proposed complete swiveling motion [1]. The CD intermediate identified in the crystallographic studies suggests that the swiveling motion of this domain in the catalytic cycle proceeds via at least one sub-step, starting from a position near either the PBD (Figure 14A) or the NBD (Figure 14C) via the intermediate conformation resolved in 5JVN to a position near the other substrate binding site [1]. Similarly, stepped movements have been described *e.g.* for the F1-ATPase rotary molecular motor [102-105] [1].
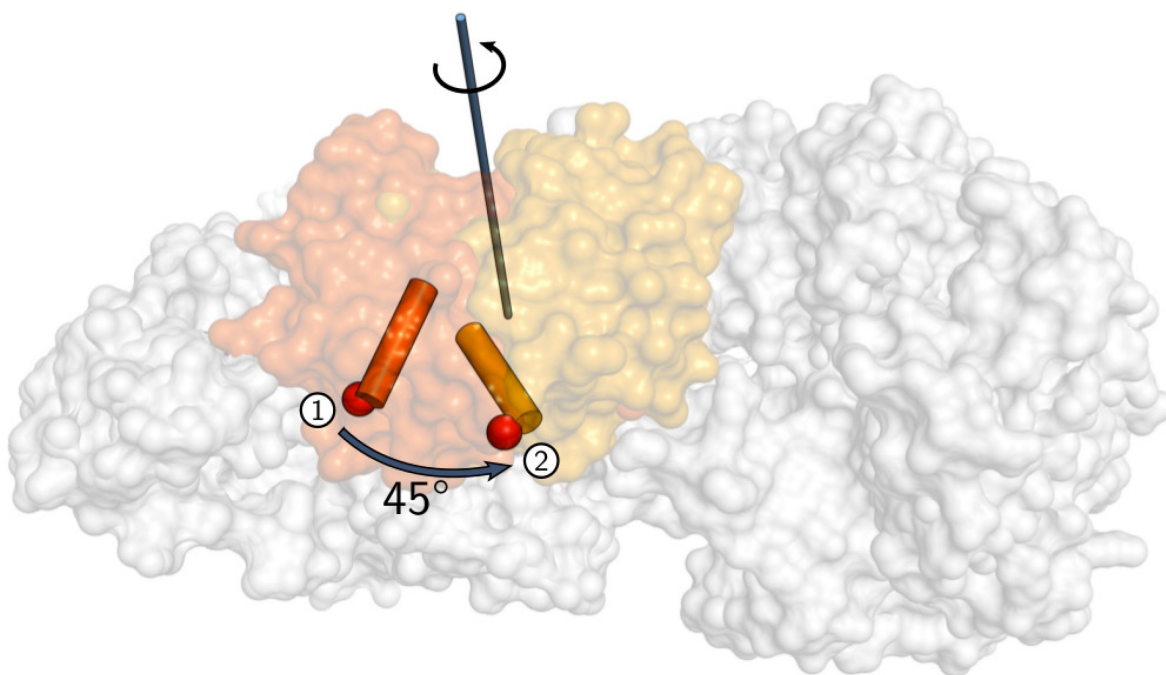
Figure 16: Stepped movement of the CD. CDs of 5JVL/C (1, orange) and 5JVN (2, yellow). Helix 20 containing the catalytic H456 is depicted as cylinder with the C_α atom of H456 shown as red sphere. The rotational axis for the transitions between states A and B is depicted as a blue arrow. The distance between the C_α atoms of the catalytic H456 is shown as a dashed line. Taken from [1].

3. Scope of this thesis

The knowledge of PPDK at the beginning of my thesis includes the separation of the reaction into two partial reactions taking place at distance reaction centers, the domain organization, and several conformations obtained by X-ray crystallography from which the swiveling domain motion model was derived (see Figure 17).

Reaction at the NBD

Enzyme + ATP + Pi ⇌ Enzyme-P + AMP +PPi

Reaction at the PBD

Enzyme-P + pyruvate ⇌ Enzyme + PEP
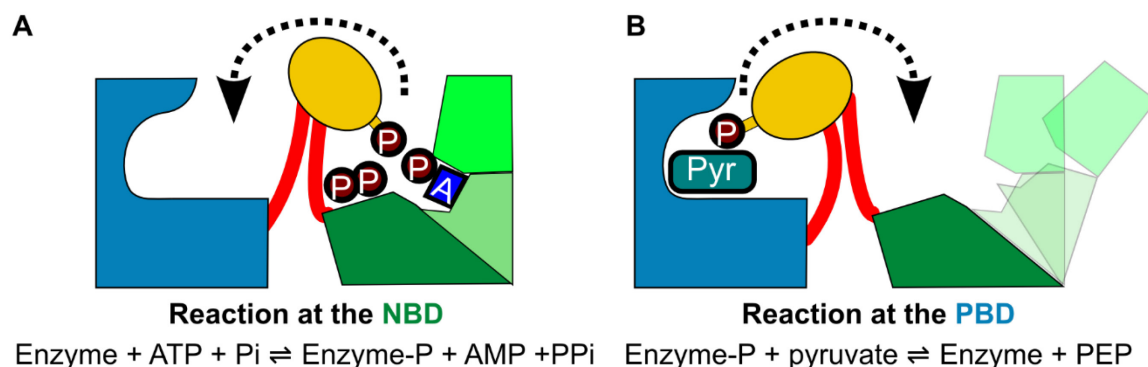
Figure 17: Model of PPDK mode of action at the beginning of this thesis. In the "swiveling domain model" from Herzberg *et al.* [83], a CD motion connects **(A)** the conformation where the partial reaction (2) at the NBD and **(B)** the conformation where the partial reaction (3) at the PBD take place. Adapted from [83, 106].

However, a comprehensive understanding of PPDK's mode of action was incomplete, as the following questions have remained unanswered:

(I)   If and how are the intramolecular motions (swiveling motion of the CD and opening-closing motion of the NBD) coupled?

(II)   Are there stable intermediate conformations of the swiveling motion?

(III)   How is the direction of the swiveling motion of the CD controlled? The CD has to move from near-NBD → near-PBD within the catalytic cycle and then move in the opposite direction from near-PBD → near-NBD to complete the cycle.

(IV)   What is the driving force of the swiveling motion of the CD?

(V)   Why is the dimer of PPDK active but not the monomer? What is the effect of the dimerization on PPDK motions?

(VI)   Are the opening-closing motions of the NBDs intermolecularly coupled?

To expand our knowledge about PPDK and find answers to these questions I investigated the energetics, dynamics and the connection of the proposed swiveling domain motion and opening-closing motions in the PPDK monomer and dimer using cluster analysis, principle component analysis, homology modeling, molecular dynamics (MD) stimulations, targeted normal mode-based geometric simulations, umbrella sampling / PMF computations, and rigidity analysis.

4. Materials and methods

4.1 Comparison of PPDK structures

All currently available structures of PPDK were obtained from the Protein Data Bank (PDB) [55] or provided by the Groth group prior to its release in the PDB (Table **2**) [1]. The structure with the PDB ID 5JVL, chain B and the NMR-derived structure of the CD only (PDB ID 2FM4) were excluded from further analysis as the NBD has not been resolved there [1]. A multiple sequence and structural alignment of the PPDK structures was generated with PROMALS3D [107] [1]. Residues resolved and common to all structures were identified from the alignment and used for further analysis [1].

4.2 Cluster analysis

Cluster analysis on all PPDK crystal structures listed in Table 2 was performed with the CPPTRAJ [108] module of the AMBER suite of programs [109] using the hierarchical agglomerative (bottom-up) algorithm [1]. As a distance measure, the best-fit Cα atom RMSD of all residues common in the PPDK structures with all domains resolved was used [1]. A maximal distance between all members of two clusters (complete linkage) of < 4 Å was used as terminating criterion for the clustering [1]. For subsequent analysis, only those PPDK structures were used that are cluster representatives in order to avoid a bias towards the number of crystal structures in the same conformational state [1].

4.3 Principle component analysis

The essential dynamics of PPDK was describe by performing principle component analysis in Cartesian space using CPPTRAJ [108] on a set of experimental structures and snapshots of MD simulations of the PPDK monomer and dimer with an aggregate length of ~10 μs and ~3.6 μs, respectively [1, 2]. In detail, an RMS-fit using the 15% least fluctuating residues was performed prior to the analysis to remove global translational and rotational motion [1, 2]. The coordinate covariance matrix was calculated for all Cα atoms [1]. The symmetric matrix was diagonalized by an orthogonal coordinate transformation, yielding the eigenvalues and eigenvectors (principle components) [1, 2]. An eigenvalue corresponds to the mean square eigenvector coordinate fluctuation (the variance) and, hence, describes how much a principal component contributes to the total coordinate fluctuations [110] [1, 2].

To analyze the locality or collectivity of motions for the domains of PPDK, the collectivity index κ described in refs. [111, 112] was calculated (eq. 10) [1, 2]:

$$\kappa = \frac{1}{N} \exp\left\{ -\sum_{i=1}^{N} \Delta\vec{r}_i^2 \log\Delta\vec{r}_i^2 \right\} \qquad (\text{eq. 10})$$

with N being the number of atoms in the domain, and $\Delta\vec{r}_i$ being the relative displacement of the principle component [1, 2]. All values of $\Delta\vec{r}_i$ were scaled consistently such that $\sum_{i=1}^{N} \Delta\vec{r}_i^2 = 1$. A value of $\kappa = 1$ indicates a mode of maximal collectivity, that is, all $\Delta\vec{r}_i$ are identical [1, 2]. Conversely, if only one atom is affected by the mode, κ reaches the minimal value of $1/N$ [1, 2].

4.4 Homology modeling of *Ft*PPDK conformational states

In cooperation with the Groth group, we work on the PPDK from *Flaveria trinervia*. As there was no crystal structure of *Ft*PPDK available at the beginning of this thesis, homology models of conformational states I, III, IV, and V were generated [1]. From each cluster (Figure 18), the crystal structures with the highest resolution and without mutations that interfere with the enzymatic activity were used as templates (*C. symbiosum* structure 1KBL and 1KC7 for state I; *Z. mays* structure 1VBH and 1VBG for state II; *F. pringlei* structure 5JVN for state IV; and *T. brucei* structure 2X0S for state V) [1]. The sequence alignment of the *Ft*PPDK sequence and the respective template sequence(s) was generated by MAFFT [113] [1]. The sequence identities are for conformational state I: 55%; conformational state III: 79%; conformational state IV: 96%; conformational state V: 54% [1]. The homology models were generated using the program MODELLER [114] in a multi-template approach, applying the dope loop model algorithm, and including ligands (if present in the crystal structure) [1]. The quality of our models was assessed for conformational state III, for which now PDB ID 5JVJ, chain A is available: RMSD$_{C\alpha}$ for NBD: 1.16 Å, for CD including the linker domain: 1.13 Å, and for the PBD: 0.93 Å [1]. Such structural deviations are close to the experimental uncertainty of the crystal structure [1].

4.5 Generation of all-atom dimeric PPDK structures

In the crystal structure 5JVJ, which exposed for the first time the PPDK dimer in the asymmetric unit, 86 residues (9 in chain A and 77 in chain B) out of in total 1748 residues have not been resolved [1, 2]. To have a complete structure of the PPDK dimer, a model containing all residues

was generated using the program MODELLER [114] [2]. For this, a sequence alignment was prepared considering the parts resolved in the crystal structure, and the missing parts by the template structure PDB ID 1VBG (NBD open) for chain A and the template structure PDB ID 5JVL (NBD closed) for chain B [2]. Additional models of the PPDK dimer with both NBDs closed, and with both NBDs open, were generated by aligning respective monomers of the 5JVJ structure to the residues of the dimer interface (see Table 7) [2].

4.6 Molecular dynamics simulations

The crystal structure of *Ft*PPDK 5JVL chain C (conformational state II) and homology models of monomeric *Ft*PPDK for conformational states I, III, IV and V as well as models of the PPDK dimer (open/open, open/closed, closed/closed, and ATP-bound closed/closed) served as input structures for MD simulations [1, 2]. Co-crystallized water was removed from the PPDK structures [1, 2]. Hydrogen atoms were added using REDUCE [115], flipping side chains of Asn, Gln, and His when appropriate [1, 2]. These model systems were placed in a truncated octahedral box of TIP3P water [116] leaving a distance of at least 11 Å between the solute and the border of the box [1, 2]. Counter ions were added to neutralize the systems [1, 2]. All MD simulations were performed with the ff99SB force field [117] using the AMBER suite of programs [109] and the GPU version of PMEMD [118] [1, 2]. Force field parameters for ATP were obtained from Meagher *et al.* [119]; force field parameters for Mg^{2+} were taken from Aqvist [120] [1, 2]. Bonds containing hydrogen atoms were constrained using the SHAKE algorithm [121], and long-range interactions were treated by the particle mesh Ewald (PME) method [122] [1, 2]. A time step of 2 fs was used [1, 2]. The system was equilibrated by, first, applying harmonic restraints to solute atom positions with force constants of at least $5\,\mathrm{kcal\,mol^{-1}\,\mathring{A}^{-2}}$ for 100 steps of steepest descent and 400 steps of conjugate gradient minimization [1, 2]. Second, the temperature of the system was raised from 100 K to 300 K in 50 ps of NVT-MD simulations [1, 2]. Third, 150 ps of NPT-MD simulations were performed to adjust the system density [1, 2]. Finally, the force constants of harmonic restraints were gradually reduced to zero during 250 ps of NVT-MD simulations [1, 2]. Production NVT-MD simulations were carried out at 300 K, using the Berendsen thermostat [123] and a coupling constant of 0.5 ps [1, 2]. Three independent replicates of MD simulations were performed for each system by spawning production runs after the thermalization at temperatures of 299.9 K, 300.0 K, and 300.1 K respectively [1, 2]. The first 2 ns of each trajectory were omitted from subsequent analyses. All unstrained MD simulations are listed in Table 3 [1, 2].

4.7 Generation of transition paths

For the potential of mean force calculations, plausible pathways of the swiveling motion and opening-closing motion have been generated using targeted normal mode-based geometric simulations by the NMSim approach [124, 125] [1, 2]. In detail, NMSim is a three-step protocol for multiscale modeling of protein conformational changes that incorporates information about preferred directions of protein motions into a geometric simulation algorithm [1, 2]. In the first step, the protein structure is coarse-grained by the software FIRST [74, 126] into rigid parts connected by flexible links [1, 2]. For this, an energy cut-off for including hydrogen bonds (and salt bridges) of -1 kcal mol^{-1} and a distance cutoff for including hydrophobic constraints of 0.35 Å were used [1, 2]. In the second step, low-frequency normal modes are computed by rigid cluster normal mode analysis (RCNMA) with a 10 Å distance cutoff for considering interactions between C_α-atoms [1, 2]. In the third step, a linear combination of the first 50 normal modes was used to bias backbone motions along the low-frequency normal modes, while the side chain motions were biased towards favored rotamer states, generating 100 conformations in 100 simulation cycles with a step size of 0.5 Å and side chain distortion of 0.3 [1, 2].

Targeted NMSim calculations for the PPDK monomer, were performed between start / end states of cluster III / cluster V, cluster III / cluster I, and cluster II / cluster I, using homology models for cluster III, V, and I, and the cluster-representative crystal structure for cluster II [1]. For the PPDK dimer, transition paths were generated by performing targeted NMSim calculations for five states, which vary in the conformation of chain B [1]. For chain A, cluster III (open NBD, CD facing PBD) was used, both, as start and end conformations for all systems, respectively [1]. The start and end conformations of chain B were in state a): cluster III (open and unbound NBD, CD facing PBD); in state b): cluster II (closed NBD, CD facing PBD); in state c) ATP-bound cluster II (closed NBD, CD facing PBD); in state d): cluster I (closed NBD, CD facing NBD) and in state e): ATP-bound cluster I (closed NBD, CD facing NBD) (see Figure 31) [2]. State c) resembles what has been found in PDB ID 5JVJ [1, 2].

4.8 Umbrella sampling / Potential of mean force computation

Free energy profiles of the swiveling motion of the CD and the opening-closing motion of the NBD were computed along the NMSim-generated transition paths by umbrella sampling [64] followed by the Weighted Histogram Analysis Method (WHAM) [65] [1, 2]. As a reaction coordinate for analyzing *Ft*PPDK's swiveling motion, the distance$_{CD-PBD}$ between H456$_{C\alpha}$ – H565$_{C\alpha}$ was used (Figure 21), as it changes monotonously between the two endpoints and

provides an intuitive measure for the progress of the swiveling movement of the CD [1, 2]. The opening-closing motion of the NBD of *Ft*PPDK was analyzed along the reaction coordinate distance$_{NBD1-NBD3}$, measured between $S215_{C\alpha} - E272_{C\alpha}$ (Figure 21) [1, 2]. These reaction coordinates represent the swiveling and opening-closing motions very well (see Figure 21) [1, 2].

1D and 2D umbrella sampling MD simulations for the PPDK monomer were performed along reaction coordinate(s) distance$_{CD-PBD}$ between 10 Å and 52 Å (9 Å and 54 Å for the 2D umbrella sampling) (and distance$_{NBD1-NBD3}$ between 26 Å and 42 Å) in intervals of 1 Å, applying harmonic potentials with a force constant of 1 kcal mol^{-1} Å^{-2} to tether the conformations to the respective reference point [1]. This resulted in 42 umbrella sampling simulations for the 1D PMFs of the swiveling motion, 17 umbrella sampling simulations for the 1D PMFs of the opening-closing motion, and 782 umbrella sampling simulations for the 2D PMF [1]. The 1D umbrella sampling simulations of the swiveling motion were performed for 42.5 ns each, excluding the first 2.5 ns from the WHAM analysis [1]. The 1D umbrella sampling simulations of the opening-closing motion were performed for 30 ns each, excluding the first 2 ns from the WHAM analysis [2]. The 2D umbrella sampling simulations were performed for 9 ns excluding the first 1 ns from the WHAM analysis [1].

Possible interference between the NBDs was investigated by performing multiple umbrella sampling/PMF calculations of the opening-closing motion of chain A for systems with chain B in states a) - e) (see above) [2]. All umbrella sampling simulations with ATP bound to the NBD of chain B also contained Mg^{2+} in the NBD of chain B [2]. Umbrella sampling MD simulations for the PPDK dimer were performed for distance$_{NBD1-NBD3}$ of chain A between 26 Å and 42 Å in intervals of 1 Å, applying harmonic potentials with a force constant of 1 kcal mol^{-1} Å^{-2} to tether the conformations to the respective reference point [2]. This resulted in 17 umbrella sampling simulations, each 20 ns long, excluding the first 1 ns from the WHAM analysis [2]. Approximately Gaussian-shaped frequency distributions were obtained for each reference point along the reaction coordinate(s), with all such distributions well overlapping (Figure 22, Figure 23, Figure 32) [1, 2]. The latter is a prerequisite for the successful application of WHAM to extract a PMF from these distributions [1, 2]. The Monte Carlo bootstrapping analysis implemented in WHAM using 200 resampling trails was applied to established the error at the reference points [1, 2].

4.9 Constraint Network Analysis

CNA was applied to investigate the effect of the dimerization on the structural stability of PPDK using three sets of coordinates extracted from one MD simulation of the PPDK dimer: coordinates of only chain A, of only chain B, and of both chains [2]. For this, the first 200 ns of the MD simulation started with the PPDK dimer in the open/closed conformation of the NBDs and equilibrated at 300.0 K was used, where the conformations of the NBDs remain particularly stable (MD simulation started from open/closed and unbound conformation no. 2 shown in Figure 28B; mean distance$_{NBD1-NBD3}$ of 38.7 Å (SEM ~1.0 Å) and 31.3 Å (SEM ~1.0 Å) for chain A and B, respectively) [2]. With CNA, constraint dilution simulations [75, 76] were performed on ensembles of 10,000 structures each, of which water molecules and counter ions had been stripped off [2]. A stepwise decrease of the energy cutoff E_{cut} for including hydrogen bonds from -0.1 kcal mol^{-1} to -6 kcal mol^{-1} in steps of 0.1 kcal mol^{-1} was performed [2].

Stability maps rc_{ij} were introduced in ref. [75] to characterize the local rigidity of a biomolecule [2]. A stability map depicts rigid contacts (*rc*) for each residue pair (*i* and *j*), represented by the Cα atom, respectively [2]. A rigid contact exists if the two residues belong to the same rigid cluster [127] [2]. Notably, this stability information is not only a qualitative by also quantitative measurement [2]. By performing constraint dilution simulations [128], each rigid contact is associated with an energy E_{cut} at which this rigid contact is lost [2].

5. Results and discussion

5.1 The swiveling and opening-closing motion investigated in the PPDK *monomer*

The following results were published in our paper Minges and Ciupka *et al.* [1], and are reproduced verbatim in subchapter 5.1:

5.1.1 The conformational space spanned by PPDK structures

To investigate the conformational space of PPDK, all available PPDK structures from the PDB were collected, combined with the newly resolved PPDK structures published in [1] (Table 2), but excluding the structure of PDB 5JVL, chain B and the NMR-derived structure of the CD only (PDB 2FM4) from further analysis as the NBD has not been resolved there [1]. Then, a multiple sequence and structural alignment of the PPDK were performed with PROMALS3D to identify the residues resolved and common to all structures from the alignment [1]. This way, the PPDK structure were clustered with respect to structural similarity (using the parameters described in chapter 4.2) (See Table 1) [1].

Table 1: Structural comparison of PPDK crystal structures by Cα-atom RMSD [a]

	Cluster I						Cluster II				Cluster III				Cluster IV	Cluster V
	1DIK [b]	2DIK	1GGO	1JDE	1KBL	1KC7	5JVJ/B	5JVL/A	5JVL/C [b]	5JVL/D	2R82	1VBG [b]	1VBH	5JVJ/A	5JVN [b]	2X0S [b]
1DIK [b]	0	0.4	0.5	0.6	0.9	0.9	10.7	9.2	10.8	10.6	11	10	9.9	10.9	8.1	5.9
2DIK	0.4	0	0.5	0.6	1	1	10.7	9.2	10.8	10.6	11	10	9.9	10.9	8.1	5.8
1GGO	0.5	0.5	0	0.7	0.9	0.9	10.6	9.2	10.7	10.5	10.9	9.9	9.8	10.8	8	5.8
1JDE	0.6	0.6	0.7	0	1.1	1.1	10.7	9.2	10.8	10.6	11	10	9.9	10.9	8.1	5.9
1KBL	0.9	1	0.9	1.1	0	0.3	10.7	9.2	10.7	10.6	10.9	10	9.9	10.8	8	5.8
1KC7	0.9	1	0.9	1.1	0.3	0	10.7	9.2	10.7	10.5	11	10	9.9	10.9	8	5.9
5JVJ/B	10.7	10.7	10.6	10.7	10.7	10.7	0	3.2	1.5	2.7	5.1	4.5	5	4.6	6.1	11.4
5JVL/A	9.2	9.2	9.2	9.2	9.2	9.2	3.2	0	2.6	2.2	5	4.4	4.7	5.2	5.6	10.8
5JVL/C [b]	10.8	10.8	10.7	10.8	10.7	10.7	1.5	2.6	0	1.4	5.1	4.8	5.3	5	6.4	11.8
5JVL/D	10.6	10.6	10.5	10.6	10.6	10.5	2.7	2.2	1.4	0	5.3	5.3	5.7	5.7	6.7	12
2R82	11	11	10.9	11	10.9	11	5.1	5	5.1	5.3	0	3	3.1	2.9	5.5	10.8
1VBG [b]	10	10	9.9	10	10	10	4.5	4.4	4.8	5.3	3	0	0.8	2.2	4.6	9.7
1VBH	9.9	9.9	9.8	9.9	9.9	9.9	5	4.7	5.3	5.7	3.1	0.8	0	2.4	4.5	9.5
5JVJ/A	10.9	10.9	10.8	10.9	10.8	10.9	4.6	5.2	5	5.7	2.9	2.2	2.4	0	4.5	10.3
5JVN [b]	8.1	8.1	8	8.1	8	8	6.1	5.6	6.4	6.7	5.5	4.6	4.5	4.5	0	7.5
2X0S [b]	5.9	5.8	5.8	5.9	5.8	5.9	11.4	10.8	11.8	12	10.8	9.7	9.5	10.3	7.5	0

[a] The Cα-atom RMSD was computed over all Cα atoms common and resolved in all PPDK structures with all domains resolved; in Å. Small RMSD are highlighted in blue, high RMSD in orange. Taken from [1].
[b] Cluster-representative structures.

