



Technische Universität München

Fakultät für Chemie

Computergestützte Biokatalyse

# **Computational multi-scale studies for control of biological activity of proteins**

Sophie L. Mader

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitzende:

Prof. Dr. Corinna Hess

Prüfer der Dissertation:

1. Prof. Dr. Ville R. I. Kaila
2. Prof. Dr. Johannes Buchner
3. Prof. Dr. Carlo Camilloni

Die Dissertation wurde am 20.04.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 26.05.2020 angenommen.



# Table of Contents

<b>Zusammenfassung</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>List of publications</b>	<b>ix</b>
<b>Author contributions</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Proteins . . . . .	1
1.2 Enzymes . . . . .	2
1.3 Computational biochemistry . . . . .	3
1.4 The molecular chaperone Hsp90 . . . . .	4
1.5 The tyrosine kinase c-Src . . . . .	5
1.6 The GTPase Rab32 in complex with the protease GtgE . . . . .	7
1.7 The dioxygenase AsqJ . . . . .	8
1.8 Biomimetic diiron catalyst . . . . .	9
<b>2 Theory</b>	<b>11</b>
2.1 Molecular mechanics . . . . .	11
2.2 Molecular dynamics simulations . . . . .	13
2.3 Principal component analysis . . . . .	14
2.4 Continuum electrostatics and $pK_a$ calculations . . . . .	15
2.5 Quantum mechanics . . . . .	16
2.5.1 Born-Oppenheimer approximation . . . . .	16
2.5.2 Hartree-Fock theory . . . . .	17
2.5.3 Density functional theory . . . . .	18
2.5.4 Basis sets . . . . .	21
2.6 Quantum chemical cluster models . . . . .	22
2.7 Continuum solvation . . . . .	23
2.8 Reaction path optimization . . . . .	23
2.9 Hybrid quantum mechanics/molecular mechanics (QM/MM) . . . . .	24
2.9.1 General concept . . . . .	24
2.9.2 QM/MM schemes and embedding . . . . .	25
2.10 Energy minimization . . . . .	26
2.11 Umbrella sampling and the weighted histogram analysis method . . . . .	27

## Table of Contents

<b>3</b>	<b>Methods</b>	<b>29</b>
3.1	General . . . . .	29
3.2	The molecular chaperone Hsp90 . . . . .	29
3.2.1	Quantum chemical and hybrid QM/MM calculations . . . . .	29
3.2.2	Classical molecular dynamics simulations . . . . .	30
3.2.3	Calculation of $pK_a$ values . . . . .	31
3.3	The tyrosine kinase c-Src . . . . .	31
3.4	The GTPase Rab32 in complex with the protease GtgE . . . . .	31
3.5	The dioxygenase AsqJ . . . . .	32
3.5.1	Quantum chemical calculations . . . . .	32
3.5.2	Hybrid QM/MM calculations . . . . .	32
3.5.3	Classical molecular dynamics simulations . . . . .	33
3.6	Biomimetic diiron catalyst . . . . .	33
<b>4</b>	<b>Results and discussion</b>	<b>35</b>
4.1	The molecular chaperone Hsp90 . . . . .	35
4.1.1	The role of Arg-32 . . . . .	35
4.1.2	The role of Lys-594 . . . . .	38
4.2	The tyrosine kinase c-Src . . . . .	40
4.3	The GTPase Rab32 in complex with the protease GtgE . . . . .	42
4.4	The dioxygenase AsqJ . . . . .	44
4.5	Biomimetic diiron catalyst . . . . .	47
<b>5</b>	<b>Conclusions</b>	<b>51</b>
	<b>Bibliography</b>	<b>53</b>
	<b>Publications</b>	<b>65</b>
	<b>Acknowledgements</b>	<b>133</b>



# Zusammenfassung

Proteine sind komplexe Biomoleküle, die zentrale Funktionen in lebenden Zellen erfüllen, wie zum Beispiel die Katalyse biochemischer Reaktionen, Signalübertragung und Transport von Molekülen und Ionen. In dieser Doktorarbeit erforschen wir die Verbindung zwischen der Struktur und Konformationsdynamik von Proteinen, und wie Enzymkatalyse durch Veränderung der Proteinkonformation gesteuert werden kann. Dazu verwenden wir molekulare Multiskalen-Simulationen und integrieren sie mit biochemischen und biophysikalischen Experimenten. Wir untersuchen die Konformationswechsel und Katalyse des ATP-abhängigen Chaperons Hsp90, der Tyrosinkinase c-Src und der an die Cysteinprotease GtgE gebundenen GTPase Rab32, sowie die Reaktivität der  $\text{Fe}^{\text{II}}$ -abhängigen Dioxygenase AsqJ und eines biomimetischen Eisenkomplexes. Wir erklären, wie Konformationsänderungen, posttranslationale Modifikationen und Ligandenbindung die katalytische Aktivität der untersuchten Proteine beeinflussen. Weiterhin identifizieren wir Aminosäuren, die die katalytische Funktion der Proteine regulieren, und prüfen den Effekt des Austauschs zentraler Aminosäuren. Unsere Ergebnisse ermöglichen einen Einblick in den molekularen Mechanismus dieser biologischen Systeme und zeigen, wie sie auf der Grundlage von kombinierten computergestützten und experimentellen Studien umgestaltet werden können.



# Abstract

Proteins are complex biological molecules that fulfill central functions in living cells, such as catalysis of biochemical reactions, signal propagation, and transport of ions and molecules. In this thesis, we investigate how protein structure links to conformational dynamics and how enzyme catalysis can be controlled by changing the protein conformation. To this end, we employ multi-scale molecular simulations and integrate them with biochemical and biophysical experiments. We study the conformational switching and catalysis of the ATP-dependent chaperone Hsp90, the tyrosine kinase c-Src, and the GTPase Rab32 bound to the cysteine protease GtgE, as well as the reactivity of the  $\text{Fe}^{\text{II}}$ -dependent dioxygenase AsqJ and of a biomimetic iron complex. We elucidate how conformational changes, post-translational modifications, and ligand binding affect the catalytic activity of the investigated proteins. Furthermore, we identify residues that control the catalytic function of the proteins, and probe the effect of substituting central amino acids. Our results provide insight into the molecular mechanism of these biological systems, and show how they can be redesigned based on combined computational and experimental studies.



## List of publications

- I. Wachtel, R.\* , Bräuning, B.\* , **Mader, S. L.**, Ecker, F., Kaila, V. R. I., Groll, M. & Itzen, A. The protease GtgE from Salmonella exclusively targets inactive Rab GTPases. *Nat. Commun.* **9**, 44:1-13 (2018).
- II. Lindsay, S., **Mader, S. L.**, Kaila, V. R. I. & Hess, C. R. C-H Oxidation by a Diiron Complex with Facially Opposing Active Sites. *ChemistrySelect* **3**, 1602-1608 (2018).
- III. **Mader, S. L.\***, Bräuer, A.\* , Groll, M. & Kaila, V. R. I. Catalytic mechanism and molecular engineering of quinolone biosynthesis in dioxygenase AsqJ. *Nat. Commun.* **9**, 1168:1-8 (2018).
- IV. Boczek, E. E.\* , Luo, Q.\* , Dehling, M.\* , Röpke, M., **Mader, S. L.**, Seidl, A., Kaila, V. R. I. & Buchner, J. Autophosphorylation activates c-Src kinase through global structural rearrangements. *J. Biol. Chem.* **294**, 13186-13197 (2019).
- V. Rehn, F.\* , Lawatscheck, J.\* , Jokisch, M.-L., **Mader, S. L.**, Qi, L., Tippel, F., Blank, B., Richter, K., Lang, K., Kaila, V. R. I. & Buchner, J. A methylated lysine is a switch point for conformational communication in the chaperone Hsp90. *Nat. Commun.* **11**, 1219:1-14 (2020).
- VI. **Mader, S. L.\***, Lopez, A.\* , Lawatscheck, J.\* , Luo, Q., Rutz, D. A., Gamiz-Hernandez, A. P., Sattler, M., Buchner, J. & Kaila, V. R. I. Conformational dynamics modulate the catalytic activity of the molecular chaperone Hsp90. *Nat. Commun.* **11**, 1410:1-12 (2020).

\*Authors contributed equally.



## Author contributions

- I. A.I. conceived research; R.W., B.B., S.L.M. and F.E. performed experiments; R.W., B.B., S.L.M., F.E., V.R.I.K., M.G., A.I. designed the experiments, analyzed and interpreted the data, and wrote the manuscript. S.L.M. performed and analyzed the MD simulations and created figures.
- II. V.R.I.K. and C.R.H. designed research; S.L., S.L.M., V.R.I.K., and C.R.H. performed research; S.L., S.L.M., V.R.I.K., and C.R.H. analyzed data; V.R.I.K. and C.R.H. wrote the paper. S.L.M. performed and analyzed electronic structure calculations including optimization of structures and reaction pathways, and plotted figures.
- III. V.R.I.K. and M.G. designed the project; S.L.M. and V.R.I.K. performed and analyzed all calculations; A.B. created mutants, purified proteins, and performed activity assays; A.B. and M.G. collected and analyzed X-ray data; V.R.I.K. and M.G. wrote the manuscript, with contributions from all authors. S.L.M. performed and analyzed DFT calculations including optimization of structures and reaction pathways, MD simulations, and QM/MM calculations, and created figures.
- IV. E.E.B., A.S., V.R.I.K., and J.B. designed the project; E.E.B. and M.D. performed biochemical and biophysical experiments; Q.L. and S.L.M. performed MD simulations; E.E.B., Q.L., M.D., M.R., S.L.M., V.R.I.K., and J.B. analyzed data; E.E.B., V.R.I.K., and J.B. wrote the paper.
- V. A.R. and J.B. designed the mutants; A.R., J.L., M.-L.J., F.T., and B.B. performed the *in vitro* and *in vivo* experiments, and A.R., J.L., and M.-L.J. analyzed the results; K.R. and J.L. performed the aUC experiments; Q.L., S.L.M., and V.R.I.K. designed, performed and analyzed the MD simulations; M.-L.J. and K.L. designed and performed amber suppression experiments; A.R., J.L., K.L., V.R.I.K. and J.B. wrote the manuscript. S.L.M. performed and analyzed MD simulations and plotted figures.
- VI. V.R.I.K. designed the project; S.L.M., Q.L., A.P.G.-H., and V.R.I.K. performed the molecular simulations; J.L. and D.A.R. performed the protein biochemistry, F.R.E.T., and fluorescence anisotropy experiments; A.L. performed NMR and SAXS experiments; S.L.M., A.L., J.L., Q.L., D.A.R., A.P.G.-H., M.S., J.B., and V.R.I.K. analysed data; V.R.I.K. wrote the manuscript with contributions from all authors. S.L.M. performed and analyzed QM/MM free energy calculations, DFT calculations, and MD simulations, and plotted figures.





# 1 Introduction

## 1.1 Proteins

Proteins are central biological macromolecules that participate in every process within living cells. They are built from amino acids that are connected by peptide bonds, and differ from each other in their amino acid sequence, also referred to as their primary structure, that is usually composed of the 20 standard amino acids. Additionally, proteins can be modified after their biosynthesis by undergoing post-translational modifications (PTMs), and they may employ further chemical cofactors to alter their function.

Proteins form complex statistical ensembles of differently populated conformations that are interconvertible on time scales between nanoseconds and hours.<sup>1</sup> Their dynamical three-dimensional structure is often linked to their biological function. Transitions between conformations take place by sampling the inherent equilibrium dynamics of a protein. Conformational transitions can also be induced by environmental changes including light absorption, enzymatic modification, ligand binding events, and change in pH or temperature. These changes can influence the equilibrium distribution of conformational states of a protein, as well as the interconversion dynamics between them.<sup>2</sup> Conformational switches are proteins that undergo a structural change upon receiving an input signal and provide a defined output response.<sup>3</sup>

The population of multiple conformational states by a protein depends on the underlying free energy landscape, formed by energy wells and barriers separating different conformations.<sup>4</sup> The relative distribution of conformations  $i$  and  $j$  depends on the free energy difference between them,  $\Delta G_{ij}$ ,

$$\frac{p_i}{p_j} = e^{-\frac{\Delta G_{ij}}{RT}}, \quad (1.1)$$

with the gas constant  $R$  and the temperature  $T$ . Accordingly, low energy regions on the landscape show larger populations than high energy regions. The rate constant  $k$  for a conformational change is determined by the free energy barrier  $\Delta G^\ddagger$  and the temperature  $T$ , empirically described by the Arrhenius equation,

$$k = Ae^{-\frac{\Delta G^\ddagger}{RT}}. \quad (1.2)$$

$A$  is a pre-exponential factor and  $R$  is the gas constant. Depending on their energetics, internal motions and conformational events in proteins span a wide range of length and time scales. Bond vibrations can occur within several femtoseconds to picoseconds, whereas loop motions take place within nano- to microseconds, while allosteric regulation and conformational rearrangements occur within micro- to milliseconds, and protein

## 1 Introduction

folding can last up to hours.<sup>5</sup>

These structural changes of proteins control their biological function.<sup>6,7</sup> Key to the coupling between structural fluctuations and functional properties is the difference in affinity of certain conformations for different binding partners.<sup>2</sup> Well-folding proteins can be specific catalysts and binders, while proteins with intrinsically disordered regions usually display a broad spectrum of conformations and ligand binding capabilities.<sup>8</sup>

Ligands influence in many cases the conformational energy landscape of proteins. Protein-ligand binding often involves conformational changes that can be connected to the intrinsic conformational dynamics of the protein. If a conformational change occurs in a protein leading to a weakly populated conformation that the ligand selectively binds to, the binding is characterized by a conformational selection mechanism; if a ligand binding event to a predominant free conformation leads to a subsequent conformational change in the protein, the binding is characterized by an induced fit mechanism.<sup>9</sup> In conformational selection, the ligand selects and stabilizes a pre-existing transient protein conformation and shifts the conformational ensemble towards it. In induced fit, the ligand induces a structural change in the protein followed by a shift in the population.<sup>10</sup>

Post-translational modifications (PTMs), where chemical groups are covalently attached to selected amino acids, are a possibility to modulate the structure and function of proteins in the cell. PTMs include amongst others phosphorylation, glycosylation, methylation, acetylation, and amidation. For example, in protein phosphorylation, a phosphate group is attached to serine, threonine, or tyrosine residues, while methylation refers to the transfer of methyl groups to arginine or lysine residues.<sup>11</sup>

The three-dimensional tertiary structure of a protein is evolutionary more conserved than its amino acid sequence.<sup>12</sup> In certain cases, protein structure and function do not change significantly over time, but the sequence evolves, leading to differing sequences of functionally similar proteins. Mutations between amino acids with similar properties do not necessarily change the overall structure and function of a protein, but they can alter the population of the pre-existing conformational states and lead to an alternate structural ensemble. In contrast, adaptive mutations allow proteins to gain new functionality.<sup>8</sup> A few amino acid substitutions and minor conformational changes in the active site of a protein can lead to drastically changed substrate specificity or enzymatic activity. Protein active sites usually mutate at a slower rate and remain better conserved compared to the remaining protein framework.<sup>13</sup>

## 1.2 Enzymes

Many proteins are enzymes that catalyze biochemical reactions and significantly enhance their reaction rates.<sup>14,15</sup> Enzyme active sites provide a preorganized environment that stabilizes the transition state of a chemical reaction more than the corresponding environment in water.<sup>16</sup> Biochemical reactions involving a single substrate  $S$  often follow Michaelis–Menten kinetics, where the enzyme and substrate form an enzyme-substrate complex that is converted to the product and the original enzyme, according to,<sup>17,18</sup>

$$v = v_{max} \frac{[S]}{K_M + [S]} = k_{cat} [E]_0 \frac{[S]}{K_M + [S]}. \quad (1.3)$$

In this equation,  $v$  is the rate of product formation,  $v_{max}$  is the maximum rate achieved at saturating substrate concentration,  $[S]$  is the substrate concentration, and  $[E]_0$  is the total initial enzyme concentration.  $K_M$  is the Michaelis constant that determines both the substrate concentration at which the reaction rate reaches half maximum ( $\frac{v_{max}}{2}$ ), as well as the affinity of the substrate for the enzyme. The rate constant for product formation from the enzyme-substrate complex is characterized by  $k_{cat}$ , whereas the  $\frac{k_{cat}}{K_M}$  ratio is the parameter that measures the specificity of an enzyme and its capacity to discriminate between substrates.<sup>15,19</sup>

Conformational changes play an important role in the function of many enzymes. For some enzymes, *e.g.* the prolyl *cis*-*trans* isomerase cyclophilin A, it has been observed that characteristic enzyme motions that occur during the catalytic cycle are already represented in the free enzyme.<sup>20</sup> In selected enzyme catalytic cycles, conformational changes have been suggested to be rate limiting for the conversion of substrates to products.<sup>21</sup>

Perhaps the most prominent example of proteins that undergo large-scale conformational changes to regulate cell processes are guanine and adenine nucleotide triphosphatases (NTPases). These ubiquitous enzymes function as conformational switches that cycle between distinct states. They are regulators fueled by nucleotide binding and hydrolysis, by splitting the nucleotide triphosphate into nucleotide diphosphate and inorganic phosphate and using the released energy for their function.<sup>2</sup>

The most universal energy carrier nucleotide is adenosine triphosphate (ATP) which provides chemical energy for reactions and processes in the cell. It is primarily produced in  $F_0F_1$ -ATP synthesis as a part of cellular respiration. ATP synthase couples the synthesis of ATP from ADP and inorganic phosphate with an electrochemical driving force generated by the pumping of protons through the coupling membrane.<sup>22</sup> Moreover, phosphoryl transfer from GTP is also used for signaling and energy transduction in the cell.<sup>23</sup>

## 1.3 Computational biochemistry

Over the past few years, developments in cryo-electron microscopy (cryo-EM) and X-ray crystallography have started to reveal multiple snapshots of increasingly large and flexible biological systems.<sup>24</sup> Although conformational substates and their rates of interconversion can nowadays be detected experimentally, it is still challenging to resolve structural transitions, owing to the low probability and short lifetime of the high-energy conformers.<sup>4</sup>

Computer simulations can capture the structure and dynamics of proteins in full atomic detail and at very fine temporal resolution in the femtosecond range, and the atomic positions can be followed for time scales of pico- to milliseconds along with the corresponding energies, provided that high-resolution structures are known as starting

## 1 Introduction

points.<sup>4,25</sup> To this end, classical and quantum chemical methods have been developed to computationally explore the structure and function of biomolecules.<sup>26</sup> Computational studies can provide detailed molecular information about proteins, and can be integrated with experimental techniques to gain complementary information.<sup>27</sup> Measurements using, *e.g.*, Nuclear Magnetic Resonance (NMR) spectroscopy, Small Angle X-ray Scattering (SAXS), or Förster Resonance Energy Transfer (FRET), allow to resolve structural and dynamic information from biological systems on different time and length scales.<sup>24</sup>

In this thesis, the catalytic activity and conformational dynamics of different biological systems are investigated using the methodology of computational biochemistry, in combination with various complementary experimental techniques. In the following, a short introduction to the studied systems is provided.

### 1.4 The molecular chaperone Hsp90

The correct folding of many proteins in the cell depends on specialized proteins named molecular chaperones, that stabilize so-called client proteins.<sup>28</sup> Many of these chaperones are known as heat shock proteins.<sup>29</sup> The heat shock protein 90 (Hsp90) is a conserved molecular chaperone that is essential for viability in eukaryotes, and is responsible for the maturation of key signalling proteins, including regulatory kinases, steroid hormone receptors, and transcription factors.<sup>30,31</sup> Hsp90 is explored as a potential anticancer target, since regulators of cancer cell signaling have also been identified as its client proteins.<sup>32</sup> For example, the oncogenic viral Src kinase v-Src, a constitutively active non-receptor tyrosine kinase, is a client of Hsp90.<sup>33</sup> Hsp90 is a 90 kDa enzyme in humans, whereas the homologous yeast Hsp82 is often used as a model system to understand the human protein.

Hsp90 functions as a homodimer, and dimerization is required for its *in vivo* function.<sup>34</sup> Both protomers forming the Hsp90 homodimer consist of an N-terminal domain (NTD), a middle domain, and a C-terminal domain (CTD) (Figure 1.1). The N-terminal domain harbors a conserved ATP binding site.<sup>35</sup> While ATP binding by both NTDs in the Hsp90 dimer was shown to be essential for its *in vivo* function, ATP hydrolysis was reported to be required only in one of the two NTDs.<sup>36</sup> The middle domain carries the binding site for client proteins and cochaperones, while the CTD is required for dimerization.<sup>31,37</sup> The dimer interface between the CTDs is formed by a pair of helices at the C-terminal end of the domain, creating a four helix bundle.<sup>38</sup> In 2012, Lys-594 (residue numbering of yeast Hsp82) in the C-terminal domain was reported to be methylated by lysine methyltransferase Smyd2, which establishes an important post-translational modification for skeletal muscle maintenance by Hsp90.<sup>39</sup>

Hsp90 undergoes ATPase-coupled structural rearrangements, in which binding of ATP to a C-terminally dimerized Hsp90 dimer leads to N-terminal dimerization and association of the NTDs with the middle domains, important for ATP hydrolysis.<sup>40</sup> In this closed form of the dimer with ATP bound to the NTD, the  $\gamma$ -phosphate of ATP forms contacts with Arg-380 (residue numbering of the yeast isoform Hsp82) from the middle domain, while the  $\alpha$ - and  $\beta$ -phosphates of ATP are bound to an octahedrally coordinated  $Mg^{2+}$

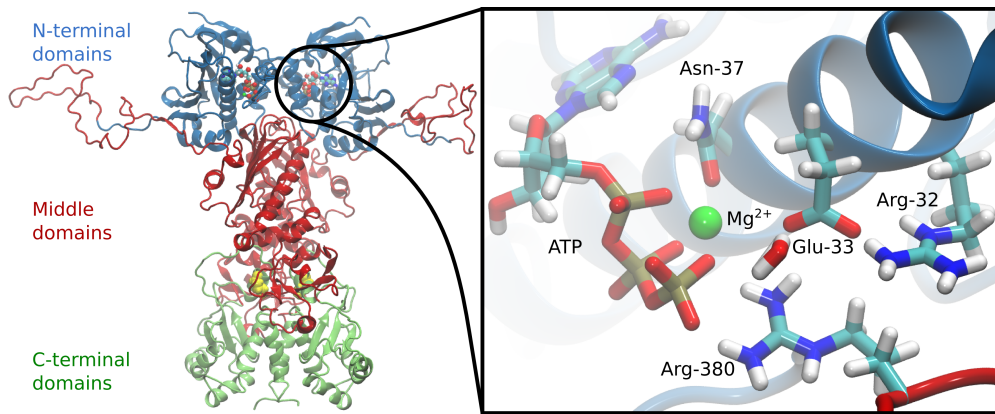


Figure 1.1: Structure of the closed conformation of the homodimeric chaperone Hsp90 from yeast, based on the crystal structure with PDB ID: 2CG9,<sup>41</sup> with N-terminal domains (NTDs) shown in blue, charged linker regions and middle domains in red, and C-terminal domains (CTDs) in green. A methylation site in the CTD, Lys-594, is shown in yellow van der Waals representation. The inset shows a relaxed model of the ATPase site formed by residues from the NTD and the middle domain. The figure is adapted from ref. [44].

ion connected to the side chain of Asn-37 (Figure 1.1).<sup>41</sup> Arg-380 is important for ATP hydrolysis since the R380A mutant does not support yeast cell growth and shows minimal ATPase activity.<sup>42</sup> Glu-33 activates as a catalytically active residue a water molecule for the ATP hydrolysis reaction.<sup>41</sup> After ATP hydrolysis, the NTDs dissociate, ADP and inorganic phosphate ( $P_i$ ) are released, and Hsp90 returns to the open conformation.<sup>37</sup> These large conformational rearrangements between open and closed conformations constitute the Hsp90 chaperone cycle and are crucial for client maturation.<sup>43</sup>

Hsp90 has a low ATPase activity with catalytic rates of  $0.1 \text{ min}^{-1}$  for human Hsp90, and *ca.*  $0.5 \text{ min}^{-1}$  for the yeast homolog Hsp82.<sup>45,46</sup> The conformational changes leading to the closed state were reported to be the rate-limiting steps of the reaction cycle.<sup>47</sup> These conformational transitions are sufficient for yeast cells to survive, even in the absence of ATP hydrolysis, as shown for the ATPase-deficient Hsp90-E33A mutant that still binds ATP.<sup>48</sup>

## 1.5 The tyrosine kinase c-Src

Protein phosphorylation is a central event in eukaryotic signaling and involves transfer of the  $\gamma$ -phosphate group of ATP to the hydroxyl group of serine, threonine, or tyrosine residues of substrate proteins by protein kinases.<sup>49</sup> c-Src, or Src, is a non-receptor protein-tyrosine kinase that regulates cellular responses to extracellular stimuli.<sup>50</sup> Dysregulation of protein kinases occurs in diseases including cancer and inflammatory disorders. The viral Src protein, v-Src, in Rous sarcoma virus, is a related protein-tyrosine

## 1 Introduction

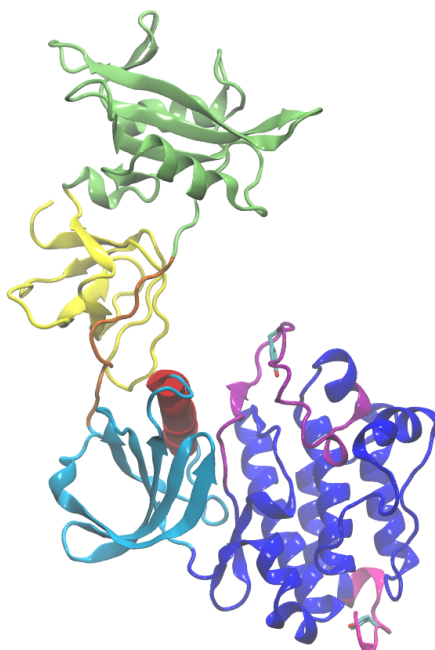


Figure 1.2: Crystal structure (PDB ID: 1Y57)<sup>53</sup> of unphosphorylated c-Src, with the SH3 domain shown in yellow, SH2 domain in green, linker in orange, N-lobe of the kinase domain in light blue,  $\alpha$ C helix in red, C-lobe of the kinase domain in dark blue, activation loop in purple, and C-terminal tail in pink. Tyr-416 in the activation loop and Tyr-527 in the C-terminal tail are shown in licorice representation. The figure is adapted from ref. [54].

kinase with increased activity because it lacks the autoinhibitory phosphorylation site of c-Src.<sup>51</sup>

From N-terminus to C-terminus, c-Src consists of a unique domain, the regulatory SH3 (Src homology 3) and SH2 domains, a linker region, a protein-tyrosine kinase domain (KD) divided into a small N-terminal lobe (N-lobe) and a large C-terminal lobe (C-lobe), and a C-terminal tail (Figure 1.2).<sup>51</sup> In all protein kinases, the N-lobe of the kinase domain includes the so-called  $\alpha$ C helix.<sup>52</sup> ATP binds between the two lobes of the kinase domain, while protein substrates bind to the so-called activation loop (A-loop) in the C-lobe of the kinase domain. The A-loop needs to be phosphorylated at Tyr-416 for full kinase activity, while phosphorylation of Tyr-527 in the C-terminal tail downregulates c-Src.<sup>49</sup>

X-ray structures of inactive c-Src with phosphorylated Tyr-527 show a compact, assembled enzyme.<sup>55–57</sup> Intramolecular interactions between the phosphorylated Tyr-527 and the SH2 domain, and between the linker region and the SH3 domain, stabilize an inhibitory conformation of the  $\alpha$ C helix and the activation loop that collapses into the active site, blocks ATP and protein substrate binding, and prevents Tyr-416 phosphorylation.<sup>49</sup>

## 1.6 The GTPase Rab32 in complex with the protease GtgE

Competitive binding of proteins to the SH2/SH3 domains and dephosphorylation of Tyr-527 lead to a conformational change in the kinase domain, enabling intermolecular autophosphorylation of Tyr-416 by a partner Src kinase, leading to activation of the enzyme.<sup>51</sup> When phosphorylated at Tyr-416, the tyrosine-phosphate group interacts with Arg-385 and Arg-409, and the conserved  $\alpha$ C-glutamate Glu-310 forms contacts with Lys-295.<sup>57</sup> The  $\alpha$ C helix adopts an active conformation, the activation loop switches to an extended conformation and moves away from the active site, and the kinase is stabilized in an open conformation that binds ATP and the protein substrate. After catalysis, the phosphorylated substrate protein and ADP are released.<sup>51</sup>

In 2005, a crystal structure of completely unphosphorylated c-Src and without the unique domain was reported, in which the SH2 domain is not bound to the C-terminal tail (Figure 1.2).<sup>53</sup> This results in a more open structure of the protein where the kinase domain, including the helix  $\alpha$ C and the activation loop, adopt an active conformation, while the SH3 domain is bound to the linker region, as observed in the inactive conformation. This crystal structure offers a starting point from which to gain molecular insight into the mechanism and the underlying structural changes within the activation loop following autophosphorylation at Tyr-416.