The resulting clusters are characterized by a closed NBD combined with the CD facing the NBD (cluster I), a closed NBD with the CD facing the PBD (cluster II), an open NBD with the CD facing the PBD (cluster III), an open NBD with the CD in between the PBD and NBD (cluster IV), and an open NBD with the CD facing the NBD (cluster V) (Figure 18) [1].
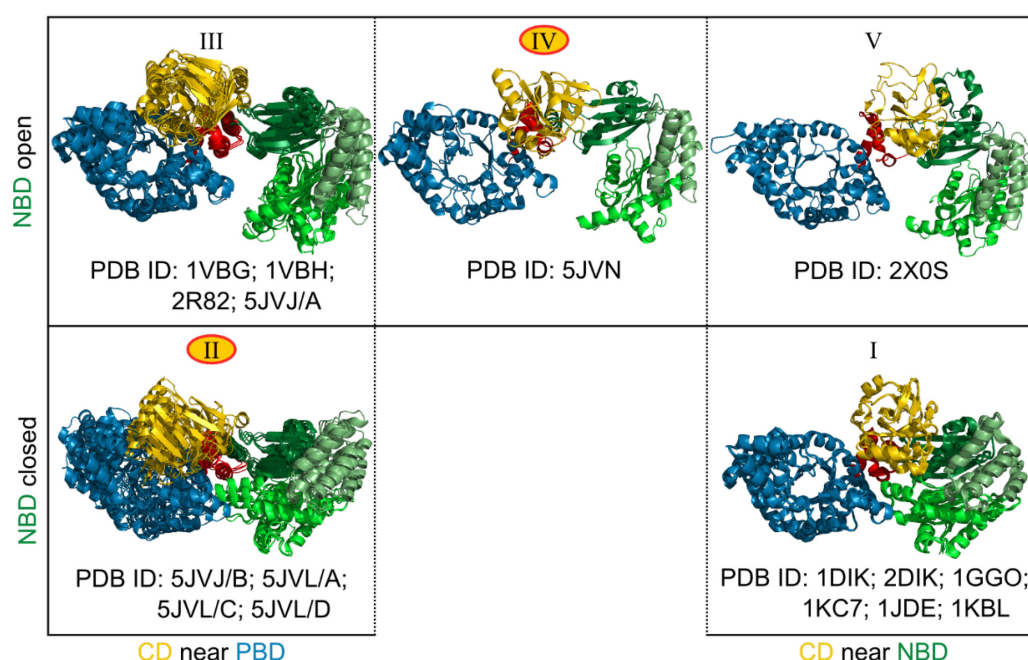
Figure 18: Clustering of all currently known PPDK structures according to Cα-atom RMSD (roman numbers), showing the CD near the PBD (left), between PBD and NBD (middle), and near the NBD (right) as well as the NBD in an open (top) or closed (bottom) conformation. Conformational states observed for the first time in *F. pringlei* and *F. trinervia* are marked with an orange circle. Adapted from [1].

5.1.2 Collective and correlated motions in monomeric PPDK

Next, a principle component analysis in Cartesian coordinate space (PCA) was performed, corroborating the existence of two predominant motions: ∼89% of the variance in the Cα atom coordinates can be explained by the first two principal components (PC) (Figure 19B) [1]. Atomic displacements along the component directions and a domain-wise index of the collectivity of motions (Figure 19A) reveal that the 1st PC predominantly characterizes a swiveling motion of the CD (Figure 19C) [1]. The 2nd PC characterizes a coordinated opening-closing motion of the NBD, with almost exclusively the first and second subdomain executing the movement (Figure 19C) [1].

Notably, the 1st PC also indicates a coupling of the swiveling motion of the CD with the opening-closing motion of the NBD as this PC leads to displacements in both domains (Figure 19A) [1]. Cross-correlations of atomic fluctuations computed from the cluster-representative crystal structures agree very well with this result (Figure 19D/upper triangle): They reveal, aside from positively correlated motions of the NBD, CD, and PBD themselves, weakly anti-correlated motions between the CD and the NBD (correlation coefficient down to -0.2), in agreement with the collective motions described by the 1st PC (Figure 19A) [1]. Particularly, subdomains NBD1 and NBD2 move towards the CD if the latter swivels from the PBD to the NBD and vice versa [1].
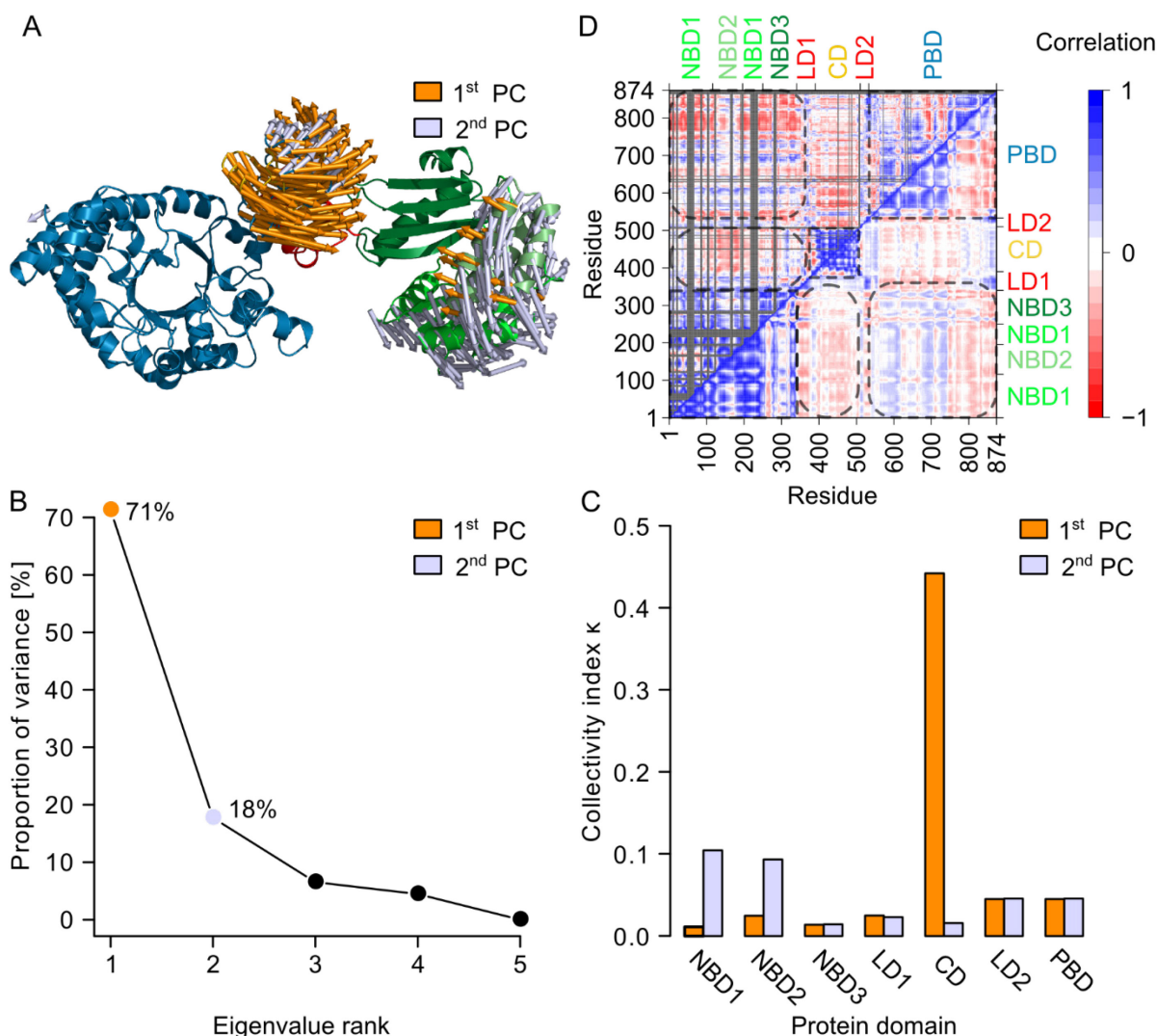
Figure 19: Motion of PPDK (A) Representation of atomic displacements along the directions of the first two principal components (1st PC: gold; 2nd PC: silver) obtained from a PCA over the representative structures of each cluster depicted in Figure 18. The amplitudes of the motions were scaled, and a cutoff for small displacements was applied for best graphical representation. **(B)** The proportion of explained variance by each principal component PC vs. the rank of the eigenvalue. 89% of the variance can be explained by the first two PCs. **(C)** Domain-wise collectivity index κ of the 1st and 2nd PC. **(D)** Cross-correlation maps of Cα atom fluctuations of the representative structures of each cluster (upper triangular) and from an aggregate MD simulation time of ~10 μs (lower triangular). The two axes on the left and at the bottom refer to residue indices according to the *F. trinervia* numbering. Positive correlations are indicated in blue, negative correlations in red (see color scale). Residues not resolved in all compared crystal structures are shown in black. Substructures of PPDK are labeled on the top and right axes. Correlated movements within domains are marked with squares, correlated movements between the domains are marked with boxes with round corners. Adapted from [1].

To exclude any bias by the small number of available crystal structures, we further explored the conformational space of *Ft*PPDK by molecular dynamics (MD) simulations of in total ~10 μs length (see chapter 5.4) [1]. Cross-correlations of atomic fluctuations computed from the MD simulations yield a qualitatively and quantitatively highly similar result compared with correlated motions computed from the cluster-representative crystal structures (Figure 19D), confirming the anti-correlated motions between the CD and the NBD (correlation coefficient down to -0.15) [1].

5.1.3 Umbrella sampling / PMF of the PPDK monomer

In order to investigate the energetics of the swiveling motion of the CD in the non-phosphorylated and phosphorylated state and the opening-closing motion of the NBD, umbrella sampling along three plausible transition paths of the domain motion followed by the calculation of the potential of mean force was performed [1]. For generating a pathway, a targeted normal mode-based geometric simulation by the NMSim approach [124] was used as NMSim efficiently generates a series of stereochemically correct conformations that lie preferentially in the subspace spanned by low-frequency normal modes [124] (

Figure 20) [1].
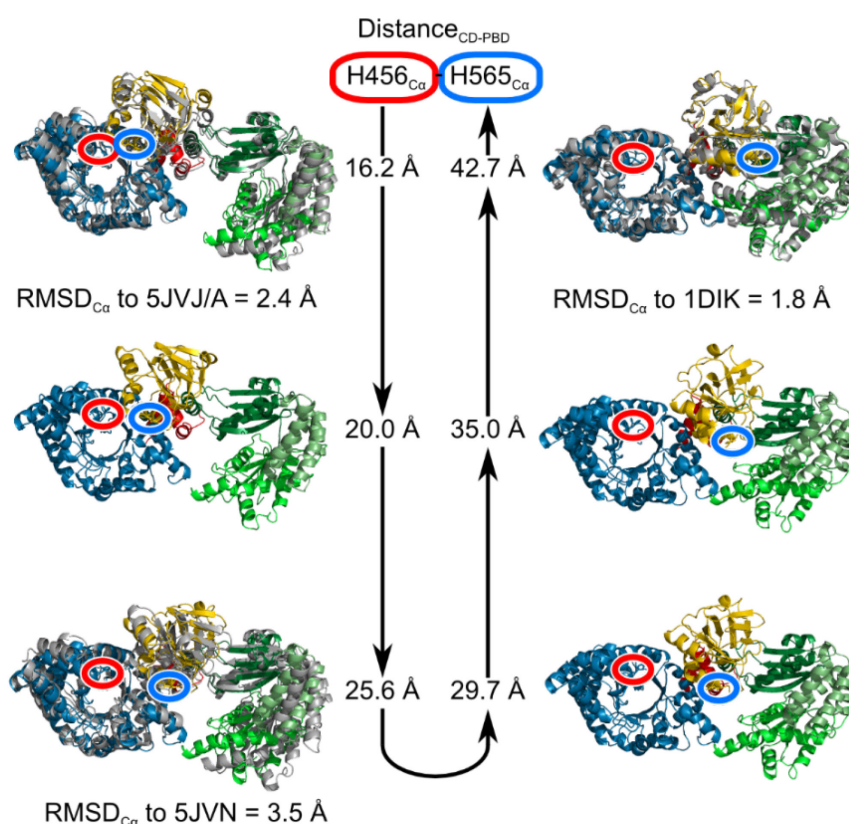
Figure 20: Transition pathway for umbrella sampling, between conformations III and I (Figure 22B/E), created by targeted simulations with NMSim starting from homology models with the template structure PDB ID 1KBL and 1KC7 towards homology models with the template structure PDB ID 1VBH and 1VBG. Intermediate conformations are shown at approximately 5 Å spacing along the reaction coordinate distance$_{CD-PBD}$ (Figure 21) (depicted next to the conformations). The generated conformations are overlaid with cluster representatives of PPDK crystal structures (depicted in gray), and the Cα-atom RMSD is given. Taken from [1].

The two endpoints and the generated intermediate states were used as reference points for subsequent umbrella sampling simulations [1]. Further reference points were obtained by MD simulations of the respective endpoint structures of 50 ns length and filtering of the trajectories for conformations in which the swiveling domain is closer to the NBD or PBD, respectively, than in the endpoint structures [1]. At each of these reference points, MD simulations for production of 40 ns length were performed after equilibration phases of 2.5 ns [1]. The

conformations were tethered to the respective reference point by a harmonic restraint potential with a force constant of 1 kcal mol^{-1} Å^{-2} [1]. This resulted in approximately Gaussian-shaped frequency distributions of the reaction coordinate values at each reference point, with all such distributions well overlapping (Figure 22) [1]. The latter is a prerequisite for the successful application of the Weighted Histogram Analysis Method (WHAM) [65] to extract a PMF from these distributions [1].

The distance between H456$_{C\alpha}$ – H565$_{C\alpha}$ (distance$_{CD-PBD}$, Figure 21A, numbering according to *F. trinervia*, see also supporting Table 2) was used as a reaction coordinate for the swiveling motion of the CD [1]. This reaction coordinate represents the swiveling motion very well (Figure 21B) [1]. This coordinate was chosen because it changes monotonously between the two end points and provides an intuitive measure for the progress of the CD movement from one active site in a binding domain to the other [1].
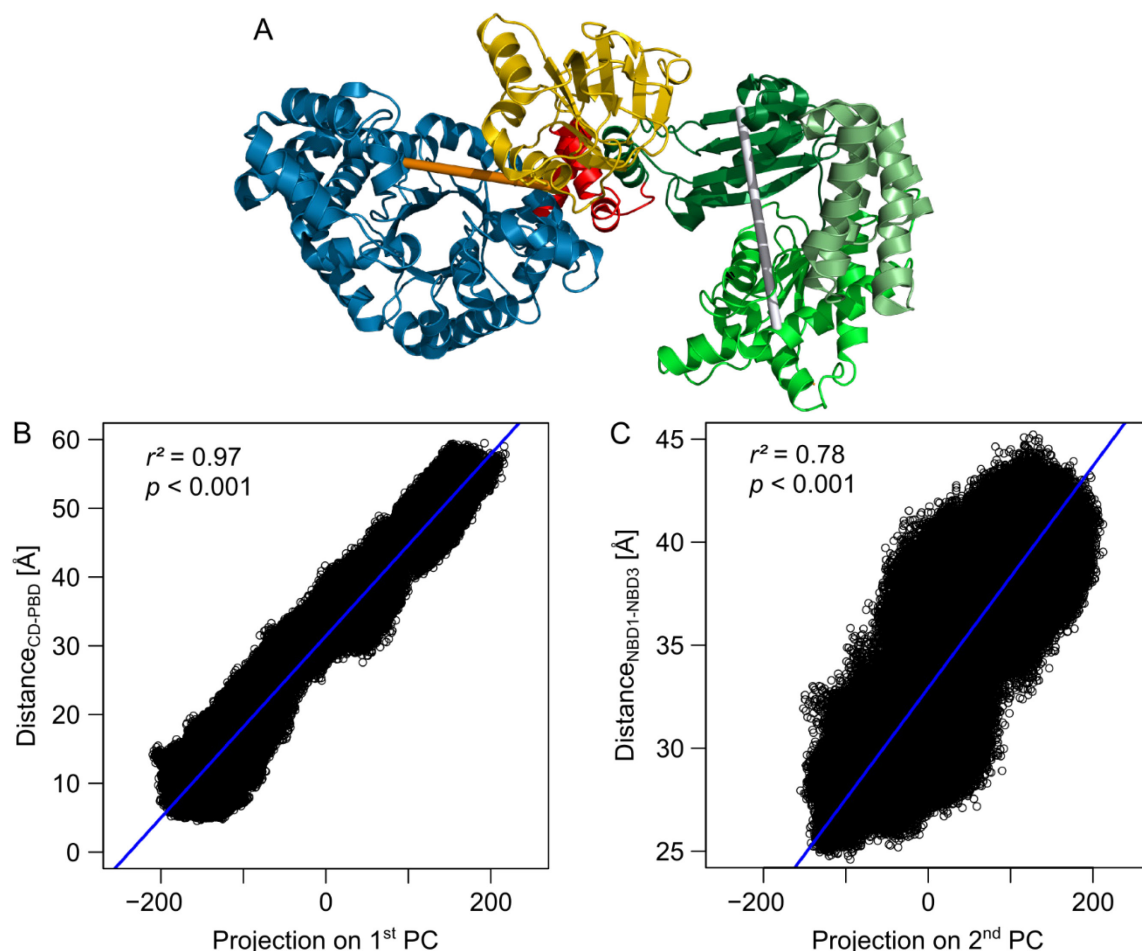
Figure 21: Reaction coordinates of the swiveling and opening-closing motion. (A) The distance$_{CD-PBD}$ (depicted as gold line) and distance$_{NBD1-NBD3}$ (depicted as silver line) between H456$_{C\alpha}$ – H565$_{C\alpha}$ and S215$_{C\alpha}$ – E272$_{C\alpha}$ (numbering according to *F. trinervia* / *pringlei*) are used as reaction coordinates for the swiveling motion of the CD and the opening-closing motion of the NBD, respectively. The depicted structure was taken from PDB ID 5JVN. The domain coloring is according to Figure 10. **(B, C)** Validation of the reaction coordinates to represent the swiveling motion of the CD and the opening-closing motion of the NBD. Scatter plots of the distances H456$_{C\alpha}$ – H565$_{C\alpha}$ **(B)** and S215$_{C\alpha}$ – E272$_{C\alpha}$ **(C)** versus projections onto the 1st **(B)** and 2nd **(C)** principle component

obtained from the PCA over the cluster representatives (Figure 18); each dot represents one conformation generated by MD simulations of ~10 μs length. The correlation line is shown in blue. Taken from [1].

Transition paths were generated between start / end states of *Ft*PPDK from cluster III / cluster V, cluster III / cluster I, and cluster II / cluster I (Figure 18, Table 1), so that the swiveling motion occurs in the presence of an open NBD (Figure 22 A/D), a closed NBD (Figure 22 C/F), or where the NBD closes when the CD approaches it and vice versa (Figure 22 B/E) [1]. A non-phosphorylated and in a phosphorylated state of the catalytic H456 was simulated [1]. While the opening-closing motion of the NBD was not restrained during the umbrella sampling, the observed sampling very well represents these paths (Figure 22) [1]. Distance$_{CD-PBD}$ was sampled in intervals of 1 Å, leading to a sufficient overlap of the sampling windows (Figure 22) [1].
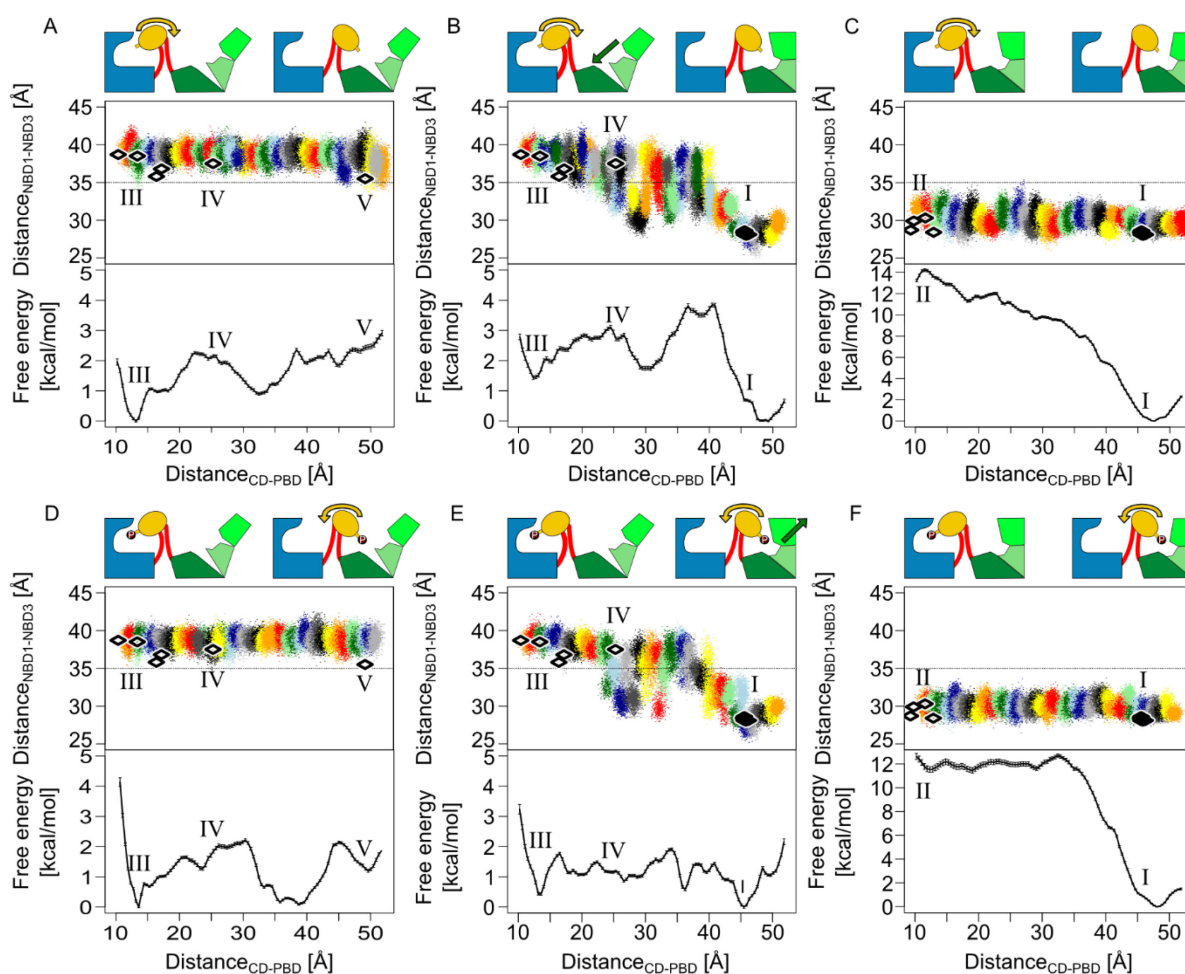
Figure 22: 1D Potential of mean force of the swiveling motion of the CD with the distance H456$_{C\alpha}$ – H565$_{C\alpha}$ used as a reaction coordinate. Results for the non-phosphorylated state are depicted in **(A-C)**, those for the state with phosphorylated H456 in **(D-F)**. **(A/D)** depict results obtained with open NBD, **(B/E)** those with a simultaneous opening-closing of the NBD, and **(C/F)** with closed NBD. At the top, schematic representations of each state at the respective endpoints of a PMF are shown. In the middle row, sampled conformations are projected onto distance$_{CD-PBD}$ and distance$_{NBD1-NBD3}$ (see Figure 21), with each color representing one MD simulation with an umbrella potential applied at a given value of distance$_{CD-PBD}$. The PMFs are depicted in the bottom row. The diamonds show projections of PPDK crystal structures in conformational states marked by Roman numbers (Figure 18) onto the plane spanned by the two reaction coordinates (using for each organism the corresponding residues to evaluate the reaction coordinates (Table 2)). Taken from [1].

The obtained PMFs (Figure 22A/B/D/E) show several remarkable characteristics [1]. First, the overall precision is high, with standard deviations for all points < 0.15 kcal mol^{-1}. Second, the free energy difference between conformational states III and I, or III and V (Figure 22) is < 2 kcal mol^{-1} (Figure 22A/B/D/E), showing that the respective higher-energy conformational state is populated to ~2% at room temperature [1]. Third, we identify a stable conformational intermediate of the CD swiveling motion, revealed by pronounced minima in the PMFs at a distance$_{CD-PBD}$ ~30 Å to 35 Å (Figure 22A/B/D/E), which is structurally highly similar to the intermediate state IV found in the *Fp*PPDK crystal structure (located at a distance$_{CD-PBD}$ ~25 Å) as shown by a Cα atom RMSD of 2.7 Å [1]. Note that the *Fp*PPDK crystal structure was not used for generating the transition path, hence, no information about conformational state IV entered the PMF calculations [1]. The presence of the intermediate state leads to a sawtooth-like PMF [1]. Fourth, the PMFs reveal for PPDK with non-phosphorylated H456 (Figure 22A/B) that the correlated movement of the CD and the NBD results in state I being favored over III by 1.4 kcal mol^{-1} (Figure 22B); in contrast, if the CD moves with the NBD remaining open, state V is disfavored over III by 2.1 kcal mol^{-1} (Figure 22A) [1]. Therefore, movement of the non-phosphorylated CD from the PBD towards the NBD is exergonic only if it is coupled to a closing motion of the NBD, which is in line with the swiveling domain model [1]. This finding corroborates cross-correlations between these motions observed from crystal structures (Figure 19D, upper triangle) and from structures obtained by MD simulations (Figure 19D, lower triangle) [1]. Fifth, the PMFs suggest that the phosphorylation state of the CD influences the preferred direction of motion of this domain: When H456 is phosphorylated, state III becomes favorable over V by 1.5 kcal mol^{-1} (Figure 22D) or is only slightly disfavored over I by 0.5 kcal mol^{-1} (Figure 22E) [1]. Compared to the non-phosphorylated state, the sawtooth-like PMF is thus tilted towards III. Therefore, with phosphorylated H456, the movement of the CD from the NBD towards the PBD is exergonic or approximately isoenergetic. This result is also in line with the swiveling domain model [83] [1]. In the phosphorylated state, coupled motions between the CD and the NBD have a smaller influence on the energetics of the conformational states [1]. We speculate that the electrostatic repulsion between the phosphate group at H456 and reaction products, including adenosine monophosphate, still bound to or being in the vicinity of the NBD fosters the CD movement towards the PBD instead [1].

PMFs calculated for the transition between conformational states I and II unexpectedly do not reveal a structural intermediate along the transition path and show a marked prevalence of state I (free energy difference > 12 kcal mol^{-1}), irrespective of the phosphorylation state of H456 (Figure 22 C/F) [1]. To provide an explanation for this observation and further details on the coupling between motions of the CD and the NBD, we computed a 2D PMF, using as reaction coordinates distance$_{CD-PBD}$ and the distance between S215$_{C\alpha}$ – E272$_{C\alpha}$ (distance$_{NBD1-NBD3}$, Figure 21, numbering according to *F. trinervia*, see also Table 2) for the non-phosphorylated PPDK [1]. The reaction coordinate distance$_{NBD1-NBD3}$ represents the opening-closing motion of the NBD very well (Figure 21) [1]. Reference points for umbrella sampling were generated using targeted constrained geometric simulations [124] between start/end states of *Ft*PPDK from cluster III / cluster I (Figure 20), with further reference points added using the conformation after 6 ns of umbrella sampling as a starting point for the next interval [1]. The sampling windows overlap well along the two reaction coordinates (Figure 23), and the 2D PMFs are qualitatively indistinguishable irrespective whether only the first half, the second half, or the complete sampling time is used for their calculation (Figure 41), strongly indicating converged results [1].
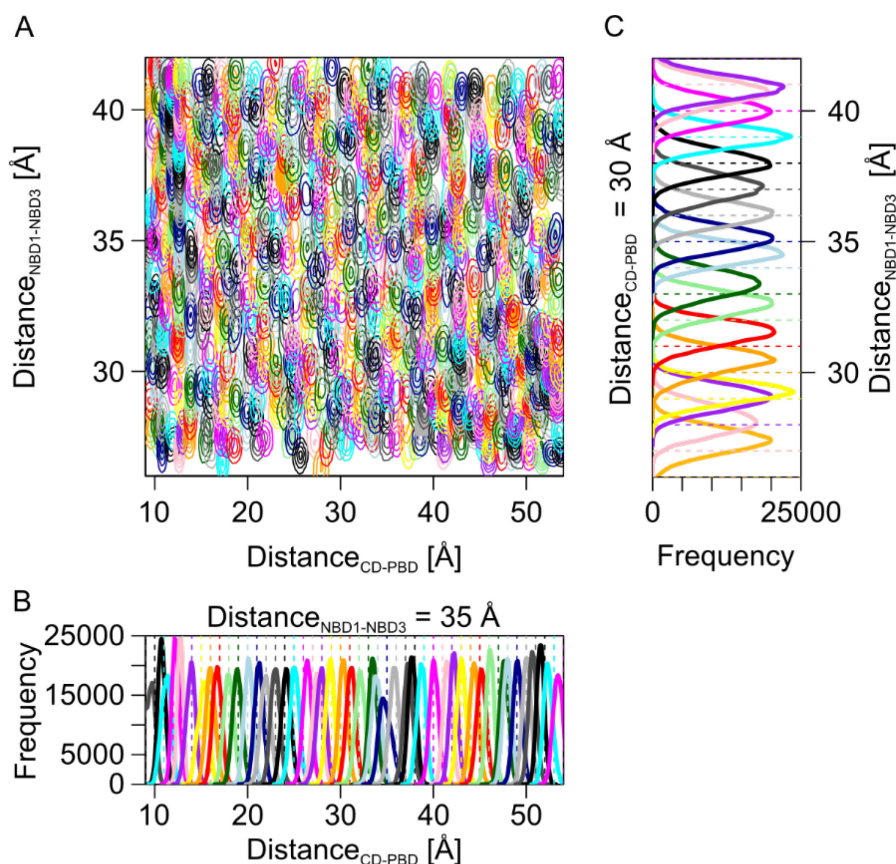
Figure 23: Umbrella sampling. Frequency distributions of the sampled distances used as reaction coordinates (Figure 21) for umbrella sampling are shown **(A)** for the 782 MD simulations covering the 2D space spanned by both reaction coordinates (contour lines are plotted at frequencies of 5000, 10000, 15000, 20000, and 25000 conformations) and **(C)** for 17 MD simulations at a restrained distance$_{CD-PBD}$ = 30 Å as well as **(B)** for 45 MD simulations at a restrained distance$_{NBD1-NBD3}$ = 35 Å. Each color represents one MD simulation. Taken from [1].
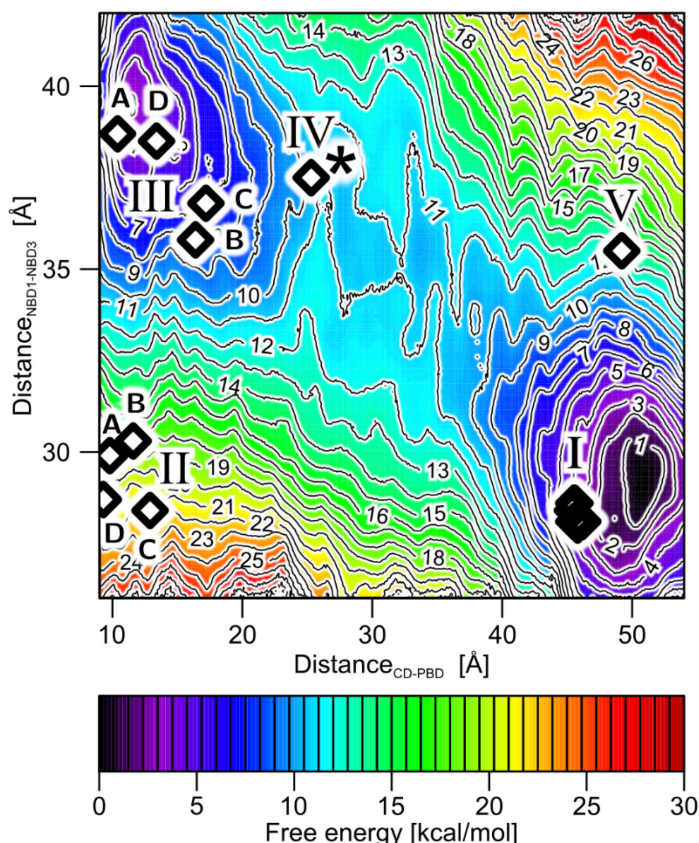
Figure 24: 2D Potential of mean force, with the distance$_{CD\text{-}PBD}$ between H456$_{C\alpha}$ – H565/$_{C\alpha}$ and the distance$_{NBD1\text{-}NBD3}$ between S215$_{C\alpha}$ – E272$_{C\alpha}$ used as reaction coordinate of the swiveling motion of the CD and the opening-closing motion of the NBD, respectively, (see Figure 21) computed for H456 in the non-phosphorylated state. The diamonds show projections of PPDK crystal structures in conformational states marked by Roman numbers (Figure 18) onto the plane spanned by the two reaction coordinates (using for each organism the corresponding residues to evaluate the reaction coordinates (Table 2)) (I: PDB 1DIK, 2DIK, 1GGO, 1JDE, 1KBL, 1KC7; II$_A$: 5JVJ/B, II$_B$: 5JVL/A, II$_C$: 5JVL/C, II$_D$: 5JVL/D; III$_A$: 2R82, III$_B$: 1VBG, III$_C$: 1VBH, III$_D$: 5JVJ/A; IV: 5JVN; V 2X0S). The star marks a shallow energy minimum close to conformation IV. Taken from [1].

The most prominent feature of the PMF is that conformational states III and I reside in or close to minima of the free energy landscape (Figure 24) [1]. Furthermore, a shallow free energy minimum is identified at distance$_{CD\text{-}PBD}$ ~28 Å and distance$_{NBD1\text{-}NBD3}$ ~38 Å (marked by a star in Figure 24), lying close to the structural intermediate from the *Fp*PPDK crystal structure (conformational state IV, Figure 18) [1]. The transition path between conformational state III and IV runs such that distance$_{NBD1\text{-}NBD3}$ remains almost constant while distance$_{CD\text{-}PBD}$ changes by ~15 Å [1]. In contrast, between conformational state IV and I, the transition paths continues in a diagonal manner, revealing a correlation between the swiveling motion of the CD (change of distance$_{CD\text{-}PBD}$ by ~25 Å) and the opening-closing motion of the NBD (change of distance$_{NBD1\text{-}NBD3}$ by ~10 Å) [1]. These characteristics agree very well with those found by the 1D PMFs, where distance$_{NBD1\text{-}NBD3}$ only starts to decrease once distance$_{CD\text{-}PBD}$ > 25 Å (Figure 22B/E) [1]. Overall, both PMF calculations thus reveal a complex, stepped swiveling motion of the CD with varying degrees of coupling to the NBD motions and proceeding via conformational state IV [1]. Moreover, the 2D PMF reveals that conformational state II is located in a region of elevated free energy (Figure 24), providing an explanation why a transition path obtained by targeted simulations between conformational states II and I results in a downhill motion towards state I (Figure 22C/F) [1].

5.1.4 Unrestrained MD simulations

To support the finding that conformational state II is located in a region of elevated free energy (Figure 24), we mapped all unrestrained MD simulations started from each of the conformational states of non-phosphorylated PPDK onto the plane spanned by distance$_{CD-PBD}$ and distance$_{NBD1-NBD3}$ (Figure 25A-E) [1]. The results reveal that for all but two of the 3 x 5 MD simulations of 600 ns length started from conformational states I, III, IV, and V, the conformational sampling occurs in the vicinity of the starting structure (Figure 25A/B/C/E); in two cases, MD simulations started from conformational state III show a conformational transition to state IV (Figure 25A) [1]. Notably, in all three MD simulations started from conformational state II, conformational transitions of PPDK either to state III or IV are observed after at most 192 ns of MD simulations (Figure 25D), strongly indicating that conformational state II is unstable [1].
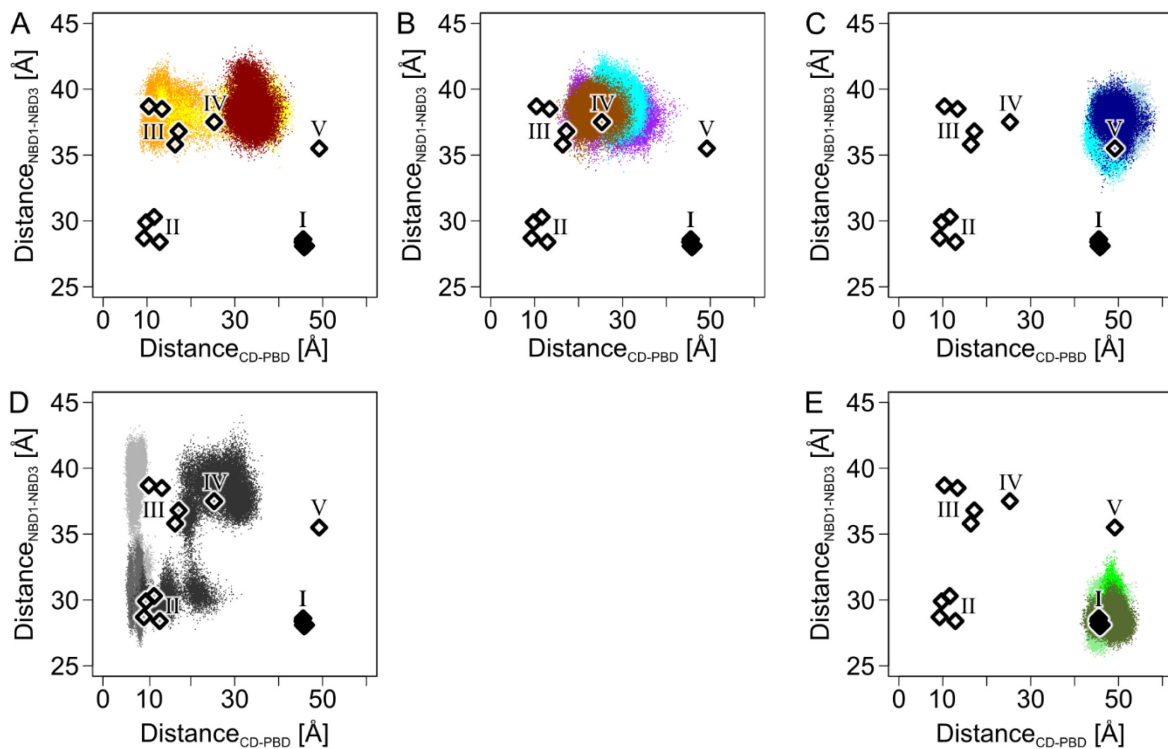
Figure 25: Unrestrained MD simulations of PPDK. Projection of conformations from MD simulations and crystal structures of PPDK (dots; each color represents one MD simulation of non-phosphorylated *Ft*PPDK, see Table 3) (using the corresponding residues for each organism (Supporting Table 2), with Roman numerals according to Figure 18) onto the reaction coordinates distance$_{CD-PBD}$ and distance$_{NBD1-NBD3}$ (Figure 21). (I: PDB ID 1DIK, 2DIK, 1GGO, 1JDE, 1KBL, 1KC7; II: 5JVJ/B, 5JVL/A, 5JVL/C, 5JVL/D; III: 2R82, 1VBG, 1VBH, 5JVJ/A; IV: 5JVN; V 2X0S). Three MD simulations of 600 ns length each were started from **(A)** conformation III, **(B)** conformation IV, **(C)** conformation V **(D)** conformation II, and **(E)** conformation I. Taken from [1].

5.1.5 Effective driving force of the swiveling motion
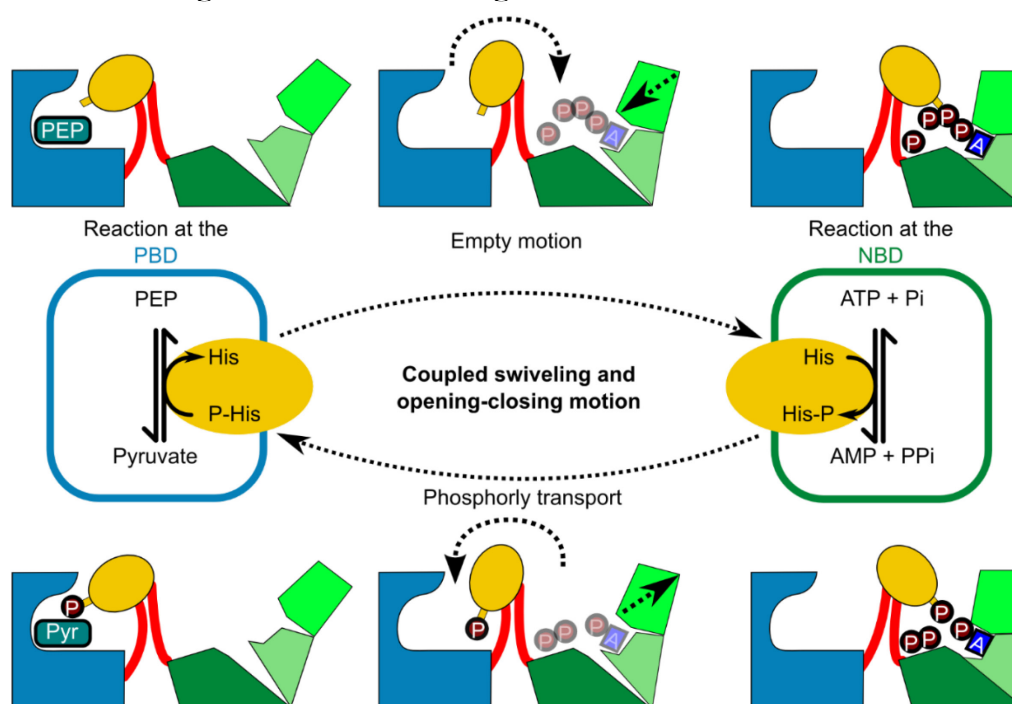
Figure 26: Simplified model of the coupled motions in the PPDK monomer. The PPDK-catalyzed reaction (middle row) involves a phosphoryl transfer via H456 of the CD (yellow) between the locations of the two reactions at the NBD (green, right side) and at the PBD (blue, left side). The dotted arrows in the cartoons in the top and bottom rows indicate motions of the CD and within the NBD. The upper row indicates the empty motion in the direction toward the NBD and the lower row the phosphoryl transport motion in the direction toward the PBD. Taken from [1].

The analysis of crystal structures, unrestrained MD simulations, and configurational free energy calculations consistently and independently revealed the existence of structural intermediates of the swiveling motions and coupled motions of the CD and the NBD for non-phosphorylated PPDK [1]. The respective 1D PMF of PPDK with non-phosphorylated H456 is sawtooth-like (Figure 22B), with barrier heights for the transition from state III to I between 1.5-2.5 kcal mol^{-1} (2.5-4.1 k_BT at T = 300 K) [1]. For the reverse transition with phosphorylated H456, the 1D PMF also displays a sawtooth-like character (Figure 22E), with barrier heights of 1.8 kcal mol^{-1} (3 k_BT at T = 300 K) and the free energy profile being tilted in favor of state III compared to the non-phosphorylated PPDK [1]. The sawtooth-like free energy profiles suggest that PPDK can exploit random thermal fluctuations for directional motion of its CD [1]. Typically, this type of profile is indicative of a Brownian ratchet mechanism [50], pioneered by Feynman [13] and Huxley [38] [1]. Barrier heights on the order of k_BT, as associated with the CD motion in the PPDK catalytic cycle, suggest that a Brownian ratchet biases fluctuations rather than rectifying them [46] [1]. To drive the respective directional motions, a non-equilibrium situation needs to be created that relaxes towards equilibrium [1]. For the non-phosphorylated PPDK, this situation is suggested to be created by binding of ATP to the NBD, as such binding leads to a closing of the NBD due to the progressive formation of ATP/NBD interactions [1]. Transmitted

via coupled motions, the NBD closing then leads to a preference for the CD to be close to the NBD, at least in the second half of the transition pathway between PBD and NBD [1]. However, we are aware that biased Brownian ratchets and power stroke motors have the same phenomenological behavior and are difficult to distinguish experimentally [46] [1]. Hence, we cannot exclude that the conformational changes observed in the NBD of our structural intermediates induce a strain in the enzyme that would directly drive the CD motion upon ATP binding, thus resulting in a power stroke mechanism [129] [1]. Overall, this situation is similar to ATP synthase and ATP-dependent molecular machines such as myosin, kinesin, or chaperonin parts, where the stepped motions in the catalytic cycle are triggered by binding of the high energy substrates PEP or ATP, respectively [105] [1]. For the phosphorylated PPDK, at least part of the non-equilibrium situation is suggested to arise from electrostatic repulsion between the phosphorylated H456 and the NBD [1]. A similar situation was created by introducing appropriately charged residues into a mutant of non-phosphorylated *Cs*PPDK, resulting in the CD being adjacent near the PBD (PDB ID 2R82) [91] [1]. In all, our analyses suggest that both changes of the binding state of PPDK and of its molecular identity ((non-)phosphorylation) contribute to the enzyme acting as a molecular switch with respect to the swiveling motion [1]. Moreover, our molecular simulation data are indicative that PPDK might employ a Brownian ratchet mechanism biasing thermal fluctuations in order to generate a net directional CD motion [1].

5.2 The alternate binding change mechanism of the PPDK *dimer*

The biologically active from of PPDK in bacteria and protozoa is considered to be a dimer, and may be a tetramer in plants [78-80] [1, 2]. The dimeric form of PPDK was exposed for the first time in the asymmetric unit with the crystal structure 5JVJ, with the dimer interface formed by the two PBDs [1]. Notably, the two monomers differs in the conformational state of the NBDs, with that of chain A being in an open and that of chain B in a closed conformation [1]. The NBD of monomer B exhibits additional electron density that might reflect a bound adenine nucleotide [1]. The distinct conformational states of the NBD of 5JVJ have led to the hypothesis that PPDK employs an alternate binding change mechanism [1] (also termed reciprocating mechanism stressing its processivity [130]) similar to ATP synthase [131, 132] or bacterial ATP-dependent DNA helicases [94] [2]. However, why PPDK is active only in as dimer and to what extent an alternate binding change mechanism could underlie this fact has remained elusive [2]. The following results were published in our paper Ciupka and Gohlke [2], and are reproduced verbatim in subchapter 5.2:

5.2.1 Native interface of the PPDK dimer

Initially, the interface found between the PBD domains in 5JVJ was investigated regarding its role for forming the dimeric biological unit of *F. trinervia* PPDK [2]. Using the PDBePISA tool [133] the analysis revealed an interface area of 2077 Å^2 , > 40 potential hydrogen bonds and salt bridges found within the interface, a ΔG *p*-value of 0.370, and a complexation significance score (CSS) of 0.514, these data imply that the interface plays an essential role in complex formation rather than it being a result of crystal packing only [2]. Mutations reported to reduce PPDK's cold-dependent dissociation and therefore inactivation [134] appear in the vicinity of the dimer interface (Figure 27A). Together with the high degree of conservation of the residues forming the dimer interface (Figure 27B), found in a multiple sequence alignment performed with MAFFT [113] on 1000 PPDK sequences obtained by BLASTp [135], this finding support the assumption that the considered dimer interface represents the native functional dimer [2].
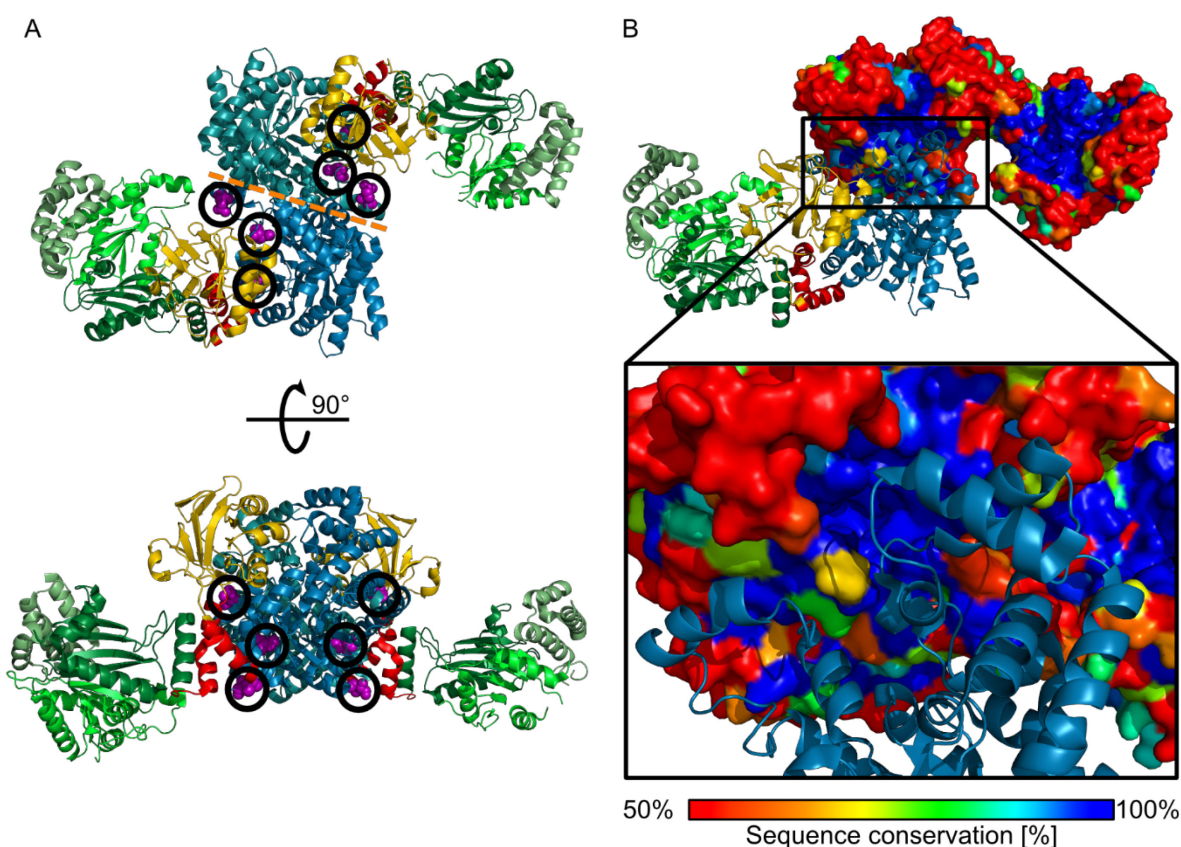
Figure 27: Validation of dimer interface. **(A)** Crystal structure PDB ID 5JVJ showing mutations (spheres in purple and circled) reported to reduce PPDK's cold-dependent dissociation and thereby inactivation [134] with the dimer interface indicated by the dashed orange line in top view (top) and side view (bottom). **(B)** All-atom model of PDB ID 5JVJ showing the degree of sequence conservation (see color scale) mapped onto the surface of chain B of the whole structure (top) and as a close-up of the dimer interface. The conservation degree was computed from a multiple sequence alignment obtained with MAFFT [113] of 1000 PPDK sequences identified by BLASTp [135] in the NCBI-NR database of non-redundant protein sequences [136]. Taken from [2].