## 1.6 The GTPase Rab32 in complex with the protease GtgE

Small GTP- (guanosine triphosphate) binding proteins, also referred to as G proteins or GTPases, exist in eukaryotes from yeast to human and constitute a superfamily consisting of more than 100 members. GTP-binding proteins regulate a wide variety of cell functions as biological timers, and play key roles also in the spatial determination of specific cell functions.<sup>58</sup>

G proteins use the binding energy of GTP to produce a conformation that permits their association with effector proteins. Upon GTP hydrolysis into GDP and inorganic phosphate, conformational rearrangements are triggered in the catalytic domain, and the effector protein is released. The GTP- and GDP-bound complexes define the active and inactive states of a G protein as a regulatory molecule.<sup>59</sup> In the activated (GTP-bound) state, G proteins interact with downstream effector proteins, which in turn trigger cellular responses.<sup>23</sup>

The Rab-family of small GTPases regulates membrane and vesicle trafficking.<sup>58,60</sup> Rab32 is involved in traffic control to lysosome-related organelles, including the delivery of enzymes and antimicrobial proteins.<sup>61</sup> Due to the pivotal role of small GTPases in intracellular signaling and the maintenance of cellular homeostasis, they are frequently targeted by bacterial pathogens to promote infection.<sup>60</sup>

The effector protein GtgE from *Salmonella typhimurium*<sup>62</sup> is a cysteine protease that proteolytically cleaves members of the Rab-family of small GTPases. Cys-45 of GtgE has been identified as the catalytically active cysteine residue for proteolysis. GtgE specifically cleaves the small GTPases Rab29, Rab32, and Rab38, thereby preventing the delivery of antimicrobial factors to the bacteria-containing vacuole that would impair *Salmonella* infection.<sup>63</sup>

## 1.7 The dioxygenase AsqJ

In 2014, a fungal  $\text{Fe}^{\text{II}}$ - and  $\alpha$ -ketoglutarate-dependent, non-heme dioxygenase was reported, which not only performs two sequential oxidation reactions on a single substrate, but also catalyzes a subsequent epoxidation reaction (Figure 1.3).<sup>64</sup> This enzyme was discovered in *Aspergillus nidulans*, where it is involved in the biosynthesis of 4'-methoxyviridicatin, a quinolone alkaloid. Quinolone alkaloids were shown to have antimicrobial activities.<sup>65</sup>

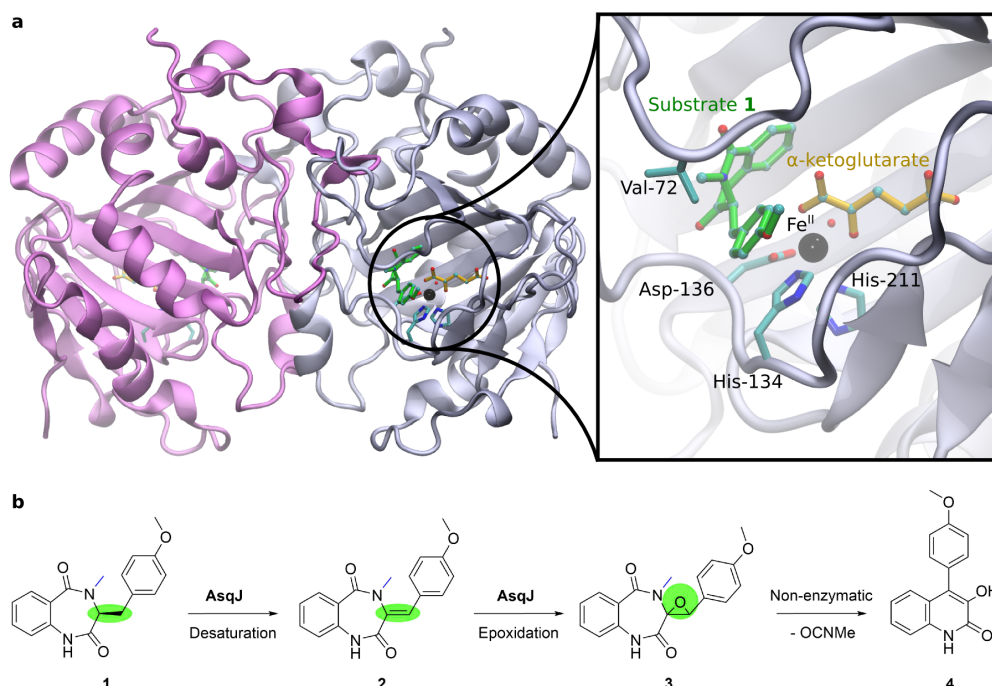


Figure 1.3: **a)** Crystal structure (PDB ID: 5DAQ) of the dioxygenase AsqJ as a homodimer with bound substrate **1** (in green).<sup>66</sup> The inset shows the active site with iron depicted as a black sphere and a crystallographic water molecule as a red dot. Active site residues are shown in licorice representation, as well as the substrate **1** (in green) and the cofactor  $\alpha$ -ketoglutarate (in yellow). **b)** Reaction sequence catalyzed by AsqJ. Desaturation of 4'-methoxycyclopeptin (**1**) yields 4'-methoxydehydrocyclopeptin (**2**), which is converted into 4'-methoxycyclopenin (**3**) by epoxidation, and finally non-enzymatically rearranged to the final product, 4'-methoxyviridicatin (**4**). The figure is adapted from ref. [67].

Under consumption of dioxygen and the  $\alpha$ -ketoglutarate ( $\alpha$ KG) cofactor, as in other  $\alpha$ KG-dependent dioxygenases,<sup>68</sup> a C-C single bond in the natural substrate of AsqJ, 4'-methoxycyclopeptin (**1**), is converted into a double bond by two sequential hydrogen atom transfer reactions, yielding 4'-methoxydehydrocyclopeptin (**2**). Subsequently,



AsqJ adds an oxygen atom to the newly created double bond, forming the epoxide 4'-methoxycyclopinin (**3**). A non-enzymatic elimination reaction of the epoxide leads to formation of the final product, 4'-methoxyviridicatin (**4**) (Figure 1.3 b).<sup>64</sup>

In 2016, high-resolution crystal structures of AsqJ with different substrates were reported.<sup>66</sup> In these structures, the metal in the active site of AsqJ is octahedrally coordinated by the characteristic motif of Fe<sup>II</sup>/ $\alpha$ KG-dependent dioxygenases, consisting of two histidine residues (His-134, His-211), one aspartate (Asp-136), an  $\alpha$ KG, and a water molecule as a sixth ligand (Figure 1.3 a).  $\alpha$ KG is stabilized by a salt bridge with Arg-223, and enhances oxygen reactivity by coordinating the iron with both its carboxylate and keto groups. The sixth coordination site that is later on used for dioxygen activation and is here initially occupied by a water molecule, is suggested to be created by substrate binding.<sup>68</sup>

Notably, the substrate forms a  $\pi$ - $\pi$  stacking interaction with His-134 (Figure 1.3 a). While a surrogate lacking the N-methyl group of **1** still binds to the enzyme, as resolved in an X-ray structure, no catalytic activity of AsqJ could be detected towards this surrogate.<sup>66</sup> Experimental and computational studies were performed to investigate the catalytic mechanism of AsqJ. It was demonstrated that AsqJ catalyzes the oxidation reactions via a high-spin (quintet spin,  $S = 2$ ) ferryl species (Fe(IV)=O),<sup>69–71</sup> as proposed earlier for non-heme iron enzymes.<sup>72</sup>

## 1.8 Biomimetic diiron catalyst

Metalloenzymes possess unique reactivities and catalyze the selective oxidation of their substrates in diverse oxidative transformations. Many of these oxidation reactions are unprecedented in synthetic organic chemistry.<sup>73</sup> Iron is involved as a cofactor in dioxygen activation and biochemical oxidation reactions in both heme as well as non-heme oxidases and oxygenases.<sup>74–76</sup> To better understand the catalytic role and mechanistic details of iron centers, we studied a biomimetic non-heme diiron catalyst that catalyzes a similar chemistry as the dioxygenase AsqJ. The cooperative employment of two iron centers offers two coordination sites for dioxygen.

Non-heme diiron clusters are found within ferritin-like four-helix-bundle protein structures, with two adjacent metal ion binding sites at a distance of *ca.* 4 Å.<sup>77</sup> One of these non-heme diiron oxidases and oxygenases is soluble methane monooxygenase, the reactive intermediate of which contains a dinuclear bis-oxo Fe<sup>IV</sup>Fe<sup>IV</sup> cluster with two bridging oxygen atoms originating from a dioxygen molecule. One oxygen atom is inserted into a C-H bond of methane, generating methanol and water.<sup>78</sup>

Mononuclear non-heme iron-dependent oxygenases were shown to utilize a Fe<sup>IV</sup>=O intermediate to catalyze oxidative reactions.<sup>79</sup> These non-heme iron enzymes have been serving as a model for the design of synthetic oxidation catalysts based on multidentate nitrogen ligands, which also use ferryl groups as reactive species.<sup>80</sup> These complexes provide chemical models to study the reactivity and intermediates of non-heme iron enzymes. Mechanistic developments in biomimetic research can improve the understanding of dioxygen activation and oxygen atom transfer reactions in nature.<sup>81</sup>



## 2 Theory

The energetics and dynamics of proteins can be modelled computationally by applying the laws of classical molecular mechanics or quantum mechanics. The following section comprises the theoretical background for the methods used in this thesis.

### 2.1 Molecular mechanics

Methods modelling biological systems with classical molecular mechanics make use of force fields, where the potential energy functions of different molecular interaction terms are parametrized to reproduce either accurate computational or experimental data. These force fields can then be used to evaluate and propagate the structure and energetics of biomolecules within a solvated environment. Traditional force fields cannot be used to model chemical reactions or tunnelling events, since electrons and quantum aspects of atoms are not explicitly modelled. In the past years, reactive force fields have been developed that are able to deal with bond breaking and formation.<sup>82</sup>

In atomistic force fields, different atom types are defined for various charge and hybridization states of each element, with unique parameters connected to each atom type. The energy terms in a molecular force field are divided into bonded and non-bonded interactions. The most common bonded energy terms comprise covalent bonds between two atoms, angles between three covalently bound atoms, dihedral angles between four covalently bound atoms, and improper torsional rotations describing out-of-plane motions of central atoms. Non-bonded interactions are modelled as electrostatic and dispersive interactions between atoms that are connected via more than three consecutive covalent bonds.

The mathematical formulation of these potential energy terms depends on the force field. The potential energy of covalent bonds and angles as well as of improper torsional rotations is usually parametrized in the form of the harmonic oscillator. Within the force field, bonds, angles, and improper torsional rotations oscillate around an equilibrium value for each specific set of atoms. These equilibrium values are determined experimentally or computationally, and harmonic force constants are assigned to each interatomic interaction. The resulting quadratic energy functions show a wrong behaviour for strongly deformed structures, but for biomolecular simulations near the equilibrium, classical force fields adequately treat the structural changes occurring in biomolecules.<sup>83</sup>

In the harmonic oscillator approximation, the potential energy of a covalent bond,  $V_b$ , between two atoms at distance  $r$  is expressed as,

$$V_b(r) = k_r (r - r_0)^2, \quad (2.1)$$

## 2 Theory

with the force constant  $k_r$  and the equilibrium bond length  $r_0$ .

The potential energy of an angle,  $V_a$ , formed by three covalently bound atoms with angle  $\theta$  is given by the harmonic potential,

$$V_a(\theta) = k_\theta (\theta - \theta_0)^2, \quad (2.2)$$

with the harmonic force constant  $k_\theta$  and the equilibrium angle  $\theta_0$ .

The potential energy for improper torsional rotations,  $V_e$ , of a group consisting of three atoms bound to a central fourth atom at the torsional angle  $\psi$  is given by,

$$V_e(\psi) = k_\psi (\psi - \psi_0)^2, \quad (2.3)$$

with the force constant  $k_\psi$  and the equilibrium torsional angle  $\psi_0$  (mostly  $0^\circ$ ).

The potential energy contribution of a dihedral angle,  $V_d$ , formed by four linearly connected atoms with dihedral angle  $\chi$ , is described by a sinusoidal term,

$$V_d(\chi) = k_\chi (1 + \cos(n\chi - \delta)), \quad (2.4)$$

where  $k_\chi$  determines the rotational barrier for the dihedral angle  $\chi$  with periodicity  $n$  and the phase shift  $\delta$ .

Non-bonded interactions usually use the Lennard-Jones potential to describe dispersive interactions between two atoms  $I$  and  $J$  that are not directly connected to each other,<sup>84</sup>

$$V_{LJ}(r_{IJ}) = \epsilon_{IJ} \left[ \left( \frac{r_{0,IJ}}{r_{IJ}} \right)^{12} - 2 \left( \frac{r_{0,IJ}}{r_{IJ}} \right)^6 \right], \quad (2.5)$$

where  $r_{0,IJ}$  is the equilibrium distance for the respective non-bonded interaction, and  $\epsilon_{IJ}$  is the well-depth of this minimum. The well-depth  $\epsilon_{IJ}$  is calculated from the parametrized atomic values  $\epsilon_I$  and  $\epsilon_J$  of the involved atom types by,

$$\epsilon_{IJ} = \sqrt{\epsilon_I \cdot \epsilon_J}. \quad (2.6)$$

The equilibrium distance  $r_{0,IJ}$  is calculated from the parametrized atomic values  $r_{0,I}$  and  $r_{0,J}$  of the involved atom types by,

$$r_{0,IJ} = \frac{1}{2} (r_{0,I} + r_{0,J}). \quad (2.7)$$

The electrostatic energy between two partially charged atoms  $I$  and  $J$  that are not directly bonded is given by the Coulomb potential,

$$V_{el}(r_{IJ}) = \frac{Q_I Q_J}{4\pi\epsilon_0 r_{IJ}}, \quad (2.8)$$

where the  $Q_i$  are the atomic partial charges, and  $\epsilon_0$  is the vacuum permittivity. The atomic partial charges are parametrized or calculated and assigned to the atoms of the

molecule.

In an additive potential energy function, all bonded and non-bonded interactions in the system of interest are summed up to give the combined force field potential,

$$\begin{aligned}
 V = & \sum_{bonds} k_r (r - r_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 \\
 & + \sum_{impropers} k_\psi (\psi - \psi_0)^2 + \sum_{dihedrals} k_\chi (1 + \cos(n\chi - \delta)) \\
 & + \sum_{non-bonded} \left( \epsilon_{IJ} \left[ \left( \frac{r_{0,IJ}}{r_{IJ}} \right)^{12} - 2 \left( \frac{r_{0,IJ}}{r_{IJ}} \right)^6 \right] + \frac{Q_I Q_J}{4\pi\epsilon_0 r_{IJ}} \right).
 \end{aligned} \tag{2.9}$$

## 2.2 Molecular dynamics simulations

In molecular dynamics (MD) simulations, the coordinates of the particles in the system of interest are propagated in time, yielding a trajectory. In classical mechanics, Newton's laws of motion can be used to describe the motion of particles in response to forces acting on them. Newton's second law states the direct proportionality of the masses  $m$  and acceleration  $\mathbf{a}$  of particles to the forces  $\mathbf{F}$  acting on them at time  $t$ ,

$$-\nabla V(\mathbf{x}(t)) = \mathbf{F}(\mathbf{x}(t)) = m \mathbf{a}(t) = m \frac{d\mathbf{v}(t)}{dt} = m \frac{d^2\mathbf{x}(t)}{dt^2}. \tag{2.10}$$

The forces  $\mathbf{F}(\mathbf{x}(t))$  are obtained as the negative gradient of the potential  $V(\mathbf{x}(t))$  that is calculated in the force field potential energy function, which depends on the coordinates  $\mathbf{x}(t)$  of the particles at time  $t$ . The acceleration  $\mathbf{a}(t)$  is the first time derivative of the velocities  $\mathbf{v}(t)$  and the second time derivative of the positions  $\mathbf{x}(t)$ .

In classical MD simulations, Newton's equation of motion is solved numerically to propagate a starting structure in time. The positions  $\mathbf{x}$  after a time step  $\Delta t$  can be calculated using the *Verlet* algorithm,<sup>85</sup>

$$\mathbf{x}(t + \Delta t) = -\mathbf{x}(t - \Delta t) + 2\mathbf{x}(t) + \mathbf{a}(t)(\Delta t)^2. \tag{2.11}$$

The acceleration  $\mathbf{a}(t)$  is calculated from the potential at time  $t$ . The Verlet algorithm is the basis for a recursive procedure for calculating all subsequent positions after  $n$  time steps. Another variant is the *leapfrog* algorithm that uses half-step velocities to update the positions. In the *velocity Verlet* algorithm, the positions and velocities after a time step  $\Delta t$  are calculated from the values of the previous time step,<sup>86</sup>

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \Delta t \mathbf{v}(t) + \frac{1}{2} \mathbf{a}(t)(\Delta t)^2, \tag{2.12}$$

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{1}{2} \Delta t (\mathbf{a}(t) + \mathbf{a}(t + \Delta t)). \tag{2.13}$$

The acceleration  $\mathbf{a}(t + \Delta t)$  is derived from the force field interaction potential using  $\mathbf{x}(t + \Delta t)$ .

## 2 Theory

The time step  $\Delta t$  has to be small enough to accurately capture the fastest motion in the system. For atomistic MD simulations, time steps of 1-2 fs are typically chosen to be able to describe bond vibrations of covalently bound hydrogen atoms.

To propagate the equations of motion, the initial configuration of the system of interest has to be prepared carefully. When simulating a protein, the starting structure of the protein (from X-ray crystallography, NMR, cryo-EM, or modelling) is usually embedded in a solvent to mimic the biological environment. Soluble proteins are solvated and neutralized in an ionic solution consisting of water and ions like sodium and chloride. To avoid surface effects and to decrease the system size, *periodic boundary conditions* (PBC) are usually employed where a central cell is replicated into all spatial directions.<sup>87</sup> This way, a particle leaving the original cell on one side will enter it from the opposite direction, whereas the central cell is propagated in the computations.

Depending on the size of the protein of interest, atomistic simulation systems can contain millions of atoms. One possibility to evaluate the long-range electrostatic energies of such large periodic macromolecular systems is the *particle mesh Ewald* (PME) technique.<sup>88</sup> In this approach, the electrostatic interaction potential is separated into short-range and long-range terms. The short-ranged potential is calculated in real space, while the long-ranged part is calculated in reciprocal space, using a fast Fourier transform on a discrete grid in space.

MD simulations can also be performed using combined quantum and classical energies and forces. The quantum mechanical energies and forces are calculated by employing, *e.g.*, the density functional theory (DFT) approach (see below), while the classical energies and forces are obtained using a force field.<sup>89</sup>

### 2.3 Principal component analysis

Classical atomistic MD simulations can nowadays sample microsecond time scales and therefore provide complementary information to experimental measurements. In addition to the dynamics of individual side chains and cofactors, concerted motions of groups of atoms within the protein are also of interest. The principal component analysis (PCA) method offers a powerful technique to identify the collective modes in the protein that contribute most significantly to the fluctuations in the protein motion.<sup>90</sup>

When PCA is applied to extract the slowest motions of a protein from the coordinates propagated in an MD simulation, the coordinates  $\mathbf{q}(t)$  of selected atoms over time are described by a linear combination of orthonormal basis vectors  $\boldsymbol{\eta}_\alpha$ , the so-called principal components (PCs),<sup>91</sup>

$$\mathbf{q}(t) \simeq \sum_{\alpha} a_{\alpha}(t) \boldsymbol{\eta}_{\alpha}. \quad (2.14)$$

The  $a_{\alpha}(t)$  are the time-dependent PCs. In this formulation, PCA transforms the correlated atomic coordinates  $\mathbf{q}(t)$  into new, uncorrelated variables, the PCs  $\boldsymbol{\eta}_{\alpha}$ . The PCs  $\boldsymbol{\eta}_{\alpha}$  are linear combinations of the originally observed atomic coordinates  $\mathbf{q}(t)$  and

constructed in such a way that the first PC,  $\eta_1$ , accounts for the largest variation in the original data. The PCs  $\eta_\alpha$  can be mapped back to the protein structure, to identify and analyze collective motion in the protein structure, so-called modes of motion. PCA is typically used to determine the smallest number of uncorrelated PCs that explain a large percentage of the total variation in the data.<sup>92</sup>

## 2.4 Continuum electrostatics and $pK_a$ calculations

Electrostatic interactions play an important role in biological systems.<sup>93</sup> Characterization of the electrostatic properties of proteins in aqueous solution can give valuable insight for understanding their structure and function. This characterization can be achieved by using continuum electrostatics calculations.

To reduce the complexity of the solvated protein, the solvent is usually modelled using implicit solvation techniques and described as a polarizable medium. The different dielectric environments of protein and solvent are identified by molecular surface regions formed by the van der Waals envelope of the biomolecule and a probe solvent molecule. The cavity inside this surface is assigned a low dielectric constant mostly around  $\epsilon = 4$  to model the protein interior, whereas the exterior region is typically assigned the dielectric constant of water,  $\epsilon = 80$ . The charge distribution of the biomolecule is usually represented as point charges located at the atomic nuclei.<sup>94</sup>

The classical treatment of electrostatic interactions in solution is based on the Poisson-Boltzmann equation (PBE), which describes the distribution of the solving electrostatic potential of solvated charged species. The PBE for a one-to-one electrolyte relates the electrostatic potential  $\phi(x)$  to the dielectric function of the system ( $\epsilon(x)$ ), the ion accessibility function ( $\bar{\kappa}^2$ ), and the charge distribution of the solute ( $f(x)$ ),<sup>95</sup>

$$-\nabla(\epsilon(x)\nabla\phi(x)) + \bar{\kappa}^2(x)\sinh(\phi(x)) = f(x). \quad (2.15)$$

Classical electrostatics can also be used to calculate electrostatic solvation free energies when a molecule is transferred from one dielectric medium to another, as the difference in electrostatic free energy of the solute in the two phases. This approach is also employed in the calculation of  $pK_a$  values of amino acids in proteins.<sup>96</sup>

The  $pK_a$  values of ionizable groups in proteins can significantly differ from the  $pK_a$  values in solution. The source of the shift lies in the interactions among ionizable groups in proteins and their desolvation energy in the hydrophobic protein interior. The  $pK_a$  value in the protein interior is computed by adding a protein induced shift,  $\Delta pK_a$ , which is calculated from electrostatic energies, to the reference  $pK_a$  value of the titratable group in solution. The calculated electrostatic energies comprise the desolvation penalty of the titratable group when entering the protein, as well as its interactions with the non-titratable and titratable groups of the protein.<sup>97</sup>

To determine the protonation states of the interacting amino acids, a Monte Carlo method can be applied to sample the combinations of side chain protonation states, and to compute the average protonation of residues as a function of pH.<sup>98</sup>

## 2.5 Quantum mechanics

Electrons and their motion cannot be modelled accurately using purely classical mechanics. Due to their small mass, the laws of classical mechanics cannot be applied to electrons and their behaviour can only be described accurately using the theory of quantum mechanics, where electrons are represented as wave functions. In this section, basic concepts of molecular quantum mechanics and electronic structure theory are presented.

### 2.5.1 Born-Oppenheimer approximation

For a stationary state of a quantum mechanical system like a molecule, the total wave function  $\Psi$  and energy  $E$  can be obtained from the non-relativistic time-independent Schrödinger equation,<sup>99</sup>

$$\hat{H}\Psi = E\Psi. \quad (2.16)$$

The so-called Hamilton operator or Hamiltonian  $\hat{H}$  contains the kinetic energy operators of electrons ( $\hat{T}_e$ ) and nuclei ( $\hat{T}_n$ ), as well as the potential energy operators for the interactions between electrons ( $\hat{V}_{ee}$ ), nuclei ( $\hat{V}_{nn}$ ), and between nuclei and electrons ( $\hat{V}_{ne}$ ),

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{ee} + \hat{V}_{nn} + \hat{V}_{ne}. \quad (2.17)$$

In principle, the motion of nuclei and electrons is coupled to each other. However, to reduce the complexity of the Schrödinger equation, it can be assumed that due to the large velocity difference between electrons and nuclei, the nuclei are stationary with respect to the motion of the electrons, and electronic and nuclear motion can be separated. Pictorially, the electrons are assumed to instantaneously follow the motion of the nuclei and relax with respect to their movement. This is the Born-Oppenheimer approximation,<sup>100</sup> which allows the construction of potential energy surfaces. The total wave function  $\Psi$  of a system is approximated by the product of the wave functions of the nuclei,  $\mathcal{X}$ , and electrons,  $\Phi$ ,

$$\Psi = \mathcal{X} \Phi. \quad (2.18)$$

The electronic wave function is relaxed for a fixed nuclei geometry. An electronic Hamiltonian is constructed by neglecting the kinetic energy of the nuclei and treating the interaction potential between the nuclei as a constant,

$$\hat{H}_{elec} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ne}. \quad (2.19)$$

In the electronic Schrödinger equation, the electronic Hamiltonian acts on the electronic wavefunction  $\Phi$  to obtain the electronic energy  $E_{elec}$ ,

$$\hat{H}_{elec}\Phi = E_{elec}\Phi. \quad (2.20)$$



### 2.5.2 Hartree-Fock theory

Due to the correlated motion of the electrons in an atom or a molecule, the Schrödinger equation cannot be solved analytically for chemical systems with more than one electron. Therefore, numerical techniques have been developed to solve the Schrödinger equation for atoms or molecules. One of these techniques is Hartree-Fock theory, which is a wave function-based electronic structure method. Within this framework, the true electronic wave function  $\Phi$  of a molecule is approximated by a single Slater determinant, composed of one-electron wave functions or molecular spin orbitals  $\phi$ .

The energy of the system is obtained from the Hartree-Fock equations,

$$\hat{F}\phi_i = \varepsilon_i\phi_i, \quad (2.21)$$

where  $\hat{F}$  is the Fock operator,  $\phi_i$  is an occupied molecular spin orbital, and  $\varepsilon_i$  is the eigenvalue that can be interpreted as the orbital energy, as formulated in Koopmans' theorem.<sup>101</sup> The Hartree-Fock equations are coupled non-linear integro-differential equations and are solved iteratively.<sup>102</sup>

The Fock operator consists of the kinetic energy of the electrons,  $\hat{T}_e$ , the interaction energy between electrons and nuclei,  $\hat{V}_{ne}$ , the Coulomb repulsion between electrons,  $\hat{J}$ , and the exchange interaction between electrons of same spin,  $\hat{K}$ ,

$$\hat{F} = \hat{T}_e + \hat{V}_{ne} + \hat{J} - \hat{K}. \quad (2.22)$$

In Hartree-Fock theory, only the averaged potential acting on each electron, generated by the other electrons, is taken into account, and results in the mean field description of the electron-electron interaction. This approach neglects electron correlation effects, which describe electrons moving in a correlated way and avoiding being too close to each other.

The total energy as a functional of a wave function,  $E[\Phi]$ , can be minimized by applying the generally formulated variational principle,

$$E[\Phi] = \frac{\langle \Phi | \hat{H} | \Phi \rangle}{\langle \Phi | \Phi \rangle} \geq E_0. \quad (2.23)$$

The variational theorem states that the energy of any trial wave function  $\Phi$  cannot be lower than the true ground-state energy  $E_0$ , and that the trial wave function is closest to the true wave function if the energy eigenvalue is stationary with respect to variations in the wave function  $\Phi$ .<sup>102</sup>

In Hartree-Fock theory, the electronic wave function  $\Phi$  of a molecule is approximated as a single Slater determinant, formed by the molecular spin orbitals  $\phi$ . Employing the LCAO (linear combination of atomic orbitals) expansion commonly used in quantum chemistry, the initially unknown molecular orbitals  $\phi_i$  are expanded in a set of known basis functions of atomic orbitals,

$$\phi_i = \sum_{\rho} c_{\rho i} \chi_{\rho}, \quad (2.24)$$

## 2 Theory

where  $\chi_\rho$  are the atomic orbitals or basis functions, and the  $c_{\rho i}$  are the coefficients of the basis functions in the linearly combined molecular orbital  $\phi_i$ . Using this approach, the constructed Fock matrix contains the interactions between all atomic orbitals,

$$F_{pq} = \langle \chi_p | \hat{F} | \chi_q \rangle. \quad (2.25)$$

Within the LCAO scheme, the energy is minimized with respect to the coefficients  $c_{\rho i}$  specifying the molecular orbitals. These atomic orbital coefficients are determined in the Roothaan-Hall equations, written as a matrix equation,

$$\mathbf{FC} = \mathbf{SC}\epsilon, \quad (2.26)$$

with the Fock matrix  $\mathbf{F}$ , the coefficient matrix  $\mathbf{C}$  containing the atomic orbital coefficients of the molecular orbitals, the overlap matrix  $\mathbf{S}$  containing the overlap elements between all atomic orbitals, and the diagonal matrix  $\epsilon$  containing the energies of the molecular orbitals.<sup>103</sup> The Roothaan-Hall equations are solved using the self-consistent field (SCF) method, where guessed starting molecular orbitals are iteratively refined until they remain unchanged.

### 2.5.3 Density functional theory

While Hartree-Fock theory uses a wave function-based approach to solve the Schrödinger equation for a molecule, a different electronic structure method was developed starting from the three-dimensional electron density of a molecule. This technique is referred to as density functional theory (DFT), and it has become one of the most important methods in molecular quantum chemistry.

#### General concept

The foundation of density functional theory (DFT) was the realization that the ground-state energy of a molecule can be calculated from its electron density,  $\rho$ . In the Hohenberg-Kohn theorems, it was shown in 1964 that the ground-state density determines the ground state properties and therefore also the Hamiltonian of a molecular system. Hohenberg and Kohn further showed that the ground-state electron density can be determined applying the variational principle.<sup>104</sup>

The electronic energy is written as a functional of the electron density, that itself is a function of the three spatial coordinates,

$$E[\rho] = T_e[\rho] + V_{ne}[\rho] + V_{ee}[\rho]. \quad (2.27)$$

$T_e[\rho]$  is the kinetic energy of the electrons,  $V_{ne}[\rho]$  is the nuclei-electron interaction term, and  $V_{ee}[\rho]$  collects the electron-electron interactions that comprise the Coulomb repulsion of the density,  $J[\rho]$ , and the non-classical electron exchange and correlation energies,  $E_X[\rho]$  and  $E_C[\rho]$ .  $V_{ne}[\rho]$  and  $J[\rho]$  can be computed exactly from the electron density, while the exact form of the other density functionals is unknown and has to be approximated.<sup>105</sup>

In 1965, Kohn and Sham defined the kinetic energy  $T_S$  of a system of non-interacting electrons,<sup>106</sup> while the remaining part of the total kinetic energy is denoted  $T_C$ . The non-interacting kinetic energy,  $T_S$ , makes up a major part of the total kinetic energy and can be obtained exactly by representing the wave function of the non-interacting system, called the Kohn-Sham wave function, by a single determinant composed of the so-called Kohn-Sham orbitals  $\phi_i$ . In this Kohn-Sham DFT formalism, the electron density  $\rho$  of the non-interacting system has been constructed to be identical to the electron density of the real, interacting system and can be calculated from the Kohn-Sham orbitals  $\phi_i$ ,

$$\rho = \sum_i \phi_i^2, \quad (2.28)$$

summing over all occupied Kohn-Sham orbitals.

After the partitioning of the kinetic energy, all unknown contributions of the energy functional are collected in the exchange-correlation functional,

$$E_{XC}[\rho] = (T_e[\rho] - T_S[\rho]) + (V_{ee}[\rho] - J[\rho]). \quad (2.29)$$

The exact form of the exchange-correlation functional is unknown, which led to the development of approximate exchange-correlation functionals.

As in Hartree-Fock theory, the Kohn-Sham molecular orbitals are guessed initially and expanded in a basis set of atomic orbitals, which are optimized by applying the self-consistent field method with the objective of finding the coefficients for the basis set expansion.

### DFT exchange-correlation functionals

The quest for the unknown DFT exchange-correlation functional has brought up various approaches and is still going on today.

Kohn and Sham suggested in 1965 a local density approximation (LDA) for a hypothetical homogeneous uniform electron gas with constant or slowly varying electron density.<sup>106</sup> This approximation is useful for metallic models, but does not apply to small molecules with an inhomogeneous electron density distribution.

To take into account inhomogeneities of the electron density, the local gradient of the electron density is included, leading to the semilocal generalized gradient approximation (GGA). Among the GGA functionals are Becke's B88X gradient correction for exchange functionals, containing one empirical parameter that was fitted to Hartree-Fock exchange energies of noble gas atoms,<sup>107</sup> and the Lee-Yang-Parr (LYP) correlation functional.<sup>108</sup> Combining Becke's B88X functional with Perdew's correlation functional P86<sup>109</sup> yields the popular BP86 GGA exchange-correlation functional.

Going beyond the GGA, the meta-generalized gradient approximation (meta-GGA) includes second derivative information on the electron density, *e.g.* in the form of the kinetic energy. One example of this approach is the TPSS functional, a so-called constraint-driven functional that was constructed non-empirically.<sup>110</sup>

## 2 Theory

Hybrid functionals replace a fraction of the local density functional exchange by the exact exchange energy of the Slater determinant of the Kohn-Sham orbitals. The parameters for the proportions of the different interaction energies are fitted semi-empirically.<sup>111</sup> Among the most used hybrid functionals in molecular quantum chemistry is the B3LYP functional that is composed of the LDA exchange and correlation functionals, Becke’s B88X gradient correction for exchange functionals, the GGA correlation functional LYP, and  $a_0 = 20\%$  exact exchange energy from Hartree-Fock,

$$E_{XC}^{B3LYP} = (1 - a_0) E_X^{LDA} + a_0 E_X^{HF} + a_X B88_X + a_C E_C^{LYP} + (1 - a_C) E_C^{LDA}. \quad (2.30)$$

The values of the three linear parameters are chosen based on fitting to experimental data of small molecules.<sup>112</sup> B3LYP has been widely used to computationally investigate metalloenzymes,<sup>113</sup> and it is the main DFT functional used in this thesis.

An example of another hybrid functional is the TPSSh functional, a one-parameter hybrid version of the TPSS functional with a smaller fraction of  $a = 10\%$  exact exchange,<sup>114</sup> with promising results in modelling bioinorganic systems,<sup>115</sup>

$$E_{XC}^{TPSSh} = a E_X^{exact} + (1 - a) E_X^{TPSS} + E_C^{TPSS}. \quad (2.31)$$

In Hartree-Fock, the exact Coulomb and exchange terms of a single electron cancel out, but this is not the case in DFT. The approximated exchange energy in DFT, especially its wrong long-range behavior, gives rise to the so-called self-interaction error because the self-interaction in the Coulomb potential is not cancelled exactly by the exchange energy. This self-interaction error leads to an artificial stabilization of states with delocalized electronic charge density, which also leads to errors in the description of charge-transfer complexes. Transition states have more delocalized electron densities compared to minima on the potential energy surface, and they can be artificially stabilized by the self-interaction error. Therefore, pure DFT functionals have a tendency to underestimate the barriers for small organic reactions.<sup>116</sup> Inclusion of exact exchange from Hartree-Fock can help to improve the calculated chemical reaction barrier heights.<sup>105</sup>

### Dispersion corrected DFT

Dispersive interactions between non-polar molecules are important in many (bio)chemical systems, especially in large proteins. Commonly used standard density functionals do not describe long-range dispersion interactions correctly. As van der Waals interactions appear even at distances where overlap of the electron densities of fragments is negligible, local exchange-correlation functionals fail to reproduce them.<sup>117</sup>

Grimme introduced an empirical atom pairwise potential as an additive correction to the self-consistent DFT energy, which improved the performance of density functionals in modelling non-covalent interactions. The dispersion correction was described in its original form as,<sup>117</sup>

$$E_{disp} = -s_6 \sum_{i,j} \frac{C_6^{(ij)}}{R_{ij}^6} f_{dmp}(R_{ij}), \quad (2.32)$$

where  $R_{ij}$  is the interatomic distance, and  $s_6$  is an empirical global scaling factor to account for the different behavior of the intermolecular potential of density functionals. The pairwise dispersion coefficients  $C_6^{(ij)}$  are calculated from atomic  $C_6$  coefficients,

$$C_6^{(ij)} = \frac{2C_6^{(i)}C_6^{(j)}}{C_6^{(i)} + C_6^{(j)}}. \quad (2.33)$$

The empirical damping function  $f_{dmp}$  determines the range of the dispersion correction. Originally, the damping function was designed to decay to zero at small distances  $R_{ij}$  such that the dispersion correction becomes negligible for typical covalent bond distances,

$$f_{dmp}(R_{ij}) = \frac{1}{1 + e^{-\alpha\left(\frac{R_{ij}}{R_{0,ij}} - 1\right)}}. \quad (2.34)$$

The atom pairwise cutoff radius  $R_{0,ij}$  is calculated as the sum of atomic van der Waals radii. With this damping function, the dispersion correction is largest for typical van der Waals distances. To improve the model, also the Becke-Johnson (BJ) approach with finite damping for small interatomic distances was implemented, that avoids repulsive interatomic forces at short distances. In both forms, the empirical additive potential is of  $-R^{-6}$  form for large intermolecular distances. In this thesis, the DFT-D3 approach was used in all DFT calculations.<sup>118</sup>

#### 2.5.4 Basis sets

In the LCAO expansion, molecular orbitals are expressed as a linear combination of atomic orbitals. Since the Schrödinger equation can be solved analytically only for one-electron species, the exact form of the atomic orbitals in a molecule is unknown and has to be approximated. Initially, Slater-type orbitals (STOs) were used to approximate atomic wave functions, which were constructed as simplified hydrogen-like wave functions and therefore decay exponentially with distance.<sup>119</sup> However, the evaluation of multicenter integrals with STOs is impractical, which led to the introduction of Gaussian-type orbitals (GTOs) of the form,<sup>120</sup>

$$\chi^{GTO}(r) \propto e^{-\alpha r^2}. \quad (2.35)$$

The computational advantage of GTOs is that the product of two Gaussians yields a new Gaussian function, which simplifies the computation of multicenter integrals. To increase the accuracy of the modelled atomic orbital, contracted Gaussian functions can be formed from a linear combination of Gaussian functions. In a so-called minimal basis set, each spatial atomic orbital in an occupied electron shell is described by a single

(contracted) basis function. To increase the accuracy of the calculations, multiple basis functions can be used to describe one spatial atomic orbital. The double zeta (DZ) basis set uses two basis functions per spatial atomic orbital, while three basis functions are used in a triple zeta (TZ) basis set to represent one spatial atomic orbital.

Since increasing the number of basis functions leads to an increase in unknown coefficients and therefore in computational cost, and chemical bonding does not affect the core shells of atoms, a strategy has been developed to combine a minimal basis set for the core shell with usually a double zeta or even a triple zeta basis set for the valence shell. Such a combined basis set is called a split-valence (SV) basis set. To model polarized atomic orbitals in molecules, polarization (P) functions representing orbitals with higher orbital quantum number are included, such as p-orbitals for the description of s-orbitals, and d-orbitals for the description of p-orbitals.

When modelling molecules with an excess negative charge leading to a diffuse electron density distribution, so-called diffuse basis functions can be included to improve the description of the outer electron density.

The Karlsruhe def2-SV(P) basis set is a split-valence basis set with two basis functions for the valence shell. Additionally, it contains a polarizing d-function for p-elements and a diffuse p-set for d-elements. The def2-SVP basis set additionally includes a polarizing p-function for hydrogen atoms and an f-function for d-elements. The def2-TZVP basis set is a split-valence basis set with three basis functions for the valence shell and polarization functions for all elements.<sup>121–123</sup> In this thesis, all final electronic energies are reported at def2-TZVP level, while structure optimizations are performed at def2-SVP level, with def2-TZVP basis sets used for metals.

## 2.6 Quantum chemical cluster models

Chemical reactions including bond breaking and formation cannot be described by classical force fields. To model these reactions, quantum mechanical methods are necessary. However, enzyme-catalyzed reactions pose the challenge that a whole solvated enzyme cannot be modelled at feasible computational cost using quantum mechanics. One approach to computationally model biochemical reactions is the use of quantum chemical cluster models (in short, QM cluster models or DFT models). In this approach, a QM cluster model of the active site of an enzyme is constructed. The model usually contains the side chains of amino acids within the first and possibly second solvation sphere of the site of interest, as well as the substrates and cofactors involved in the biochemical reaction.

The three-dimensional atomic coordinates can be taken from resolved X-ray, NMR, cryo-EM, or modelled structures, or a classically relaxed MD-model thereof. Amino acid side chains are usually included up through the  $\beta$ -carbon with bound hydrogens, while the backbone is seldom explicitly modelled. The unsaturated terminal carbon atoms are saturated with hydrogen atoms and kept fixed to their original positions during geometry optimization (coordinate-locking scheme) to mimic steric constraints of the protein framework. The enzyme surrounding can be modelled as a homogeneous

polarizable medium with a dielectric constant of  $\epsilon = 4$  for proteins.<sup>124</sup>

QM cluster models can be used to test molecular mechanisms of enzymatic reactions. Structures and energetics of intermediates and transition states of biochemical reaction sequences can be identified and characterized. Obtained reaction barriers can be compared to experimental reaction rates using transition state theory. Nowadays, QM cluster models consisting of up to 300 atoms can be constructed, and are typically treated with DFT,<sup>125</sup> but even small quantum mechanical models can capture the main features of chemical reaction mechanisms.<sup>113</sup>

## 2.7 Continuum solvation

Solvation effects play an important role in chemistry, which is why the screening energy of molecules in a solvating dielectric medium should be included in quantum chemical calculations. The interior of a protein is usually assumed to have a low dielectric constant around  $\epsilon = 4$ ,<sup>124</sup> while a high dielectric medium such as water has a relative dielectric permittivity of around  $\epsilon = 80$ .<sup>126</sup>

To take into account the dielectric screening effects in specific solvents without modelling the surroundings explicitly, dielectric continuum solvation models have been developed where the solute molecule is embedded in a dielectric continuum of permittivity  $\epsilon$ . This way, statistically averaged information on the solvent effect are taken into account, by modelling the solute in a cavity surrounded by a macroscopic homogeneous continuum medium with a suitable dielectric constant. The homogeneous dielectric medium is polarized by the charge distribution of the solute which leads to a surface charge distribution at the cavity interface.<sup>127</sup>

In the Conductor-like Screening Model (COSMO), the dielectric continuum is approximated by a scaled conductor, leading to the boundary conditions of vanishing total potential for an infinite conductor. The solute forms a real-shape cavity within the dielectric continuum which is defined by the van der Waals surface of the solute. Such molecular-shaped cavities require a numerical treatment of the polarization of the continuum, unlike spherical or ellipsoidal cavities.<sup>126</sup> The COSMO model has been employed in this thesis for all DFT calculations.

## 2.8 Reaction path optimization

In a chemical reaction, reactants are transformed into products with different geometry or electronic structure, with both reactants and products being stable chemical species. Parameters describing the chemical coordinate transformation between reactants and products are referred to as reaction coordinate. On an energy surface of a chemical reaction, reactants and products are located in separated energy minima. Different pathways between reactants and products can be found on the energy surface and connected with different energy barriers. The connecting path with the lowest energy barrier is called the minimum energy path (MEP). The transition state (TS) of the reaction corresponds

## 2 Theory

to the highest point in energy along the identified reaction coordinate between reactants and products, and is located at a stationary first-order saddle point.

Chemical reaction barriers are directly linked to the experimental reaction rate constant  $k$  via transition state theory (TST) and the Eyring relation,<sup>128</sup>

$$k = \kappa \frac{k_B T}{h} e^{-\frac{\Delta G^\ddagger}{RT}}, \quad (2.36)$$

with the Boltzmann constant  $k_B$ , Planck’s constant  $h$ , and the gas constant  $R$ .  $\kappa$  is the transmission coefficient that is often assumed to equal unity. The transmission coefficient reflects the possible reduction of the reaction rate by activated complexes recrossing the transition state and returning to the reactants without reacting to the product.  $\Delta G^\ddagger$  is the free energy difference between reactant and transition state (TS) species. The Eyring relation implies that reactions with large barriers ( $\Delta G^\ddagger$ ) will proceed more slowly than reactions with small barriers.

To optimize a discretized minimum energy path and to find an approximate geometry of the transition state, so-called chain-of-state methods can be applied. These methods require optimized geometries of reactant and product states (“double-ended” methods). The two structures are subsequently transformed into each other, leading to an ensemble of replicas and their corresponding energies. The simplest chain-of-state method is the linear synchronous transit (LST) method. It linearly interpolates between the start and end structures without further optimization. This approximation is usually employed to obtain an initial guess of the reaction path.<sup>129</sup> Starting from the linear interpolation, the structures of each point along the reaction path are iteratively minimized, except for certain restrained internal coordinates that describe the main differences between reactant and product structures and keep the structures distributed along the path. The reactant and product structures are fixed during the minimization. The reaction path is optimized self-consistently, by approximating the underlying potential energy surface.<sup>130</sup>

## 2.9 Hybrid quantum mechanics/molecular mechanics (QM/MM)

While QM cluster models accurately describe the enzyme active site, they do not include the protein surroundings that may stabilize the active site through steric and electrostatic interactions and accelerate convergence of the calculations.<sup>131</sup> To explicitly incorporate the whole solvated enzyme into the computational model, hybrid quantum mechanics/classical mechanics (QM/MM) methods can be used which will be introduced in this section.

### 2.9.1 General concept

Hybrid quantum mechanics/molecular mechanics (QM/MM) calculations were first introduced in 1976.<sup>132</sup> In the modern variant of QM/MM methods, the active site of an



## 2.9 Hybrid quantum mechanics/molecular mechanics (QM/MM)

enzyme forms the QM region and is treated with quantum mechanics, *e.g.* DFT, while the remaining part of the enzyme is modelled using a force field and forms the MM region. The QM region usually includes side chains of catalytically important amino acids and substrates and cofactors involved in the enzymatic reaction. Amino acid side chains are usually divided between C $\alpha$  and C $\beta$  atoms, with the side chain included in the QM region, while the backbone is part of the MM region. For the QM calculations, the dangling bonds of the  $\beta$ -carbons can be saturated by additional link atoms, usually modelled as hydrogen atoms, as in the QM cluster approach.<sup>133</sup>

### 2.9.2 QM/MM schemes and embedding

QM/MM energies can be computed employing both subtractive and additive energy schemes.<sup>89</sup> In the subtractive QM/MM scheme, the MM energy of the whole system is calculated using a force field (FF), and then the force field energy of the QM region is subtracted and replaced by the QM energy of the QM region,

$$E_{QM/MM}^{subtractive} = E_{FF}^{total} - E_{FF}^{QM} + E_{QM}^{QM}. \quad (2.37)$$

In this approach, no explicit coupling terms between QM and MM regions are calculated, they are implicitly contained in the MM calculation of the total system. For the MM calculations, force field parameters for the QM region are required.

In the additive QM/MM scheme, the total QM/MM energy of the system is calculated as,

$$E_{QM/MM}^{additive} = E_{QM}^{QM} + E_{FF}^{MM} + E_{coupling}. \quad (2.38)$$

In this approach, the energy of the MM region is calculated using a force field (FF), and the energy of the QM region is calculated using a quantum mechanical method. Additionally, the interactions between the two regions are calculated explicitly in the coupling term  $E_{coupling}$  in three different embedding schemes.<sup>89</sup>

In *mechanical embedding*, an MM charge model has to be developed for the QM region beforehand. These charges are then used to calculate the electrostatic interactions between QM and MM regions depending on the atomic positions, while the charges are not updated during the calculations.

Using *electronic embedding* or *electrostatic embedding*, the QM region is polarized by the MM region during the calculations. The QM calculation is performed by incorporating the MM point charges in the QM Hamiltonian and calculating the interactions of the QM electrons  $i$  and QM nuclei  $a$  with the MM point charges  $j$ ,

$$\hat{H}_{coupling}^{electronic} = - \sum_{i,j} \frac{q_j}{r_{ij}} + \sum_{a,j} \frac{q_j q_a}{r_{aj}}. \quad (2.39)$$

In this approach, the electronic structure of the QM region is polarized by the charge distribution of its environment. The charges of the QM region do not have to be parametrized in advance, while the MM partial charges are usually available in the

widely used biomolecular force fields.

In *polarized embedding*, both QM and MM regions are polarizing each other. This requires a polarizable force field where the atomic partial charges are not fixed, which can be achieved using atomic multipoles, induced point dipoles, Drude oscillators, fluctuating charges, atomic polarizabilities, or virtual interaction sites.<sup>134</sup>

When using electrostatic or polarized embedding, an overpolarization of the QM density by the MM charges in close proximity can occur. To avoid this, MM point charges in the link region are usually deleted or redistributed among the closest atoms.<sup>133</sup>

## 2.10 Energy minimization

Equilibrium geometries of molecules correspond to minima on the potential energy surfaces of molecules. At the energy minimum, the gradient is zero with a positive curvature in all directions, *i.e.* the Hessian is positive definite,  $\frac{d^2 E}{dq^2} > 0$ . Since it is not feasible to compute the whole potential energy surface of a molecule, algorithms have been developed to identify energy minima, corresponding to stable geometries of molecules.

The simplest energy minimization algorithm is the method of *steepest descent*,<sup>135,136</sup> a first-order iterative optimization algorithm. In this procedure, a step is taken in the direction of the lowest energy, opposite to the energy gradient from the current point,  $\mathbf{g}_c$ ,

$$\mathbf{s} = -\alpha_c \mathbf{g}_c. \quad (2.40)$$

The step size  $\alpha_c$  is determined by *line search* in each iteration. Exact line search gives the exact minimum along the steepest descent direction, whereas in inexact line search, a point of lower energy along the steepest descent direction is identified.

The *conjugate gradient* technique provides another useful optimization method.<sup>137</sup> This iterative algorithm makes use of the previous history of minimization steps as well as the current gradient to determine the next step. The first search direction is the initial gradient, as in steepest descent. Subsequently, the previous search direction is added to the current gradient to eliminate repeated minimization along the same directions and to remain near the minimum in the previous search direction.<sup>138</sup>

A second-order minimization algorithm is the *Newton-Raphson* method, an adopted version of which has been used in QM/MM optimization in this thesis. In contrast to steepest descent and conjugate gradients methods, both of which use first-order derivatives, the Newton-Raphson method uses second-order derivative information to locate the minimum of the energy function. The idea behind the Newton-Raphson method is to iteratively approximate the energy function by a local second-order model, and then move to the minimum of this local quadratic model. The quadratic function is given by a Taylor expansion of the energy truncated at second order. The displacement step  $\mathbf{s}$  to the minimum of the constructed local quadratic approximation to the potential energy surface is given by,<sup>139</sup>

$$\mathbf{s} = -\mathbf{H}_c^{-1} \mathbf{g}_c, \quad (2.41)$$

calculated from the inverse Hessian  $\mathbf{H}_c$  and the gradient  $\mathbf{g}_c$  at the current point. When  $\mathbf{H}$  is the exact Hessian, the step is referred to as Newton step. When an approximate Hessian is used, the step is called a quasi-Newton step.

In the *restricted second-order model*, the quadratic model is applied only within a trust region around the current point, defined in every iteration by a trust radius. In these iterative *trust region methods*, the step taken is towards the minimum of the restricted quadratic model. It is either the Newton step towards the stationary point of the restricted second-order model, if it lies inside the trust region, or a step to the lowest point on the boundary of the trust region.<sup>140</sup>

## 2.11 Umbrella sampling and the weighted histogram analysis method

The calculation of free energy differences is important to obtain barriers and thermodynamic driving forces of biochemical reactions.<sup>141</sup> Umbrella sampling can provide a free energy profile along a reaction coordinate.<sup>142,143</sup> In this method, bias potentials are applied to ensure efficient sampling along the whole reaction coordinate from reactants to products. The reaction coordinate is divided into windows, each of which is sampled in a restrained MD simulation. Often, harmonic bias potentials are applied for the restrained sampling. From the sampled distribution along the reaction coordinate, the free energy change in each window can be calculated. Unbiased free energy curves  $A_i(x)$  can be obtained from the biased simulation distribution  $P_i^{biased}(x)$  and the bias potential  $w_i(x)$  in window  $i$ ,

$$A_i(x) = -k_B T \ln \left( P_i^{biased}(x) \right) - w_i(x) - k_B T \ln \left\langle e^{-\frac{w_i(x)}{k_B T}} \right\rangle. \quad (2.42)$$

The last term is the free energy shift associated with introducing the bias potential and connecting the free-energy curves of the different windows. This term is initially unknown.

The different windows can be combined by using *e.g.* the weighted histogram analysis method (WHAM).<sup>144,145</sup> WHAM optimizes the links between the windows and calculates an estimate of the free energy profile by taking into account all biased simulations. This method requires sufficient sampling overlap between the different windows. The global unbiased distribution  $P^{unbiased}(x)$  is calculated by combining the unbiased probability distributions of the individual windows which are weighted by a factor  $p_i(x)$ ,

$$P^{unbiased}(x) = \sum_i p_i(x) P_i^{unbiased}(x). \quad (2.43)$$

The weights  $p_i(x)$  are iteratively optimized to minimize the statistical error of the total unbiased probability distribution  $P^{unbiased}(x)$ .



## 3 Methods

### 3.1 General

Quantum chemical calculations were performed using TURBOMOLE,<sup>146</sup> and it was interfaced with CHARMM by a Python module for QM/MM calculations.<sup>147,148</sup>

Classical atomistic molecular dynamics simulations were performed at  $T = 310$  K using NAMD.<sup>149</sup> Periodic boundary conditions were applied using cubic water boxes, and friction was introduced employing Langevin dynamics. Van der Waals interactions and direct electrostatics were truncated at a cutoff distance of 12 Å. Long-range electrostatics were calculated using the particle mesh Ewald approach,<sup>150</sup> with full electrostatics evaluated every second time step. Constant pressure was maintained by Nosé-Hoover Langevin piston pressure control.<sup>151,152</sup>

Structures and simulations were visualized and analyzed using VMD.<sup>153</sup>

### 3.2 The molecular chaperone Hsp90

#### 3.2.1 Quantum chemical and hybrid QM/MM calculations

QM cluster and QM/MM models of the Hsp90 ATPase site were constructed using an MD-relaxed model of the X-ray structure with PDB ID: 4IVG from *Danio rerio*.<sup>154</sup> The QM region contained ATP,  $\text{Mg}^{2+}$ , the side chains of Arg-32, Glu 33, Asn-37, Ser-99, and Arg-380 (residue numbering of yeast Hsp82), and six water molecules. The backbone of residues 118-124 was added to the QM region in the QM cluster models only. The QM region was treated by dispersion corrected DFT (B3LYP-D3) with def2-SVP and def2-TZVP (for Mg) basis sets.<sup>112,118</sup>

The QM cluster models were embedded in a continuum polarizable medium with  $\varepsilon = 4$  using COSMO.<sup>126</sup> Terminal carbons of included side chains, as well as C $\alpha$  atoms of included backbone, were kept fixed. QM cluster reaction pathways were obtained employing the Woelfling method, a chain-of-state method.<sup>130</sup> Vibrational and entropic corrections were calculated from the numerical Hessian of the QM cluster models using the B3LYP-D3 functional with def2-SVP/def2-TZVP (Mg) basis sets and a polarizable medium with  $\varepsilon = 4$ .<sup>155</sup> The distance between the C $\zeta$  atom of Arg-32 and the C $\delta$  atom of Glu-33 in the QM cluster models was scanned in the reactant, transition state, and product geometries of the optimized reaction path, by shifting the whole Arg-32 side chain.

QM/MM models comprising *ca.* 17 000 atoms were extracted from the crystal structure with PDB ID: 4IVG after solvation in a water box with 150 mM NaCl, as a sphere

### 3 Methods

of 30 Å radius around ATP. Link atoms were inserted between the C $\alpha$  and C $\beta$  atoms of the side chains included in the QM region. The MM region was modelled using the CHARMM36 force field, and was fixed in QM/MM reaction pathway optimizations.<sup>156</sup> Individual residues were removed from the QM region of the optimized reaction pathway and energies were recalculated using the QM cluster approach to obtain residue contributions. Umbrella sampling simulations consisted of 27 ps QM/MM MD simulations per free energy profile at  $T = 310$  K, using a 1 fs integration time step. Harmonic bias potentials were applied on the reaction coordinate from reactants to products with force constants of 100 or 500 kcal mol<sup>-1</sup> Å<sup>-2</sup>, with atoms more than 15 Å from the QM region kept fixed. Free energy profiles were calculated from the umbrella sampling simulations by applying the weighted histogram analysis method (WHAM) using a convergence criterion of 0.00001 kcal mol<sup>-1</sup>.<sup>157</sup> The statistical error of the obtained free energy profiles was evaluated employing the bootstrap error analysis of WHAM, using ten Monte Carlo trial steps. Rate constants  $k$  were computed from the Eyring equation,

$$k = \kappa \frac{k_B T}{h} e^{-\frac{\Delta G^\ddagger}{RT}}, \quad (3.1)$$

with the reflection coefficient  $\kappa = 1$ , Boltzmann’s constant  $k_B$ , the temperature  $T = 310$  K, Planck’s constant  $h$ , the free energy barrier  $\Delta G^\ddagger$  of the free energy profiles, and the gas constant  $R$ .

#### 3.2.2 Classical molecular dynamics simulations

Classical atomistic molecular dynamics (MD) simulations were conducted of the yeast Hsp90 dimer in the closed state with bound ATP, built from the X-ray structure with PDB ID: 2CG9.<sup>41</sup> Missing amino acid regions in the crystal structure were added using MODELLER.<sup>158</sup> The final construct comprised residues 1-677 in both protomers and was solvated in a water box with 100 mM NaCl, yielding a model consisting of *ca.* 303 000 atoms. Full-length dimeric models of wild type Hsp90 and the R32A variant were propagated for 250 ns, and models with a mono-methylated Lys-594 as well as with the K594A/E/I/R substitutions for 200 ns, using the CHARMM27 force field.<sup>159</sup>

A monomeric model consisting of N-terminal and middle (M) domains and with bound ATP was constructed from the crystal structure with PDB ID: 4IVG from *Danio rerio*.<sup>154</sup> Models of the isolated N-terminal domain of yeast Hsp90 in the apo state and with bound ADP and ATP were built from the crystal structure with PDB ID: 1AMW.<sup>35</sup> These monomeric models were solvated in water boxes with 150 mM NaCl, leading to systems consisting of *ca.* 72 000 atoms (NM-construct) and *ca.* 77 000 atoms (NTD). The NM-construct was simulated for 250 ns using the CHARMM27 force field,<sup>159</sup> while the NTD models were simulated using the CHARMM36 force field for 1000 ns each.<sup>156</sup>

In the classical MD simulations, a 2 fs integration timestep was employed. Water molecules were kept rigid using the SETTLE algorithm, and all distances involving covalently bound hydrogen atoms were constrained using the SHAKE algorithm, eliminating bond motion for hydrogens.<sup>160,161</sup>

The Arg-32/Glu-33 distances presented are measured between their side chains, specifically between the C $\zeta$  atom of Arg-32 and the C $\delta$  atom of Glu-33.

### 3.2.3 Calculation of $pK_a$ values

$pK_a$  values were estimated from Poisson-Boltzmann continuum electrostatics calculations, with the average protonation of residues determined using a Monte Carlo method to sample  $2^N$  protonation states, where  $N = 514$  is the number of protonatable residues in the full-length Hsp90 model. Poisson-Boltzmann calculations were performed using the Adaptive Poisson-Boltzmann Solver (APBS), and protonation states were sampled using Karlsberg+.<sup>95,97,162</sup> The protein was modelled by partial charges located at the atomic nuclei, and by an inhomogeneous dielectric continuum using  $\epsilon = 4$ . Solvating water was modelled by employing a homogeneous dielectric continuum with a dielectric constant of  $\epsilon = 80$ . The surface between protein and water was determined by employing the molecular surface routine of APBS with a probe radius of 1.4 Å, modelling implicit KCl salt with a concentration of 100 mM.  $pK_a$  values were probed every ns of the 250 ns MD simulation of the dimeric wild type Hsp90 model.