5.2.2 Structural dynamics of dimeric PPDK

Next, all-atom structures of PPDK dimers from *F. trinervia* were generated with the two NBDs in unbound closed/closed, open/closed, or open/open conformations, as well as in ATP-bound closed/closed conformations (see Materials and Methods for details) [2]. The dimers served as starting structures for three independent MD simulations of 300 ns length each (Table 3), respectively, resulting in total twelve MD simulations (Figure 28) [2].
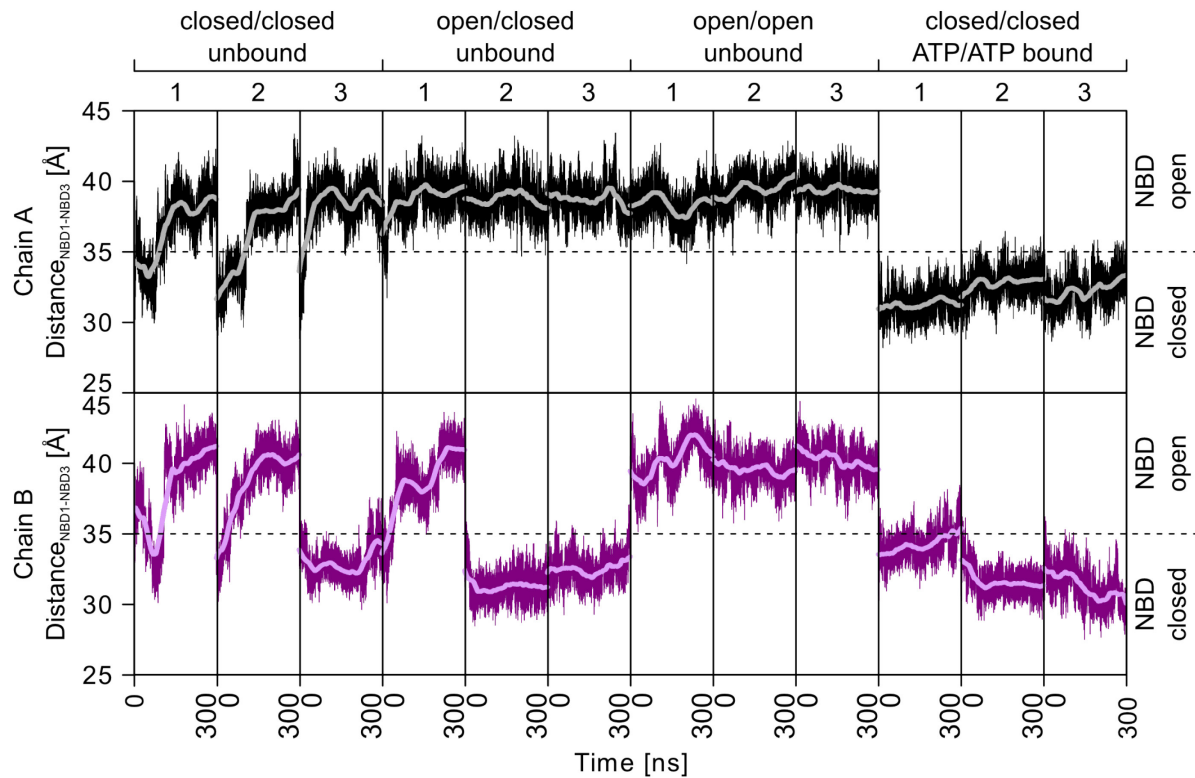
Figure 28: Time course of the opening-closing motions of the NBDs of chain A (black line) and chain B (deep purple line) measured by the respective distance$_{NBD1-NBD3}$ between S215$_{C\alpha}$ – E272$_{C\alpha}$. From the conformations labeled on the top axis, three replica MD simulations each were started of 300 ns length, with the mean over an interval of 60 ns shown in gray (for chain A) and light magenta (for chain B). Adapted from [2].

The MD simulations revealed a pronounced tendency for unbound, closed NBD domains to open (six out of nine cases) with an opening usually occurring during the first 150 ns of a trajectory [2]. In contrast, open NBD conformations remained open during the simulation time (nine out of nine cases), and ATP-bound closed conformations remained closed (six out of six cases), although it cannot be excluded that this finding is due to too short simulation times, particularly in the last case [2]. Still, with respect to the question of an alternate binding change mechanism, ten out of twelve MD simulations do not indicate a preference for one NBD of the PPDK dimer to be in a conformation different from the other, given identical binding states of both domains [2]. Note that the CDs stayed at their starting positions facing the PBD in all MD simulations, in contrast to the observation from our study on monomeric PPDK [2]. The latter observation might result from MD simulation times that are at least two-fold longer than those here [2].

Further, a PCA in Cartesian space was performed on coordinates derived from all twelve MD simulations [2]. 66% of the variance in the motions of the PPDK dimers can be characterized by the first two principal components (PC), with the 1st and 2nd PC explaining ~34%, and ~32%, respectively (Figure 29B) [2]. According to atomic displacements along PC directions (Figure 29A), both the 1st and 2nd PC primarily reflect the opening-closing motions of both NBDs, which are predominantly executed by the first and second subdomains (NBD1 and NBD2), as indicated by the (sub)domain-wise collectivity index (Figure 29C) [2]. Interestingly, while the 1st PC describes a symmetric motion of both NBDs (i.e., both open, or close, simultaneously), the 2nd describes an anti-symmetric one (i.e., one opens, the other closes) (Figure 29A) [2]. Considering the almost equal proportion of variance both PCs characterize, this finding does not support that the opening-closing motion of one NBD depends in a defined manner on the motion of the other NBD [2].
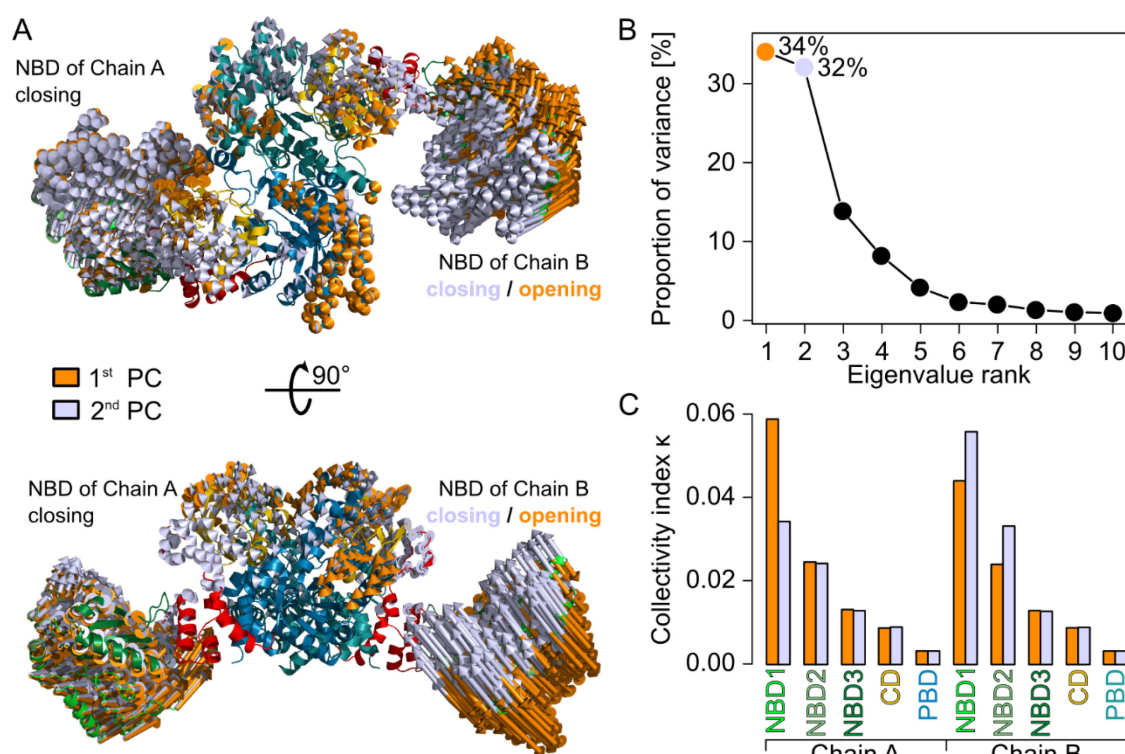
Figure 29: Principle component analysis of the PPDK dimer, computed from MD simulations of an aggregate simulation time of 3.6 µs, depicting the 1st and 2nd principal component (PC) in gold and silver, respectively. **(A)** Representation of atomic displacements along the directions of the 1st and 2nd PC on the crystal structure PDB ID 5JVJ. The amplitudes of the motions were scaled, and a cutoff for small displacements was applied for best graphical representation. **(B)** The proportion of explained variance is depicted *versus* the number of the principal component (PC) ranked by its eigenvalue. 66% of the variance can be explained by the first two PCs. **(C)** Domain-wise collectivity index κ of the 1st and 2nd PC. Adapted from [2].

Moreover, cross-correlation analysis of Cα atom fluctuations, computed to identify a possible coupling between the opening-closing motions, reveals, first, positively correlated motions of the PPDK domains (NBDs, CDs, and PBDs) themselves, and, second, in part positively and in part weakly negatively correlated motions between the PBDs (Figure 30A); the positively correlated motions occur in those regions of the PBDs that form the dimer interface (Figure 30B) [2]. Third, motions of the NBD of one chain are weakly negatively correlated to those of the PBD of the other chain, which might reflect the close distances between these domains [2]. However, motions between the NBD of one chain and the PBD of the same chain are largely uncorrelated, as also found for monomeric PPDK (Figure 19) [2]. Finally, motions between the two NBDs are also uncorrelated (mean correlation coefficient of 0.1), implying that the opening-closing motion of one NBD is not directly dependent on the opening-closing motion of the other NBD [2].
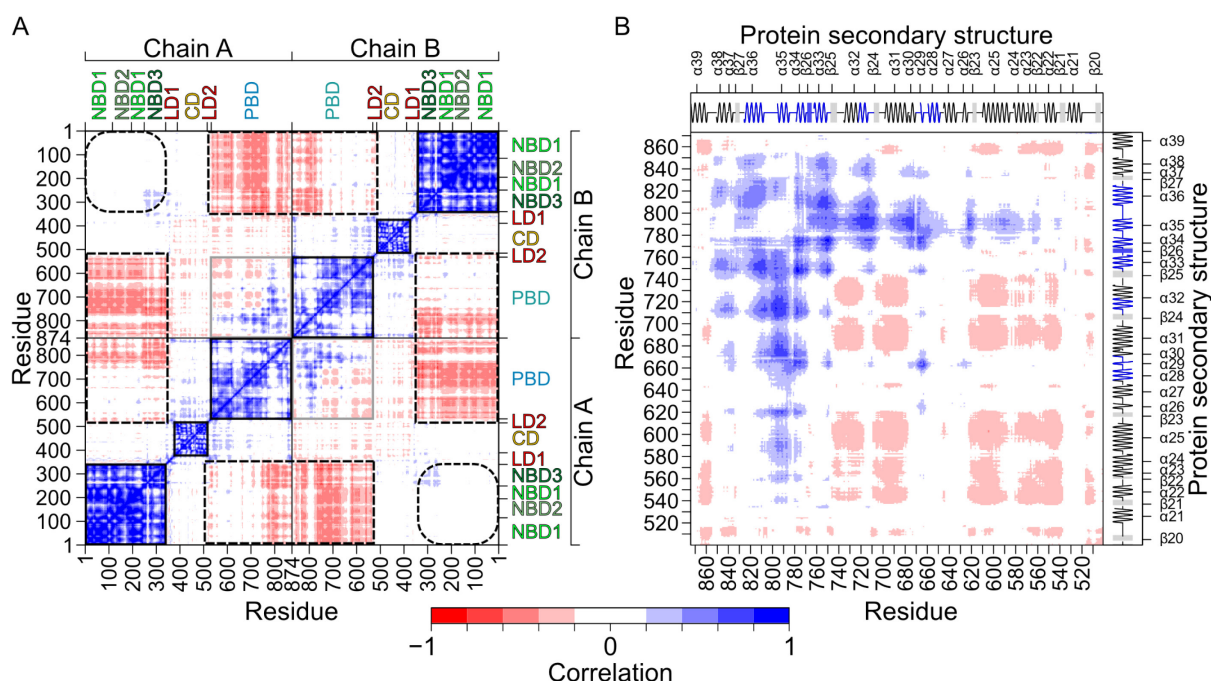
Figure 30: Cross-correlations map of Cα atom fluctuations of the PPDK dimer, performed on snapshots from MD simulations. **(A)** For the whole structure with the residue numbering labeled on left and at the bottom, and the substructures labeled on the top and right axes. Positive correlations are indicated in blue, negative correlations in red (see color scale). Correlated motions within and between domains are marked with squares with solid and dashed line, respectively. Squares with round corners indicate that no correlated motions between the two NBDs were found. **(B)** Blow-up of the region displaying intermolecular correlations between the PBDs of chains A and B, with the secondary structures as predicted by DSSP [137] in black/gray and the interface residues in blue labeled on the top and right axes. Adapted from [2].

5.2.3 Energetics of the opening-closing motions

To complement the unrestrained MD simulations, the PMF of the opening-closing motion of the NBD of chain A was computed, using umbrella sampling along the $distance_{NBD1-NBD3}$ between $S215_{C\alpha} - E272_{C\alpha}$ of chain A as a reaction coordinate. This reaction coordinate had been shown to represent the opening-closing motion of the NBD very well (Figure 21) [2]. From the PMF of the PPDK monomer, we expected the open conformation of PPDK being energetically preferred for the conformation of the CD facing the PBD [2]. To analyze how the conformational and binding states of chain B modulate the PMF, umbrella sampling simulations were performed for five states of chain B (Figure 31): With the CD facing the PBD and the NBD being unbound and a) open or b) closed; c) with the CD facing the PBD and the NBD being closed and ATP-bound; with the CD facing the NBD and the NBD being closed and d) unbound or e) ATP-bound. State c) resembles what has been found in PDB ID 5JVJ [2].
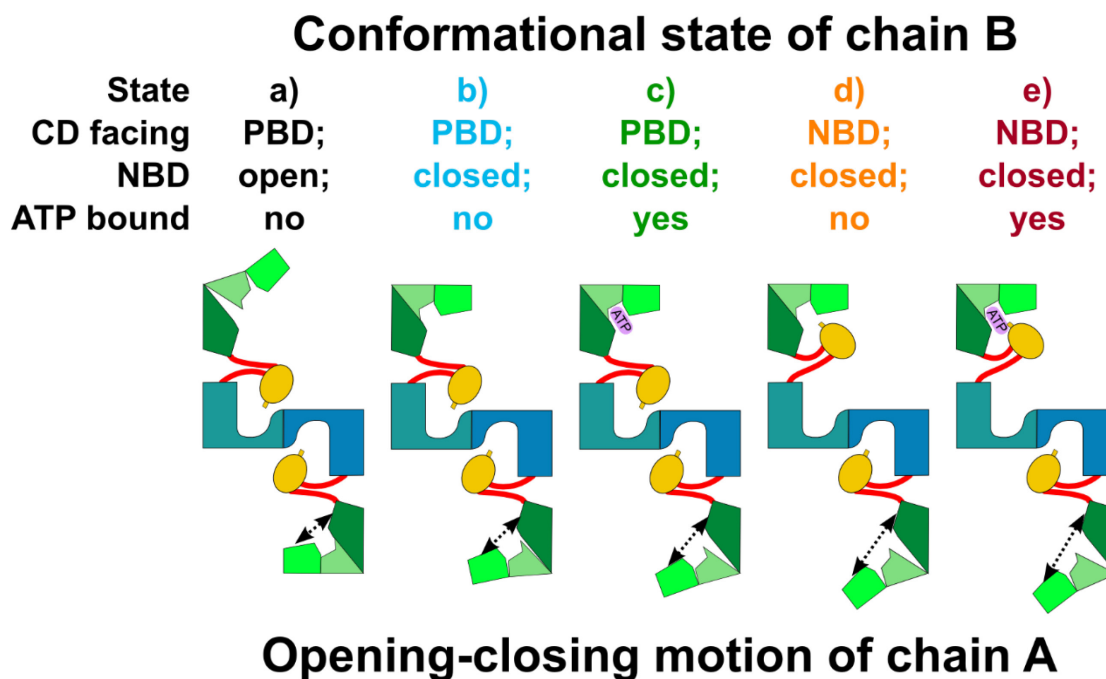
Figure 31: Umbrella sampling of PMF computations setup for dimeric PPDK. Schematic representation of states to validate the alternate binding change mechanism with umbrella sampling / PMF computations. Adapted from [2].

Intermediate states between the open and closed NBD conformations of chain A were generated by targeted simulations with the NMSim approach [124, 125]. The umbrella windows display sufficient overlap regarding the frequency distribution of values for the reaction coordinate (Figure 32) [2]. Furthermore, although the motions of the CDs, the NBD of chain B and ATP were not restrained, their conformational and binding states remain similar to the starting conditions (Figure 32A-D) [2]. Finally, the PMFs using ½ or all sampled data differ by at most ~2 kcal mol^{-1} (Figure 32E), suggesting converged results [2].
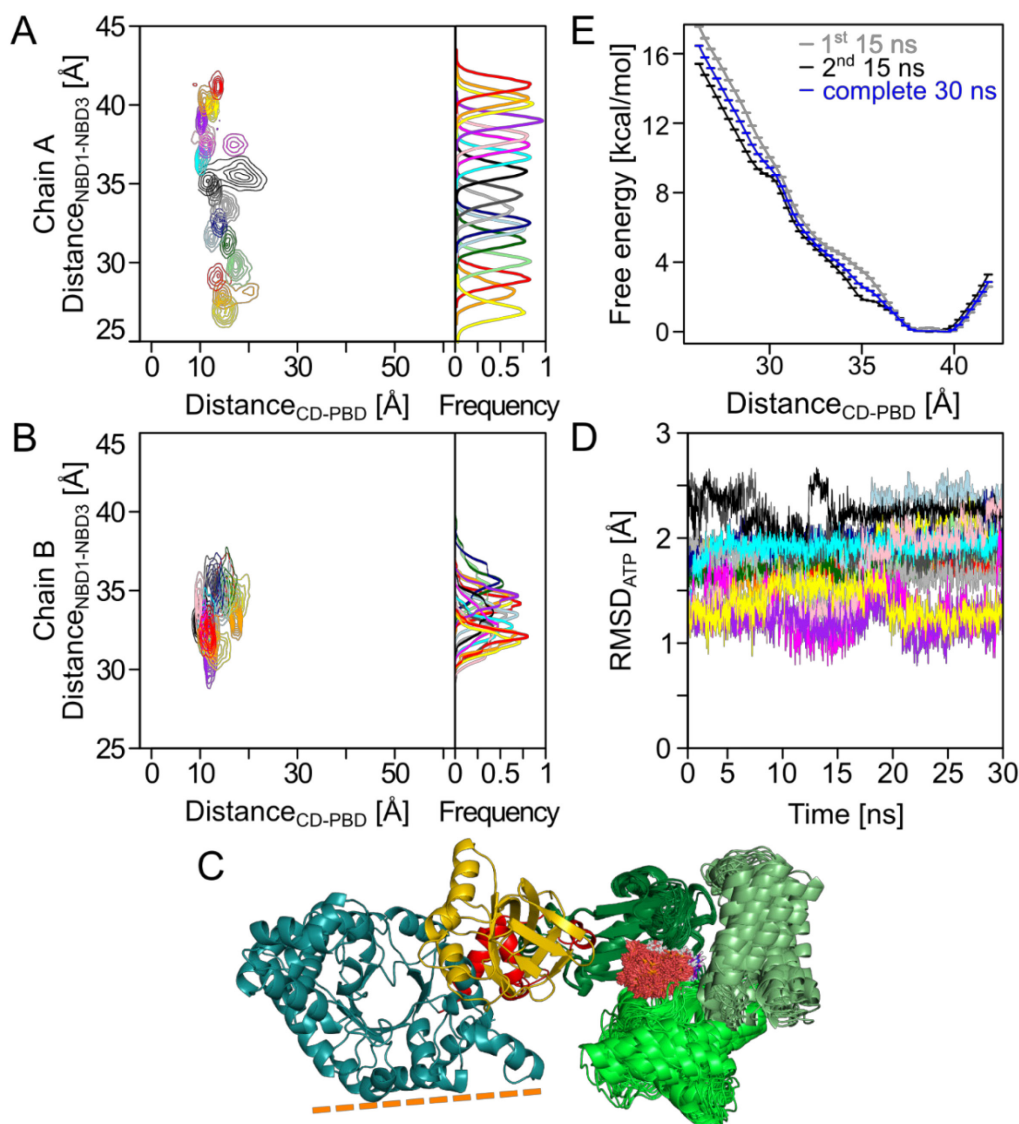
Figure 32: Overlap of umbrella sampling simulations, exemplarily shown for state c) (Figure 31) and chain A **(A)** and B **(B)** in terms of frequency distributions of the distance$_{NBD1-NBD3}$ between S215$_{C\alpha}$ – E272$_{C\alpha}$, (used as a reaction coordinate in the case of chain A), and distance$_{CD-PBD}$ between H565$_{C\alpha}$ – H456$_{C\alpha}$, which was not restrained during the simulations. The umbrella sampling simulations display well-overlapping frequency distributions of distance$_{NBD1-NBD3}$ for chain A. **(C)** Overlay of the NBD and ATP coordinates to an all-atom model of PDB ID 5JVJ. For clarity, only chain B is displayed; the dimer interface is indicated by the orange line; the state of the opening-closing motion is shown for each umbrella sampling window by an overlay of the closest-to-the-average structures of the NBD. The ATP coordinates are shown for every 20 ps for all umbrella sampling simulations of state c. **(D)** RMSD of ATP within the binding site with respect to the starting structure. **(E)** Convergence of the PMF, computed for the first 15 ns (gray), the second 15 ns (black), and the complete 30 ns (blue) simulation time per window. Taken from [2].