## 3.3 The tyrosine kinase c-Src

Full-length c-Src models with and without phosphorylation at Tyr-416, and with and without bound ATP, were built from the crystal structure with PDB ID: 1Y57, containing residues 82-533.<sup>53</sup> Models without the regulatory domains comprised residues 251-533 and contained the linker region, the kinase domain, and the C-terminal tail. The models were embedded in a water box with 100 mM NaCl, leading to a system size of *ca.* 100 000 atoms for full-length constructs, and *ca.* 56 000 atoms for kinase domain (KD) constructs.

Classical atomistic molecular dynamics simulations of the different models were performed for 1 microsecond each, using an integration time step of 2 fs and the CHARMM27 force field.<sup>159</sup> The geometry of water molecules was kept rigid using the SETTLE algorithm, and all distances to covalently bound hydrogen atoms were constrained using the SHAKE algorithm.<sup>160,161</sup>

The extent of the activation loop was estimated as the average backbone distance between Asp-413-O and Thr-417-N, Asn-414-O and Ala-418-N, and Glu-415-O and Arg-419-N, as performed previously.<sup>163</sup>

## 3.4 The GTPase Rab32 in complex with the protease GtgE

Classical atomistic molecular dynamics (MD) simulations were conducted starting from the crystallized Rab32:GDP:GtgE<sub>C45A</sub>-complex, after substituting Ala-45<sub>GtgE</sub> by a cysteine to model the wild type protease. The complex was solvated in a water box containing 150 mM NaCl, yielding a simulation setup of *ca.* 99 000 atoms. The wild type Rab32:GtgE complex was simulated with bound GDP or GTP for 190 ns each, and the

### 3 Methods

complex with the F88<sub>GtgE</sub> variant was simulated with bound GDP or GTP for 30 ns each, employing the CHARMM36 force field and a 1 fs integration time step at  $T = 310$  K.<sup>156</sup> ProDy was used for principal component analysis (PCA) of the motion of C $\alpha$  atoms.<sup>164</sup>

## 3.5 The dioxygenase AsqJ

### 3.5.1 Quantum chemical calculations

QM cluster models consisting of 190–194 atoms were built from the crystal structure with PDB ID: 5DAQ, replacing the crystallized nickel by iron.<sup>66</sup> The QM cluster models comprised the iron-oxo species (or iron and water), the substrate,  $\alpha$ -ketoglutarate (or succinate and CO<sub>2</sub>), two water molecules, and protein residues Asn-70, Met-118, Gln-131, His-134, Asp-136, Asn-157, Leu-159, Ile-205, His-211, Arg-223, and Thr-227. Only the amino acid side chains starting from C $\beta$  atoms were included in the models, except for Asn-70, starting from C $\gamma$ , and Arg-223, starting from C $\delta$ . Also the backbone connection of Pro-132/Leu-133 and Thr-209/Ile-210 was added to the models. Terminal carbons as well as backbone carbon and nitrogen atoms were saturated with hydrogens and kept fixed to their crystal structure position. The QM cluster models were embedded in a continuum polarizable medium with  $\epsilon = 4$  using COSMO.<sup>126</sup> Structures were optimized using dispersion-corrected DFT with the B3LYP-D3 functional, and def2-SVP/def2-TZVP (Fe) basis sets.<sup>112,118</sup> Reaction pathways optimizations were performed by employing the Woelfling method in Turbomole.<sup>130</sup> Electronic energies of all intermediates in the quintet spin state were obtained using the B3LYP-D3 functional with def2-TZVP basis sets and a dielectric constant of  $\epsilon = 4$ , and their Mulliken spin populations and electron spin densities were analyzed with regard to electronic configurations.<sup>165</sup> Vibrational and entropic contributions were estimated from the numerical Hessian using the B3LYP-D3 functional with def2-SV(P) basis sets and a dielectric constant of  $\epsilon = 4$ .<sup>155</sup> Strain energies of the substrate were calculated from restrained optimization of the substrate optimized in vacuum towards the optimized geometry in the QM cluster models. NCIPLOT was employed to compute non-covalent interaction surfaces.<sup>166</sup>

### 3.5.2 Hybrid QM/MM calculations

Combined quantum and molecular mechanical (QM/MM) MD simulations of wild type AsqJ with bound methylated and non-methylated substrate, before and after the first PCET reaction, were conducted using the B3LYP-D3 functional with def2-SVP/def2-TZVP (Fe) basis sets and the CHARMM36 force field.<sup>156</sup> The QM region contained the side chains of His-134, Asp-136, and His-211, succinate, CO<sub>2</sub>, the substrate, and the Fe<sup>IV</sup>=O or Fe<sup>III</sup>-OH species. Link atoms were inserted between the C $\alpha$  and C $\beta$  carbons of the amino acids included in the QM region. The protein monomer was solvated in a water box with 100 mM NaCl, consisting of *ca.* 59 000 atoms in total. Each state was propagated for 3 ps at  $T = 310$  K, using an integration time step of 1 fs.



### 3.5.3 Classical molecular dynamics simulations

Classical atomistic molecular dynamics (MD) simulations were conducted of the wild type AsqJ with bound methylated and non-methylated substrate, and for the *in silico* V72I, V72K, and F139I mutants with bound non-methylated substrate, before the first PCET reaction. The models were built using the crystal structure with PDB ID: 5DAQ, and simulated for 15 ns using a 1 fs integration time step.<sup>66</sup> The protein was modelled using the CHARMM36 force field, combined with force field parameters from CGenFF and from numerical Hessians calculated with Turbomole, and with restrained electrostatic potential (RESP) charges calculated with NWChem using the B3LYP functional and def2-TZVP basis set.<sup>167,168</sup>

## 3.6 Biomimetic diiron catalyst

QM cluster models of the biomimetic diiron catalyst consisting of 125 atoms were built from the X-ray structure, and prepared with one ferryl ( $\text{Fe}^{\text{IV}}=\text{O}$ ) species with one cyclohexene molecule next to it, and one  $\text{Fe}^{\text{II}}$  with a bound acetonitrile ligand. Models consisting of 120 atoms were constructed with two ferryl species and one cyclohexene substrate. Structures of the intermediates of epoxide and allylic alcohol formation were optimized using dispersion-corrected DFT with the TPSSh-D3BJ functional, and def2-SVP/def2-TZVP (Fe) basis sets.<sup>114,169</sup> The models were embedded in a dielectric continuum with a dielectric constant of  $\epsilon = 38$  with COSMO.<sup>126</sup> Reaction pathway optimizations for the  $\text{Fe}^{\text{II}}\text{Fe}^{\text{IV}}=\text{O}$  model were performed by employing the Woelfling method in Turbomole.<sup>130</sup> Reaction pathways for the  $(\text{Fe}^{\text{IV}}=\text{O})_2$  model were approximated by restrained optimization of the optimized reaction pathway of the  $\text{Fe}^{\text{II}}\text{Fe}^{\text{IV}}=\text{O}$  model. Electronic energies of the optimized intermediates were calculated using the TPSSh-D3BJ functional with def2-TZVP basis sets and a dielectric constant of  $\epsilon = 38$ , and analysis of Mulliken populations and spin densities in Turbomole yielded information about electronic configurations.<sup>165</sup>



## 4 Results and discussion

The articles included in this publication-based thesis investigate the role of protein conformation with regard to their biological function in the molecular chaperone Hsp90, the tyrosine kinase c-Src, the GTPase Rab32, and the dioxygenase AsqJ, and show how single residue substitutions can have a pronounced effect on protein structure and function by changing their conformation. To learn more about the reactivity of iron as a cofactor in substrate oxidation as found in the dioxygenase AsqJ, we also studied a biomimetic iron catalyst.

The studies were performed in collaboration with different groups at the chemistry department of the Technical University of Munich, mostly within the collaborative research center SFB1035. The collaborating groups were led by Johannes Buchner (publications IV, V, VI), Michael Groll (publication I, III), Aymelt Itzen (publication I), Corinna Hess (publication II), Kathrin Lang (publication V), and Michael Sattler (publication VI) (*cf.* “List of publications”).

In this section, the central results of the included papers are briefly summarized.

### 4.1 The molecular chaperone Hsp90

We studied the conformational dynamics of the molecular chaperone Hsp90, and how conformational and electrostatic transitions are propagated within the Hsp90 dimer. In this regard, we identified residues Arg-32 in the N-terminal domain and Lys-594 in the C-terminal domain as switch points for the regulation of Hsp90 structure and function, and studied the effect of point mutations of the identified switch points, as well as of post-translational modification of Lys-594. To gain complementary insight into the conformational changes in Hsp90, we integrated molecular simulations with biochemical and biophysical experiments.

#### 4.1.1 The role of Arg-32

ATP hydrolysis in the Hsp90 chaperone is an important, yet not completely understood process, that we studied in collaboration with the Buchner and Sattler groups (Mader *et al.* Conformational dynamics modulate the catalytic activity of the molecular chaperone Hsp90. *Nat. Commun.* **11**, 1410:1-12 (2020); publication VI).<sup>44</sup>

We observed that in the crystal structures of the closed Hsp90 dimer from yeast, the catalytically important Glu-33 residue forms an ion pair with the neighbouring Arg-32 residue. To investigate the conformational dynamics of this ion pair that is possibly relevant for catalysis, we performed classical atomistic molecular dynamics (MD) simulations for 250 ns of full-length dimeric yeast Hsp90, as well as of a monomeric model of

#### 4 Results and discussion

Hsp90 from *Danio rerio* without the C-terminal domain. Furthermore, we studied the monomeric N-terminal domain (NTD) of yeast Hsp90 in the apo state and with bound ADP and ATP, respectively, in 1  $\mu$ s MD simulations.

In the MD simulation of full-length dimeric Hsp90, the Arg-32/Glu-33 ion pair was observed to open up, leading to contact formation between Arg-32 and residues of the middle domain, while the ion pair stays closed in the simulation of the NM-domain model (Figure 4.1 a). Also in the NTD model with bound ATP, the Arg-32/Glu-33 ion pair opens up, while it stays mostly closed in the apo form and with bound ADP. These findings imply that conformational changes in this ion pair are modulated by the ligand state, and that ion pair opening as observed in the full-length Hsp90 simulations can be triggered by ligand exchange.

The observed ion pair dynamics are supported by NMR experiments of the NTD in different ligand states, where a NOE (Nuclear Overhauser effect) cross peak between the Arg-32 H $\epsilon$  and the Glu-33 H $\beta$  atoms hints at a closed ion pair in the apo form.  $^1\text{H}$ - $^{15}\text{N}$  HSQC (heteronuclear single quantum coherence) spectra of the NTD show that the chemical environment of the Arg-32 side chain differs, when comparing the apo NTD and the domain with bound ADP and ATP, respectively (Figure 4.1 b). This indicates that Arg-32 adopts different conformations depending on the ligand state. Our MD simulations suggest that the different conformations of Arg-32 are connected to breaking of the ion pair with Glu-33 upon ATP binding.

To assess the influence of the conformation of the Arg-32 residue on the ATPase activity of Hsp90, we calculated the free energy profile for ATP hydrolysis in the wild type Hsp90, using a high-resolution X-ray structure from *Danio rerio*. In our models, ATP hydrolysis occurs by semi-concerted deprotonation of a water molecule by Glu-33 and nucleophilic attack of the hydroxide at the  $\gamma$ -phosphate of ATP, leading to phosphate cleavage and formation of ADP and inorganic phosphate. With the Arg-32/Glu-33 ion pair closed, the free energy barrier is larger than 30 kcal mol $^{-1}$ , and the profile does not show a clear product minimum for phosphate cleavage in the active site, indicating an unstable conformation.

With the Arg-32/Glu-33 ion pair open, the free energy barrier is reduced to *ca.* 25 kcal mol $^{-1}$ , indicating that the conformation of Arg-32 modulates the reaction barrier for ATP hydrolysis. Opening of the ion pair increases the p $K_a$  value of Glu-33, therefore facilitating deprotonation of the catalytic water molecule during ATP hydrolysis. To model a fully open conformation of the ion pair, we replaced Arg-32 by an alanine and recalculated the free energy profile, yielding a barrier of *ca.* 18 kcal mol $^{-1}$ , suggesting that the computationally identified R32A mutation lowers the reaction barrier for ATP hydrolysis in Hsp90 (Figure 4.1 c).

To probe the influence of the Arg-32 residue on the ATP hydrolysis reaction experimentally, we performed ATPase measurements of the Hsp90-R32A variant, the results of which show that the ADP release rate is only marginally increased (Figure 4.1 d). Since the conformational transition from the open to the closed state is the rate-limiting step in the Hsp90 cycle, a possibly faster intrinsic ATPase rate might not be detected using the employed ADP release assay.

#### 4.1 The molecular chaperone Hsp90

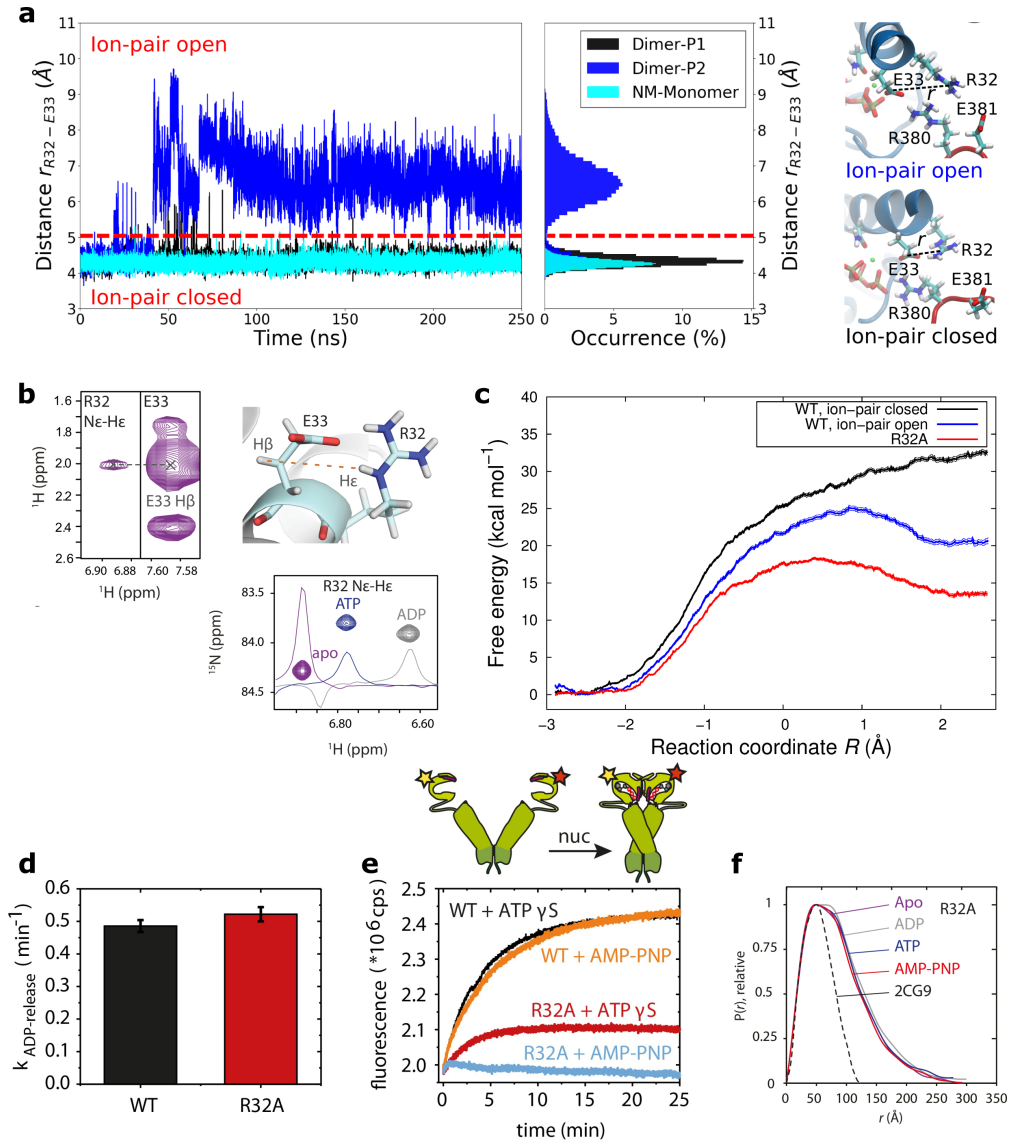


Figure 4.1: **a)** Arg-32/Glu-33 side chain distance in MD simulations of full-length dimeric Hsp90 and a monomeric NM-domain model. **b)** NOE cross-peak between Arg-32 and Glu-33 in the apo NTD (left panel), and HSQC signals of Arg-32 in the apo NTD and with bound ADP and ATP (bottom panel). **c)** Free energy profiles for ATP hydrolysis in Hsp90 with the Arg-32/Glu-33 ion pair closed and open, as well as for the R32A variant. **d)** Rate constants for global ADP release in wild type Hsp90 and the R32A variant. **e)** FRET signal for N-terminal dimerization of the wild type (WT) Hsp90 and the R32A variant, induced by the ATP analogues ATP $\gamma$ S and AMP-PNP. **f)** Pair distance distributions calculated from SAXS profiles of Hsp90-R32A in different ligand states, as well as of the X-ray structure of the compact WT Hsp90 dimer with PDB ID: 2CG9. The figure is adapted from ref. [44].

To investigate the effect of the R32A mutation on the global structure of Hsp90, we performed FRET and SAXS experiments of Hsp90-R32A in different ligand states. Both techniques indicate that the R32A variant does not adopt the fully closed conformation to the same extent as the wild type Hsp90 (Figure 4.1 e, f). The R32A mutant seems to form only a partially N-terminally closed dimer with the slowly hydrolyzable ATP-analogue ATP $\gamma$ S, and no closed conformation could be detected with the non-hydrolyzable ATP-analogue AMP-PNP (ATP[ $\beta,\gamma$ -NH]).

Since the formation of the closed conformation is crucial for ATP hydrolysis to occur in Hsp90, the decreased reaction barrier and therefore faster ATP hydrolysis reaction in Hsp90-R32A could be compensated by the lower population of the closed state, leading to an overall unchanged ADP release rate. The unchanged ATPase activity of Hsp90-R32A at a lower population of the closed state suggests that catalysis could be decoupled from conformation in the R32A variant.

Taken together, this study shows that the reaction barrier for a biochemical reaction can be modulated by the conformation of a nearby residue. In this specific case, conformational changes of an arginine residue decrease the reaction barrier by electrostatic tuning of the  $pK_a$  value of a catalytic glutamate. Removal of the positive charge from the system and substitution by an alanine further lowers the catalytic barrier. Due to the complex chaperone cycle of Hsp90, an acceleration of the ATP hydrolysis reaction could not be observed directly in the employed ADP release assay showing an unchanged global ATPase activity of the R32A variant. However, FRET and SAXS experiments indicate that the R32A mutation changes the global conformation of the Hsp90 dimer and leads to a lower population of the N-terminally dimerized Hsp90. Since this is the catalytically active conformation in the wild type Hsp90, this shift in the conformational ensemble could compensate for a faster intrinsic ATP hydrolysis rate.

### 4.1.2 The role of Lys-594

To understand how post-translational modifications may affect the conformational dynamics of Hsp90, we probed the effect of mono-methylation of Lys-594 (K594metK) as well as of the K594A/E/I/R mutations in collaboration with the Buchner and Lang groups (Rehn *et al.* A methylated lysine is a switch point for conformational communication in the chaperone Hsp90. *Nat. Commun.* **11**, 1219:1-14 (2020); publication V).<sup>170</sup>

Substituting Lys-594 by neutral (A, I) and negatively charged (E) residues results in a decrease in binding to the co-chaperone Aha1 and the inhibitor Sba1 as well as in ATP turnover, while the K594R variant behaves similar to the wild type protein. The effect on ATP turnover is unexpected since the mutations are carried out in the C-terminal domain, *ca.* 70 Å away from the ATPase site in the N-terminal domain (NTD). Methylation of Lys-594 results in significantly slower N-terminal dimerization upon addition of the nucleotide ATP $\gamma$ S, which is observed similarly in the K594I variant.

To probe how post-translational modification and mutations of Lys-594 influence the dynamics of Hsp90, we performed classical atomistic molecular dynamics (MD) simulations of the full-length yeast Hsp90 dimer without modifications and with mono-

#### 4.1 The molecular chaperone Hsp90

methylation at Lys-594 (K594metK), as well as for the *in silico* K594A/E/I/R variants.

In the MD simulations of the wild type Hsp90, Lys-594 strongly interacts with Glu-590 and Asp-659, therefore inhibiting a salt bridge between Glu-590 and Arg-591, whereas the methylation increases the distance between Lys-594 and Glu-590 and Asp-659, enhancing the interaction between Glu-590 and Arg-591 in the K594metK variant (Figure 4.2 a, c). Our simulations suggest that methylation of Lys-594 affects electrostatic interactions between nearby charged residues and alters their conformation, which changes the network of salt bridges and may propagate further to the middle and N-terminal domains.

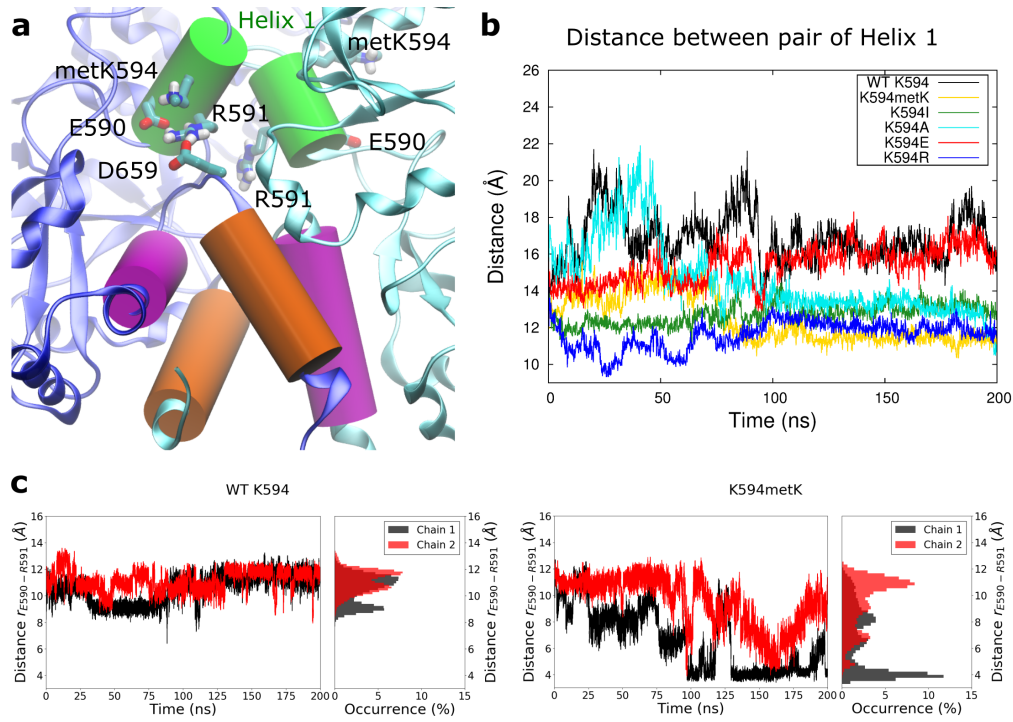


Figure 4.2: **a)** Close-up of the C-terminal dimerization interface in Hsp90 after 200 ns of MD simulation with a methylated Lys-594 (metK594), with Helix 1 shown in green and two other  $\alpha$ -helices at the dimer interface in purple and orange. **b)** Distance between the centers of mass of the dimeric pair of Helix 1 during 200 ns of MD simulation of wild type Hsp90 and with different substitutions of Lys-594. **c)** Side chain distance between Glu-590 and Arg-591 in both protomers (chains 1 and 2) of the wild type (WT) Hsp90 dimer and of the K594metK variant. The figure is adapted from ref. [170].

With a methylated Lys-594, the distance between the centers of mass of a pair of  $\alpha$ -helices involved in dimeric interactions in the C-terminal domain<sup>41</sup> (here referred to as Helix 1, comprised by residues 587-595), decreases in comparison to the wild type (WT) Hsp90. The distance between the centers of mass of the dimeric pair of Helix 1 behaves

similar in the methyllysine (K594metK) and K594I models, supporting the notion that K594I could mimic the effect of Lys-594 methylation (Figure 4.2 b).

The dynamics of the Glu-590/Arg-591 ion pair in the MD simulations of Hsp90-K594I show a few closing events, approaching the closed form observed in the simulations with a methylated Lys-594, while the ion pair is not formed in the wild type Hsp90 simulation (Figure 4.2 c). Together with the experimentally observed similarities between the K594I and K594metK variants in N-terminal dimerization, this suggests that the methyllysine residue can be mimicked by an isoleucine. While the methyllysine side chain still possesses a positively charged ammonium group, the methylation donates electron density to the nitrogen, spreading out the local positive charge. Despite the increased size of the methyllysine side chain, the isoleucine residue seems to have similar chemical properties.

Overall, this study shows that certain residues in large proteins like Hsp90 have a crucial role for the overall structure and function. At the position of Lys-594, the positive charge seems to be important for ligand binding, N-terminal dimerization, and ATPase activity, since the K594R variant behaves similar to the wild type Hsp90. The effects of substitution and post-translational modification of Lys-594 upon the ATPase activity are especially astonishing, as the residue is located in the C-terminal domain. Methylation of Lys-594 leads to rearrangement of local neighbouring ion pairs, and also changes the conformational ensemble of the Hsp90 dimer by impairing N-terminal dimerization, as observed for the R32A variant. Surprisingly, the effects of methylation and substitution by an isoleucine are similar, indicating similar properties of methyllysine and isoleucine residues.

## 4.2 The tyrosine kinase c-Src

To probe the influence of phosphorylation of Tyr-416 in c-Src, we compared the autophosphorylated enzyme to non-phosphorylated c-Src in collaboration with the Buchner group (Boczek *et al.* Autophosphorylation activates c-Src kinase through global structural rearrangements. *J. Biol. Chem.* **294**, 13186-13197 (2019); publication IV).<sup>54</sup>

Kinase activity measurements show that Tyr-416 phosphorylation accelerates protein substrate phosphorylation by c-Src, while a reduced  $K_M$  value for ATP indicates a higher affinity for ATP and leads to an increase in catalytic efficiency ( $\frac{k_{cat}}{K_M}$ ). Biophysical analysis and hydrogen/deuterium exchange mass spectrometry (H/DX-MS) indicate that the kinase domain adopts a more rigid conformation upon phosphorylation, while the regulatory SH2 and SH3 domains seem to become more dynamic. Without phosphorylation at Tyr-416, the kinase domain seems to interact with the regulatory SH2 and SH3 domains.

To gain understanding of the molecular effects of Tyr-416 phosphorylation in c-Src, we performed classical atomistic MD simulations of unphosphorylated (FLU) as well as phosphorylated (FLP) full-length c-Src, and of corresponding models without the regulatory SH2 and SH3 domains (KDU and KDP constructs).

In the FLP and KDP states, the negatively charged phosphorylated Tyr-416 (pTyr-416) in the activation loop (A-loop) interacts with three arginine residues in the C-lobe



of the kinase domain, Arg-385, Arg-409, and Arg-419. Arg-419 of the A-loop also forms contacts with Glu-305 at the N-terminus of the  $\alpha$ C helix in the N-lobe of the kinase domain, thereby bridging between pTyr-416 and Glu-305 and connecting the N-terminus of the  $\alpha$ C helix to pTyr-416 in the A-loop (Figure 4.3 a). This could form the molecular basis for the observed increased rigidity of the kinase domain in phosphorylated c-Src.

To analyze the conformation of the A-loop, we calculated the average backbone distance between the six neighbouring residues of Tyr-416 for all simulated constructs. In FLP and KDP, this average distance is *ca.* 2 Å larger than in FLU and KDU and stays constant over the simulation time, indicating that the phosphoryl group stabilizes the neighbouring residues of the A-loop in an extended conformation. In the FLU and KDU simulations, this average distance fluctuates considerably, showing that the conformation of the A-loop is more flexible without phosphorylation at Tyr-416 (Figure 4.3 b).

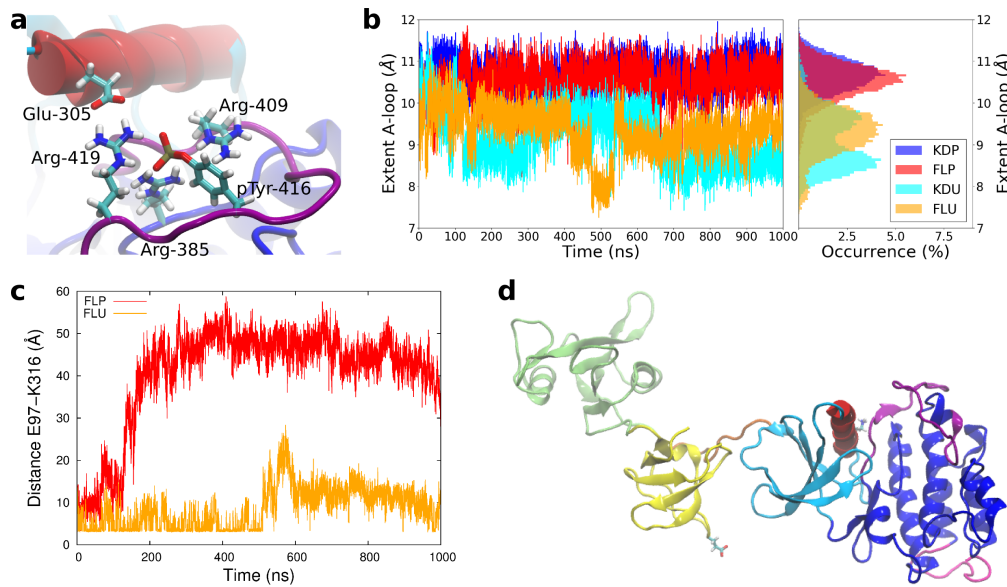


Figure 4.3: **a)** The phosphorylated Tyr-416 (pTyr-416) in the activation loop (in purple) interacts with three arginine residues located in the activation loop and in the C-lobe of the kinase domain (in dark blue). Arg-419 also contacts Glu-305 of the  $\alpha$ C helix (in red) in the N-lobe of the kinase domain. **b)** Extent of the activation loop (A-loop), calculated as the average backbone distance between the six neighbouring residues of Tyr-416. **c)** Distance between Glu-97 of the SH3 domain and Lys-316 of the  $\alpha$ C helix in the kinase domain observed in MD simulations of the FLP (in red) and FLU (in orange) constructs. **d)** Snapshot of the FLP MD simulation, with Glu-97 and Lys-316 shown in licorice representation. The figure is adapted from ref. [54].