Several points stand out in the PMFs (Figure 33B) [2]. First, for all three systems, the global minimum is found at distance$_{NBD1-NBD3}$ ~39 Å, demonstrating that the unbound NBD with the CD facing the PBD prefers the open conformation, as also indicated by the unrestrained MD simulations of the PPDK dimer [2]. This result is also in line with PMF computations for the non-phosphorylated PPDK monomer (Figure 24), the respective distance$_{NBD1-NBD3}$ of all currently available X-ray structures with an unbound NBD and the CD facing the PBD (PDB ID: 2R82 (distance$_{NBD1-NBD3}$ = 38.7 Å), 5JVJ (38.5 Å), 1VBG (35.8 Å), and 1VBH (36.8 Å) (Supporting Table 2)), and the suggested catalytic mechanism [2]. Second, the PMFs are qualitatively indistinguishable irrespective of whether chain B is in state a) - c), corroborating the above implications that the opening-closing motion of one NBD does not directly depend on the conformational and binding state of the NBD of the other chain [2].
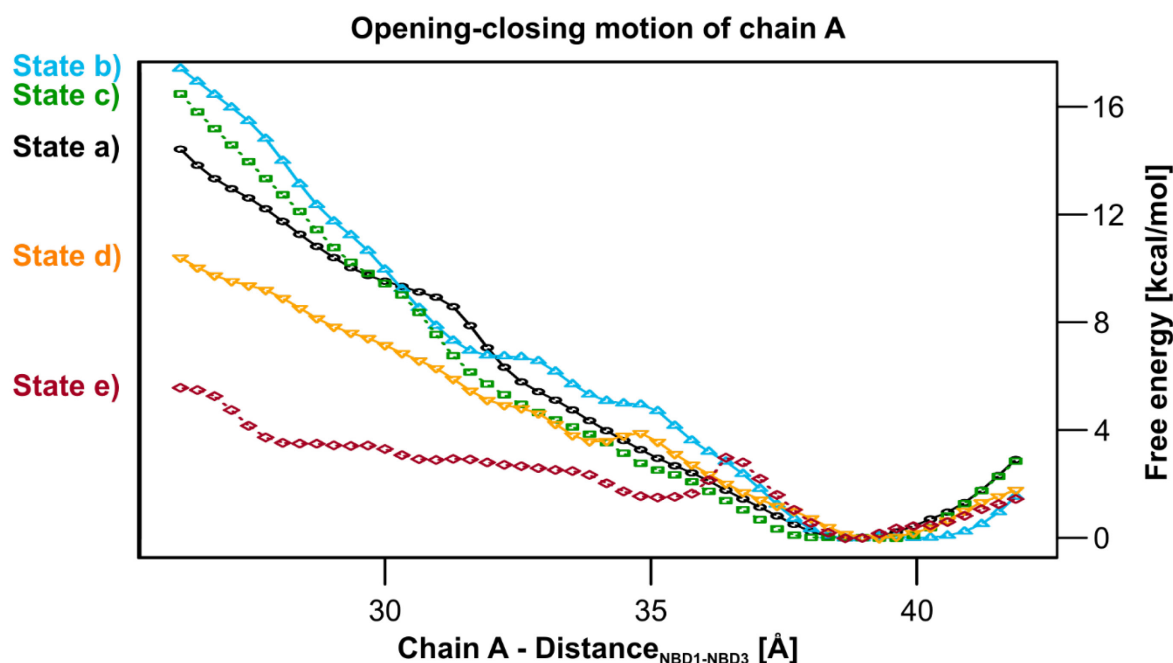
Figure 33: Potential of mean force of the opening-closing motion. PMF of the opening-closing motion of the NBD of chain A, with chain B having the CD facing the PBD with the NBD in the open, unbound conformation (a, black circle), closed, unbound conformation (b, blue triangle point up), and in the closed, ATP-bound conformation (c, green square), as well as the CD facing the NBD with the NBD in closed, unbound conformation (d, orange triangle point down) and in closed, ATP-bound conformation (e, orange dotted line). The distance$_{NBD1-NBD3}$ between Ser215$_{C\alpha}$ – Glu272$_{C\alpha}$ was used as a reaction coordinate. The standard deviation for all data points is < 0.02 kcal mol^{-1} computed by a bootstrapping procedure (see material and methods). Adapted from [2].

5.2.4 Changes in structural stability due to PPDK dimer formation

As an independent way to analyze mechanical coupling within the PPDK dimer structure, CNA [72] was performed on coordinates of chain A, chain B, or the dimer, which were all extracted from one MD simulation of the PPDK dimer with a conformation as observed in the X-ray structure (PDB ID 5JVJ) (see material and methods) [2].

In CNA, the molecular system is represented as a constraint network, which is analyzed using rigidity theory [73, 138], revealing rigid clusters and flexible links in between [2]. By performing a constraint dilution simulation [76], a stability map [128] is obtained that reports on the hierarchy of structural stability of the molecular system [2]. It does so in terms of the energy along a constraint dilution trajectory at which a rigid contact rc_{ij} between a pair of amino acids (i, j) vanishes; a rigid contact is present as long as the two residues belong to the same rigid cluster [2]. The difference stability map calculated as rc_{ij}(Dimer) – rc_{ij}(Monomer, chain A) – rc_{ij}(Monomer, chain B) then reports on the influence on structural stability due to dimerization (Figure 34) [2]. Here, only additional interactions across the interface affect the structural stability but not conformational changes, as the conformations of the dimer and the two monomer chains were extracted from one MD trajectory [2]. Note that this analysis allows revealing long-range aspects to rigidity percolation, that is, whether a region is flexible or rigid may depend on structural details that are faraway [139] [2].
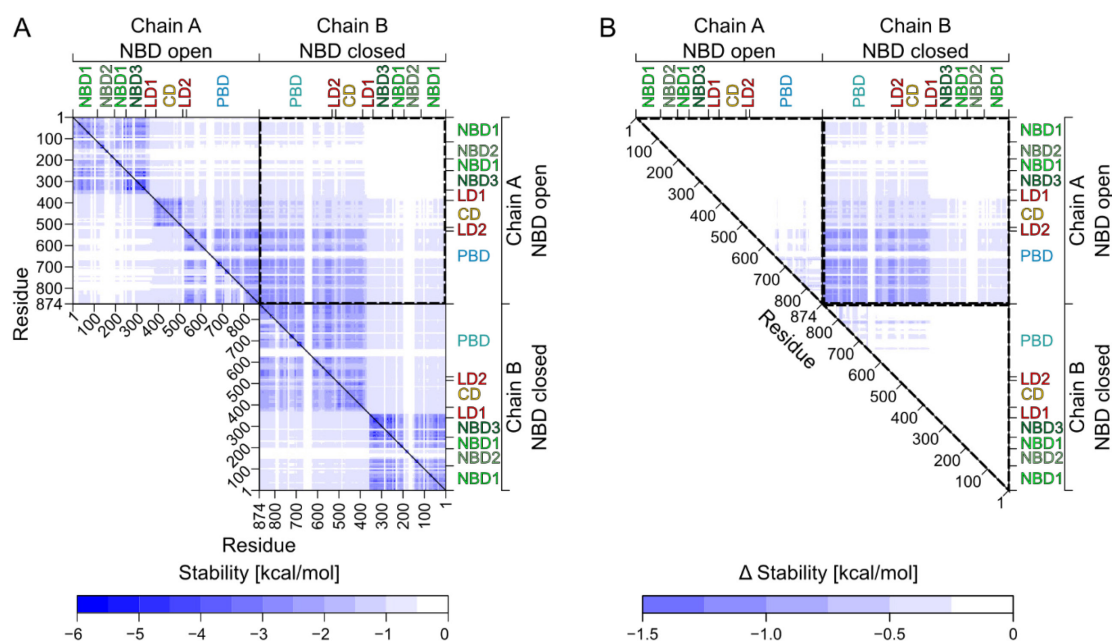
Figure 34: Differences in the structural stability of monomeric and dimeric PPDK, computed with CNA. **(A)** Stability maps depicting the energy E_{cut} (see color scale at the bottom) at which a rigid contact is lost between two residues along a constraint dilution trajectory of the PPDK dimer (upper triangle), the monomer with open NBD (Chain A, top lower triangle), and the monomer with closed NBD (Chain B, bottom lower triangle). **(B)** Difference stability map calculated as rc_{ij}(Dimer) – rc_{ij}(Monomer, chain A) – rc_{ij}(Monomer, chain B) for all residues pairs in the upper triangle and for neighboring (≤ 5 Å apart) residues in the lower triangle. The residue numbering is given on the left and on the bottom, and the proteins domains labeled on the top and on the right. Squares with round corners indicate were contacts between the NBDs would have been. Adapted from [2].

The intermolecular part of the difference stability map was projected onto the PPDK dimer structure in terms of struts connecting respective residue pairs; the color of the struts indicates to what extent a rigid contact was stabilized due to formation of the PPDK dimer from the monomers (Figure 35) [2]. Dimerization does not only increase structural stability of both PBDs, as may have been expected because of additional interactions across the interface (Figure 35G), but also that of the CD of one chain (Figure 35D/H), or the NBD of one chain (Figure 35A/I), due to the presence of the PBD of the other chain, respectively, as may have been inferred from the close distances between the respective domains [2]. Effects on structural stability due to dimerization are also observed for residue pairs on the NBD of one chain and the CD of the other chain (Figure 35B/F) [2]. This finding clearly reflects the long-range aspect to rigidity percolation [139], as the distance between both domains is at least 50 Å but additional constraints due to dimerization are only placed to model (short-range) noncovalent interactions (i.e., hydrogen bonds, salt bridges, and hydrophobic interactions between the monomers) [2]. Notably, however, no change in structural stability due to dimerization is observed for pairs of residues located on either NBD (Figure 35C) [2].
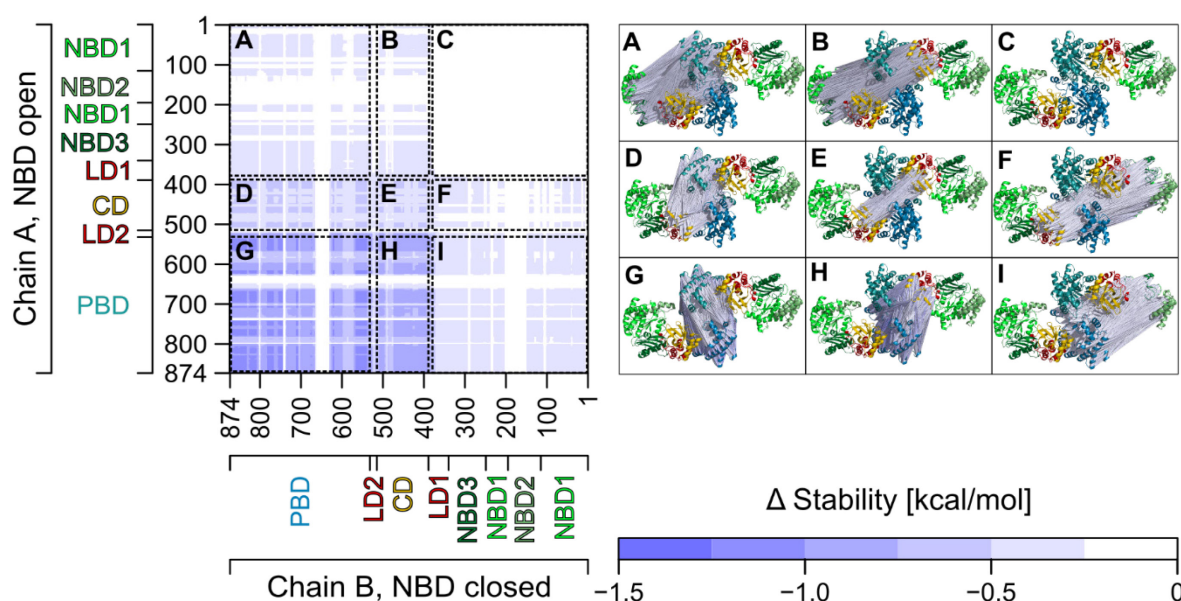
Figure 35: Effect of the dimerization on the stability of PPDK as computed by CNA. Stability map (left) of the intermolecular differences in rigid contacts with secondary structures and residue numbering given on the left and on the bottom and projected difference in the stability of rigid contacts (right) onto Cα atoms of the PPDK structure 5JVJ (missing atoms were added by homology modeling). Rigid contacts that are more stable in the dimer than in the monomer are indicated in darker blue colors (see color scale at the bottom). Between **(A)** NBD and PBD, **(B)** NBD and CD, **(C)** NBD and NBD, **(D)** CD and PBD, **(E)** CD and CD, **(F)** CD and NBD, **(G)** PBD and PBD, **(H)** PBD and CD, and **(I)** PBD and NBD of chain A and chain B, respectively. Adapted from [2].

5.2.5 Crystal packing environments of the NBDs in PDB ID 5JVJ

Finally, the structure PDB ID 5JVJ was analyzed in order to validate the presence of the two distinct conformations of the NBD (open/closed) in the crystal [2]. The PDBePISA tool [133] reveals six symmetry-related interfaces involving the NBD (Table 6, interfaces given in italic), forming two different crystal packing environments for the NBDs (Table 8) [2]. We next probed if the observed crystal packing would also be possible with the respective alternative conformational state of the NBD by superimposing an all-residue model of the open conformation of the NBD onto the closed conformation, and *vice versa*, using the coordinates of the third NBD subdomain (aa 244–340) for overlaying [2].
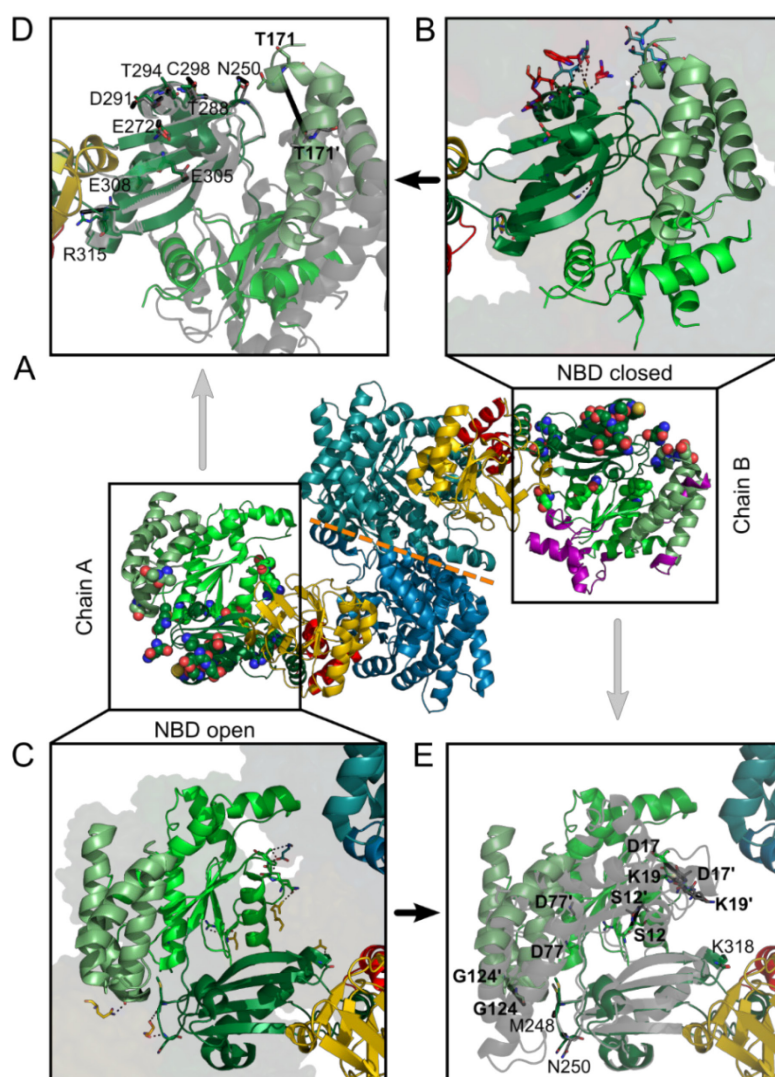
Figure 36: Crystal contacts involving the NBDs. of the PPDK structure PDB ID 5JVJ. **(A)** Cartoon representation of the PPDK dimer (including all atoms) with the PBD colored blue, the CD colored yellow, the NBD depicted by three different greens, the dimer interface indicated by the dashed orange line, and residues forming crystal contacts shown as spheres. Residues not resolved in the structure were modeled and are displayed in purple. **(B, C)** Detailed view on chain B (closed NBD, **(B)**) and chain A (open NBD, **(C)**) with residues forming crystal contacts shown as sticks. **(D, E)** Superposition of the open and closed conformation of the NBD of chain A **(D)** and B **(E)**, respectively. The respective alternative NBD conformation is shown in gray, with the distance of respective atoms involved in forming crystal contacts shown as black lines. Published in [2].

Excluding crystal contacts where the positions of the interacting atoms in the alternative NBD conformations differ by > 1 Å, or where positions of interacting atoms can be maintained by sidechain rotations or motions of flexible loops, the analysis reveals that 12 crystal contacts for chain A (open conformation) and 2 crystal contacts for chain B (closed conformation) would break if the respective alternative conformation of the NBD were located there, rather than the one found in the crystal structure (Figure 36 and Table 8, labeled in bold) [2]. While this analysis is qualitative only, it suggests that the crystal packing contributes favorably to, if not fosters, the occurrence of conformationally different states of the NBDs in the asymmetric unit [2]. Notable support for the influence of crystal packing on the conformational state of that PPDK structure arises from the fact that for chain B with a closed and likely adenine nucleotide-bound NBD, the CD is located close to the PBD, rather than the NBD [2]. This arrangement is unexpected in view of the enzyme mechanism (Figure 26) and leads to an energetically unfavorable conformational state of that chain (see Figure 24) [2].

5.2.6 Coupling of the opening-closing motions mediated by the CDs

Although we did not find evidence for a *direct* coupling of the opening-closing motion or the structural stability of the two NBDs, the fact that PPDK dissociates in a cold-dependent manner and then becomes inactive [134] leads to the question as to why only the PPDK dimer is active [2]. Previously, we showed for monomeric PPDK that the CD motion and the opening-closing motion of the NBD is coupled [1] [2]. Above, we showed by CNA that the structural stability of the CD and the NBD of one monomer is influenced by the presence of another monomer (Figure 35A/B/F/I) [2]. Taken together, it is tempting to hypothesize that PPDK *dimerization influences the opening-closing motions of the NBDs*, and that this influence is *mediated via the CDs of both chains* (Figure 37) [2].
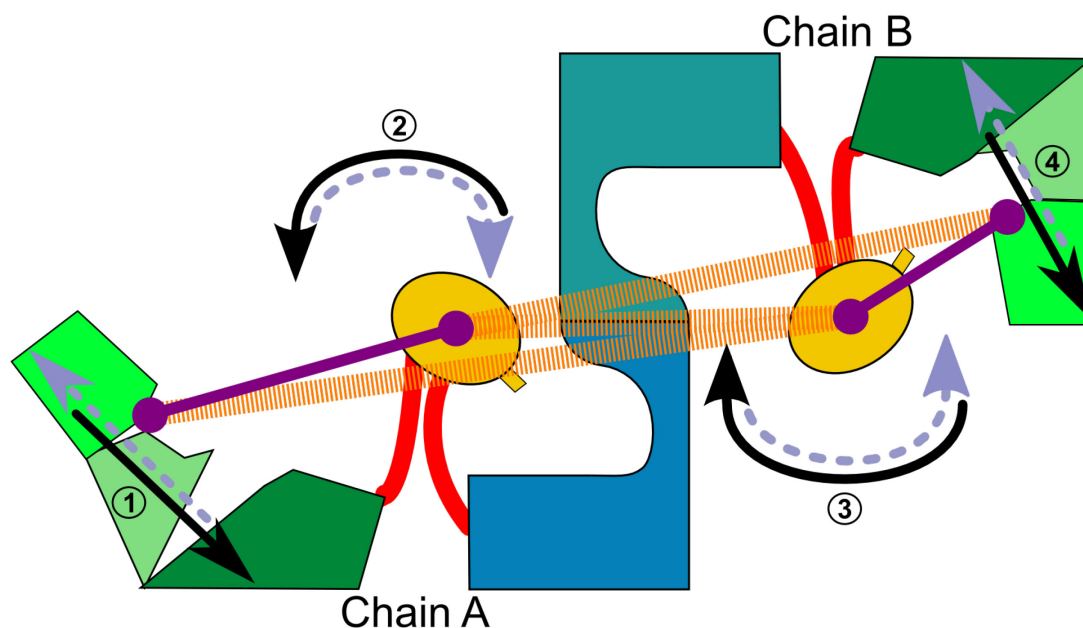
Figure 37: Suggested coupling of motions and structural stability in the PPDK dimer. Intramolecular coupling of the swiveling motion of the CD with the opening-closing motion of the NBD (indicated by deep purple bars) was demonstrated for the PPDK monomer (see chapter 5.5). The results from the study on the PPDK dimer furthermore indicate that PPDK dimerization influences the opening-closing motion of an NBD *mediated via the CD of the other chain*. Mutual increases in structural stability due to dimerization as revealed by CNA are indicated by orange dashed lines. Our findings lend support to the hypothesis of a binding change mechanism in the PPDK dimer: When the NBD of chain A is closing (1), the CD of chain A swivels towards the NBD (2), the CD of chain B swivels towards the PBD (3), and the NBD of chain B is opening (4) (illustrated by black arrows) and *vice versa* (illustrated by gray arrows). Taken from [2].

Initial support for this hypothesis is found by comparing the PMFs of the opening-closing motion of the NBD, when the CD is near the PBD, in the monomeric state ($\Delta G_{\text{open} \rightarrow \text{closed}}$ ~20 kcal mol^{-1}; see Figure 24) and in the dimeric state (with the CD of the other chain facing the PBD = state a), $\Delta G_{\text{open} \rightarrow \text{closed}}$ ~16 kcal mol^{-1}; Figure 33) [2]. The comparison indicates that the closing of the NBD is less energetically unfavorable in the dimeric state, which could explain why the PPDK dimer is the active form [2]. Further support for this hypothesis arises from additional PMF computations of the opening-closing motion of the NBD (chain A) of dimeric PPDK, in which the CD of the same chain faces the PBD but that of chain B faces the NBD (states d) and e), Figure 33). The PMF of state d) reveals a lower free energy for the closing of the NBD of chain A (by ~6 kcal mol^{-1} compared to states a) - c) at distance$_{\text{NBD1-NBD3}}$ ~26 Å; Figure 33) [2]. The ATP-bound conformation of chain B (state e)), which is expected to stabilize the conformation of the NBD and CD of chain B, fosters this effect in that the free energy for the closing of the NBD of chain A is now lower by ~10 kcal mol^{-1} (at distance$_{\text{NBD1-NBD3}}$ ~26 Å; Figure 33) compared to states a) - c) [2]. Taken together with our previous results (chapter 5.5), these findings suggest that the conformational states of the CDs of *both* chains in the PPDK dimer affect the energetics of the opening-closing motions of the two NBDs [2]. However, a thorough validation of this hypothesis would require

PMF computations considering both distances$_{\text{NBD1-NBD3}}$ as well as two reaction coordinates characterizing the swiveling motion of each CD simultaneously, resulting in a 4D-PMF [2]. Given the computational burden we faced when performing 2D-PMF calculations on monomeric PPDK, such calculations are beyond the scope of the present work [2].

Summary, significance, and perspectives

The aim of this thesis was to analyze the internal motion of PPDK in order to resolve PPDK's mode of action. This task was subdivided into two manageable parts.

First, my calculations revealed an intramolecular coupling between the CD motion and the opening-closing motion of the NBD in the PPDK *monomer* using all at that time available PPDK structures, molecular dynamics (MD) simulations and configurational free energy (potential of mean force (PMF)) computations. Further, the unrestrained MD simulations and 1D and 2D PMF calculations predicted that the swiveling motion proceeds via a conformational intermediate where the CD is located roughly in between the PBD and NBD. The crystal structure of PPDK from *Flaveria trinervia* independently confirmed this, until then, unknown intermediate (PDB ID 5JVJ) [1]. The molecular simulations and crystallographic data furthermore indicated that PPDK might employ a Brownian ratchet mechanism biasing thermal fluctuations in order to generate a net directional CD motion [1].

Secondly, analyses were performed on the PPDK *dimer* validating an alternate binding change mechanism proposed based on the crystal structure PDB ID 5JVJ. Results from unbiased MD simulations, cross-correlation analysis, PCA, PMF computations, and CNA did not provide evidence for a *direct* coupling of the opening-closing motion or the structural stability of the two NBDs in dimeric PPDK [2]. However, I provided results that support the hypothesis that PPDK dimerization does influence the opening-closing motion of the NBDs, and that this influence is mediated via the CDs of both chains [2]. Such an influence would be a prerequisite for an alternate binding change mechanism to occur (Figure 37) and could result in an overall mechanism of dimeric PPDK as displayed in Figure 38 [2]. To the best of our knowledge, this is the first time that a possible explanation has been suggested as to why only dimeric PPDK is active [2]. The suggestion would add another example to the rare occurrences of asymmetric organization of protein structures, and, as in other cases [130, 140], the asymmetry would serve a special role, here enzyme activation [2].

Besides further comprehensive PMF computations, it should be interesting to investigate the structural dynamics and function of dimeric PPDK by single molecule spectroscopy, applying, *e.g.*, Förster resonance energy transfer measurements, in order to map the time-scales of exchange and the pathways between conformational states [141] [2].
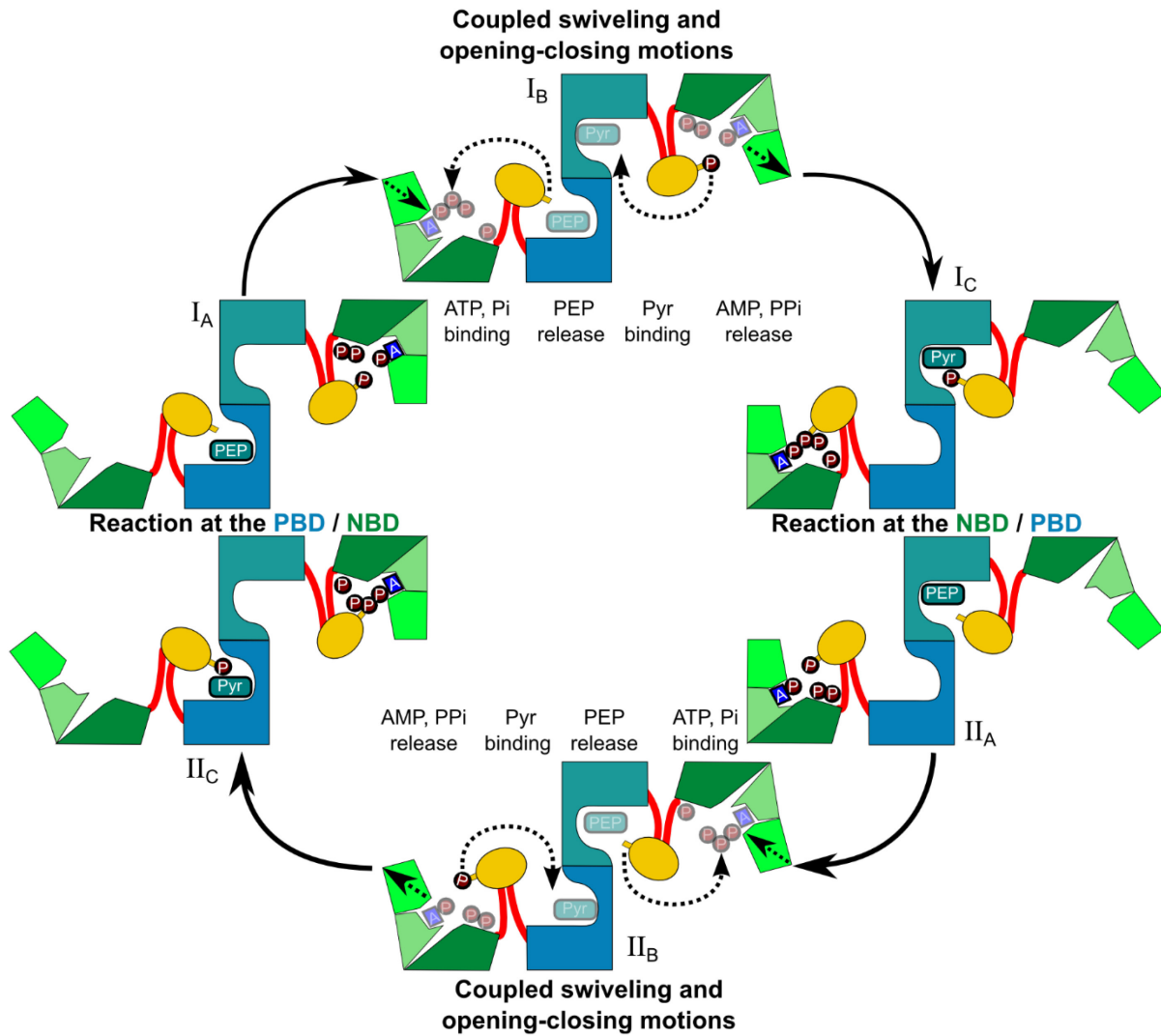
Figure 38: Proposed mechanism of dimeric PPDK. The mechanism proposed for PPDK involves the following steps: (II$_C$→I$_A$) Transfer of a phosphoryl group from ATP to H456 of the CD of chain B. (I$_B$) ATP binding to the NBD of chain A leads to the closing of this NBD, which is coupled with the CD motion of chain A swiveling from the PBD to the NBD, and the CD motion of chain B swiveling from the NBD to the PBD; the latter CD transports the phosphoryl group from the NBD to the PBD of chain B. (I$_C$→II$_A$) The phosphoryl group bound to H456 of the CD of chain B is transferred to pyruvate at the PBD of chain B to generate PEP. Steps II$_{A-C}$ are identical to I$_{A-C}$ but now the functional roles of chains A and B are exchanged. Between II$_C$ and I$_A$ as well as I$_C$ and II$_A$, reactions (1) and (2) mentioned in the Background section take place. Taken from [2].

Appendix

Supplementary Figures
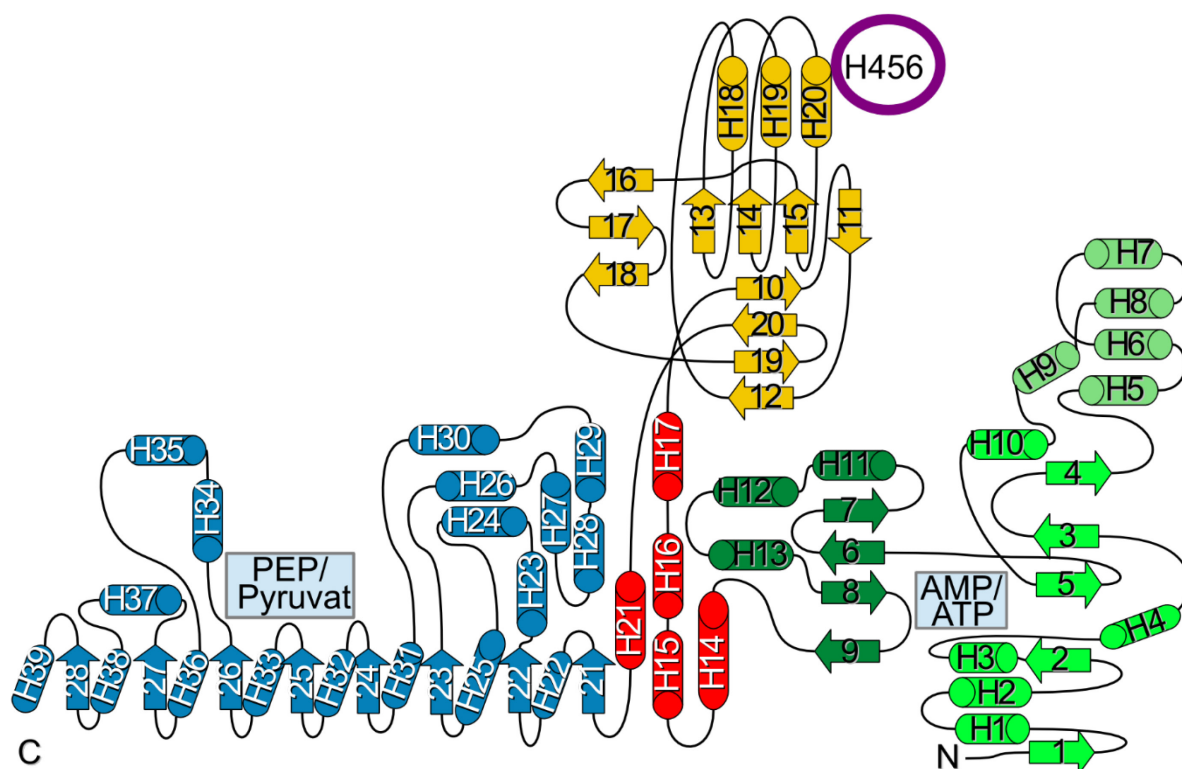
Figure 39: Topology representation of the PPDK structure, with CD (colored yellow) containing the catalytic histidine (circled), connected via LD (colored red) to PBD (colored blue) and to the NBD (with the three subdomains depicted by three different greens). Helices and sheets are numbered and represented as tubes and arrows, respectively. Taken from [1].

Nucleotide-Binding Domain (NBD)
- NBD1 Subdomain: N-Terminal Subdomain
- NBD2 Subdomain: Four-helix bundle insert
- NBD3 Subdomain: C-Terminal Subdomain

β	Pos	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	α
β1		K/K	K/K	R/R	V/V	F/F	T/T	F/F	G/G	K/K	G/G	
	10	R/R	S/S	E/E	G/G	N/N	R/K	D/D	M/M	K/K	S/S	α1
	20	L/L	L/L	G/G	G/G	K/K	G/G	A/A	N/N	L/L	A/A	α2
	30	E/E	M/M	S/A	S/S	I/I	G/G	L/L	S/S	V/V	P/P	
β2	40	P/P	G/G	L/L	T/T	I/I	S/S	T/T	E/E	A/A	C/C	α3
	50	E/E	E/E	Y/Y	Q/Q	Q/Q	N/N	G/G	K/K	S/K	L/L	
	60	P/P	P/P	G/G	L/L	W/W	D/D	E/E	I/I	S/L	L/E	α4
	70	G/G	L/L	D/R	Y/Y	V/V	Q/Q	K/K	E/E	M/M	S/S	
	80	A/A	S/S	L/L	G/G	D/D	P/P	S/S	K/K	P/P	L/L	
β3	90	L/L	L/L	S/S	V/V	R/R	S/S	G/G	A/A	A/A	I/I	
	100	S/S	M/M	P/P	G/G	M/M	M/M	D/D	T/T	V/V	L/L	
β4	110	N/N	L/L	G/G	L/L	N/N	D/D	E/E	V/V	V/V	A/A	α5
	120	G/G	L/L	A/A	G/G	K/K	S/S	G/G	A/A	R/R	F/F	α6
	130	A/A	Y/Y	D/D	S/S	Y/Y	R/R	R/R	F/F	L/L	D/D	
	140	M/M	F/F	G/G	N/N	V/V	V/V	M/M	G/G	I/I	P/P	
	150	H/X	S/S	L/L	F/F	D/D	E/E	K/K	L/L	E/E	Q/E	α7
	160	M/M	K/K	A/A	E/E	K/K	G/G	I/V	H/X	L/L	D/D	
	170	T/T	D/D	L/L	T/T	A/A	A/A	D/D	L/L	K/K	D/D	
	180	L/L	V/V	E/E	K/Q	Y/Y	K/K	N/N	V/V	Y/Y	V/V	α8
	190	E/E	A/A	K/K	G/G	E/E	K/K	F/F	P/P	T/T	D/D	
	200	P/P	K/K	K/K	Q/Q	L/L	E/E	L/L	A/A	V/V	N/N	α9
	210	A/A	V/V	F/F	D/D	S/S	W/W	D/D	S/S	P/P	R/R	α10
β5	220	A/A	N/N	A/A	K/K	Y/Y	R/R	S/S	I/I	N/N	Q/Q	
	230	T/T	G/G	L/L	K/K	G/G	T/T	A/A	V/V	N/N	I/I	
	240	Q/Q	S/C	M/M	V/V	F/F	G/G	N/N	M/M	G/G	N/N	
β6	250	T/T	S/S	G/G	T/T	G/G	V/V	L/L	F/F	T/T	R/R	
	260	N/N	P/P	S/S	T/T	G/G	E/E	K/K	K/K	L/L	Y/Y	
β7	270	G/G	E/E	F/F	L/L	I/V	N/N	A/A	Q/Q	G/G	E/E	α11
	280	D/D	V/V	V/V	A/A	G/G	I/I	R/R	T/T	P/P	E/E	
	290	D/D	L/L	G/A	T/T	M/M	E/E	T/T	C/C	M/M	P/P	α12
	300	E/E	A/A	Y/Y	K/R	E/E	L/L	V/V	E/E	N/N	C/C	α13
	310	E/K	I/I	L/L	E/E	R/R	H/X	Y/Y	K/K	D/D	M/M	
β8	320	M/M	D/D	I/I	E/E	F/F	T/T	V/V	Q/Q	E/E	N/N	
β9	330	R/R	L/L	W/W	M/M	L/L	Q/Q	C/C	R/R	T/T	G/G	

Linker Domain (LD1)

β	Pos	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	α
	340	K/K	R/R	T/T	G/G	K/K	G/G	A/A	V/V	R/R	I/I	α14
	350	A/A	V/V	D/D	M/M	V/V	N/N	E/E	G/G	L/L	I/I	
	360	D/D	T/T	R/R	T/T	A/A	I/I	K/K	R/R	V/V	E/E	α15
	370	T/T	Q/Q	H/X	L/L	D/D	Q/Q	L/L	L/L	H/X	P/P	α16
	380	Q/Q	F/F	E/E	A/A	N/N	P/P	S/S	A/A	Y/Y	K/K	α17

Central Domain (CD)

β	Pos	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	α
β10 β11 β12	390	H/X	V/V	V/V	A/A	T/T	G/G	L/L	P/P	A/A	S/S	
	400	P/P	G/G	A/A	A/A	V/V	G/G	Q/Q	V/V	C/V	F/F	
	410	S/S	A/A	E/E	D/D	A/A	E/E	T/T	W/W	H/X	A/A	α18
β13	420	Q/Q	G/G	K/K	S/S	A/A	I/I	L/L	V/V	R/R	T/T	
	430	E/E	T/T	S/S	P/P	E/E	D/D	V/V	G/G	G/G	M/M	α19
β14	440	H/X	A/A	A/A	A/A	G/G	I/I	L/L	T/T	A/A	R/R	
	450	G/G	G/G	M/M	T/T	S/S	H/X	A/A	A/A	V/V	V/V	α20
β15	460	A/A	R/R	G/G	W/W	G/G	K/K	C/C	C/C	V/V	S/S	
β16	470	G/G	C/C	A/A	D/D	I/I	R/R	V/V	N/N	D/D	D/D	
β17 β18	480	M/M	K/K	I/V	F/L	T/T	I/I	G/G	D/D	R/R	V/V	
β19	490	I/I	K/K	E/E	G/G	D/D	W/W	L/L	S/S	L/L	N/N	
β20	500	G/G	T/S	T/T	G/G	E/E	V/V	I/I	L/L	G/G	K/K	

Linker Domain (LD2)

| β	Pos	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	α
	510	Q/Q	L/L	L/L	A/A	P/P	P/P	A/A	M/M	S/S	N/N	
	520	D/D	L/L	E/E	I/T	F/F	M/M	S/S	W/W	A/A	D/D	α21
β21	530	Q/Q	A/V	R/R	R/R	L/L	K/K	V/V	M/M	A/A	N/N	

PEP/Pyruvate-Binding Domain (PBD)

β	Pos	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	α
	540	A/A	D/D	T/T	P/P	N/N	D/D	A/A	L/L	T/T	A/A	α22
β22	550	R/R	N/N	N/N	G/G	A/A	Q/Q	G/G	I/I	G/G	L/L	α23
	560	C/C	R/R	T/T	E/E	H/X	M/M	F/F	F/F	A/A	S/S	
	570	D/D	E/E	R/R	I/I	K/K	A/A	V/V	R/R	K/K	M/M	α24
	580	I/I	M/M	A/A	V/V	T/T	P/P	E/E	Q/Q	R/R	K/K	
	590	V/A	A/A	L/L	D/D	L/L	L/L	L/L	P/P	Y/Y	Q/Q	α25
	600	R/R	S/S	D/D	F/F	E/E	G/G	I/I	F/F	R/R	A/A	
β23	610	M/M	D/D	G/G	L/L	P/P	V/V	T/T	I/I	R/R	L/L	
	620	L/L	D/D	P/P	P/P	L/L	H/X	E/E	F/F	L/L	P/P	α26
	630	E/E	G/G	D/D	L/L	E/E	H/H	I/I	V/V	N/N	E/E	α27
	640	L/L	A/A	V/V	D/D	T/T	G/G	M/M	S/S	A/E	D/D	α28
	650	E/E	I/I	Y/Y	S/S	K/K	I/I	E/E	N/K	L/L	S/S	
	660	E/E	V/V	N/N	P/P	M/M	L/L	G/G	F/F	R/R	G/G	α29
	670	C/C	R/L	L/L	G/G	I/I	S/S	Y/Y	P/P	E/E	L/L	α30
	680	T/T	E/E	M/M	Q/Q	V/V	R/R	A/A	I/I	F/F	Q/Q	α31
	690	A/A	A/A	V/V	S/S	M/M	T/N	N/N	Q/Q	G/G	V/V	
β24	700	T/T	V/V	I/I	P/P	E/E	I/I	M/M	V/V	P/P	L/L	
	710	V/V	G/G	T/T	P/P	Q/Q	E/E	L/L	R/R	H/X	Q/Q	
	720	I/I	S/G	V/V	I/I	R/R	G/G	V/V	A/A	A/A	N/N	α32
	730	V/V	F/F	A/A	E/E	M/M	G/G	V/L	T/T	L/M	E/D	
β25	740	Y/Y	K/K	V/V	G/G	T/T	M/M	I/I	E/E	I/I	P/P	
	750	R/R	A/A	A/A	L/L	I/I	A/A	E/E	E/E	I/I	G/A	α33
β26	760	K/K	E/E	A/A	D/E	F/F	F/F	S/S	F/F	G/G	T/T	α34
	770	N/N	D/D	L/L	T/T	Q/Q	M/M	T/T	F/F	G/G	Y/Y	
	780	S/S	R/R	D/D	D/D	V/V	G/G	K/K	F/F	L/L	Q/Q	α35
	790	I/I	Y/Y	L/L	A/S	Q/Q	G/G	I/I	L/L	Q/Q	H/X	
	800	D/D	P/P	F/F	E/E	V/V	I/L	D/D	Q/Q	K/K	G/G	
	810	V/V	G/G	Q/Q	L/L	I/I	K/K	M/M	A/A	T/T	E/E	α36
β27	820	K/K	G/G	R/R	A/A	A/A	N/N	P/P	S/N	L/L	K/K	
	830	V/V	G/G	I/I	C/C	G/G	E/E	H/X	G/G	G/G	E/E	α37
	840	P/P	S/S	S/S	V/V	A/A	F/F	F/F	D/D	G/G	V/V	α38
β28	850	G/G	L/L	D/D	Y/Y	V/V	S/S	C/C	S/S	P/P	F/F	
	860	R/R	V/V	P/P	I/I	A/A	R/R	L/L	A/A	A/A	A/A	α39
	870	Q/Q	V/V	I/V	V/V							

Figure 40: PPDK sequence from *F. trinervia* / *F. pringlei*. One-letter code of the amino acid sequence of *F. trinervia* (first letter) and *F. pringlei* (second letter) with protein (sub)domains and background indicated by colored boxes and the secondary structure by background colors (blue for helices and yellow for sheet) as well as labeled on the right and left, respectively.
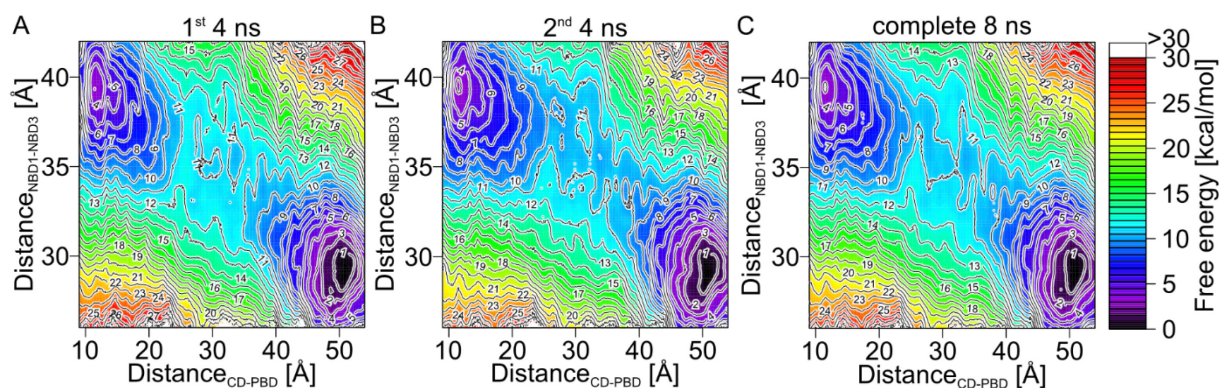
Figure 41: Convergence of the 2D PMF calculation of non-phosphorylated *Ft*PPDK. 2D PMFs computed using umbrella sampling along the two reaction coordinates (Figure 21) are shown for **(A)** only the 1st half of the sampling time, **(B)** the 2nd half, and **(C)** the complete sampling time of 8 ns per window. Taken from [1].

Supplementary Tables

Table 2: All currently known PPDK structures [a]

Cluster [b]	Organism	PDB ID	Comment	Distance CD-PBD [d]	Distance NBD1-NBD3 [e]	Reference
I	*C. symbiosum*	1DIK	-	45.5	28.1	[83]
I	*C. symbiosum*	2DIK	R337A	45.8	28.0	[142]
I	*C. symbiosum*	1GGO	T453A	45.5	28.6	[143]
I	*C. symbiosum*	1JDE	K22A	46.2	28.1	[101]
I	*C. symbiosum*	1KBL	-	45.6	28.6	[144]
I	*C. symbiosum*	1KC7	PEP analog bound	45.4	28.4	[144]
II	***F. trinervia***	**5JVJ**	**Chain B; PEP bound; CD and NBD incomplete**	**11.6**	**30.3**	**[1]**
II	***F. trinervia***	**5JVL**	**Chain A; PEP and ATP analog bound**	**12.9**	**28.4**	**[1]**
II	***F. trinervia***	**5JVL**	**Chain C; PEP and ATP analog bound**	**9.8**	**29.9**	**[1]**
II	***F. trinervia***	**5JVL**	**Chain D; PEP and ATP analog bound**	**9.3**	**28.7**	**[1]**
III	*C. symbiosum*	2R82	R219E/E271R/S262D	10.4	38.7	[144]
III	*Z. mays*	1VBG	-	16.4	35.8	[90]
III	*Z. mays*	1VBH	PEP bound	17.2	36.8	[90]
III	***F. trinervia***	**5JVJ**	**Chain A; PEP bound**	**13.4**	**38.5**	**[1]**
IV	***F. pringlei***	**5JVN**	**PEP and ATP analog bound**	**25.3**	**37.5**	**[1]**
V	*T. brucei*	2X0S	-	49.2	35.5	[89]
-[c]	***F. trinervia***	**5JVL**	**Chain B; PEP bound; NBD not resolved**	**15.5**	**-**	**[1]**
-[c]	*C. symbiosum*	2FM4	NMR-derived structure of the CD	-	-	[145]

[a] Crystal structures described in Minges and Ciupka *et al.* [1] are shown in bold. Table taken from [1].

[b] The cluster number relates to Figure 18 with cluster I (closed NBD with CD facing NBD); cluster II (closed NBD with CD facing PBD); cluster III (open NBD with CD facing PBD); cluster IV (open NBD with CD in between PBD and NBD); and cluster V (open NBD with CD facing NBD). Within each cluster, structures are sorted in chronological order.

[c] Structure not used for analysis due to missing domains.

[d] Measured between C_α atoms of residues 455-564 (*C. symbiosum*); 456-565 (*F. trinervia*); 456-565 (*F. pringlei*); 482-589 (*T. brucei*); 458-567 (*Z. mays*); in Å.

[e] Measured between C_α atoms of residues 214-271 (*C. symbiosum*); 215-272 (*F. trinervia*); 215-272 (*F. pringlei*); 218-275 (*T. brucei*); 217-274 (*Z. mays*); in Å.

Table 3: MD simulations of PPDK [a]

Conformation chain A	Conformation chain B	Repetitions	Simulation length [b]
I: NBD closed, CD facing NBD	-	3	600
II: NBD closed, CD facing PBD	-	3	600
III: NBD open, CD facing PBD	-	3	600
III: NBD open, CD facing PBD	-	1	1500
III: NBD open, CD facing PBD	-	1	700
IV: NBD open, CD between PBD and NBD	-	3	600
V: NBD open, CD facing NBD	-	3	600
NBD closed, CD facing PBD	NBD closed, CD facing PBD	3	300
NBD closed, CD facing PBD	NBD open, CD facing PBD	3	300
NBD open, CD facing PBD	NBD open, CD facing PBD	3	300
NBD closed, ATP bound, CD facing PBD	NBD closed, ATP bound, CD facing PBD	3	300

[a] Table adapted from [1, 2].
[b] In ns.

Table 4: Umbrella sampling simulations of the PPDK monomer [a]

System	Windows	Simulation length [b]
NBD open (non- phosphorylated)	42	42.5 (1785)
NBD closed while the CD approached and vice versa (non-phosphorylated)	42	42.5 (1785)
NBD closed (non- phosphorylated)	42	42.5 (1785)
NBD open (phosphorylated)	42	42.5 (1785)
NBD closed while the CD approached and vice versa (phosphorylated)	42	42.5 (1785)
NBD closed (phosphorylated)	42	42.5 (1785)
2D	45 * 17 (782) [c]	9 (7035)

[a] Table taken from [1].

[b] for each window (Total simulation time for all windows), in ns.

[c] For $distance_{CD-PBD}$ and $distance_{NBD1-NBD3}$ respectively (total number of windows).

Table 5: Umbrella sampling simulations of PPDK dimer [a]

Dimer arrangement	Conformation chain B	Windows	Simulation length [b]
a)	CD facing PBD, NBD open, unbound	17	30 (510)
b)	CD facing PBD, NBD closed, unbound	17	30 (510)
c)	CD facing PBD, NBD closed, ATP bound	17	30 (510)
d)	CD facing NBD, NBD closed, unbound	17	30 (510)
e)	CD facing NBD, NBD closed, ATP bound	17	30 (510)

[a] Taken from [2].

[b] For each window (Total simulation time for all windows), in ns.