The MD simulations of FLP show a high flexibility of the regulatory SH2 and SH3 domains, leading to the SH3 domain detaching from the kinase domain. Specifically, Glu-97 of the SH3 domain loses its contact with Lys-316 of the  $\alpha$ C helix in the N-lobe

of the kinase domain, while this interaction persists during the MD simulation of FLU (Figure 4.3 c, d). This is consistent with the cross-talk between the kinase domain and the regulatory SH2 and SH3 domains in FLU identified in the H/DX-MS experiments.

Generally, this study shows that a structural modification like phosphorylation can have significant effects on the local and global protein structure and activity. Especially the introduction of a charged group, like a phosphoryl group, rearranges the network of charged residues, as also observed after methylation of Lys-594 in Hsp90. These local rearrangements propagate through the protein and lead to loss of inter-domain interactions and therefore to a change of the global conformation, as was also the case for the R32A and K594metK variants of Hsp90.

### 4.3 The GTPase Rab32 in complex with the protease GtgE

To understand molecular principles of ligand-triggered conformational switching, we studied the small GTPase Rab32 with bound GDP or GTP in complex with the cysteine protease GtgE from *Salmonella Typhimurium*, in collaboration with the Itzen and Groll groups (Wachtel *et al.* The protease GtgE from *Salmonella* exclusively targets inactive Rab GTPases. *Nat. Commun.* **9**, 44:1-13 (2018); publication I).<sup>171</sup>

Investigation of Rab32-cleavage by GtgE in biochemical experiments indicates that only the inactive, GDP-bound Rab32, but not the active GTP-bound form, can be proteolytically cleaved by GtgE. To characterize the molecular interactions between the Rab32 and GtgE proteins, the atomic structure of the Rab32:GDP:GtgE<sub>C45A</sub> complex was resolved at a resolution of 2.3 Å using X-ray crystallography (Figure 4.4 a). For the structure determination of the complex, the catalytically inactive GtgE<sub>C45A</sub> variant showing a high affinity for Rab32:GDP with a  $K_D$  of *ca.* 100 nM was used.

To gain insight into the nucleotide-state specificity of GtgE towards the inactive Rab32:GDP, we performed classical atomistic molecular dynamics (MD) simulations based on the resolved crystal structure of the Rab32:GDP:GtgE complex. Additionally to an MD simulation of the GDP-bound complex, we set up a second MD simulation by *in silico* exchanging GDP by GTP in the nucleotide binding site of Rab32. This MD simulation of the GTP-bound complex aims at resolving the first steps of GtgE dissociation from the active Rab32:GTP.

During *ca.* 200 ns of MD simulation, the GTP triphosphate group pushes the side chain of Glu-86 towards Arg-87, and the distance between the C $\alpha$  atoms of Phe-88<sub>Rab32</sub> (Phe-88<sub>R</sub>) and Lys-194<sub>GtgE</sub> (Lys-194<sub>G</sub>) is increased in the GTP-bound complex, indicating that Phe-88<sub>R</sub> could inhibit complex formation with bound GTP (Figure 4.4 b, c).

To probe the selectivity of GtgE towards GDP-bound Rab32, we purified the F88G<sub>Rab32</sub> variant and analyzed its proteolysis by GtgE. Although the catalytic activity of GtgE towards the F88G<sub>R</sub> variant decreases compared to the wild type Rab32, GtgE cleaves the F88G<sub>R</sub> variant both bound to GDP as well as to GTP and thereby loses its specificity for GDP-bound Rab32 (Figure 4.4 d). Phe-88<sub>R</sub> seems to contribute to the discrimination between GDP-bound, inactive and GTP-bound, active Rab32 by GtgE.

### 4.3 The GTPase Rab32 in complex with the protease GtgE

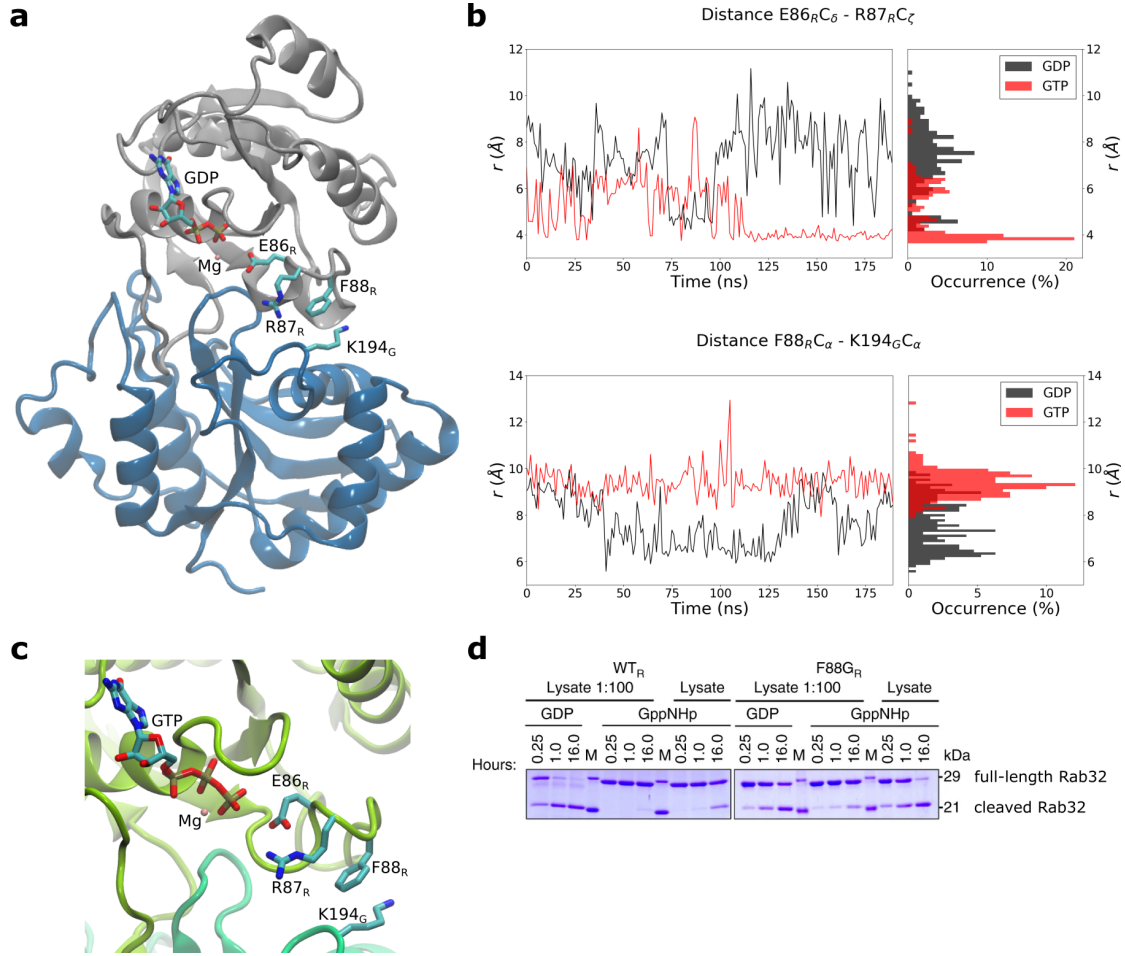


Figure 4.4: **a)** Crystal structure of the Rab32:GDP:GtgE complex with Rab32 shown in gray and GtgE shown in blue. **b)** Side chain distance between Glu-86<sub>R</sub> and Arg-87<sub>R</sub> of Rab32 (top), and backbone distance between Phe-88<sub>R</sub> and Lys-194<sub>G</sub> (bottom) in MD simulations with bound GDP (in black) and GTP (in red). **c)** Close-up of the nucleotide binding site in Rab32 after *ca.* 200 ns of MD simulation of the Rab32-GtgE complex with bound GTP. Rab32 is shown in light green and GtgE in dark green. **d)** Proteolysis of wild type Rab32 (WT<sub>R</sub>) and the F88G<sub>R</sub> variant with bound GDP or the non-hydrolyzable GTP analogue GppNHp by GtgE contained in cleared cell lysate. The figure is adapted from ref. [171].

Altogether, this study shows that different ligands can stabilize different conformations of a protein, as also observed for the isolated N-terminal domain of Hsp90. In the GTPase Rab32, exchange of GDP by GTP triggers formation of a nearby ion pair and reshapes an interaction site with the protease GtgE, thereby inhibiting complex formation and proteolytic cleavage. We identified Phe-88<sub>R</sub> at the complex interface as a discriminatory

factor for GtgE between GDP- and GTP-bound Rab32. The nucleotide-state specificity of GtgE towards Rab32 implicates that Rab32 has to be inactivated during *Salmonella* infection for quantitative cleavage by GtgE.

## 4.4 The dioxygenase AsqJ

To understand how conformational dynamics may determine catalytic activity in enzymes, we studied the dioxygenase AsqJ in collaboration with the Groll group (Mader *et al.* Catalytic mechanism and molecular engineering of quinolone biosynthesis in dioxygenase AsqJ. *Nat. Commun.* **9**, 1168:1-8 (2018); publication III).<sup>67</sup>

As in other  $\text{Fe}^{\text{II}}$ / $\alpha$ -ketoglutarate-dependent dioxygenases,<sup>68</sup> the catalytic cycle of AsqJ starts by dioxygen binding, completing the octahedral coordination sphere of  $\text{Fe}^{\text{II}}$ , and leading to formation of an  $\text{Fe}^{\text{III}}$ -superoxo species. The iron-bound dioxygen molecule is split, leading to formation of a ferryl species with one of the oxygen atoms. The other oxygen atom is inserted into the  $\alpha$ -ketoglutarate ( $\alpha$ KG) cofactor that is subsequently decarboxylated, forming succinate. Both  $\alpha$ KG and succinate are stabilized in the active site by a salt bridge with Arg-223.

After the decarboxylation of  $\alpha$ KG, the ferryl group can flip towards the former coordination site of the carboxyl group of  $\alpha$ KG and is now rotated towards the substrate, from where it abstracts two hydrogen atoms in two proton-coupled electron transfer (PCET) reactions. After release of the formed water and succinate molecules, binding of new dioxygen and  $\alpha$ KG species to AsqJ initiates the second reaction sequence, at the end of which the ferryl transfers its oxygen onto the desaturated intermediate, yielding an epoxide.

We optimized the intermediates of the two reaction sequences with both the methylated and non-methylated substrates using DFT, and calculated the corresponding DFT free energy profiles in the quintet spin state. Interestingly, the obtained energy barriers and reaction free energies are similar for both substrates. To investigate why methylation of the substrate is required for catalytic turnover by AsqJ, we performed geometry optimizations of both the methylated and non-methylated substrate with and without the protein surroundings, respectively. To compare the substrate geometries, we measured the dihedral angle  $\theta$  of the optimized substrates before the first PCET step (Figure 4.5).

While the methylated substrate adopts a similar conformation in the active site of AsqJ as without the protein surroundings (dihedral angle of  $-27^\circ$  vs.  $-22^\circ$ ), the non-methylated surrogate is strained by the protein to a dihedral angle of  $-52^\circ$  from its native conformation with a dihedral angle of  $-72^\circ$  (Figure 4.5). Consequentially, the methylated substrate can adopt almost its native conformation in the binding pocket of AsqJ, whereas the absence of the N-methyl group changes the equilibrium geometry of the substrate by  $50^\circ$ . Because of the altered equilibrium geometry, the non-methylated surrogate has to adopt a strained conformation in the enzyme active site, where the measured dihedral angle differs by  $20^\circ$  different from the equilibrium geometry of the non-methylated surrogate.

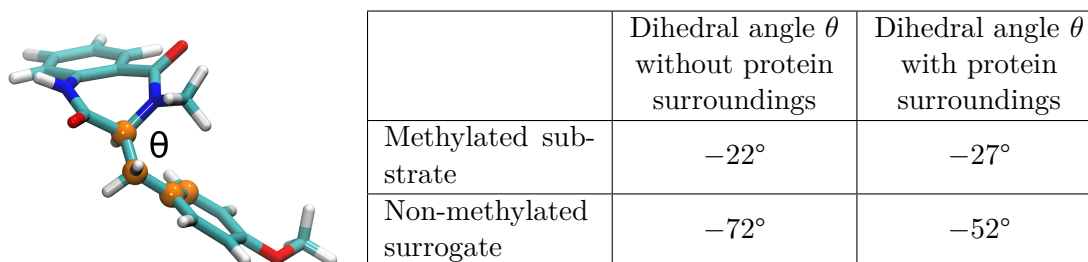


Figure 4.5: Dihedral angle  $\theta$  formed by the atoms marked in orange, measured in the methylated and non-methylated substrates optimized with and without the enzyme surroundings before the first PCET reaction.

The difference in equilibrium geometries of the methylated and non-methylated substrates affects their stabilization in the active site of AsqJ. While the methylated substrate forms a  $\pi$ - $\pi$  stacking interaction with His-134 (Figure 4.6 a), the altered geometry of the non-methylated surrogate does not allow for this stabilizing dispersive interaction to be formed to the same extent in our DFT models (Figure 4.6 c). Also in QM/MM MD simulations of AsqJ after the first PCET reaction, a larger distance between the non-methylated surrogate and His-134 is observed as compared to the methylated substrate (Figure 4.6 b). These findings suggest that the non-methylated surrogate could dissociate from the active site during the PCET reactions.

Analysis of the QM/MM simulations of AsqJ with explicitly modelled protein surroundings reveals that the side chain of Val-72 reaches behind the substrate/His-134 interaction axis. This position therefore offers the possibility to push the non-methylated surrogate closer to the His-134 side chain, enhancing their  $\pi$ - $\pi$  interaction and stabilizing the non-methylated surrogate in the active site. *In silico* replacement of Val-72 by an isoleucine residue indeed re-establishes the dispersive interactions between the non-methylated surrogate and His-134 in both QM cluster models and classical atomistic MD simulations (Figure 4.6 c, d).

To test the catalytic activity of the computationally designed V72I variant towards non-methylated substrate surrogates, we performed reverse-phase high-performance liquid chromatography-mass spectrometry (HPLC-MS) activity assays. The experiments demonstrate that the desaturated reaction intermediate **2d** obtained from the non-methylated surrogate analogue **1d** is already formed after 20 s for the V72I mutant. These results are in stark contrast to the wild type AsqJ, where we observe formation of **2d** with maximum intensity only after 30 min (Figure 4.6 e). Formation of the epoxide **3d**, however, was not observed for the mutant V72I, indicating that the mutated enzyme cannot successfully proceed into the epoxidation reaction cycle after the desaturation steps. High-resolution crystal structures of the V72I variant with different substrates and with bound  $\alpha$ -ketoglutarate show a high similarity to the crystal structure of the wild type AsqJ (Figure 4.6 f).

## 4 Results and discussion

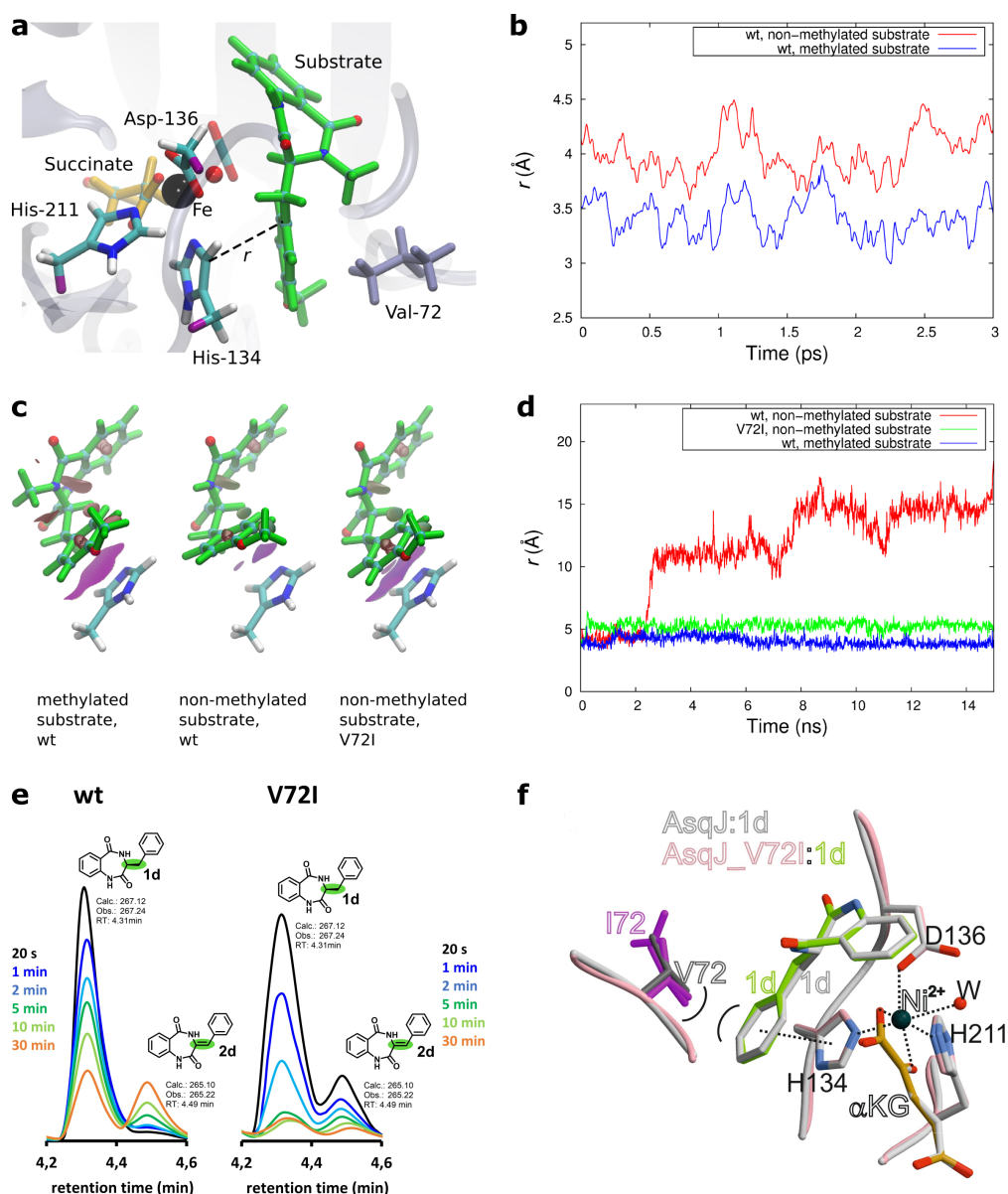


Figure 4.6: **a)** AsqJ active site from QM/MM simulations before the first PCET step, indicating the distance  $r$  between His-134 and the substrate. **b)** Distance  $r$  in QM/MM MD simulations of AsqJ with methylated and non-methylated substrates after the first PCET reaction. **c)** Non-covalent interaction surfaces (in purple) between His-134 and methylated and non-methylated substrates in QM cluster models of wild type (wt) AsqJ and the V72I variant before the first PCET step. **d)** Distance  $r$  in classical MD simulations of wt AsqJ and the V72I mutant with methylated and non-methylated substrates before the first PCET reaction. **e)** Desaturation of the non-methylated substrate analogue **1d** to **2d** by wt AsqJ and the V72I variant in HPLC-MS-coupled activity assays. **f)** Overlaid crystal structures of wt AsqJ and the mutant AsqJ\_V72I with bound **1d**. The figure is adapted from ref. [67].

This study shows that a small modification of an enzyme active site can drastically increase the catalytic activity towards certain substrates. While the binding site of the wild type AsqJ is optimized for its natural, methylated substrate, the V72I mutation helps to stabilize the non-methylated substrate in the active site by enhancing the  $\pi$ - $\pi$  interaction with His-134. After identification of the  $\pi$ - $\pi$  interaction as a crucial factor for the stabilization of the natural, methylated substrate, we engineered AsqJ for turnover of non-methylated surrogates by stabilizing the surrogate conformation that enables its dispersive interaction with His-134.

## 4.5 Biomimetic diiron catalyst

The oxidative chemistry of non-heme,  $\text{Fe}^{\text{II}}$ -dependent dioxygenases like AsqJ involves conformational changes in their active site that are important for the catalysis to occur. To understand a similar Fe-mediated catalytic process, we studied a biomimetic diiron catalyst developed by the Hess group (Lindsay *et al.* C-H Oxidation by a Diiron Complex with Facially Opposing Active Sites. *ChemistrySelect* **3**, 1602-1608 (2018); publication II).<sup>172</sup>

This catalyst is a binuclear iron complex with two identical pentadentate pyridine-based ligands and two acetonitrile ligands (Figure 4.7 a), with both  $\text{Fe}^{\text{II}}$  centers in the high-spin state (quintet spin,  $S = 2$ ) and an internuclear distance of 7.4 Å. Reactivity studies were performed to assess cyclohexene oxidation by the biomimetic diiron catalyst, using iodosylbenzene (PhIO) as the oxidant that can create a ferryl ( $\text{Fe}^{\text{IV}}=\text{O}$ ) species by O-atom transfer to ferrous iron ( $\text{Fe}^{\text{II}}$ ). The experiments indicate that allylic C-H bond oxidation of cyclohexene by the biomimetic diiron catalyst is kinetically favoured over epoxidation of the C=C double bond.

After crystallization of the biomimetic diiron catalyst, the spin state of the supposedly catalytically active ferryl species was examined in models of the catalyst with one ferryl ( $\text{Fe}^{\text{II}}\text{Fe}^{\text{IV}}=\text{O}$ ) and with two ferryl ( $(\text{Fe}^{\text{IV}}=\text{O})_2$ ) species, respectively, by employing density functional theory (DFT). In both models, the triplet state of the ferryl species is *ca.* 8 kcal mol<sup>-1</sup> more stable than the quintet state, as observed for other non-heme ferryl complexes.<sup>173</sup> The total energy of the diiron catalyst does not significantly vary upon inverting the orientation of the unpaired electron spins of one iron species, suggesting that the electron spins of the two iron centers do not communicate with each other (Figure 4.7 b).

We calculated the energetics for the formation of the epoxide and allylic alcohol products from cyclohexene by the binuclear catalyst in the  $\text{Fe}^{\text{II}}\text{Fe}^{\text{IV}}=\text{O}$  configuration (Figure 4.7 c). Oxygen atom transfer from the ferryl in the triplet state to the cyclohexene substrate with its double bond oriented towards the iron, results in epoxide formation with a barrier of *ca.* 11 kcal mol<sup>-1</sup>. Concerted oxygen atom insertion into the allylic C-H bond of cyclohexene pointing towards the ferryl yields the allylic alcohol, cyclohexenol, with a barrier of *ca.* 7 kcal mol<sup>-1</sup>. The lowest barrier is therefore obtained for allylic C-H hydroxylation leading to cyclohexenol formation, consistent with the experimental data.

## 4 Results and discussion

We obtain a barrier of *ca.* 16 kcal mol<sup>-1</sup> for allylic hydrogen abstraction from cyclohexene with the allylic methylene group pointing towards the iron, leading to formation of Fe<sup>III</sup>-OH and a cyclohexene radical. Subsequent backtransfer of the hydroxyl group proceeds without barrier. Thus, for allylic alcohol formation, concerted oxygen insertion seems to be preferred over hydrogen abstraction, as the reaction barrier for hydrogen abstraction from cyclohexene is significantly higher.

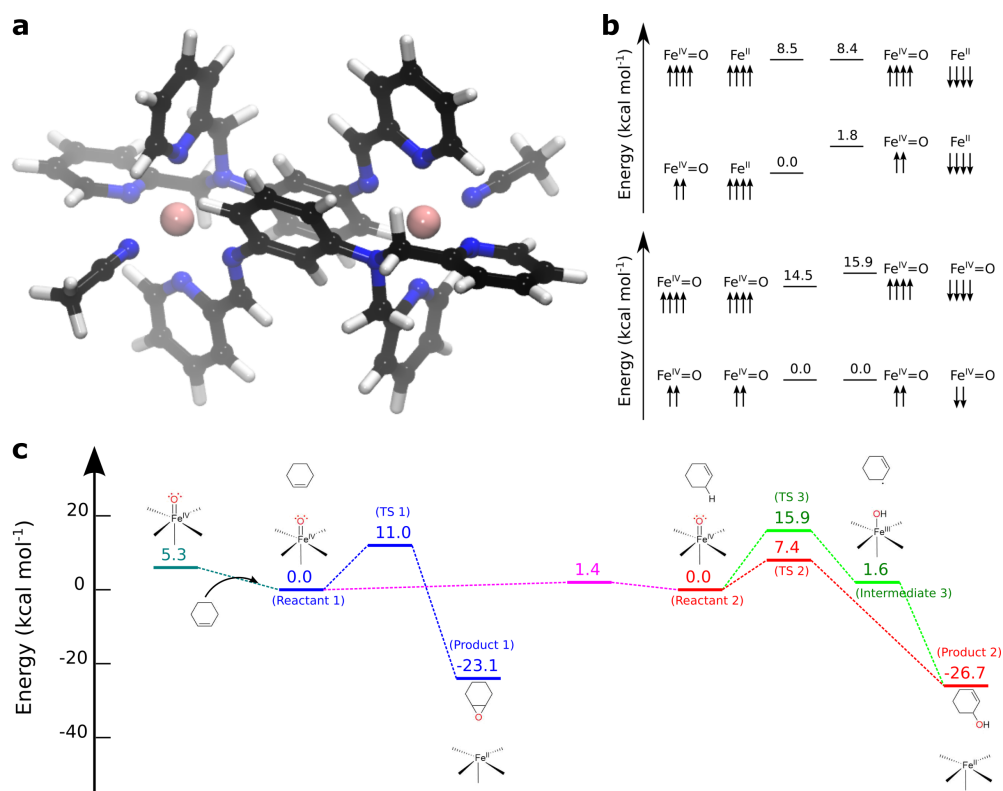


Figure 4.7: **a)** Structure of the biomimetic diiron catalyst with two identical pentadentate pyridine-based ligands and two acetonitrile ligands. **b)** Electronic energies of the binuclear complex with one Fe<sup>IV</sup>=O (top) and two Fe<sup>IV</sup>=O species (bottom) with triplet and quintet spin, respectively. **c)** Energy diagram of cyclohexene oxidation by the biomimetic diiron catalyst. Cyclohexene binding is colored in cyan, epoxidation in blue, transition between two cyclohexene binding conformations in pink, concerted allylic oxygen atom insertion in red, and stepwise allylic hydroxylation in green. The figure is adapted from ref. [172].

There are a few differences between the diiron catalyst and the dioxygenase AsqJ. While the dioxygenase AsqJ employs a high-spin ferryl in the quintet spin state ( $S = 2$ ) to catalyze its reaction sequences, the electronic ground state of the ferryl species in the biomimetic diiron catalyst was found to be the triplet state ( $S = 1$ ). Furthermore, AsqJ



activates dioxygen using  $\alpha$ KG as a cofactor. In the experimental reactivity studies of the biomimetic diiron catalyst, the ferryl species was created by using iodosylbenzene as the oxidant, which means that the diiron compound acts as an oxo-transfer agent in the oxidation reactions and does not activate dioxygen itself.

Compared to pyridine-based mononuclear non-heme complexes, the binuclear catalyst displays a lower catalytic activity. While it successfully activates cyclohexene, the diiron catalyst produces a mixture of epoxide and allylic oxidation products and is unable to activate the stronger C-H bonds of cyclohexane. The substrate of AsqJ is also not cyclohexane, but AsqJ selectively catalyzes subsequent desaturation and epoxidation of the substrate.

Presumably, one of the main differences between the biomimetic diiron catalyst and the dioxygenase AsqJ is the fact that the active site of AsqJ seems to be optimized for its natural substrate that is stabilized there by a  $\pi$ - $\pi$  stacking interaction. Also, the ferryl is positioned precisely towards the substrate after release of carbon dioxide from  $\alpha$ KG. The biomimetic diiron catalyst, on the other hand, offers a few sites to bind cyclohexene by dispersive interactions, but the specific conformation of the substrate cannot be controlled. This could be the reason for the rather unselective and low reactivity of the diiron catalyst.



## 5 Conclusions

In this thesis, we investigated several biological systems using multi-scale molecular simulations, combined with experimental techniques. We studied conformational changes, and how residue substitutions, post-translational modifications, and ligand binding perturb the conformational dynamics of proteins and affect their catalytic activity.

Both Arg-32 in the ATPase site as well as Lys-594 in the C-terminal domain were found to be important switch points in the chaperone Hsp90. Arg-32 modulates electrostatically the catalytic barrier for ATP hydrolysis by conformational switching between the N- and M-domains. In the R32A variant, the global conformation of the Hsp90 dimer is shifted towards a more open state, while the ADP release rate remains unchanged. Modifications of Lys-594 reduce the ATPase activity and, like the R32A substitution, impair formation of the N-terminally closed Hsp90 dimer. MD simulations indicate that methylation of Lys-594 rearranges neighbouring ion pairs, and suggest that the methyllysine residue can be mimicked by an isoleucine. Overall, both Arg-32 and Lys-594 play an important role in maintaining the catalytically active conformation of the Hsp90 dimer.

Phosphorylation of Tyr-416 in the kinase c-Src is crucial to fully activate the enzyme. In MD simulations, the phosphorylation event leads to conformational rearrangements of nearby charged residues and to the loss of inter-domain interactions. Similar to the R32A substitution and modifications of Lys-594 in Hsp90, phosphorylation of Tyr-416 changes the local charged network and stabilizes the protein in an open conformation.

In the Rab32:GtgE complex, we identified Phe-88<sub>R</sub> as a residue that contributes to the discrimination between GDP- and GTP-bound Rab32 by the protease GtgE. The F88G<sub>R</sub> mutation results in a loss of the GtgE preference for GDP-bound Rab32. GTP-binding to the wild type Rab32 triggers local conformational changes that result in an inhibitory conformation of Phe-88, while the Gly-88 residue does not prevent GtgE binding.

We were able to substantially accelerate the desaturation reaction of a non-natural substrate surrogate of the non-heme, Fe<sup>II</sup>-dependent dioxygenase AsqJ, by enhancing a dispersive  $\pi$ - $\pi$  stacking interaction in the enzyme active site. The V72I substitution stabilizes the reactive conformation of the surrogate in the catalytic pocket and thereby increases its catalytic turnover.

The reactivity of non-heme iron centers was also studied in a biomimetic diiron complex that catalyzes the oxidation of cyclohexene via allylic oxidation or epoxidation, with the first pathway preferred over the latter, but is not able to oxidize cyclohexane. Its rather low reactivity and selectivity could be related to the missing ability to stabilize the substrate in a reactive conformation, as achieved in the dioxygenase AsqJ.

Taken together, by integrating molecular simulations with experimental methods, we showed that local modifications in proteins can lead to significant changes in the global protein conformation and may control their enzymatic activity.



# Bibliography

1. Laskowski, R. A., Gerick, F. & Thornton, J. M. The structural basis of allosteric regulation in proteins. *FEBS Lett.* **583**, 1692–1698 (2009).
2. Grant, B. J., Gorfe, A. A. & McCammon, J. A. Large conformational changes in proteins: signaling and other functions. *Curr. Opin. Struct. Biol.* **20**, 142–147 (2010).
3. Ha, J.-H. & Loh, S. N. Protein Conformational Switches: From Nature to Design. *Chem. Eur. J.* **18**, 7984–7999 (2012).
4. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
5. Skjærven, L., Reuter, N. & Martinez, A. Dynamics, flexibility and ligand-induced conformational changes in biological macromolecules: a computational approach. *Future Med. Chem.* **3**, 2079–2100 (2011).
6. Seo, M. H., Park, J., Kim, E., Hohng, S. & Kim, H. S. Protein conformational dynamics dictate the binding affinity for a ligand. *Nat. Commun.* **5**, 3724 (2014).
7. Henzler-Wildman, K. A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T., Pozharski, E., Wilson, M. A., Petsko, G. A., Karplus, M., Hübner, C. G. & Kern, D. Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838–844 (2007).
8. Mannige, R. V. Dynamic New World: Refining Our View of Protein Structure, Function and Evolution. *Proteomes* **2**, 128–153 (2014).
9. Hammes, G. G., Chang, Y.-C. & Oas, T. G. Conformational selection or induced fit: A flux description of reaction mechanism. *Proc. Natl. Acad. Sci. USA* **106**, 13737–13741 (2009).
10. Paul, F. & Weikl, T. R. How to Distinguish Conformational Selection and Induced Fit Based on Chemical Relaxation Rates. *PLOS Comput. Biol.* **12**, 1–17 (2016).
11. Duan, G. & Walther, D. The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. *PLOS Comput. Biol.* **11**, 1–23 (2015).
12. Illergård, K., Ardell, D. H. & Elofsson, A. Structure is three to ten times more conserved than sequence—A study of structural response in protein cores. *Proteins: Struct., Funct., Bioinf.* **77**, 499–508 (2009).
13. Catazaro, J., Caprez, A., Swanson, D. & Powers, R. Functional Evolution of Proteins. *Proteins: Struct., Funct., Bioinf.* **87**, 492–501 (2019).
14. Radzicka, A. & Wolfenden, R. A Proficient Enzyme. *Science* **267**, 90–93 (1995).

## Bibliography

15. Wolfenden, R. & Snider, M. J. The Depth of Chemical Time and the Power of Enzymes as Catalysts. *Acc. Chem. Res.* **34**, 938–945 (2001).
16. Warshel, A., Sharma, P. K., Kato, M., Xiang, Y., Liu, H. & Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **106**, 3210–3235 (2006).
17. Michaelis, L. & Menten, M. L. Die Kinetik der Invertinwirkung. *Biochem. Z.* **49**, 333–369 (1913).
18. Fersht, A. R. Catalysis, Binding and Enzyme-Substrate Complementarity. *Proc. R. Soc. Lond., B, Biol. Sci.* **187**, 397–407 (1974).
19. Cornish-Bowden, A. One hundred years of Michaelis–Menten kinetics. *Perspect. Sci.* **4**, 3–9 (2015).
20. Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E. & Kern, D. Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117–121 (2005).
21. Goodey, N. M. & Benkovic, S. J. Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* **4**, 474–482 (2008).
22. Junge, W. & Nelson, N. ATP Synthase. *Annu. Rev. Biochem.* **84**, 631–657 (2015).
23. Carvalho, A. T., Szeler, K., Vavitsas, K., Åqvist, J. & Kamerlin, S. C. Modeling the mechanisms of biological GTP hydrolysis. *Arch. Biochem. Biophys.* **582**, 80–90 (2015).
24. Orellana, L. Large-Scale Conformational Changes and Protein Function: Breaking the in silico Barrier. *Front. Mol. Biosci.* **6**, 117 (2019).
25. Hollingsworth, S. A. & Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **99**, 1129–1143 (2018).
26. Saura, P., Röpke, M., Gamiz-Hernandez, A. P. & Kaila, V. R. I. in *Biomolecular Simulations: Methods and Protocols* (eds Bonomi, M. & Camilloni, C.) 75–104 (Springer New York, New York, NY, 2019).
27. Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M. & Kern, D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913–916 (2007).
28. Hartl, F. U. Molecular chaperones in cellular protein folding. *Nature* **381**, 571–580 (1996).
29. Hartl, F. U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16**, 574–581 (2009).
30. Picard, D. Heat-shock protein 90, a chaperone for folding and regulation. *Cell. Mol. Life Sci.* **59**, 1640–1648 (2002).
31. Abu-Farha, M., Lanouette, S., Elisma, F., Tremblay, V., Butson, J., Figeys, D. & Couture, J.-F. Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2. *J. Mol. Cell Biol.* **3**, 301–308 (2011).

32. Jaeger, A. M. & Whitesell, L. HSP90: Enabler of Cancer Adaptation. *Annu. Rev. Cancer Biol.* **3**, 275–297 (2019).
33. Boczek, E. E., Reefschräger, L. G., Dehling, M., Struller, T. J., Häusler, E., Seidl, A., Kaila, V. R. I. & Buchner, J. Conformational processing of oncogenic v-Src kinase by the molecular chaperone Hsp90. *Proc. Natl. Acad. Sci. USA* **112**, E3189–E3198 (2015).
34. Wayne, N. & Bolon, D. N. Dimerization of Hsp90 Is Required for in Vivo Function: Design and Analysis of Monomers and Dimers. *J. Biol. Chem.* **282**, 35386–35395 (2007).
35. Prodromou, C., Roe, S., O’Brien, R., Ladbury, J. E., Piper, P. W. & Pearl, L. H. Identification and Structural Characterization of the ATP/ADP-Binding Site in the Hsp90 Molecular Chaperone. *Cell* **90**, 65–75 (1997).
36. Mishra, P. & Bolon, D. N. Designed Hsp90 Heterodimers Reveal an Asymmetric ATPase-Driven Mechanism In Vivo. *Mol. Cell* **53**, 344–350 (2014).
37. Schopf, F. H., Biebl, M. M. & Buchner, J. The HSP90 chaperone machinery. *Nat. Rev. Mol. Cell Biol.* **18**, 345–360 (2017).
38. Pearl, L. H. & Prodromou, C. Structure and Mechanism of the Hsp90 Molecular Chaperone Machinery. *Annu. Rev. Biochem.* **75**, 271–294 (2006).
39. Donlin, L. T., Andresen, C., Just, S., Rudensky, E., Pappas, C. T., Kruger, M., Jacobs, E. Y., Unger, A., Zieseniss, A., Dobenecker, M.-W., Voelkel, T., Chait, B. T., Gregorio, C. C., Rottbauer, W., Tarakhovsky, A. & Linke, W. A. Smyd2 controls cytoplasmic lysine methylation of Hsp90 and myofilament organization. *Genes Dev.* **26**, 114–119 (2012).
40. Cunningham, C. N., Krukenberg, K. A. & Agard, D. A. Intra- and Intermonomer Interactions Are Required to Synergistically Facilitate ATP Hydrolysis in Hsp90. *J. Biol. Chem.* **283**, 21170–21178 (2008).
41. Ali, M. M., Mark Roe, S., Vaughan, C. K., Meyer, P., Panaretou, B., Piper, P. W., Prodromou, C. & Pearl, L. H. Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440**, 1013–1017 (2006).
42. Meyer, P., Prodromou, C., Hu, B., Vaughan, C., Roe, S. M., Panaretou, B., Piper, P. W. & Pearl, L. H. Structural and Functional Analysis of the Middle Segment of Hsp90: Implications for ATP Hydrolysis and Client Protein and Cochaperone Interactions. *Mol. Cell* **11**, 647–658 (2003).
43. Taipale, M., Jarosz, D. F. & Lindquist, S. HSP90 at the hub of protein homeostasis: Emerging mechanistic insights. *Nat. Rev. Mol. Cell Biol.* **11**, 515–528 (2010).
44. Mader, S. L., Lopez, A., Lawatscheck, J., Luo, Q., Rutz, D. A., Gamiz-Hernandez, A. P., Sattler, M., Buchner, J. & Kaila, V. R. I. Conformational dynamics modulate the catalytic activity of the molecular chaperone Hsp90. *Nat. Commun.* **11**, 1410 (2020).

## Bibliography

45. McLaughlin, S. H., Smith, H. W. & Jackson, S. E. Stimulation of the weak ATPase activity of human Hsp90 by a client protein. *J. Mol. Biol.* **315**, 787–798 (2002).
46. Panaretou, B., Prodromou, C., Roe, S. M., O'Brien, R., Ladbury, J. E., Piper, P. W. & Pearl, L. H. ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo. *EMBO J.* **17**, 4829–4836 (1998).
47. Hessling, M., Richter, K. & Buchner, J. Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90. *Nat. Struct. Mol. Biol.* **16**, 287–293 (2009).
48. Zierer, B. K., Rübbelke, M., Tippel, F., Madl, T., Schopf, F. H., Rutz, D. A., Richter, K., Sattler, M. & Buchner, J. Importance of cycle timing for the function of the molecular chaperone Hsp90. *Nat. Struct. Mol. Biol.* **23**, 1020–1028 (2016).
49. Huse, M. & Kuriyan, J. The Conformational Plasticity of Protein Kinases. *Cell* **109**, 275–282 (2002).
50. Brown, M. T. & Cooper, J. A. Regulation, substrates and functions of src. *Biochim. Biophys. Acta Rev. Cancer* **1287**, 121–149 (1996).
51. Roskoski, R. Src protein-tyrosine kinase structure, mechanism, and small molecule inhibitors. *Pharmacol. Res.* **94**, 9–25 (2015).
52. Taylor, S. S., Keshwani, M. M., Steichen, J. M. & Kornev, A. P. Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **367**, 2517–2528 (2012).
53. Cowan-Jacob, S. W., Fendrich, G., Manley, P. W., Jahnke, W., Fabbro, D., Liebetanz, J. & Meyer, T. The Crystal Structure of a c-Src Complex in an Active Conformation Suggests Possible Steps in c-Src Activation. *Structure* **13**, 861–871 (2005).
54. Boczek, E. E., Luo, Q., Dehling, M., Röpke, M., Mader, S. L., Seidl, A., Kaila, V. R. I. & Buchner, J. Autophosphorylation activates c-Src kinase through global structural rearrangements. *J. Biol. Chem.* **294**, 13186–13197 (2019).
55. Xu, W., Harrison, S. C. & Eck, M. J. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* **385**, 595–602 (1997).
56. Williams, J. C., Weijland, A., Gonfloni, S., Thompson, A., Courtneidge, S. A., Superti-Furga, G. & Wierenga, R. K. The 2.35 Å crystal structure of the inactivated form of chicken src: a dynamic molecule with multiple regulatory interactions. *J. Mol. Biol.* **274**, 757–775 (1997).
57. Xu, W., Doshi, A., Lei, M., Eck, M. J. & Harrison, S. C. Crystal Structures of c-Src Reveal Features of Its Autoinhibitory Mechanism. *Mol. Cell* **3**, 629–638 (1999).
58. Takai, Y., Sasaki, T. & Matozaki, T. Small GTP-Binding Proteins. *Physiol. Rev.* **81**, 153–208 (2001).
59. Sprang, S. R. G Protein Mechanisms: Insights from Structural Analysis. *Annu. Rev. Biochem.* **66**, 639–678 (1997).



60. Aktories, K. Bacterial protein toxins that modify host regulatory GTPases. *Nat. Rev. Microbiol.* **9**, 487–498 (2011).
61. Spanò, S. & Galán, J. E. A Rab32-Dependent Pathway Contributes to Salmonella Typhi Host Restriction. *Science* **338**, 960–963 (2012).
62. Ho, T. D., Figueroa-Bossi, N., Wang, M., Uzzau, S., Bossi, L. & Slauch, J. M. Identification of GtgE, a Novel Virulence Factor Encoded on the Gifsy-2 Bacteriophage of Salmonella enterica Serovar Typhimurium. *J. Bacteriol.* **184**, 5234–5239 (2002).
63. Kohler, A. C., Spanò, S., Galán, J. E. & Stebbins, C. E. Structural and enzymatic characterization of a host-specificity determinant from *Salmonella*. *Acta Crystallogr. D* **70**, 384–391 (2014).
64. Ishikawa, N., Tanaka, H., Koyama, F., Noguchi, H., Wang, C. C. C., Hotta, K. & Watanabe, K. Non-Heme Dioxygenase Catalyzes Atypical Oxidations of 6,7-Bicyclic Systems To Form the 6,6-Quinolone Core of Viridicatin-Type Fungal Alkaloids. *Angew. Chem. Int. Ed.* **53**, 12880–12884 (2014).
65. Heeb, S., Fletcher, M. P., Chhabra, S. R., Diggle, S. P., Williams, P. & Cámara, M. Quinolones: from antibiotics to autoinducers. *FEMS Microbiol. Rev.* **35**, 247–274 (2011).
66. Bräuer, A., Beck, P., Hintermann, L. & Groll, M. Structure of the Dioxygenase AsqJ: Mechanistic Insights into a One-Pot Multistep Quinolone Antibiotic Biosynthesis. *Angew. Chem. Int. Ed.* **55**, 422–426 (2016).
67. Mader, S. L., Bräuer, A., Groll, M. & Kaila, V. R. I. Catalytic mechanism and molecular engineering of quinolone biosynthesis in dioxygenase AsqJ. *Nat. Commun.* **9**, 1168 (2018).
68. Hausinger, R. P. Fe(II)/ $\alpha$ -Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Crit. Rev. Biochem. Mol. Biol.* **39**, 21–68 (2004).
69. Chang, W.-c., Li, J., Lee, J. L., Cronican, A. A. & Guo, Y. Mechanistic Investigation of a Non-Heme Iron Enzyme Catalyzed Epoxidation in (-)-4'-Methoxycyclophenin Biosynthesis. *J. Am. Chem. Soc.* **138**, 10390–10393 (2016).
70. Su, H., Sheng, X., Zhu, W., Ma, G. & Liu, Y. Mechanistic Insights into the Decoupled Desaturation and Epoxidation Catalyzed by Dioxygenase AsqJ Involved in the Biosynthesis of Quinolone Alkaloids. *ACS Catal.* **7**, 5534–5543 (2017).
71. Song, X., Lu, J. & Lai, W. Mechanistic insights into dioxygen activation, oxygen atom exchange and substrate epoxidation by AsqJ dioxygenase from quantum mechanical/molecular mechanical calculations. *Phys. Chem. Chem. Phys.* **19**, 20188–20197 (2017).
72. Krebs, C., Galonić Fujimori, D., Walsh, C. T. & Bollinger, J. M. Non-Heme Fe(IV)–Oxo Intermediates. *Acc. Chem. Res.* **40**, 484–492 (2007).

## Bibliography

73. Bruijninx, P. C. A., van Koten, G. & Klein Gebbink, R. J. M. Mononuclear non-heme iron enzymes with the 2-His-1-carboxylate facial triad: recent developments in enzymology and modeling studies. *Chem. Soc. Rev.* **37**, 2716–2744 (2008).
74. Costas, M., Mehn, M. P., Jensen, M. P. & Que, L. Dioxygen Activation at Mononuclear Nonheme Iron Active Sites: Enzymes, Models, and Intermediates. *Chem. Rev.* **104**, 939–986 (2004).
75. Kovaleva, E. G. & Lipscomb, J. D. Versatility of biological non-heme Fe(II) centers in oxygen activation reactions. *Nat. Chem. Biol.* **4**, 186–193 (2008).
76. Poulos, T. L. Heme enzyme structure and function. *Chem. Rev.* **114**, 3919–3962 (2014).
77. Krebs, C., Bollinger, J. M. & Booker, S. J. Cyanobacterial alkane biosynthesis further expands the catalytic repertoire of the ferritin-like ‘di-iron-carboxylate’ proteins. *Curr. Opin. Chem. Biol.* **15**, 291–303 (2011).
78. Banerjee, R., Proshlyakov, Y., Lipscomb, J. D. & Proshlyakov, D. A. Structure of the key species in the enzymatic oxidation of methane to methanol. *Nature* **518**, 431–434 (2015).
79. Martinez, S. & Hausinger, R. P. Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases. *J. Biol. Chem.* **290**, 20702–20711 (2015).
80. Que, L. The Road to Non-Heme Oxoferryls and Beyond. *Acc. Chem. Res.* **40**, 493–500 (2007).
81. Nam, W. High-Valent Iron(IV)–Oxo Complexes of Heme and Non-Heme Ligands in Oxygenation Reactions. *Acc. Chem. Res.* **40**, 522–531 (2007).
82. Senftle, T. P., Hong, S., Islam, M. M., Kylasa, S. B., Zheng, Y., Shin, Y. K., Junkermeier, C., Engel-Herbert, R., Janik, M. J., Aktulga, H. M., Verstraelen, T., Grama, A. & Van Duin, A. C. The ReaxFF reactive force-field: Development, applications and future directions. *Npj Comput. Mater.* **2**, 15011 (2016).
83. MacKerell Jr., A. D. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **25**, 1584–1604 (2004).
84. Jones, J. E. & Chapman, S. On the determination of molecular fields. I. From the variation of the viscosity of a gas with temperature. *Proc. R. Soc. Lond. A* **106**, 441–462 (1924).
85. Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **159**, 98–103 (1967).
86. Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **76**, 637–649 (1982).
87. Alder, B. J. & Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* **31**, 459–466 (1959).

88. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An  $N\log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
89. Senn, H. M. & Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Edit.* **48**, 1198–1229 (2009).
90. Hayward, S. & Go, N. Collective Variable Description of Native Protein Dynamics. *Annu. Rev. Phys. Chem.* **46**, 223–250 (1995).
91. Balsera, M. A., Wriggers, W., Oono, Y. & Schulten, K. Principal Component Analysis and Long Time Protein Dynamics. *J. Phys. Chem.* **100**, 2567–2572 (1996).
92. Teodoro, M. L., Phillips, G. N. & Kavraki, L. E. Understanding Protein Flexibility through Dimensionality Reduction. *J. Comput. Biol.* **10**, 617–634 (2003).
93. Shurki, A., Štrajbl, M., Schutz, C. N. & Warshel, A. Electrostatic Basis for Bioenergetics. *Method. Enzymol.* **380**, 52–84 (2004).
94. Bashford, D. & Karplus, M. pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* **29**, 10219–10225 (1990).
95. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041 (2001).
96. Honig, B. & Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **268**, 1144–1149 (1995).
97. Kieseritzky, G. & Knapp, E.-W. Optimizing pKA computation in proteins with pH adapted conformations. *Proteins: Struct., Funct., Bioinf.* **71**, 1335–1348 (2008).
98. Beroza, P., Fredkin, D. R., Okamura, M. Y. & Feher, G. Protonation of interacting residues in a protein by a Monte Carlo method: application to lysozyme and the photosynthetic reaction center of Rhodobacter sphaeroides. *Proc. Natl. Acad. Sci. USA* **88**, 5804–5808 (1991).
99. Schrödinger, E. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Phys. Rev.* **28**, 1049–1070 (1926).
100. Born, M. & Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys. (Leipzig)* **389**, 457–484 (1927).
101. Koopmans, T. Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den einzelnen Elektronen eines Atoms. *Physica* **1**, 104–113 (1934).
102. Echenique, P. & Alonso, J. L. A mathematical and computational review of Hartree–Fock SCF methods in quantum chemistry. *Mol. Phys.* **105**, 3057–3098 (2007).
103. Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **23**, 69–89 (1951).
104. Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **136**, B864–B871 (1964).

## Bibliography

105. Mardirossian, N. & Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **115**, 2315–2372 (2017).
106. Kohn, W. & Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **140**, A1133–A1138 (1965).
107. Becke, A. D. Density-Functional Exchange-Energy Approximation With Correct Asymptotic Behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
108. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **37**, 785–789 (1988).
109. Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B* **33**, 8822–8824 (1986).
110. Tao, J., Perdew, J. P., Staroverov, V. N. & Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **91**, 146401 (2003).
111. Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **98**, 1372–1377 (1993).
112. Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **98**, 11623–11627 (1994).
113. Blomberg, M. R. A., Borowski, T., Himo, F., Liao, R.-Z. & Siegbahn, P. E. M. Quantum Chemical Studies of Mechanisms for Metalloenzymes. *Chem. Rev.* **114**, 3601–3658 (2014).
114. Staroverov, V. N., Scuseria, G. E., Tao, J. & Perdew, J. P. Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *J. Chem. Phys.* **119**, 12129–12137 (2003).
115. Jensen, K. P. Bioinorganic Chemistry Modeled with the TPSSh Density Functional. *Inorg. Chem.* **47**, 10357–10365 (2008).
116. Lundberg, M. & Siegbahn, P. E. M. Quantifying the effects of the self-interaction error in DFT: When do the delocalized states appear? *J. Chem. Phys.* **122**, 224103 (2005).
117. Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **25**, 1463–1473 (2004).
118. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A Consistent and Accurate ab initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
119. Slater, J. C. Atomic Shielding Constants. *Phys. Rev.* **36**, 57–64 (1930).
120. Boys, S. F. Electronic Wave Functions. I. A General Method of Calculation for the Stationary States of Any Molecular System. *Proc. R. Soc. Lond. A* **200**, 542–554 (1950).

121. Schäfer, A., Horn, H. & Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **97**, 2571–2577 (1992).
122. Schäfer, A., Huber, C. & Ahlrichs, R. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.* **100**, 5829–5835 (1994).
123. Weigend, F. & Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **7**, 3297–3305 (2005).
124. Siegbahn, P. E. M. & Himo, F. The Quantum Chemical Cluster Approach for Modeling Enzyme Reactions. *Comput. Mol. Sci.* **1**, 323–336 (2011).
125. Himo, F. Recent Trends in Quantum Chemical Modeling of Enzymatic Reactions. *J. Am. Chem. Soc.* **139**, 6780–6786 (2017).
126. Klamt, A. & Schüürmann, G. Cosmo - a New Approach To Dielectric Screening in Solvents With Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc., Perkin Trans. 2*, 799–805 (1993).
127. Miertuš, S., Scrocco, E. & Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **55**, 117–129 (1981).
128. Eyring, H. The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **3**, 107–115 (1935).
129. Halgren, T. A. & Lipscomb, W. N. The Synchronous-Transit Method for Determining Reaction Pathways and Locating Molecular Transition States. *Chem. Phys. Lett.* **49**, 225–232 (1977).
130. Plessow, P. Reaction Path Optimization without NEB Springs or Interpolation Algorithms. *J. Chem. Theory Comput.* **9**, 1305–1310 (2013).
131. Sumowski, C. V. & Ochsenfeld, C. A Convergence Study of QM/MM Isomerization Energies with the Selected Size of the QM Region for Peptidic Systems. *J. Phys. Chem. A* **113**, 11734–11741 (2009).
132. Warshel, A. & Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **103**, 227–249 (1976).
133. Lin, H. & Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **117**, 185–199 (2006).
134. Jing, Z., Liu, C., Cheng, S. Y., Qi, R., Walker, B. D., Piquemal, J.-P. & Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annu. Rev. Biophys.* **48**, 371–394 (2019).
135. Bixon, M. & Lifson, S. Potential functions and conformations in cycloalkanes. *Tetrahedron* **23**, 769–784 (1967).

## Bibliography

136. Wiberg, K. B. A Scheme for Strain Energy Minimization. Application to the Cycloalkanes. *J. Am. Chem. Soc.* **87**, 1070–1078 (1965).
137. Fletcher, R. & Reeves, C. M. Function minimization by conjugate gradients. *Comput. J.* **7**, 149–154 (1964).
138. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
139. Schlegel, H. B. Geometry optimization. *WIREs Comput. Mol. Sci.* **1**, 790–809 (2011).
140. Sorensen, D. C. Newton’s Method with a Model Trust Region Modification. *SIAM J. Numer. Anal.* **19**, 409–426 (1982).
141. Kästner, J. Umbrella sampling. *WIREs Comput. Mol. Sci.* **1**, 932–942 (2011).
142. Torrie, G. M. & Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* **28**, 578–581 (1974).
143. Torrie, G. & Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).
144. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
145. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* **16**, 1339–1350 (1995).
146. Ahlrichs, R., Bär, M., Häser, M., Horn, H. & Kölmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **162**, 165–169 (1989).
147. Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caffisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M. & Karplus, M. CHARMM: the Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
148. Riahi, S. & Rowley, C. N. The CHARMM-TURBOMOLE Interface for Efficient and Accurate QM/MM Molecular Dynamics, Free Energies, and Excited State Properties. *J. Comput. Chem.* **35**, 2076–2086 (2014).
149. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L. & Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

150. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
151. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).
152. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).
153. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **14**, 33–38 (1996).
154. Lavery, L. A., Partridge, J. R., Ramelot, T. A., Elnatan, D., Kennedy, M. A. & Agard, D. A. Structural Asymmetry in the Closed State of Mitochondrial Hsp90 (TRAP1) Supports a Two-Step ATP Hydrolysis Mechanism. *Mol. Cell* **53**, 330–343 (2014).
155. Deglmann, P., Furche, F. & Ahlrichs, R. An efficient implementation of second analytical derivatives for density functional methods. *Chem. Phys. Lett.* **362**, 511–518 (2002).
156. Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M. & MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone  $\phi$ ,  $\psi$  and Side-Chain  $\chi_1$  and  $\chi_2$  Dihedral Angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
157. Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **91**, 275–282 (1995).
158. Šali, A., Potterton, L., Yuan, F., van Vlijmen, H. & Karplus, M. Evaluation of comparative protein modeling by MODELLER. *Proteins: Struct., Funct., Bioinf.* **23**, 318–326 (1995).
159. MacKerell Jr., A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D. & Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
160. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
161. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
162. Rabenstein, B. & Knapp, E.-W. Calculated pH-Dependent Population and Protonation of Carbon-Monoxo-Myoglobin Conformers. *Biophys. J.* **80**, 1141–1150 (2001).

## Bibliography

163. Meng, Y. & Roux, B. Locking the Active Conformation of c-Src Kinase through the Phosphorylation of the Activation Loop. *J. Mol. Biol.* **426**, 423–435 (2014).
164. Bakan, A., Meireles, L. M. & Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **27**, 1575–1577 (2011).
165. Mulliken, R. S. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *J. Chem. Phys.* **23**, 1833–1840 (1955).
166. Contreras-García, J., Johnson, E. R., Keinan, S., Chaudret, R., Piquemal, J.-P., Beratan, D. N. & Yang, W. NCIPLOT: A Program for Plotting Noncovalent Interaction Regions. *J. Chem. Theory Comput.* **7**, 625–632 (2011).
167. Vanommeslaeghe, K., Raman, E. P. & MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **52**, 3155–3168 (2012).
168. Valiev, M., Bylaska, E., Govind, N., Kowalski, K., Straatsma, T., Dam, H. V., Wang, D., Nieplocha, J., Apra, E., Windus, T. & de Jong, W. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **181**, 1477–1489 (2010).
169. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32**, 1456–1465 (2011).
170. Rehn, A., Lawatscheck, J., Jokisch, M.-L., Mader, S. L., Luo, Q., Tippel, F., Blank, B., Richter, K., Lang, K., Kaila, V. R. I. & Buchner, J. A methylated lysine is a switch point for conformational communication in the chaperone Hsp90. *Nat. Commun.* **11**, 1219 (2020).
171. Wachtel, R., Bräuning, B., Mader, S. L., Ecker, F., Kaila, V. R. I., Groll, M. & Itzen, A. The protease GtgE from Salmonella exclusively targets inactive Rab GTPases. *Nat. Commun.* **9**, 44 (2018).
172. Lindsay, S., Mader, S. L., Kaila, V. R. I. & Hess, C. R. C-H Oxidation by a Diiron Complex with Facially Opposing Active Sites. *ChemistrySelect* **3**, 1602–1608 (2018).
173. Mukherjee, G., Lee, C. W. Z., Nag, S. S., Alili, A., Cantú Reinhard, F. G., Kumar, D., Sastri, C. V. & de Visser, S. P. Dramatic rate-enhancement of oxygen atom transfer by an iron(IV)-oxo species by equatorial ligand field perturbations. *Dalton Trans.* **47**, 14945–14957 (2018).