Table 6: Analysis of the PPDK interfaces by PDBePISA [a]

Structure 1			Structure 2		Interface					
Range	^iN res [b]	Symmetry operation	Range	^iN res [b]	Interface area [c]	Δ^iG [d]	Δ^iG P-value	N HB [e]	N SB [f]	CSS [g]
B PBD	**56**	**x, y, z**	**A PBD**	**53**	**2077.4**	**-17.4**	**0.370**	**34**	**10**	**0.541**
B LD NBD3	*36*	*-x-1/2, y-1/2, -z*	*B LD NBD3*	*42*	*1180*	*-10.6*	*0.344*	*13*	*3*	*0*
A PBD CD	44	-x, y, -z+1	A PBD CD	43	1145.8	-6.6	0.443	10	0	0
A CD	*19*	*-x+1/2, y-1/2, -z+1*	*A NBD2-3*	*23*	*440.2*	*-2.4*	*0.500*	*4*	*2*	*0*
A CD	*12*	*x, y-1, z*	*A NBD1, 3*	*14*	*364.4*	*-0.9*	*0.578*	*3*	*5*	*0*
B PBD	*8*	*x, y-1, z*	*B NBD1*	*8*	*203.2*	*-2.2*	*0.403*	*4*	*0*	*0*
B PBD	4	-x, y, -z	B PBD	4	161.3	-1.8	0.590	2	0	0
A PBD	*3*	*x-1/2, y-1/2, z*	*A NBD2*	*5*	*102.3*	*-2.1*	*0.615*	*0*	*0*	*0*
B NBD2	*4*	*x-1/2, y-1/2, z*	*B PBD*	*3*	*94.8*	*-1.1*	*0.501*	*0*	*0*	*0*

[a] The dimer interface of PPDK is shown in bold; interfaces involving the NBD are shown in italics. Taken from [2].
[b] Number of interface residues.
[c] In Å^2.
[d] Solvation free energy gain upon formation of the interface, in kcal mol^{-1}.
[e] Number of hydrogen bonds.
[f] Number of salt bridges.
[g] Complexation significance score.

Table 7: Conservation of dimer interface residues [a]

Interface residues[b]	Interacting atoms[c]	ASA [d]	BSA [e]		$\Delta^i G$ [f]	Conservation[g]
E657	O	71.8	1.11	|	-0.01	27.6
N658	C_α, C, **O**, $O_{\delta 1}$, $N_{\delta 2}$	111.4	28.85	|||	-0.28	25.8
L659	C_α, O	20.5	0.50	|	0.01	98.9
S660	**O**	82.6	8.83	||	-0.1	55.6
E661	C_α, C_β, C_γ, C_δ, $O_{\varepsilon 2}$	61.4	29.90	|||||	0.29	100.0
V662	**N**, O, C_β, $C_{\gamma 1}$, $C_{\gamma 2}$	143.9	64.79	|||||	0.49	67.2
N663	N, C_α, O, C_β, C_γ, **$O_{\delta 1}$**, **$N_{\delta 2}$**	97.3	96.86	|||||||||	-0.71	100.0
P664	C_g, C_δ	66.4	29.97	|||||	0.48	100.0
M665	C_β, C_γ, S_δ	48.1	14.38	|||	0.25	100.0
L666	C, **O**, C_β, C_γ, $C_{\delta 1}$, $C_{\delta 2}$	102.4	102.23	|||||||||	1.17	80.5
G667	C_α, C, O	18.3	18.30	|||||||||	0.29	100.0
F668	N, C_α, O, C_β, C_γ, $C_{\delta 1}$, $C_{\delta 2}$, $C_{\varepsilon 1}$, $C_{\varepsilon 2}$, C_ζ	66.1	49.05	|||||||	0.75	74.2
R672	C_a, O, C_β, C_γ, C_δ, N_ε, C_ζ, **$N_{\eta 1}$**, **$N_{\eta 2}$**	77.0	75.36	|||||||||	-0.6	100.0
I675	O, $C_{\gamma 2}$, $C_{\delta 1}$	47.0	46.40	|||||||||	0.58	52.7
S676	C_α, O, C_β, O_γ	61.6	45.12	|||||||	0.09	42.2
Y677	C_ζ	95.8	0.16	|	0	85.8
P678	C_γ	43.2	2.18	|	0.03	99.7
L710	$C_{\delta 2}$	0.8	0.17	|||	0	99.9
G712	C_α	8.4	4.16	|||||	0.07	63.7
T713	C_β, $O_{\gamma 1}$, $O_{\gamma 2}$	38.8	34.94	|||||||||	0.47	29.2
Q715	C, C_β, C_γ, C_δ, $O_{\varepsilon 1}$, $N_{\varepsilon 2}$	102.9	57.64	|||||	0.54	44.4
E716	N, C_α, C_γ, C_δ, $O_{\varepsilon 1}$, **$O_{\varepsilon 2}$**	31.3	29.04	|||||||||	0.1	100.0
H719	C_β, $C_{\delta 2}$, **$N_{\eta 2}$**	80.2	9.59	|||||	-0.26	29.1
I749	$C_{\gamma 1}$, $C_{\gamma 2}$, $C_{\delta 1}$	22.8	2.25	|||||||||	0.36	52.3
P750	C_β, C_γ, C_δ	33.0	2.62	|||||||||	0.52	99.7
R751	**N**, C_α, C_β, C_γ, C_δ, **N_ε**, C_ζ, **$N_{\eta 2}$**	116.3	95.03	||||||||	-0.73	99.9
L754	O, C_β, $C_{\delta 1}$, $C_{\delta 2}$	89.1	87.55	|||||||||	1.2	81.9
I755	$C_{\gamma 2}$	29.4	24.53	|||||||||	0.39	40.3
E758	C_δ, **$O_{\varepsilon 1}$**, **$O_{\varepsilon 2}$**	80.1	27.93	||||	-0.24	62.1
Q775	O, C_γ, $O_{\varepsilon 1}$	5.6	5.64	|||||||||	-0.01	100.0
M776	C_α, O, C_β, C_γ, S_δ, C_ε	38.7	38.73	|||||||||	0.34	57.8
T777	C_α, **O**, $C_{\gamma 2}$	52.5	52.48	||||||||	0.14	88.1
F778	C_α, C, **O**, C_γ, $C_{\delta 2}$, $C_{\varepsilon 1}$, $C_{\varepsilon 2}$, C_ζ	63.0	57.65	||||||||	0.38	75.4
G779	N, C_α, **C**	64.3	64.25	|||||||||	0.23	99.0
T780	C_α, C, $C_{\delta 1}$, $C_{\varepsilon 1}$, O_η	24.3	21.30	|||||||	0.16	38.7
S781	**N**, C_α, C_β, **O_γ**	43.0	42.93	|||||||||	0.32	100.0
R782	C_γ	88.6	0.16	|	0	100.0
D783	C_β, C_γ, $O_{\delta 1}$, **$O_{\delta 2}$**	114.1	32.05	|||	0.15	100.0
D784	C, O , C_β, C_γ, **$O_{\delta 1}$**, **$O_{\delta 2}$**	76.3	66.65	|||||||||	-0.25	99.9
V785	C_α	12.6	0.33	|	0.01	80.1
K787	O, C_β, C_γ, C_δ, C_ε, **N_ζ**	168.0	104.09	||||||	-0.77	54.4
F788	C_β, C_γ, $C_{\delta 1}$, $C_{\delta 2}$, $C_{\varepsilon 1}$, $C_{\varepsilon 2}$, C_ζ	60.9	59.74	|||||||||	0.96	99.2
I791	C_β, $C_{\gamma 2}$, $C_{\delta 1}$	109.5	44.84	|||||	0.72	26.0
T792	$C_{\delta 1}$, $C_{\varepsilon 1}$, $C_{\varepsilon 2}$, C_ζ, **O_η**	68.2	68.24	|||||||||	-0.08	99.8
E795	O, C_β, C_δ, **$O_{\varepsilon 1}$**, $N_{\varepsilon 2}$	132.2	34.68	|||	-0.34	59.6
G796	C, O	43.4	10.50	|||	-0.06	57.8
I797	C_α, O, $C_{\gamma 1}$, $C_{\gamma 2}$, $C_{\delta 1}$	97.6	92.70	|||||||||	1.11	79.2
L798	$C_{\delta 2}$	22.0	9.37	|||||	0.15	43.1

Q799	$N_{\varepsilon 2}$	130.5	10.35	|	-0.12	62.7
D807	C_γ, **$O_{\delta 1}$**, **$O_{\delta 2}$**	32.2	14.73	|||||	-0.17	99.6
L809	C_α, C, O, C_β, C_γ	112.5	40.42	||||	0.49	28.3
G810	N, C_α, C, O	25.6	25.60	|||||||||	0.34	100.0
E813	C_β, C_γ, C_δ, $O_{\varepsilon 1}$, $N_{\varepsilon 2}$	106.3	56.61	|||||	-0.57	44.6
L814	C_γ, $C_{\delta 1}$, $C_{\delta 2}$	22.7	21.61	|||||||||	0.35	98.0
M817	S_δ, C_ε	49.4	20.22	|||||	0.35	44.8
K821	N_ζ	66.2	11.32	||	-0.42	40.7

[a] Taken from [2].

[b] Interface residues, predicted by PDBePISA.

[c] Residues forming hydrogen bonds and salt bridges are labeled in bold and blue, respectively.

[d] Accessible surface area, in Å^2, predicted by PDBePISA.

[e] Buried surface area, in Å^2, with each vertical bar corresponding to 10% of total solvent-accessible surface area buried, predicted by PDBePISA.

[f] Solvent energy effect, in kcal mol^{-1}, predicted by PDBePISA.

[g] Degree of conservation, from a multiple sequence alignment obtained with MAFFT [113] of 1000 PPDK sequences identified by BLASTp [135] in the NCBI-NR database of non-redundant protein sequences [136].

Table 8: Crystal contacts involving the NBD in PDB ID 5JVJ [a]

Symmetry interface	Interacting atom	HB/ SB	Interacting atom at the NBD	Distance to overlay with alternative NBD conformation [b]
-x-1/2,y-1/2,-z	B-ASP 853 OD1	HB	B-ASN 250 ND2	3.6 (located on loop)
-x-1/2,y-1/2,-z	B-GLN 531 O	HB	B-THR 288 N	0.4
-x-1/2,y-1/2,-z	B-TRP 528 O	HB	B-THR 288 OG1	0.5
-x-1/2,y-1/2,-z	B-THR 362 OG1	HB	B-THR 288 OG1	0.5
-x-1/2,y-1/2,-z	B-ILE 873 O	HB	B-ASP 291 N	0.9
-x-1/2,y-1/2,-z	B-ILE 873 O	HB	B-THR 294 OG1	1.3
-x-1/2,y-1/2,-z	B-ASP 848 O	HB	B-CYS 298 SG	1.6
-x-1/2,y-1/2,-z	B-ASP 848 OD1	HB	B-CYS 298 SG	1.6
-x-1/2,y-1/2,-z	B-LYS 510 N	HB	B-GLU 308 OE1	3.6 (side chain rotated)
-x-1/2,y-1/2,-z	B-LYS 510 NZ	HB	B-GLU 305 OE2	0.8
-x-1/2,y-1/2,-z	B-LYS 510 NZ	SB	B-GLU 305 OE2	0.8
-x-1/2,y-1/2,-z	**B-LYS 536 NZ**	**HB**	**B-THR 171 O**	**15.5**
-x-1/2,y-1/2,-z	**B-GLN 556 NE2**	**HB**	**B-ASP 172 OD1**	**16.8**
-x-1/2,y-1/2,-z	B-ALA 387 O	HB	B-ARG 315 NH1	5.7 (side chain rotated)
-x-1/2,y-1/2,-z	B-ARG 363 NE	SB	B-GLU 272 OE2	0.8
-x-1/2,y-1/2,-z	B-ARG 363 NH1	SB	B-ASP 291 OD1	0.8
-x+1/2,y-1/2,-z+1	A-SER 390 OG	HB	A-MET 248 O	0.8
-x+1/2,y-1/2,-z+1	**A-GLN 511 NE2**	**HB**	**A-GLY 124 O**	**8.4**
-x+1/2,y-1/2,-z+1	**A-GLU 290 OE1**	**HB/SB**	**A-LYS 77 NZ**	**11.4**
-x+1/2,y-1/2,-z+1	**A-GLU 290 OE2**	**SB**	**A-LYS 77 NZ**	**11.4**
-x+1/2,y-1/2,-z+1	A-SER 390 O	HB	A-ASN 250 N	1.1 (located on loop)
x,y-1,z	**A-ARG 476 NH2**	**HB/SB**	**A-ASP 17 OD2**	**8.9**
x,y-1,z	A-ASP 480 OD1	SB	A-LYS 318 NZ	0.6
x,y-1,z	**A-ASP 488 OD1**	**HB**	**A-SER 12 N**	**3.0**
x,y-1,z	**A-ASP 488 OD1**	**SB**	**A-ARG 11 NE**	**9.0**
x,y-1,z	**A-ASP 488 OD2**	**HB/SB**	**A-ARG 11 NE**	**9.0**
x,y-1,z	**A-ASP 488 OD2**	**SB**	**A-ARG 11 NH2**	**10.1**
x,y-1,z	**B-ASP 650 OD2**	**HB**	**A-LYS 19 N**	**9.8**
x,y-1,z	**B-ASP 650 OD2**	**HB**	**A-SER 20 N**	**11.7**
x,y-1,z	**B-ASP 650 OD2**	**HB**	**A-SER 20 OG**	**15.3**
x,y-1,z	**B-ASP 650 N**	**HB**	**A-SER 20 OG**	**15.3**

[a] As predicted by PDBePISA. Crystal contacts that are unique for an arrangement with chain A (open) and chain B (closed) are shown in bold. Dotted lines separate the crystal contact effecting the NBD of chain A and the crystal contact effecting the NBD of chain B. Taken from [2].

[b] Distance of interacting atoms after superposition with the all-atom model of the alternative NBD conformation (see main text for details and Figure 36), in Å.

Paper contributions

[1] Structural intermediates and directionality of the swiveling motion of pyruvate phosphate dikinase

Authors: Minges, A.*, Ciupka, D.*, Winkler, C., Höppner, A., Gohlke, H.[#], Groth, G.[#]

Author Contributions: G.G. and H.G. designed research; A.M., C.W., A.H. and D.C. performed research; A.M., D.C., A.H., H.G. and G.G. analyzed data; A.M., D.C., A.H., H.G. and G.G. wrote the paper. A.M. and D.C. contributed equally to this work.

Published in: Scientific Reports 2017, 7, 45389.

Impact factor (2015): 5.228

[2] On the potential alternate binding change mechanism in a dimeric structure of pyruvate phosphate dikinase.

Authors: Ciupka, D., Gohlke, H.

Author Contributions: D.C. performed research. D.C. and H.G. designed research, analyzed data, and wrote the paper.

Published in: Scientific Reports 2017, 7, 8020.

Impact factor (2015): 5.228

* Both authors contributed equally to the respective work.
[#] These authors share senior authorship.

References

1. Minges, A.R.M., Ciupka, B.D., Winkler, C., Höppner, A., Gohlke, H., and Groth, G., *Structural intermediate and directionality of the swiveling motion of PPDK.* **Scientific Reports, 2017**. 7: p. 45389,

2. Ciupka, B.D. and Gohlke, H., *On the potential alternate binding change mechanism in a dimeric structure of Pyruvate Phosphate Dikinase.* **Scientific Reports, 2017**. 7: p. 8020,

3. Goethe, J.W., *Faust : Eine Tragödie.* **1808**: Cotta, Tübingen, Available from: Berlin-Brandenburgische Akademie der Wissenschaften, Germany http://resolver.staatsbibliothek-berlin.de/SBB00018B1700000000

4. Goodsell, D.S., *The Machinery of Life.* **2009**: Springer Science & Business Media,

5. Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H., and Schreiber, G., *The molecular architecture of protein–protein binding sites.* **Current Opinion in Structural Biology, 2007**. 17(1): p. 67-76,

6. McCammon, J., *Protein dynamics.* **Reports on Progress in Physics, 1984**. 47(1): p. 1,

7. Vale, R.D. and Milligan, R.A., *The way things move: looking under the hood of molecular motor proteins.* **Science, 2000**. 288(5463): p. 88-95,

8. Watson, H., *The stereochemistry of the protein myoglobin.* **Progress in Stereochemistry, 1969**. 4(299): p. 5,

9. Fermi, G., Perutz, M., Shaanan, B., and Fourme, R., *The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution.* **Journal of Molecular Biology, 1984**. 175(2): p. 159-174,

10. McCammon, J.A., Gelin, B.R., and Karplus, M., *Dynamics of folded proteins.* **Nature, 1977**. 267(5612): p. 585,

11. Gerstein, M. and Krebs, W., *A database of macromolecular motions.* **Nucleic Acids Research, 1998**. 26(18): p. 4280-4290,

12. Adcock, S.A. and McCammon, J.A., *Molecular dynamics: survey of methods for simulating the activity of proteins.* **Chemical Reviews, 2006**. 106(5): p. 1589,

13. Feynman, R.P., Leighton, R.B., and Sands, M., *The feynman lectures on physics; vol. i.* **American Journal of Physics, 1965**. 33(9): p. 750-752,

14. Henzler-Wildman, K. and Kern, D., *Dynamic personalities of proteins.* **Nature, 2007**. 450(7172): p. 964-972,

15. Frauenfelder, H. and McMahon, B., *Energy landscape and fluctuations in proteins.* **Annalen der Physik, 2000**. 9(9-10): p. 655-667,

16. Chowdhury, D., *Stochastic mechano-chemical kinetics of molecular motors: a multidisciplinary enterprise from a physicist's perspective.* **Physics Reports, 2013**. 529(1): p. 1-197,

17. Seyler, S.L. and Beckstein, O., *Sampling large conformational transitions: adenylate kinase as a testing ground.* **Molecular Simulation, 2014**. 40(10-11): p. 855-877,

18. Chen, Z., Li, Y., Chen, E., Hall, D.L., Darke, P.L., Culberson, C., Shafer, J.A., and Kuo, L.C., *Crystal structure at 1.9-A resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases.* **Journal of Biological Chemistry, 1994**. 269(42): p. 26344-26348,

19. Beaulieu, P.L., Wernic, D., Abraham, A., Anderson, P.C., Bogri, T., Bousquet, Y., Croteau, G., Guse, I., Lamarre, D., and Liard, F., *Potent HIV protease inhibitors containing a novel (hydroxyethyl) amide isostere.* **Journal of Medicinal Chemistry, 1997**. 40(14): p. 2164-2176,

20. Hanson, M.A., Cherezov, V., Griffith, M.T., Roth, C.B., Jaakola, V.-P., Chien, E.Y.T., Velasquez, J., Kuhn, P., and Stevens, R.C., *A specific cholesterol binding site is*

established by the 2.8 Å structure of the human β 2-adrenergic receptor. **Structure, 2008**. 16(6): p. 897-905,

21. Rasmussen, S.G.F., DeVree, B.T., Zou, Y., Kruse, A.C., Chung, K.Y., Kobilka, T.S., Thian, F.S., Chae, P.S., Pardon, E., and Calinski, D., *Crystal Structure of the β2Adrenergic Receptor-Gs protein complex.* **Nature, 2011**. 477(7366): p. 549,

22. Sobti, M., Smits, C., Wong, A.S.W., Ishmukhametov, R., Stock, D., Sandin, S., and Stewart, A.G., *Cryo-EM structures of the autoinhibited E. coli ATP synthase in three rotational states.* **eLife, 2016**. 5: p. e21598,

23. Gerstein, M., Lesk, A.M., and Chothia, C., *Structural mechanisms for domain movements in proteins.* **Biochemistry, 1994**. 33(22): p. 6739-6749,

24. Koshland, D., *Application of a theory of enzyme specificity to protein synthesis.* **Proceedings of the National Academy of Sciences, 1958**. 44(2): p. 98-104,

25. Gerstein, M., Anderson, B.F., Norris, G.E., Baker, E.N., Lesk, A.M., and Chothia, C., *Domain closure in lactoferrin: Two hinges produce a see-saw motion between alternative close-packed interfaces.* **Journal of Molecular Biology, 1993**. 234(2): p. 357-372,

26. Taylor, D., Cawley, G., and Hayward, S., *Quantitative method for the assignment of hinge and shear mechanism in protein domain movements.* **Bioinformatics, 2014**. 30(22): p. 3189-3196,

27. Astumian, R.D. and Hänggi, P., *Brownian motors.* **Physics Today, 2007**,

28. Hoffmann, P.M., *How molecular motors extract order from chaos (a key issues review).* **Reports on Progress in Physics, 2016**. 79(3): p. 032601,

29. Gajewski, E., Steckler, D.K., and Goldberg, R.N., *Thermodynamics of the hydrolysis of adenosine 5'-triphosphate to adenosine 5'-diphosphate.* **Journal of Biological Chemistry, 1986**. 261(27): p. 12733-12737,

30. van Gunsteren, W.F. and Berendsen, H.J., *Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry.* **Angewandte Chemie International Edition in English, 1990**. 29(9): p. 992-1023,

31. Frauenfelder, H., Sligar, S.G., and Wolynes, P.G., *The energy landscapes and motions of proteins.* **Science, 1991**. 254(5038): p. 1598,

32. Whitford, P.C., Altman, R.B., Geggier, P., Terry, D.S., Munro, J.B., Onuchic, J.N., Spahn, C.M.T., Sanbonmatsu, K.Y., and Blanchard, S.C., *Dynamic views of ribosome function: energy landscapes and ensembles*, in *Ribosomes*. **2011**, Springer. p. 303-319.

33. Ansari, A., Berendzen, J., Bowne, S.F., Frauenfelder, H., Iben, I., Sauke, T.B., Shyamsunder, E., and Young, R.D., *Protein states and proteinquakes.* **Proceedings of the National Academy of Sciences, 1985**. 82(15): p. 5000-5004,

34. Balzani, V., Credi, A., Raymo, F.M., and Stoddart, J.F., *Artificial molecular machines.* **Angewandte Chemie International Edition, 2000**. 39(19): p. 3348-3391,

35. Wagoner, J.A. and Dill, K.A., *Molecular Motors: Power Strokes Outperform Brownian Ratchets.* **The Journal of Physical Chemistry B, 2016**. 120(26): p. 6327-6336,

36. Kolomeisky, A.B. and Fisher, M.E., *Molecular motors: a theorist's perspective.* **Annual Review of Physical Chemistry, 2007**. 58: p. 675-695,

37. Ballardini, R., Balzani, V., Credi, A., Gandolfi, M.T., and Venturi, M., *Artificial molecular-level machines: which energy to make them work?* **Accounts of Chemical Research, 2001**. 34(6): p. 445-455,

38. Huxley, A., *A hypothesis for the mechanism of contraction of muscle.* **Progress in Biophysics and Biophysical Chemistry, 1957**. 7: p. 255-318,

39. Reimann, P., *Brownian motors: noisy transport far from equilibrium.* **Physics Reports, 2002**. 361(2): p. 57-265,

40. Maxwell, D.C., *Theory of Heat.* **1877**, Available from: https://books.google.de/books?id=EJbWDAEACAAJ

41.   Smoluchowski, M., *Experimental proof of regular thermodynamic conflicting molecular phenomenons.* **Zeitschrift für Physik 1912**. 13: p. 1069-1080,

42.   Astumian and Dean, R., *Thermodynamics and kinetics of a Brownian motor.* **Science, 1997**. 276(5314): p. 917-922,

43.   Ajdari, A. and Prost, J., *Drift induced by a spatially periodic potential of low symmetry-pulsed dielectrophoresis.* **Comptes Rendus de l'Académie des Sciences - Series II, 1992**. 315(13): p. 1635-1639,

44.   Magnasco, M.O., *Forced thermal ratchets.* **Physical Review Letters, 1993**. 71(10): p. 1477,

45.   Simon, S.M., Peskin, C.S., and Oster, G.F., *What drives the translocation of proteins?* **Proceedings of the National Academy of Sciences, 1992**. 89(9): p. 3770-3774,

46.   Wang, H. and Oster, G., *Ratchets, power strokes, and molecular motors.* **Applied Physics A, 2002**. 75(2): p. 315-323,

47.   Huxley, A.F. and Simmons, R.M., *Proposed mechanism of force generation in striated muscle.* **Nature, 1971**. 233(5321): p. 533-538,

48.   Cordova, N.J., Ermentrout, B., and Oster, G.F., *Dynamics of single-motor molecules: the thermal ratchet model.* **Proceedings of the National Academy of Sciences, 1992**. 89(1): p. 339-343,

49.   Oster, G. and Wang, H., *How protein motors convert chemical energy into mechanical work.* **Molecular Motors, 2003**: p. 205-227,

50.   Howard, J., *Motor proteins as nanomachines: the roles of thermal fluctuations in generating force and motion*, in *Biological Physics*. **2011**, Springer. p. 47-59.

51.   Lau, B., Kedem, O., Schwabacher, J., Kwasnieski, D., and Weiss, E.A., *An introduction to ratchets in chemistry and biology.* **Materials Horizons, 2017**. 4: p. 310-318,

52.   Geislinger, B., Darnell, E., Farris, K., and Kawai, R. *Are motor proteins power strokers, Brownian motors or both?* in *SPIE Third International Symposium on Fluctuations and Noise*. 2005. International Society for Optics and Photonics.

53.   Maslow, A.H., *The Psychology of Science a Reconnaissance.* **1966**,

54.   Grant, B.J., Gorfe, A.A., and McCammon, J.A., *Large conformational changes in proteins: signaling and other functions.* **Current Opinion in Structural Biology, 2010**. 20(2): p. 142-147,

55.   Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, J., EF, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M., *The Protein Data Bank. A computer-based archival file for macromolecular structures.* **European Journal of Biochemistry, 1977**. 80(2): p. 319-324,

56.   Smyth, M.S. and Martin, J.H.J., *x Ray crystallography.* **Journal of Clinical Pathology, 2000**. 53(1): p. 8,

57.   Karplus, M. and McCammon, J.A., *Molecular dynamics simulations of biomolecules.* **Nature Structural & Molecular Biology, 2002**. 9(9): p. 646-652,

58.   Karplus, M., *Molecular dynamics of biological macromolecules: a brief history and perspective.* **Biopolymers, 2003**. 68(3): p. 350-358, http://dx.doi.org/10.1002/bip.10266

59.   Kirkwood, J.G., *Statistical mechanics of fluid mixtures.* **The Journal of Chemical Physics, 1935**. 3(5): p. 300-313,

60.   van Gunsteren, W.F., Daura, X., and Mark, A.E., *Computation of free energy.* **Helvetica Chimica Acta, 2002**. 85(10): p. 3113-3129,

61.   Roux and Benoît, *The calculation of the potential of mean force using computer simulations.* **Computer Physics Communications, 1995**. 91(1-3): p. 275-282,

62.   Kästner, J., *Umbrella sampling.* **Wiley Interdisciplinary Reviews: Computational Molecular Science, 2011**. 1(6): p. 932-942,

63. Beveridge, D.L. and DiCapua, F., *Free energy via molecular simulation: applications to chemical and biomolecular systems.* **Annual Review of Biophysics and Biophysical Chemistry, 1989**. 18(1): p. 431-492,

64. Torrie, G.M. and Valleau, J.P., *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling.* **Journal of Computational Physics, 1977**. 23(2): p. 187-199,

65. Kumar, S., Rosenberg, J.M., Bouzida, D., Swendsen, R.H., and Kollman, P.A., *The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method.* **Journal of Computational Chemistry, 1992**. 13(8): p. 1011-1021,

66. Shen, J. and McCammon, J.A., *Molecular dynamics simulation of superoxide interacting with superoxide dismutase.* **Chemical Physics, 1991**. 158(2-3): p. 191-198,

67. Crouzy, S., Woolf, T.B., and Roux, B., *A molecular dynamics study of gating in dioxolane-linked gramicidin A channels.* **Biophysical Journal, 1994**. 67(4): p. 1370-1386,

68. Woolf, T.B. and Roux, B., *Conformational flexibility of o-phosphorylcholine and o-phosphorylethanolamine: a molecular dynamics study of solvation effects.* **Journal of the American Chemical Society, 1994**. 116(13): p. 5916-5926,

69. Haydock, C., Sharp, J.C., and Prendergast, F.G., *Tryptophan-47 rotational isomerization in variant-3 scorpion neurotoxin. A combination thermodynamic perturbation and umbrella sampling study.* **Biophysical Journal, 1990**. 57(6): p. 1269-1279,

70. Torrie, G.M. and Valleau, J.P., *Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid.* **Chemical Physics Letters, 1974**. 28(4): p. 578-581,

71. Mezei, M., *Adaptive umbrella sampling: Self-consistent determination of the non-Boltzmann bias.* **Journal of Computational Physics, 1987**. 68(1): p. 237-248,

72. Pfleger, C., Rathi, P.C., Klein, D.L., Radestock, S., and Gohlke, H., *Constraint Network Analysis (CNA): a Python software package for efficiently linking biomacromolecular structure, flexibility, (thermo-)stability, and function.* **Journal of Chemical Information and Modeling, 2013**. 53(4): p. 1007-1015,

73. Hermans, S., Pfleger, C., Nutschel, C., Hanke, C., and Gohlke, H., *Rigidity theory for biomolecules: Concepts, software, and applications.* **WIREs Computational Molecular Science 2017**. DOI: 10.1002/wcms.1311

74. Jacobs, D.J., Rader, A.J., Kuhn, L.A., and Thorpe, M.F., *Protein flexibility predictions using graph theory.* **Proteins, 2001**. 44(2): p. 150-165,

75. Radestock, S. and Gohlke, H., *Protein rigidity and thermophilic adaptation.* **Proteins, 2011**. 79(4): p. 1089-1108,

76. Radestock, S. and Gohlke, H., *Constraint Network Analysis: Exploiting the Link Between Protein Rigidity and Thermostability.* **Engineering in Life Sciences, 2008**. 8: p. 507-522,

77. Chastain, C.J., Failing, C.J., Manandhar, L., Zimmerman, M.A., Lakner, M.M., and Nguyen, T.H.T., *Functional evolution of C4 pyruvate, orthophosphate dikinase.* **Journal of Experimental Botany, 2011**: p. err058,

78. Hatch, M., *Regulation of C4 photosynthesis: factors affecting cold-mediated inactivation and reactivation of pyruvate, Pi dikinase.* **Functional Plant Biology, 1979**. 6(6): p. 607-619,

79. Sugiyama, T., *Purification, molecular, and catalytic properties of pyruvate phosphate dikinase from the maize leaf.* **Biochemistry, 1973**. 12(15): p. 2862-2868,

80. Shirahashi, K., Hayakawa, S., and Sugiyama, T., *Cold lability of pyruvate, orthophosphate dikinase in the maize leaf.* **Plant Physiology, 1978**. 62(5): p. 826-830,

81. Andrews, T.J. and Hatch, M.D., *Properties and mechanism of action of pyruvate, phosphate dikinase from leaves.* **Biochemical Journal, 1969**. 114(1): p. 117-125,

82. Varela-Gómez, M., Moreno-Sánchez, R., Pardo, J.P., and Perez-Montfort, R., *Kinetic mechanism and metabolic role of pyruvate phosphate dikinase from Entamoeba histolytica.* **Journal of Biological Chemistry, 2004**. 279(52): p. 54124-54130, http://dx.doi.org/10.1074/jbc.M401697200

83. Herzberg, O., Chen, C.C., Kapadia, G., McGuire, M., Carroll, L.J., Noh, S.J., and Dunaway-Mariano, D., *Swiveling-domain mechanism for enzymatic phosphotransfer between remote reaction sites.* **Proceedings of the National Academy of Sciences 1996**. 93(7): p. 2652-2657,

84. Hatch, M.D. and Slack, C.R., *A new enzyme for the interconversion of pyruvate and phosphopyruvate and its role in the C4 dicarboxylic acid pathway of photosynthesis.* **Biochemical Journal, 1968**. 106(1): p. 141-146,

85. Reeves, R.E., *A new enzyme with the glycolytic function of pyruvate kinase.* **Journal of Biological Chemistry, 1968**. 243(11): p. 3202-3204,

86. Evans, H.J. and Wood, H.G., *The mechanism of the pyruvate, phosphate dikinase reaction.* **Proceedings of the National Academy of Sciences 1968**. 61(4): p. 1448-1453,

87. Milner, Y. and Wood, H.G., *Steady state and exchange kinetics of pyruvate, phosphate dikinase from Propionibacterium shermanii.* **Journal of Biological Chemistry, 1976**. 251(24): p. 7920-7928,

88. Carroll, L.J., Xu, Y., Thrall, S.H., Martin, B.M., and Dunaway-Mariano, D., *Substrate binding domains in pyruvate phosphate dikinase.* **Biochemistry, 1994**. 33(5): p. 1134-1142,

89. Cosenza, L.W., Bringaud, F., Baltz, T., and Vellieux, F.M.D., *The 3.0 A resolution crystal structure of glycosomal pyruvate phosphate dikinase from Trypanosoma brucei.* **Journal of Molecular Biology, 2002**. 318(5): p. 1417-1432,

90. Nakanishi, T., Nakatsu, T., Matsuoka, M., Sakata, K., and Kato, H., *Crystal structures of pyruvate phosphate dikinase from maize revealed an alternative conformation in the swiveling-domain motion.* **Biochemistry, 2005**. 44(4): p. 1136-1144,

91. Lim, K., Read, R.J., Chen, C.C.H., Tempczyk, A., Wei, M., Ye, D., Wu, C., Dunaway-Mariano, D., and Herzberg, O., *Swiveling domain mechanism in pyruvate phosphate dikinase.* **Biochemistry, 2007**. 46(51): p. 14845-14853,

92. Qi, X., Lin, W., Ma, M., Wang, C., He, Y., He, N., Gao, J., Zhou, H., Xiao, Y., and Wang, Y., *Structural basis of rifampin inactivation by rifampin phosphotransferase.* **Proceedings of the National Academy of Sciences, 2016**: p. 201523614,

93. Weiße, R.H.J., Faust, A., Schmidt, M., Schönheit, P., and Scheidig, A.J., *Structure of NDP-forming Acetyl-CoA synthetase ACD1 reveals a large rearrangement for phosphoryl transfer.* **Proceedings of the National Academy of Sciences, 2016**. 113(5): p. E519-E528,

94. Korolev, S., Hsieh, J., Gauss, G.H., Lohman, T.M., and Waksman, G., *Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP.* **Cell, 1997**. 90(4): p. 635-647,

95. Wong, I. and Lohman, T.M., *A Two-Site Mechanism for ATP Hydrolysis by the Asymmetric Rep Dimer P2S As Revealed by Site-Specific Inhibition with ADP− AlF4.* **Biochemistry, 1997**. 36(11): p. 3115-3125,

96. Teplyakov, A., Lim, K., Zhu, P.-P., Kapadia, G., Chen, C.C., Schwartz, J., Howard, A., Reddy, P.T., Peterkofsky, A., and Herzberg, O., *Structure of phosphorylated enzyme I, the phosphoenolpyruvate: sugar phosphotransferase system sugar translocation signal protein.* **Proceedings of the National Academy of Sciences, 2006**. 103(44): p. 16218-16223,

97.   Mildvan, A.S. and Gupta, R.K., *[15] Nuclear relaxation measurements of the geometry of enzyme-bound substrates and analogs.* **Methods in Enzymology, 1978**. 49: p. 322-359,

98.   Knowles, J.R., *Enzyme-catalyzed phosphoryl transfer reactions.* **Annual Review of Biochemistry, 1980**. 49(1): p. 877-919,

99.   Esser, L., Wang, C.R., Hosaka, M., Smagula, C.S., Südhof, T.C., and Deisenhofer, J., *Synapsin I is structurally similar to ATP-utilizing enzymes.* **European Molecular Biology Organization Journal, 1998**. 17(4): p. 977-984,

100.   Novak, B.R., Moldovan, D., Waldrop, G.L., and de Queiroz, M.S., *Behavior of the ATP grasp domain of biotin carboxylase monomers and dimers studied using molecular dynamics simulations.* **Proteins: Structure, Function, and Bioinformatics, 2011**. 79(2): p. 622-632,

101.   Ye, D., Wei, M., McGuire, M., Huang, K., Kapadia, G., Herzberg, O., Martin, B.M., and Dunaway-Mariano, D., *Investigation of the catalytic site within the ATP-grasp domain of Clostridium symbiosum pyruvate phosphate dikinase.* **Journal of Biological Chemistry, 2001**. 276(40): p. 37630-37639,

102.   Yoshida, M., Muneyuki, E., and Hisabori, T., *ATP synthase—a marvellous rotary engine of the cell.* **Nature Reviews Molecular Cell Biology, 2001**. 2(9): p. 669-677,

103.   Kabaleeswaran, V., Puri, N., Walker, J.E., Leslie, A.G.W., and Mueller, D.M., *Novel features of the rotary catalytic mechanism revealed in the structure of yeast F1 ATPase.* **European Molecular Biology Organization Journal, 2006**. 25(22): p. 5433-5442,

104.   Adachi, K., Oiwa, K., Nishizaka, T., Furuike, S., Noji, H., Itoh, H., Yoshida, M., and Kinosita, K., *Coupling of rotation and catalysis in F 1-ATPase revealed by single-molecule imaging and manipulation.* **Cell, 2007**. 130(2): p. 309-321,

105.   Yasuda, R., Noji, H., Yoshida, M., Kinosita, K., and Itoh, H., *Resolution of distinct rotational substeps by submillisecond kinetic analysis of F1-ATPase.* **Nature, 2001**. 410(6831): p. 898-904,

106.   McGuire, M., Carroll, L.J., Yankie, L., Thrall, S.H., Dunaway-Mariano, D., Herzberg, O., Jayaram, B., and Haley, B.H., *Determination of the nucleotide binding site within Clostridium symbiosum pyruvate phosphate dikinase by photoaffinity labeling, site-directed mutagenesis, and structural analysis.* **Biochemistry, 1996**. 35(26): p. 8544-8552, http://dx.doi.org/10.1021/bi960275k

107.   Pei, J., Kim, B.-H., and Grishin, N., *PROMALS3D: a tool for multiple protein sequence and structure alignments.* **Nucleic Acids Research, 2008**. 36(7): p. 2295-300,

108.   Roe, D.R. and Cheatham III, T.E., *PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data.* **Journal of Chemical Theory and Computation, 2013**. 9(7): p. 3084-3095,

109.   Case, D.A., Cheatham III, T.E., Darden, T., Gohlke, H., Luo, R., Merz, J.K.M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J., *The Amber biomolecular simulation programs.* **Journal of Computational Chemistry, 2005**. 26 (6)(16): p. 1668-1688

110.   Hayward, S. and De Groot, B.L., *Normal modes and essential dynamics.* **Molecular Modeling of Proteins, 2008**: p. 89-106,

111.   Ahmed, A., Villinger, S., and Gohlke, H., *Large-scale comparison of protein essential dynamics from molecular dynamics simulations and coarse-grained normal mode analyses.* **Proteins: Structure, Function, and Bioinformatics, 2010**. 78(16): p. 3341-3352,

112.   Brüschweiler, R., *Collective protein dynamics and nuclear spin relaxation.* **Journal of Chemical Physics, 1995**. 102(8): p. 3396-3403,

113. Katoh, K., Misawa, K., Kuma, K.-i., and Miyata, T., *MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.* **Nucleic Acids Research, 2002**. 30(14): p. 3059-3066,

114. Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.-Y., Pieper, U., and Sali, A., *Comparative protein structure modeling using MODELLER.* **Current Protocols in Protein Science, 2007**. Chapter 2: p. Unit 2.9,

115. Word, J.M., Lovell, S.C., Richardson, J.S., and Richardson, D.C., *Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation.* **Journal of Molecular Biology, 1999**. 285(4): p. 1735-1747,

116. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L., *Comparison of simple potential functions for simulating liquid water.* **Journal of Chemical Physics, 1983**. 79(2): p. 926-935,

117. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C., *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* **Proteins, 2006**. 65(3): p. 712-725,

118. Salomon-Ferrer, R., Götz, A.W., Poole, D., Le Grand, S., and Walker, R.C., *Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald.* **Journal of Chemical Theory and Computation 2013**. 9(9): p. 3878-3888,

119. Meagher, K.L., Redman, L.T., and Carlson, H.A., *Development of polyphosphate parameters for use with the AMBER force field.* **Journal of Computational Chemistry, 2003**. 24(9): p. 1016-1025,

120. Aqvist, J., *Ion-water interaction potentials derived from free energy perturbation simulations.* **Journal of Physical Chemistry, 1990**. 94(21): p. 8021-8024,

121. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H.J., *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* **Journal of Chemical Physics, 1977**. 23(3): p. 327-341,

122. Cheatham III, T.E., Miller, J., Fox, T., Darden, T., and Kollman, P., *Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins.* **Journal of the American Chemical Society, 1995**. 117(14): p. 4193-4194,

123. Berendsen, H.J., van Postma, J., van Gunsteren, W.F., DiNola, A., and Haak, J., *Molecular dynamics with coupling to an external bath.* **Journal of Chemical Physics, 1984**. 81(8): p. 3684-3690,

124. Ahmed, A., Rippmann, F., Barnickel, G., and Gohlke, H., *A normal mode-based geometric simulation approach for exploring biologically relevant conformational transitions in proteins.* **Journal of Chemical Information and Modeling, 2011**. 51(7): p. 1604-1622,

125. Krüger, D.M., Ahmed, A., and Gohlke, H., *NMSim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins.* **Nucleic Acids Research, 2012**. 40: p. W310-W316,

126. Thorpe, M.F., Lei, M., Rader, A.J., Jacobs, D.J., and Kuhn, L.A., *Protein flexibility and dynamics using constraint theory.* **Journal of Molecular Graphics and Modelling, 2001**. 19(1): p. 60-69,

127. Rathi, P.C., Jaeger, K.-E., and Gohlke, H., *Structural Rigidity and Protein Thermostability in Variants of Lipase A from Bacillus subtilis.* **PLoS ONE, 2015**. 10(7): p. e0130289,

128. Pfleger, C., Radestock, S., Schmidt, E., and Gohlke, H., *Global and local indices for characterizing biomolecular flexibility and rigidity.* **Journal of Computational Chemistry, 2013**. 34(3): p. 220-233,

129. Eisenberg, E. and Hill, T.L., *A cross-bridge model of muscle contraction.* **Progress in Biophysics & Molecular Biology, 1978**. 33: p. 55-82,

130. Goodsell, D.S. and Olson, A.J., *Structural symmetry and protein function.* **Annual Review of Biophysics, 2000**. 29(1): p. 105-153,

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

132. Abrahams, J.P., Leslie, A.G., Lutter, R., and Walker, J.E., *Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria.* **Nature, 1994**. 370(6491): p. 621-628,

133. Krissinel, E. and Henrick, K., *Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute.* **Journal of Molecular Biology, 2007**. 372: p. 774-797,

134. Usami, S., Ohta, S., Komari, T., and Burnell, J.N., *Cold stability of pyruvate, orthophosphate dikinase of Flaveria brownii.* **Plant Molecular Biology, 1995**. 27(5): p. 969-980,

135. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J., *Basic local alignment search tool.* **Journal of Molecular Biology, 1990**. 215(3): p. 403-410,

136. NCBI-Resource-Coordinators, *Database resources of the National Center for Biotechnology Information.* **Nucleic Acids Research, 2013**. 41(Database issue): p. D8,

137. Kabsch, W. and Sander, C., *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* **Biopolymers, 1983**. 22(12): p. 2577-2637,

138. Jacobs and Thorpe, *Generic rigidity percolation: The pebble game.* **Physical Review Letters, 1995**. 75(22): p. 4051-4054,

139. Thorpe, M.F., Jacobs, D.J., and Djordjevic, B.R., *Generic Rigidity Percolation.* **Condensed Matter Theories, 1996**. 11: p. 407-424,

140. Swapna, L.S., Srikeerthana, K., and Srinivasan, N., *Extent of structural asymmetry in homodimeric proteins: prevalence and relevance.* **PloS one, 2012**. 7(5): p. e36688,

141. Dimura, M., Peulen, T.O., Hanke, C.A., Prakash, A., Gohlke, H., and Seidel, C.A.M., *Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems.* **Current Opinion in Structural Biology, 2016**. 40: p. 163-185,

142. McGuire, M., Huang, K., Kapadia, G., Herzberg, O., and Dunaway-Mariano, D., *Location of the phosphate binding site within Clostridium symbiosum pyruvate phosphate dikinase.* **Biochemistry, 1998**. 37(39): p. 13463-13474,

143. Wei, M., Li, Z., Ye, D., Herzberg, O., and Dunaway-Mariano, D., *Identification of domain-domain docking sites within Clostridium symbiosum pyruvate phosphate dikinase by amino acid replacement.* **Journal of Biological Chemistry, 2000**. 275(52): p. 41156-41165,

144. Herzberg, O., Chen, C.C.H., Liu, S., Tempczyk, A., Howard, A., Wei, M., Ye, D., and Dunaway-Mariano, D., *Pyruvate site of pyruvate phosphate dikinase: crystal structure of the enzyme-phosphonopyruvate complex, and mutant analysis.* **Biochemistry, 2002**. 41(3): p. 780-787,

145. Lin, Y., Lusin, J.D., Ye, D., Dunaway-Mariano, D., and Ames, J.B., *Examination of the structure, stability, and catalytic potential in the engineered phosphoryl carrier domain of pyruvate phosphate dikinase.* **Biochemistry, 2006**. 45(6): p. 1702-1711,

Acknowledgments

Curriculum vitae

Personenbezogene Informationen

Name: **Bartholomäus Daniel Ciupka**
Geburtsdatum / -ort: 26.04.1987 / Ruda Śląska, Polen

Akademische Ausbildung

10/2012 – 09/2017   **Studium der Biologie**
Promotion, Heinrich-Heine-Universität Düsseldorf
- Doktorarbeit "Functional state modulation in PPDK connected to long-range structural coupling and multimerization",
Institut der Pharmazeutischen und Medizinischen Chemie
Studienschwerpunkt: _Aufklärung, Modellierung, und Vorhersage der Funktion und Interaktion von Biomolekülen_
  - Molekül Dynamik Simulationen
  - Freie Energie Berechnung
  - Homologie-Modellierung
  - Rigidität Analysis

09/2010 – 10/2012   **M.Sc.**, Forschungszentrum Jülich / Heinrich-Heine-Universität Düsseldorf
- Masterarbeit "NMR based structure of the SAM domain of SLY1",
ICS-6 Strukturbiochemie / Institut für Physikalische Biologie
Studienschwerpunkt: _Strukturanalyse von biologischen Makromolekülen_
  - Kernspinresonanzspektroskopie (NMR)
  - Röntgenkristallographie

09/2007 – 09/2010   **B.Sc.**, Heinrich-Heine-Universität Düsseldorf
- Bachelorarbeit "Cloning strategies for variable tagging and localization analysis of proteins in protozoans",
Institut für Molekulare Evolution
Studienschwerpunkt: _Techniken der Molekularbiologie_
  - Klonierung
  - Proteinreinigung

Berufserfahrung

06/2015 – 12/2016   **Wissenschaftlicher Mitarbeiter**,
Institut für Pharmazeutischen und Medizinischen Chemie
Tätigkeitsfeld: _Lehre für Pharmaziestudenten_ des 7. Semesters
  - Praktikumsbetreuung
  - Vorlesungen
  - Entwurf von Klausuren

07/2011 – 10/2012   **Wissenschaftliche Hilfskraft**, ICS-6 Strukturbiochemie
Tätigkeitsfeld: _Wissenschaftliche Forschung und Präsentation_
  - Auswertung von NMR Ergebnissen
  - Präsentation vor wissenschaftlichem Publikum

Auslandserfahrung

10/2015 – 12/2015   **Studentenaustauschprogramm (MCBR4)**,
Modeling of Chemical and Biological (Re)Activity,
Indian Institute of Chemical Technology (IICT), Hyderabad, **Indien**,
Schwerpunkt: _Quantenmechanischen (QM) Methoden zur Simulation niedermolekularen Verbindung_
  - Quantenmechanik Simulation
  - Hybrid Quantenmechanik / Molekulare Mechanik Simulation

Besondere Kenntnisse und Fähigkeiten

Sprachkenntnisse:
- Deutsch (Muttersprache)
- Englisch (Verhandlungsicher)
- Polnisch (Grundkenntnisse, guter passiver Wortschatz)

Programmiersprachen:
- Shell
- R
- Perl
- C++
- Python

Softwarekenntnisse:
- AMBER
- AutoDock
- Pymol, Maestro
- Gaussion

Fortbildung

- Hybrid Quantenmechanik / Molekulare Mechanik Kurs,
  Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr
- MCBR4 Winterschule zur Theorie der Elektronenstruktur, Ruprecht-Karls-Universität Heidelberg

Veröffentlichungen

Minges, A.*, Ciupka, D.*, Winkler, C., Höppner, A., Gohlke, H., Groth, G.
Structural intermediates and directionality of the swiveling motion of Pyruvate Phosphate Dikinase.
Scientific reports. 2017, 7, 45389.

Ciupka, D., Gohlke, H.
On the potential alternate binding change mechanism in a dimeric structure of pyruvate phosphate dikinase,
Scientific reports. 2017, 7, 8020.

* geteilte Erstautorenschaft

Poster Präsentationen

Ciupka, D., Minges, A., Groth, G., Gohlke, H.
Potential of mean force calculations predict a stable conformational intermediate in the mechanism of pyruvate phosphate dikinase (PPDK).
NIC Symposium 2016, Jülich, Germany, February 11 - February 12, **2016.**

Ciupka, D., Minges, A., Groth, G., Gohlke, H.
Potential of mean force calculations predict a stable conformational intermediate in the mechanism of pyruvate phosphate dikinase (PPDK).
4th Indo-German Meeting on Modeling of Chemical and Biological Reactivity (MCBR4) Scientific Meeting, Heidelberg, Germany, February 23 - February 25, **2015.**

Ciupka, D., Minges, A., Groth, G., Gohlke, H.
MD simulations and free energy calculations in order to understand the swiveling mechanism of the kinase PPDK.
Workshop on Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, April 11 - April 12, **2014.**

Ciupka, D., Lecher, J., Stoldt, M., Thiagarajan, P., Beer-Hammer, S., König, B., Willbold, D.
Homodimer structure of the SAM domain of SLY1 solved by solution NMR.
Annual Meeting of the German Biophysical Society (DGfB), Gottingen, Germany, September 23 - September 26, **2012.**