# Publications

The research contained in this thesis was originally published in the following journals and is material from:

- I. Wachtel, R., Bräuning, B., Mader, S. L., Ecker, F., Kaila, V. R. I., Groll, M. & Itzen, A.: The protease GtgE from *Salmonella* exclusively targets inactive Rab GTPases. *Nature Communications*, published 2018, 9, 44:1-13. *Nature Research*, Springer Nature. Reproduced with permission.
- II. Lindsay, S., Mader, S. L., Kaila, V. R. I. & Hess, C. R.: C-H Oxidation by a Diiron Complex with Facially Opposing Active Sites. *ChemistrySelect*, published 2018, 3, 1602-1608. Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.
- III. Mader, S. L., Bräuer, A., Groll, M. & Kaila, V. R. I.: Catalytic mechanism and molecular engineering of quinolone biosynthesis in dioxygenase AsqJ. *Nature Communications*, published 2018, 9, 1168:1-8. *Nature Research*, Springer Nature. Reproduced with permission.
- IV. Boczek, E. E., Luo, Q., Dehling, M., Röpke, M., Mader, S. L., Seidl, A., Kaila, V. R. I. & Buchner, J.: Autophosphorylation activates c-Src kinase through global structural rearrangements. *Journal of Biological Chemistry*, published 2019, 294, 13186-13197. *American Society for Biochemistry and Molecular Biology*. Reproduced with permission.
- V. Rehn, F., Lawatscheck, J., Jokisch, M.-L., Mader, S. L., Qi, L., Tippel, F., Blank, B., Richter, K., Lang, K., Kaila, V. R. I. & Buchner, J.: A methylated lysine is a switch point for conformational communication in the chaperone Hsp90. *Nature Communications*, published 2020, 11, 1219:1-14. *Nature Research*, Springer Nature. Reproduced with permission.
- VI. Mader, S. L., Lopez, A., Lawatscheck, J., Luo, Q., Rutz, D. A., Gamiz-Hernandez, A. P., Sattler, M., Buchner, J. & Kaila, V. R. I.: Conformational dynamics modulate the catalytic activity of the molecular chaperone Hsp90. *Nature Communications*, published 2020, 11, 1410:1-12. *Nature Research*, Springer Nature. Reproduced with permission.

The following final published journal articles are reproduced in accordance with the reuse permission granted by Springer Nature:

“Authors have the right to reuse their article’s Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution.”

## ARTICLE

DOI: 10.1038/s41467-018-03442-2

OPEN

# Catalytic mechanism and molecular engineering of quinolone biosynthesis in dioxygenase AsqJ

Sophie L. Mader <sup>1</sup>, Alois Bräuer<sup>1</sup>, Michael Groll <sup>1</sup> & Ville R.I. Kaila <sup>1</sup>

The recently discovered Fe<sup>II</sup>/α-ketoglutarate-dependent dioxygenase AsqJ from *Aspergillus nidulans* stereoselectively catalyzes a multistep synthesis of quinolone alkaloids, natural products with significant biomedical applications. To probe molecular mechanisms of this elusive catalytic process, we combine here multi-scale quantum and classical molecular simulations with X-ray crystallography, and in vitro biochemical activity studies. We discover that methylation of the substrate is essential for the activity of AsqJ, establishing molecular strain that fine-tunes  $\pi$ -stacking interactions within the active site. To rationally engineer AsqJ for modified substrates, we amplify dispersive interactions within the active site. We demonstrate that the engineered enzyme has a drastically enhanced catalytic activity for non-methylated surrogates, confirming our computational data and resolved high-resolution X-ray structures at 1.55 Å resolution. Our combined findings provide crucial mechanistic understanding of the function of AsqJ and showcase how combination of computational and experimental data enables to rationally engineer enzymes.

<sup>1</sup>Center for Integrated Protein Science Munich (CIPSM), Department Chemie, Technische Universität München, Lichtenbergstraße 4, 85748 Garching, Germany. These authors contributed equally: Sophie L. Mader, Alois Bräuer. Correspondence and requests for materials should be addressed to M.G. (email: [michael.groll@tum.de](mailto:michael.groll@tum.de)) or to V.R.I.K. (email: [ville.kaila@ch.tum.de](mailto:ville.kaila@ch.tum.de))

The non-heme Fe<sup>II</sup>/α-ketoglutarate-dependent dioxygenase AsqJ from *Aspergillus nidulans* (Fig. 1a) is an exceptional enzyme that activates dioxygen and stereoselectively catalyzes a C–C bond desaturation and epoxidation reaction<sup>1,2</sup>. AsqJ converts its natural substrate, 4'-methoxycycloheptin (1), to a quinolone alkaloid, 4'-methoxyviridicatin (4) (Fig. 1b). It is remarkable that a single enzyme can catalyze such large-scale chemical transformations in a one-pot multistep reaction. The characteristic 4-arylquinolin-2(1H)-one structure of 4 is found in a variety of quinolone alkaloids<sup>3–5</sup>, which are compounds with promising antibacterial and antitumor activities<sup>6</sup>. In addition to the central role of non-heme Fe<sup>II</sup>/α-ketoglutarate-dependent oxygenases in many metabolic pathways<sup>7–9</sup>, elucidation of the catalytic mechanism of AsqJ is also important due to the promising application of different quinol alkaloids as potential drug candidates.

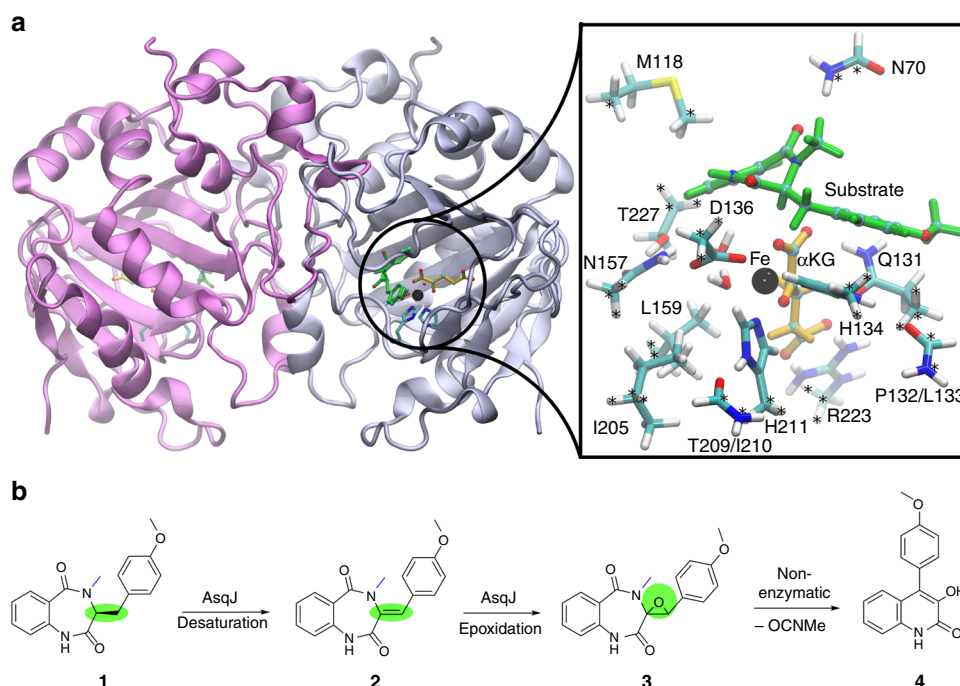
Watanabe and co-workers<sup>1</sup> initially discovered AsqJ as the enzyme responsible for the biosynthesis of the quinolone alkaloid 4 in *Aspergillus nidulans* (Fig. 1). The crystal structure of the Ni<sup>II</sup>-substituted AsqJ was recently resolved at 1.7 Å resolution, revealing the molecular architecture of the dioxygenase in presence of different reaction intermediates<sup>2</sup>. The active site of AsqJ forms a funnel like reaction chamber, located at the interface between antiparallel β-strands. The substrate interacts with the active site metal, which is ligated by His-211, His-134, Asp-136, as well as by the C-2 keto group and C-1 carboxylate of α-ketoglutarate (αKG), and a crystallographic water molecule in an octahedral coordination sphere. Moreover, His-134 forms a π-stacking interaction with the substrate that may further play an important role for the substrate binding and the catalytic activity.

Although the catalytic cycle of AsqJ has remained unclear, it was suggested that the enzyme generates a highly oxidizing ferryl species by splitting dioxygen, similar to many other non-heme iron enzymes<sup>10–12</sup>. AsqJ inserts one of the oxygen atoms into αKG, yielding succinate upon decarboxylation<sup>13</sup>. The resulting

ferryl is further employed to catalyze desaturation of the C–C single bond, yielding the intermediate 4'-methoxydehydrocycloheptin (2), characterized in high-performance liquid chromatography–mass spectrometry (HPLC/MS) experiments<sup>2</sup>. Recent experimental and computational studies<sup>14–16</sup> also support that the high-spin ferryl, Fe<sup>IV</sup>=O, intermediate could indeed be the catalytically active species. In subsequent steps of the reaction cycle, binding of a second oxygen molecule is likely to induce epoxidation of the double bond of 2 leading to formation of the epoxide 4'-methoxycycloheptin (3), which in turn undergoes a non-enzymatic re-arrangement and elimination reaction to form the final product 4 (Fig. 1b). Interestingly, it was also observed that AsqJ can effectively catalyze the desaturation reaction only when the substrate is methylated at the N4 position (see Supplementary Fig. 1 for substrate labeling), whereas upon removal of the N4-methyl group, the epoxide 3 was not formed<sup>2</sup>.

From an evolutionary perspective, this finding is in line with the N4-methylation activity of the non-ribosomal peptide synthetase (NRPS) AsqK producing 1<sup>17</sup>. Nevertheless, from a chemical point of view this finding is unexpected, since the methyl group is located three bonds apart from the reacting atoms in the substrate.

In order to probe the catalytic mechanics of AsqJ and to rationally engineer an enzyme that can catalyze chemical transformations of non-methylated substrates, we employ here an integrated computational and experimental approach. We derive the energetics and molecular structures of putative catalytic intermediates from multi-scale quantum and classical molecular simulations, which can provide powerful methodologies to study structure, dynamics, and energetics of complex (bio)chemical reactions on a wide range of timescales and spatial resolutions<sup>18–21</sup>. The computational work is combined with site-directed mutagenesis experiments, in vitro activity measurements by HPLC/MS, and structural characterization by X-ray crystallography.



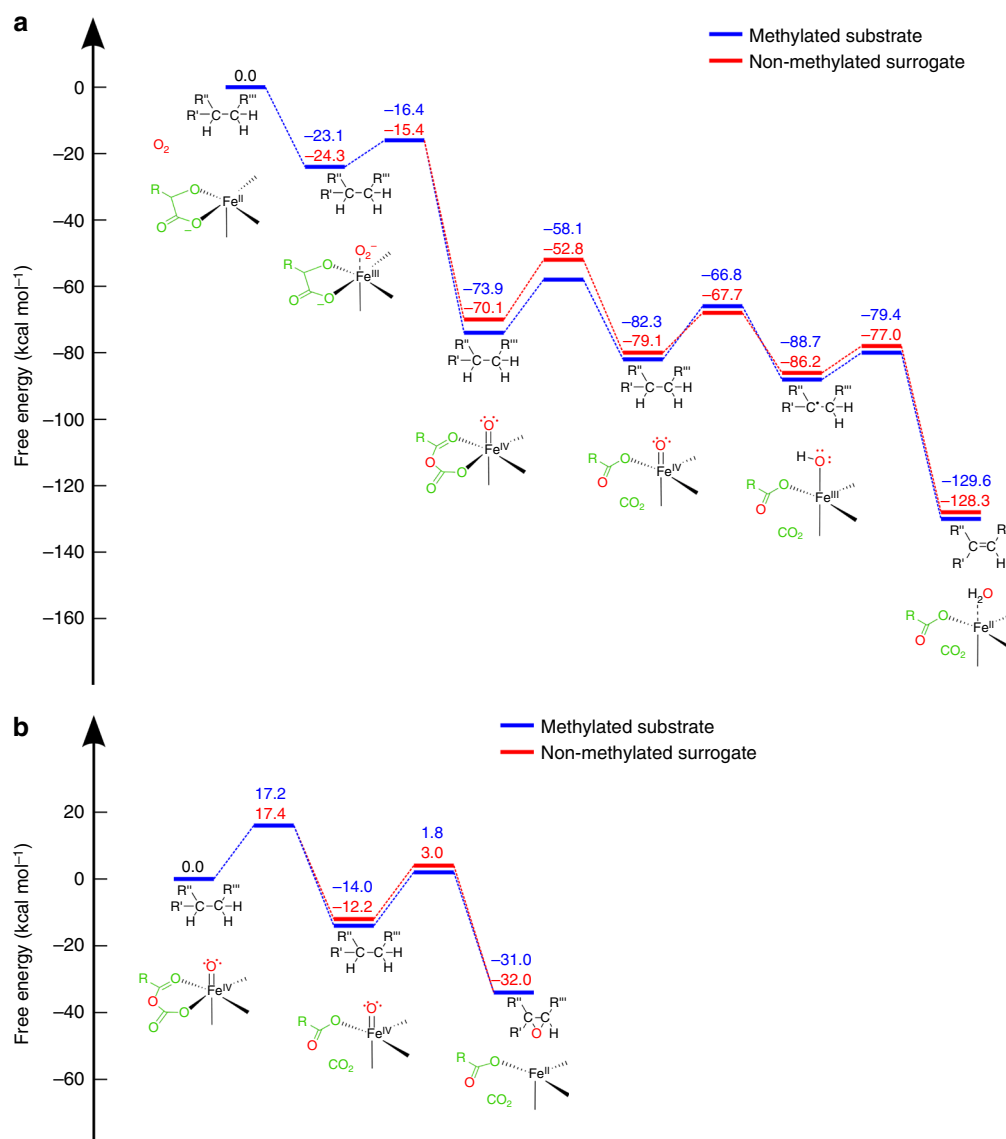
**Fig. 1** The structure and active site of AsqJ. **a** The dimeric protein structure of AsqJ (PDB ID: 5DAQ; one subunit in the asymmetric unit), depicting the location of the active site. The inset shows a DFT model of the active site of AsqJ, comprising 194 atoms. Atoms marked with asterisk are kept fixed in the structure optimizations. **b** Reaction sequence catalyzed by AsqJ. Substrate **1** is the AsqK-produced NRPS product of anthranilic acid and Tyr(OMe); the N4-methyl group is derived from S-adenosylmethionine<sup>15</sup>

## Results

**Energetics and mechanism of AsqJ.** In order to probe the energetics and structure of the reaction catalyzed by AsqJ for its natural substrate **1**, we performed quantum chemical density functional theory (DFT) calculations on active site enzyme models of AsqJ (Fig. 1a, inset). The catalytic cycle of AsqJ is initiated by binding of dioxygen to the Fe<sup>II</sup> active site. In order to accommodate O<sub>2</sub>, a crystallographic water molecule observed in the Ni<sup>II</sup>-substituted structure was removed and the structures were re-optimized. The optimized structural model suggests that substitution of Ni<sup>II</sup> with Fe<sup>II</sup> results only in minor structural changes (Supplementary Fig. 2). The DFT free energies along the quintet pathway indicate that dioxygen binding is exergonic by ca. 23 kcal mol<sup>-1</sup>. Thus, formation of the ferric-superoxide species, Fe<sup>III</sup>-O<sub>2</sub><sup>•-</sup>, occurs spontaneously, similar as in many other iron enzymes (Fig. 2a, see Supplementary Figs. 6 and 7 for the triplet spin state)<sup>22,23</sup>. For the insertion of the oxygen atom into αKG, we obtain a barrier of ca. 7 kcal mol<sup>-1</sup> with a transition state that has a peroxo-like character (Fig. 2a). For the subsequent

decarboxylation step, a barrier of ca. 16 kcal mol<sup>-1</sup> has to be overcome, for which the products CO<sub>2</sub> and succinate are stabilized by hydrogen bonds with a crystallographic water molecule and Gln-131, respectively (Supplementary Fig. 3, 4). Moreover, the DFT calculations suggest that entropic effects do not considerably alter the reaction profiles, except for the oxygen binding and decarboxylation steps, for which we obtain *TΔS* contributions at *T* = 298 K of ca. 6 kcal mol<sup>-1</sup> (destabilizing) and 4 kcal mol<sup>-1</sup> (stabilizing), respectively (Fig. 2a, Supplementary Fig. 5a).

The calculations suggest that the CO<sub>2</sub> release could be coupled to a flip of the ferryl group towards the substrate, which also enables the reorientation of the CO<sub>2</sub>-stabilizing water molecule (Supplementary Movie 1). Notably, the achieved state is ca. 8 kcal mol<sup>-1</sup> more favorable than the oxygen-inserted αKG state, and its configuration leads to an increase in the  $\pi$ -stacking interaction between the substrate and His-134, decreasing the distances between the groups from 3.7 Å to 3.5 Å. In the subsequent step, the oxidizing power of the ferryl is employed to thermodynamically drive two proton-coupled electron transfer



**Fig. 2** Calculated catalytic cycle and energetics of AsqJ. **a** The first part of the reaction cycle, comprising oxygen activation, decarboxylation of αKG, and two subsequent PCET reaction steps. **b** The second part of the reaction involving epoxidation of the substrate. All energies refer to free energies calculated in the quintet state at B3LYP-D3/def2-TZVP/ $\epsilon = 4$  level of theory with vibrational and entropic corrections at the B3LYP-D3/def2-SV(P)/ $\epsilon = 4$  level. Barriers are obtained from reaction pathway optimizations (Methods). See also Supplementary Movies 1 and 2, and Supplementary Figs. 3 and 9 for the optimized intermediate structures. Free energy profiles along the triplet surface are shown in Supplementary Figs. 6 and 7

(PCET) reactions, for which we obtain reaction barriers of ca. 16 kcal mol<sup>-1</sup> and 9 kcal mol<sup>-1</sup>, for the first and second transfer reactions, respectively (Fig. 2a). The first PCET process results in a radical on C3 that delocalizes mainly on the C3–C2 bond (Supplementary Fig. 8a), whereas no significant spin density is observed on the proton during the transfer from C3 to the iron ligand. We therefore conclude that the proton and electron move along different reaction channels between the substrate and the redox-active metal (Supplementary Fig. 8b).

After the desaturation has taken place, the second part of the reaction cycle is initiated by binding of new αKG and O<sub>2</sub> molecules, which lead to a similar decarboxylation and rotation of the oxo-ferryl bond, with a barrier of ca. 17 kcal mol<sup>-1</sup> (Fig. 2b). We further find that the resulting ferryl species could attack the double bond with a barrier of 16 kcal mol<sup>-1</sup>, resulting in the subsequent formation of the epoxide and Fe<sup>II</sup> (Fig. 2b, Supplementary Movie 2).

Taken together, the overall free energy profile is strongly exergonic, and the putative reactions have kinetically feasible barriers of ca. 7–17 kcal mol<sup>-1</sup>, which is consistent with the accumulation of the product on the seconds timescales in HPLC/MS experiments (see below). These barriers are also consistent with recent computational studies<sup>15,16</sup>, suggesting that the hydrogen abstraction and epoxidation reactions have barriers between 4 and 22 kcal mol<sup>-1</sup>.

**Substrate methylation is central for catalysis.** Next, we aimed to probe why AsqJ does not catalyze desaturation of the substrate analog lacking a methyl group at the N4 position. To this end we recalculated the reaction pathway with the non-methylated surrogate (Fig. 2, Supplementary Fig. 5, 9). Interestingly, we obtain similar energetics as for the natural substrate for most steps of the catalytic cycle. However, our computed binding affinities between the substrate and the protein, as well as the molecular strain energy stored within the substrates, suggest that there are significant differences between the N4-methylated and non-methylated educts (Table 1). We find that for the two PCET reactions, the methylated substrate undergoes marginal conformational changes around its C3–C1'–C1'–C2' dihedral angle relative to its optimized structure in gas phase (Table 1, Supplementary Fig. 10). In stark contrast, for the non-methylated surrogate, we observe that the central C3–C1'–C1'–C2' dihedral angle is twisted in the active site pocket by ca. 10–30° from its structure in gas phase (Table 1, Supplementary Fig. 10). According to the DFT calculations, these structural rearrangements introduces strain energy up to ca. 8 kcal mol<sup>-1</sup> in the surrogate, which in turn weakens the π–π stacking interaction between the surrogate's methoxy-phenyl group and His-134 (Table 1). This prediction is further supported by calculations of

non-covalent interaction surfaces within the active site pocket (Fig. 3a), and it is also consistent with reduction in the binding affinities between the substrate and the active site during the PCET reactions (Table 1). These findings thus suggest that the stabilization of non-methylated surrogates in AsqJ is impaired during the PCET reactions.

We got interested in how this π-stacking interaction is influenced by the dynamics of the enzyme, and therefore performed hybrid quantum mechanics/classical mechanics (QM/MM) molecular dynamics (MD) simulations of the states with the native substrate and non-methylated surrogate after the initial PCET process. Consistent with the results from the DFT models, our computations reveal that the distance between the His-134 and the methoxy-phenyl ring of the substrate drastically increases from 3.4 to 4.0 Å from the native substrate to the non-methylated surrogate in the QM/MM MD simulations (Fig. 3b, c). Hence, our simulations support that the molecular strain within the non-methylated substrate might lead to its dissociation from the binding pocket during the PCET steps.

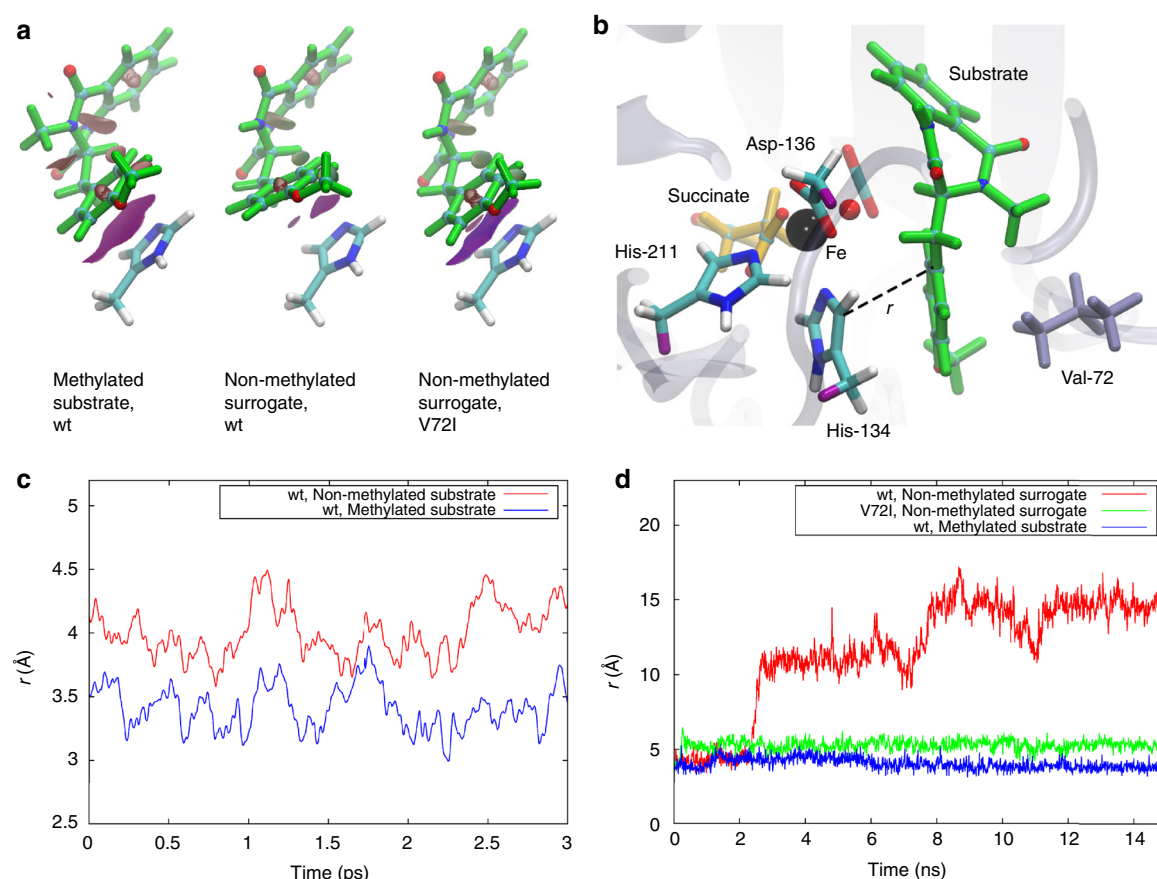
**Engineering AsqJ to catalyze non-methylated surrogate.** In the QM/MM MD trajectories, we observe that Val-72, located on a loop surrounding the substrate-binding pocket, forms dispersive interactions with the methoxy-phenyl of the ligand and His-134 (Fig. 3b). In order to strengthen the π–π interactions between the non-methylated surrogate and the protein, we replaced the Val-72 with a somewhat larger isoleucine residue, followed by new DFT calculations and classical MD simulations. Interestingly, the DFT models suggest that the V72I replacement indeed strengthens the interaction between the protein and the methoxy-phenyl ring of the non-methylated substrate analog prior to the first PCET reaction by decreasing the distances between the aromatic systems from 4.1 to 3.8 Å (Fig. 3a). The MD simulations also indicate an enhanced stabilization of the non-methylated surrogate in the active site of V72I mutant (Fig. 3c). For comparison, we also studied the effect of the V72K, V72L, and F139I substitutions to increase the interaction between His-134 and the substrate from the opposite side. However, whereas the V72K and V72L mutants did not improve the π-stacking interactions (Supplementary Fig. 12), the F139I variant showed some minor increase in the dispersive interactions between His-134 and the surrogate.

In our next approach, we probed the engineered enzyme experimentally. To this end, we cloned, expressed, and purified the V72I-, V72K-, and F139I-mutated AsqJ enzymes. The catalytic activities of these mutants were determined for the native substrate **1**, cyclopeptide **1b** (Supplementary Fig. 13b), as well as for demethylcyclopeptide **1d** (Fig. 4) and compared to the wild-type (wt) AsqJ by using a reverse-phase HPLC/MS-coupled

Table 1 Binding affinities, strain energies, and dihedral angles (C3–C1'–C1'–C2') for the substrates before and after the first and second PCET steps						
	Methylated substrate before first PCET	Non-methylated surrogate before first PCET	Methylated substrate after first PCET	Non-methylated surrogate after first PCET	Methylated substrate after second PCET	Non-methylated surrogate after second PCET
Binding affinity [kcal mol <sup>-1</sup> ]	T: -38.9 Q: -37.7	T: -32.7 Q: -38.0	T: -31.4 Q: -33.1	T: -30.9 Q: -32.6	T: -35.9 Q: -33.5	T: -31.2 Q: -30.5
Strain energy [kcal mol <sup>-1</sup> ]	T: 0.2 Q: 0.2	T: 7.6 Q: 2.4	T: 0.1 Q: 0.1	T: 0.8 Q: 0.8	T: 0.0 Q: 0.0	T: 5.9 Q: 0.5
Dihedral angle gas-phase [deg]	-22	-72	-42	-71	-22	-20
Dihedral angle protein [deg]	T: -27 Q: -27	T: -39 Q: -52	T: -46 Q: -46	T: -59 Q: -59	T: -19 Q: -22	T: -52 Q: -11

T and Q refer to structures optimized in the triplet and quintet states, respectively.



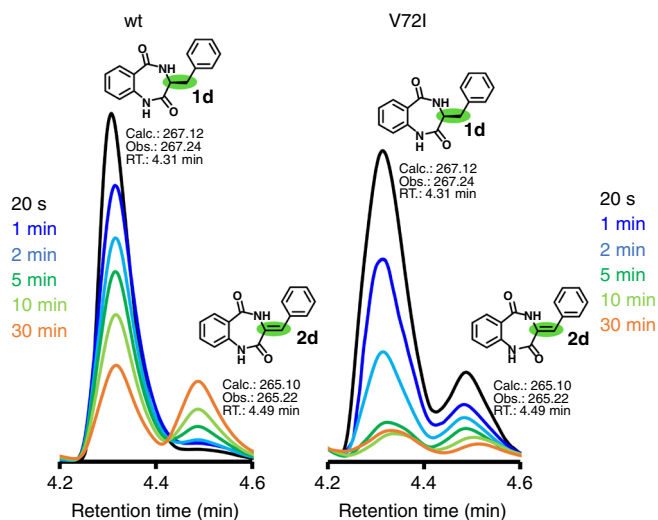


**Fig. 3** Engineering AsqJ to catalyze turnover of modified surrogates. **a** Non-covalent interaction densities (purple surfaces) between His-134 and the methylated substrate/non-methylated surrogate for wt and V72I-mutant AsqJ prior to the first PCET reaction. **b** Snapshot from a QM/MM MD simulation of AsqJ. The QM region is shown in colored licorice and link atoms are depicted in purple, and Val-72 is shown in gray licorice. **c** The stacking distance ( $r$ ) between His-134 (CD2) and the substrate (C1') in presence of the methylated substrate and the non-methylated surrogate after the first PCET reaction. **d** The stacking distance ( $r$ ) in presence of the methylated substrate and the non-methylated surrogate for wt and V72I-mutant AsqJ before the first PCET reaction. The simulations were performed using both QM/MM MD simulations (**c**) and classical atomistic MD simulations (**d**) to explore different timescales

activity assay, and X-ray crystallography. The wt and V72I-, V72K-, F139I-mutated AsqJ variants turned out to have similar turnover rates for the methylated substrates **1** and **1b** (Supplementary Fig. 13a, b), suggesting that these alterations do not lead to a significant decrease in the enzyme activity. In the following, we analyzed how the engineered enzymes affect the turnover of the non-methylated substrate analog **1d**. Remarkably, the HPLC/MS activity assays depict that the desaturated reaction intermediate, **2d**, which follows the two PCET reactions of **1d** (Figs. 1b, 2a), is already formed at 20 s for V72I, in stark contrast to the wt AsqJ for which we observe the product peak with maximum intensity after 30 min (Fig. 4). In contrast, the F139I mutant only displays a minor increase in activity, whereas the turnover of V72K for **1d** is similar to the wt enzyme as proposed by our MD simulations (Supplementary Fig. 14). Notably, formation of the epoxide **3d** was not observed for the designed mutant V72I, implying that the mutated enzyme is capable of activating C–H bonds of **1d**, while it cannot effectively proceed into the second part of the reaction cycle. On the other hand, starting with the synthesized desaturated intermediate **2** of the methylated natural substrate **1**, the activity assays show similar turnover for the wt AsqJ and the V72I variant (Supplementary Fig. 13c, d). These results led us to conclude that the substrate methylation is also a fundamental prerequisite for the epoxidation reaction. These findings are supported by the QM/MM MD

simulations, which indicate that stabilization of the non-methylated reaction intermediate **2d** inside the V72I variant is rather weak (Supplementary Fig. 15). Therefore, the following epoxidation, which further requires binding and activation of a new oxygen molecule and release of CO<sub>2</sub>, is prevented since the residence time of **2d** at the active center is probably far too short to successfully complete the oxidoreduction. However, elucidating the molecular principles for these steps and engineering an enzyme that would further catalyze the epoxidation reaction is outside the scope of the present work.

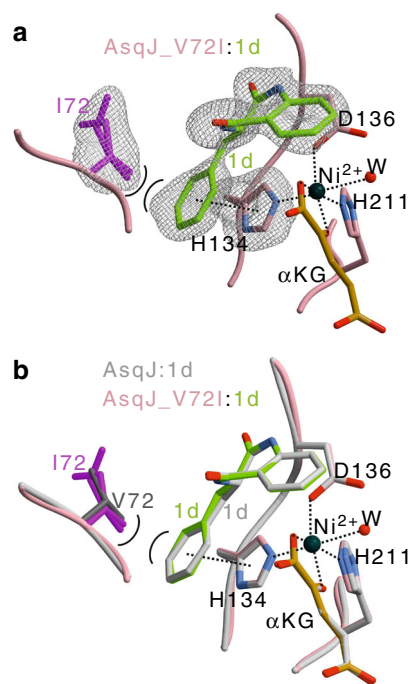
Experimental structural insight into the function of AsqJ\_V72I was obtained by crystallographically analyzing the mutant in complex with **1** (1.55 Å resolution,  $R_{\text{Free}} = 16.8\%$ ), **1b** (1.65 Å,  $R_{\text{Free}} = 17.4\%$ ), and **1d** (1.75 Å,  $R_{\text{Free}} = 19.3\%$ ), as well as determining the structure of V72K bound to **1b** (1.55 Å resolution,  $R_{\text{Free}} = 19.4\%$ , PDB ID 6EOZ). Our calculations predict that stabilization of these ligands in the active site depends on pronounced  $\pi$ -stacking interactions, which are most important for the initiation of the PCET reactions. Interestingly, the crystal structure of V72I in presence of **1d** reveals that the distance between Ile-72 and C4' of the non-methylated analog is decreased by 0.2 Å compared to wt AsqJ:**1d** complex (PDB ID 5DAX, Fig. 5). Notably, despite the local conformational changes at the active site, all other parts of the mutant structure remain unchanged with a root-mean square deviation (RMSD) of 0.15 Å



**Fig. 4** Activity measurements of wt and engineered AsqJ. Turnover of the wt and V72I-mutated AsqJ for the non-methylated substrate analog (**1d**) from reverse-phase HPLC/MS-coupled activity assays in the presence of  $\text{Fe}^{\text{II}}$ ,  $\alpha\text{KG}$ , oxygen, and ascorbic acid. The absorbance at  $\lambda = 280$  nm is shown for **1d** and the desaturated reaction intermediate **2d** (see Supplementary Fig. 13 for further characterization). The reaction progress was analyzed after 20 s (black trace), 1 min (dark blue), 2 min (light blue), 5 min (dark green), 10 min (light green), and 30 min (orange), showing that **2d** forms already after 20 s in V72I

for the protein backbone atoms. These experimental insights perfectly match to our initial computational refinements (see above). Moreover, a close inspection of the electron density map depicts an alternative conformation rotation of the Ile-72-side chain, positioning the introduced isobutyl group in an inward and outward orientation (Fig. 4). In addition, the B-factor of **1d** is increased from 28.7 (AsqJ-wt) to 35.3  $\text{\AA}^2$  (AsqJ-V72I) (Supplementary Table 3, Supplementary Fig. 16). This observation accounts for an improved mobility of the phenyl group of **1d** around the C3–C1' bond in the AsqJ\_V72I, and is caused by the two alternative conformations of Ile-72. Consistently, an extra lobe of electron density around the phenyl ring of **1d** in the AsqJ\_V72I structure indicates that the aromatic moiety of the ligand is quite flexible (Fig. 5a). Thus, our findings suggest that the stable  $\pi$ -stacking interactions of **1d** with His-134 lead to a prolonged residence time of the ligand at the active site, facilitating the desaturation reaction. Taken together, our findings suggest that substrate binding in AsqJ is linked with a conformational rearrangement of the residues in the loop surrounding the specificity pocket, and stabilization of the different conformational states might provide a promising starting point for future studies to rationally engineer AsqJ to catalyze the turnover of modified substrates.

In conclusion, the molecular mechanism of an exceptional one-pot multistep dioxygenase AsqJ from *A. nidulans* was studied by combined molecular simulations, biochemical in vitro activity assays, mutagenesis, and X-ray crystallography. The enzyme catalyzes a stepwise desaturation and epoxidation reaction of 4'-methoxycyclopeptin to the quinolone alkaloid, 4'-methoxyviridicatin by highly exergonic and kinetically feasible proton-coupled transfer PCET reactions. We observed that methylation of the substrate at the N4 position is important for stabilizing dispersive interactions within the active site during the PCET processes. Combining various orthogonal methodologies such as multi-scale computational chemistry, mutagenesis, in vitro activity assays, and protein crystallography, revealed that an



**Fig. 5** Crystal structure of the AsqJ\_V72I mutant. **a**) Close-up of the active site in complex with **1d** (shown in green). The amino acids engaged in ligand binding are depicted as sticks and labeled by the one letter code,  $\alpha$ -ketoglutarate ( $\alpha\text{KG}$ ) shown in orange, and  $\text{Ni}^{\text{II}}$  is shown as a black sphere. The electron density depicts a  $2F_o - F_c$  map with I72, H134, and **1d** omitted for phasing. I72 contains two alternative conformations with an occupancy of ~50% (PDB ID: 5OA8). An extra lobe of electron density around the phenyl ring of **1d** indicates increased mobility of the phenyl ring. **b**) Structure of the active site superposition of AsqJ\_V72I:**1d** with the wt AsqJ:**1d** (shown in gray; PDB ID: 5DAX).  $\pi$ -stacking and coordination of the metal atom are indicated by black-dotted lines. Note that the surrogate **1d** adopts identical positions in both crystal structures, while there are only marginal shifts in the distances of I72 coordinating the aromatic moiety of the ligand

engineered AsqJ\_V72I mutant is able to rapidly catalyze desaturation of the non-methylated surrogate, which is in stark contrast to wt AsqJ. We could show that the smallest possible V72I replacement in a loop surrounding the active site drastically improved turnover of demethylcyclopeptin **1d**, and that insertion of one methyl group to the protein compensates for the lack of one *N*-methyl group in the substrate. In conclusion, the achieved results provide an important starting point to understand molecular mechanisms of multifunctional enzymes, and to rationally engineer such systems for the synthesis of natural products with important biomedical applications.

## Methods

**DFT models.** Quantum chemical DFT cluster models with 190–194 atoms were constructed based on our recently crystallized  $\text{Ni}^{\text{II}}$ -substituted model of AsqJ (PDB ID: 5DAQ)<sup>2</sup>. The QM models comprised in addition to the iron-oxo species the substrate,  $\alpha\text{KG}$  (or  $\text{CO}_2$  + succinate), two crystal water molecules, and amino acid residues Asn-70, Met-118, Gln-131, His-134, Asp-136, Asn-157, Leu-159, Ile-205, His-211, Arg-223, and Thr-227 (Fig. 1a, inset). We note that, inclusion of Val-72 in the QM models did not considerably change the relative energies between the reaction intermediates (Supplementary Table 1). The amino acids were cut between the Ca and C $\beta$  atoms, except for Asn-70, Thr-227, and Arg-223, which were cut between C $\beta$ /C $\gamma$  and C $\gamma$ /C $\delta$  atoms, respectively. We also included the backbone of residues Pro-132/Leu-133 and Thr-209/Ile-210. The terminal carbon atoms were saturated with hydrogen atoms, and fixed during structure optimization to account for steric effects of the protein environment (Fig. 1a, inset). Geometry optimizations were performed at dispersion-corrected DFT level<sup>24</sup>, using the B3LYP-D3 functional<sup>25,26</sup>, and def2-SVP (light elements) and def2-TZVP basis sets (Fe).

The protein surroundings were treated as a polarizable medium with a dielectric constant of  $\epsilon = 4$  using the COSMO model<sup>27,28</sup>. Reaction pathways were optimized using the Woelfling method<sup>29</sup>, a chain-of-states method that is related to the nudged-elastic band<sup>30</sup> and zero-temperature string methods<sup>31</sup>. The final energetics for all states were computed at the B3LYP-D3/def2-TZVP/ $\epsilon = 4$  level, and electronic configurations were analyzed from Mulliken populations and spin density distributions. The reaction pathways were optimized for both the triplet and quintet spin states, as also supported by recent computational studies suggesting that several spin states may contribute in the catalysis steps of AsqJ<sup>15,16</sup>. The absolute electronic energies for the quintet states were ca. 5–10 kcal mol<sup>-1</sup> more stable than for the triplet states, whereas the structural and relative energy differences between the states were marginal. Vibrational and entropic corrections were obtained at the B3LYP-D3/def2-SV(P)/ $\epsilon = 4$  level by estimating the molecular Hessian numerically<sup>32</sup>. Furthermore, we studied the isolated substrate in different states by both complete and restrained optimizations at the B3LYP-D3/def2-SVP level. Strain energies were obtained from restrained optimizations of the relaxed substrates to their protein geometries. All distances and free energies discussed in the main text and shown in the figures were extracted from the optimized QM models. All QM calculations were performed using TURBOMOLE v. 6.6<sup>33</sup>, and non-covalent interaction regions were calculated using NCIPLOT<sup>34</sup>.

**Combined quantum mechanical and molecular mechanical methods.** Hybrid QM/MM MD simulations of AsqJ with the substrate with and without the methyl group at the N4 position, prior and after the first PCET reaction, were performed at the B3LYP-D3/CHARMM36<sup>35</sup> level using def2-SVP (light elements) and def2-TZVP (Fe) basis sets. The QM region comprised protein residues His-134, Asp-136 and His-211, the succinate/CO<sub>2</sub> moiety, in addition to the ferryl (Fe<sup>IV</sup> = O<sup>2-</sup>) or ferric (Fe<sup>III</sup>-OH<sup>-</sup>) metal center and the substrate. Link atoms were introduced between the C $\beta$  and C $\alpha$  atoms of the protein residues, and the monomeric protein model was embedded in a water box with 18,089 water molecules, and neutralized by 100 mM NaCl. The complete QM/MM system, comprising 58,896 atoms, was simulated for 3 ps for each state at  $T = 310$  K, and by using a 1 fs integration time step. No positional restraints were applied in the QM/MM MD simulations. All QM/MM calculations were performed using TURBOMOLE v. 6.6 linked together with CHARMM<sup>36,37</sup>.

**Classical MD simulations.** Classical MD simulations of the wt and V72I, V72K, and F139I in silico mutants were performed for 15 ns at  $T = 310$  K using a 1 fs time step, and treating the long-range electrostatics using the Particle Mesh Ewald (PME) approach<sup>38</sup>. We employed the CHARMM36 force field, together with force field parameters for the active site models based on CGenFF<sup>39</sup> and by calculated restrained electrostatic potential (RESP) charges for the substrates and small active site models at the B3LYP/def2-TZVP level obtained with NWChem<sup>40</sup>. Classical MD simulations were performed using NAMD<sup>41</sup>, and visual molecular dynamics was used for visualization analysis<sup>42</sup>.

**Site-directed mutagenesis experiments and protein purification.** Mutations in the wt *asqJ* gene of the fungus *A. nidulans* FGSC A4 (pET28bSUMOAnAsqJ, GenBank: XP\_682496.1)<sup>2</sup> were introduced via the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies) using oligonucleotides listed in Supplementary Table 2. Bacterial culture and protein purification were performed according to published procedures<sup>2</sup>. The protein solutions were buffer exchanged to 20 mM Tris hydrochloride, pH 7.4, containing 100 mM NaCl and 2 mM dithiothreitol, and were concentrated to a concentration of 20 mg/mL using a 10k Amicon Ultra Centrifugal Filter Device (Millipore, Billerica, MA), flash frozen with liquid nitrogen and stored at  $-80^\circ\text{C}$ .

**HPLC/MS experiments.** Reaction mixtures containing 33  $\mu\text{M}$  purified AsqJ\_wt (AsqJ\_V72I, AsqJ\_V72K, or, AsqJ\_F139I), 200  $\mu\text{M}$  of the synthesized substrate<sup>2</sup>, 100  $\mu\text{M}$  FeSO<sub>4</sub>, 2.5 mM  $\alpha$ -ketoglutarate, 4 mM ascorbic acid, and 50 mM Tris hydrochloride were incubated at  $30^\circ\text{C}$  for 20 s, 1 min, 2 min, 5 min, 10 min, 30 min, and 60 min. The reaction was stopped by adding 10% (v/v) of 3 M trichloroacetic acid, and samples were centrifuged at  $10,000 \times g$  for 10 min. All traces were monitored at 280 nm using a Dionex UltiMate 3000 HPLC system coupled with a Thermo LCQ fleet in combination with a Waters 1525 binary HPLC pump, X-Bridge Prep C18 column (5  $\mu\text{m}$ ,  $10 \times 250$  mm), Waters 2998 PDA detector and Waters Fraction Collector III (Waters).

**X-ray crystallography.** Mutant AsqJ proteins were crystallized as previously described for the wt AsqJ<sup>2</sup>. Crystals were grown at  $20^\circ\text{C}$  using the sitting drop vapor diffusion method. Drops contained a 1:1 mixture of protein solution (20 mg ml<sup>-1</sup> protein, 1 mM  $\alpha$ -ketoglutarate as well 2 mM **1**, **1b**, or **1d**) and reservoir solution (100 mM Tris hydrochloride, pH 7.8–8.2, 1.0 M LiBr, 27–30% PEG6000). Crystals were cryoprotected by addition of cryobuffer (100 mM Tris hydrochloride, pH 7.8, 500 mM LiBr, 30% PEG400). Diffraction data were collected at the beamline X06SA at the Paul Scherrer Institute, SLS, Villigen, Switzerland ( $\lambda = 1.0\text{\AA}$ ). Evaluation of reflection intensities and data reduction were performed with the program package XDS<sup>43</sup>. Positioning of the initial model was carried out with the coordinates of the AsqJ\_wt (PDB ID: 5DAQ) by rigid body refinements

(REFMAC5<sup>44</sup>). MAIN<sup>45</sup> and COOT<sup>46</sup> were used to build models. TLS (Translation/Libration/Screw) refinements finally yielded excellent  $R_{\text{work}}$  and  $R_{\text{free}}$  as well as RMSD bond and angle values (Supplementary Table 3). Note, the asymmetric unit of the crystals contains one subunit, which by crystallographic symmetry forms the physiological dimer, with two identical active sites.

**Data availability.** The coordinates, proven to have good stereochemistry from the Ramachandran plots, were deposited at the RCSB Protein Data Bank under the accession codes 5OA4 (AsqJ\_V72I:1), 5OA7 (AsqJ\_V72I:1b), 5OA8 (AsqJ\_V72I:1d), and 6EOZ (AsqJ\_V72K:1b). All other data is available from the authors upon reasonable request.

Received: 9 July 2017 Accepted: 13 February 2018

Published online: 21 March 2018

## References

- Ishikawa, N. et al. Non-heme dioxygenase catalyzes atypical oxidations of 6,7-bicyclic systems to form the 6,6-quinolone core of viridicatin-type fungal alkaloids. *Angew. Chem. Int. Ed.* **53**, 12880–12884 (2014).
- Bräuer, A., Beck, P., Hintermann, L. & Groll, M. Structure of the dioxygenase AsqJ: mechanistic insights into a one-pot multistep quinolone antibiotic biosynthesis. *Angew. Chem. Int. Ed.* **55**, 422–426 (2016).
- Scherlach, K. & Hertweck, C. Discovery of aspoquinolones A-D, prenylated quinoline-2-one alkaloids from *Aspergillus nidulans*, motivated by genome mining. *Org. Biomol. Chem.* **4**, 3517–3520 (2006).
- Uchida, R. et al. Yaequinolones, new insecticidal antibiotics produced by *Penicillium* sp. FKI-2140. *J. Antibiot.* **59**, 646–651 (2006).
- An, C.-Y. et al. 4-Phenyl-3,4-dihydroquinolone derivatives from *Aspergillus nidulans* MA-143, an endophytic fungus isolated from the mangrove plant *Rhizophora stylosa*. *J. Nat. Prod.* **76**, 1896–1901 (2013).
- Ahmed, A. & Daneshmandi, M. Nonclassical biological activities of quinolone derivatives. *J. Pharm. Pharm. Sci.* **15**, 52–72 (2012).
- Hausinger, R. P. Fe(II)/ $\alpha$ -ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* **39**, 21–68 (2004).
- Müller, I. et al. Crystal structure of the alkylsulfatase AtsK: insights into the catalytic mechanism of the Fe(II)  $\alpha$ -ketoglutarate-dependent dioxygenase superfamily. *Biochemistry* **43**, 3075–3088 (2004).
- Flashman, E. & Schofield, C. J. The most versatile of all reactive intermediates? *Nat. Chem. Biol.* **3**, 86–87 (2007).
- Krebs, C., Fujimori, D. G., Walsh, C. T. & Bollinger, J. M. Jr. Non-heme Fe (IV)-oxo intermediates. *Acc. Chem. Res.* **40**, 484–492 (2007).
- Bugg, T. D. H. Oxygenases: mechanisms and structural motifs for O<sub>2</sub> activation. *Curr. Opin. Chem. Biol.* **5**, 550–555 (2001).
- Price, J. C., Barr, E. W., Glass, T. E., Krebs, C. & Bollinger, J. M. Jr. Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine: $\alpha$ -ketoglutarate dioxygenase (TauD). *J. Am. Chem. Soc.* **125**, 13008–13009 (2003).
- Bugg, T. D. H. Dioxygenase enzymes: catalytic mechanisms and chemical models. *Tetrahedron* **59**, 7075–7101 (2003).
- Chang, W., Li, J., Lee, J. L., Cronican, A. A. & Guo, Y. Mechanistic investigation of a non-heme iron enzyme catalyzed epoxidation in (–)-4'-methoxycyclopentenolone biosynthesis. *J. Am. Chem. Soc.* **138**, 10390–10393 (2016).
- Song, X., Lu, J. & Lai, W. Mechanistic insights into dioxygen activation, oxygen atom exchange and substrate epoxidation by AsqJ dioxygenase from quantum mechanical/molecular mechanical calculations. *Phys. Chem. Chem. Phys.* **19**, 20188–20197 (2017).
- Su, H., Sheng, X., Zhu, W., Ma, G. & Liu, Y. Mechanistic insights into the decoupled desaturation and epoxidation catalyzed by dioxygenase AsqJ involved in the biosynthesis of quinolone alkaloids. *ACS Catal.* **7**, 5534–5543 (2017).
- Walsh, C. T., Haynes, S. W., Ames, B. D., Gao, X. & Tang, Y. Short pathways to complexity generation: fungal peptidyl alkaloid multicyclic scaffolds from anthranilate building blocks. *ACS Chem. Biol.* **8**, 1366–1382 (2013).
- Supekar, S., Gamiz-Hernandez, A. P. & Kaila, V. R. I. A protonated water cluster as a transient proton-loading site in cytochrome *c* oxidase. *Angew. Chem. Int. Ed.* **55**, 11940–11944 (2016).
- Ugur, I., Rutherford, A. W. & Kaila, V. R. I. Redox-coupled substrate water reorganization in the active site of photosystem II—the role of calcium in substrate water delivery. *Biochim Biophys Acta* **1857**, 740–748 (2016).
- Sharma, V. et al. Redox-induced activation of the proton pump in the respiratory complex I. *Proc. Natl Acad. Sci. USA* **112**, 11571–11576 (2015).
- Luo, Q., Boczek, E. E., Wang, Q., Buchner, J. & Kaila, V. R. I. Hsp90 dependence of a kinase is determined by its conformational landscape. *Sci. Rep.* **7**, 43996 (2017).



22. Solomon, E. I. et al. Geometric and electronic structure/function correlations in non-heme iron enzymes. *Chem. Rev.* **100**, 235–349 (2000).
23. Sahu, S. & Goldberg, D. P. Activation of dioxygen by iron and manganese complexes: a heme and nonheme perspective. *J. Am. Chem. Soc.* **138**, 11410–11428 (2016).
24. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132**, 154104 (2010).
25. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
26. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).
27. Klamt, A. & Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. 2*, 799–805 (1993).
28. Schäfer, A., Klamt, A., Sattel, D., Lohrenz, J. C. W. & Eckert, F. COSMO implementation in TURBOMOLE: extension of an efficient quantum chemical code towards liquid systems. *Phys. Chem. Chem. Phys.* **2**, 2187–2193 (2000).
29. Plessow, P. Reaction path optimization without NEB springs or interpolation algorithms. *J. Chem. Theory Comput.* **9**, 1305–1310 (2013).
30. Henkelman, G. & Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **113**, 9978–9985 (2000).
31. Weinan, E., Ren, W. & Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **66**, 052301 (2002).
32. Deglmann, P. & Furche, F. Efficient characterization of stationary points on potential energy surfaces. *J. Chem. Phys.* **117**, 9535–9538 (2002).
33. Ahlrichs, R., Bär, M., Häser, M., Horn, H. & Kölmel, C. Electronic structure calculations on workstation computers: the program system turbomole. *Chem. Phys. Lett.* **162**, 165–169 (1989).
34. Contreras-García, J. et al. NCIPLOT: a program for plotting noncovalent interaction regions. *J. Chem. Theory Comput.* **7**, 625–632 (2011).
35. Best, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_2$  dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
36. Brooks, B. R. et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
37. Riahi, S. & Rowley, C. N. The CHARMM-TURBOMOLE interface for efficient and accurate QM/MM molecular dynamics, free energies, and excited state properties. *J. Comput. Chem.* **35**, 2076–2086 (2014).
38. Essmann, U. et al. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
39. Vanommeslaeghe, K. et al. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
40. Valiev, M. et al. NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **181**, 1477–1489 (2010).
41. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
42. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
43. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
44. Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2184–2195 (2004).
45. Turk, D. MAIN software for density averaging, model building, structure refinement and validation. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1342–1357 (2013).
46. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

## Acknowledgements

This work was supported by the German Research Foundation (SFB1035 project B12 to V.R.I.K., SFB749 to M.G.) and by the 1974-01 TUM-KAUST agreement on selective C–H bond activation (A.B.). The authors are thankful for the computing time provided by the Leibniz Rechenzentrum. We thank the staff of the beamline X06SA at the Paul Scherrer Institute, SLS, Villigen (Switzerland) for assistance during data collection.

## Author contributions

V.R.I.K. and M.G. designed the project; S.L.M. and V.R.I.K. performed and analyzed all calculations; A.B. created mutants, purified proteins, and performed activity assays; A.B. and M.G. collected and analyzed X-ray data; V.R.I.K. and M.G. wrote the manuscript, with contributions from all authors.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-03442-2>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018






ARTICLE



<https://doi.org/10.1038/s41467-020-15050-0>

OPEN

# Conformational dynamics modulate the catalytic activity of the molecular chaperone Hsp90

Sophie L. Mader <sup>1,5</sup>, Abraham Lopez <sup>1,2,5</sup>, Jannis Lawatscheck<sup>1,5</sup>, Qi Luo<sup>1,3</sup>, Daniel A. Rutz <sup>1</sup>, Ana P. Gamiz-Hernandez<sup>1,4</sup>, Michael Sattler<sup>1,2</sup>, Johannes Buchner <sup>1</sup> & Ville R.I. Kaila <sup>1,4</sup>✉

The heat shock protein 90 (Hsp90) is a molecular chaperone that employs the free energy of ATP hydrolysis to control the folding and activation of several client proteins in the eukaryotic cell. To elucidate how the local ATPase reaction in the active site couples to the global conformational dynamics of Hsp90, we integrate here large-scale molecular simulations with biophysical experiments. We show that the conformational switching of conserved ion pairs between the N-terminal domain, harbouring the active site, and the middle domain strongly modulates the catalytic barrier of the ATP-hydrolysis reaction by electrostatic forces. Our combined findings provide a mechanistic model for the coupling between catalysis and protein dynamics in Hsp90, and show how long-range coupling effects can modulate enzymatic activity.

<sup>1</sup>Center for Integrated Protein Science Munich at the Department of Chemistry, Technical University of Munich, Lichtenbergstrasse 4, D85748 Garching, Germany. <sup>2</sup>Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, Neuherberg 85764, Germany. <sup>3</sup>Soft Matter Research Center and Department of Chemistry, Zhejiang University, Hangzhou 310027, China. <sup>4</sup>Department of Biochemistry and Biophysics, Stockholm University, SE-10691 Stockholm, Sweden. <sup>5</sup>These authors contributed equally: Sophie L. Mader, Abraham Lopez, Jannis Lawatscheck. ✉email: [ville.kaila@dbb.su.se](mailto:ville.kaila@dbb.su.se)

The heat shock protein 90 (Hsp90) is an ATP-dependent molecular chaperone that controls protein maturation and folding, in addition to central regulatory functions of the eukaryotic cell<sup>1–3</sup>. Stringent clients of Hsp90 include 60% of the human kinome and the chaperone is therefore of significant biomedical interest. This 90 kDa, highly conserved, homodimeric enzyme comprises an N-terminal domain (NTD), a middle domain (M-domain), and a C-terminal domain (CTD) (Fig. 1)<sup>4</sup>. During its chaperone cycle, Hsp90 undergoes large-scale conformational changes from an open inactive state to a closed active state, which induce association of the M-domains and NTDs<sup>5,6</sup>. The closing exposes important client-binding regions, as shown in recent cryo-EM structures of Hsp90 in complex with co-chaperones and clients<sup>7,8</sup>. The ATP-binding site of Hsp90 is located in the NTD, and it is completed by residues from the M-domain, including an arginine (Arg-380 in yeast Hsp90) that forms contacts with the  $\gamma$ -phosphate of ATP<sup>9</sup>. ATP-binding triggers large conformational changes in the global Hsp90 structure, stabilizing the compact closed conformation of the enzyme<sup>9</sup>.

The thermodynamic driving force for this global conformational change originates from the tightly coupled slow ATPase reaction on  $\sim 0.1$ – $1$  min<sup>−1</sup> timescales<sup>10</sup>. Elnatan et al.<sup>11</sup> further showed that ATP hydrolysis in the two protomers is sequential and deterministic in the mitochondrial Hsp90, TRAP1, with different hydrolysis rates for each protomer. After the first hydrolysis step, the Hsp90 dimer was suggested to undergo a flip into a closed asymmetric state, followed by a second hydrolysis event. Site-directed mutagenesis experiments identified several residues involved in the ATPase reaction: Glu-33 of the yeast Hsp90 is important for the ATP-hydrolysis reaction, as the E33A variant is able to bind ATP, but not to hydrolyse it<sup>10,12,13</sup>. Moreover, the R380A variant has a substantially lowered ATPase activity in comparison with the wild-type (WT) enzyme<sup>4,13</sup> and N37A prevents Hsp90 from binding ATP due to loss of the catalytically important Mg<sup>2+</sup> ion (Fig. 1)<sup>12</sup>. Furthermore, as for many other ATPases<sup>14–16</sup>, a water molecule is required for the ATP-hydrolysis reaction<sup>9</sup>. However, despite enormous efforts to study Hsp90 by structural, biochemical, biophysical, and computational approaches in the last decades<sup>17–22</sup>, the specific functions as well as the overall catalytic mechanism and its coupling to the conformational state of Hsp90 still remain unclear.

To probe the energetics, dynamics, and molecular mechanism of the ATP-hydrolysis reaction and its coupling to the conformational changes in Hsp90, we integrate here multi-scale computational simulations with biophysical experiments. We probe the coupling between the conformational dynamics of

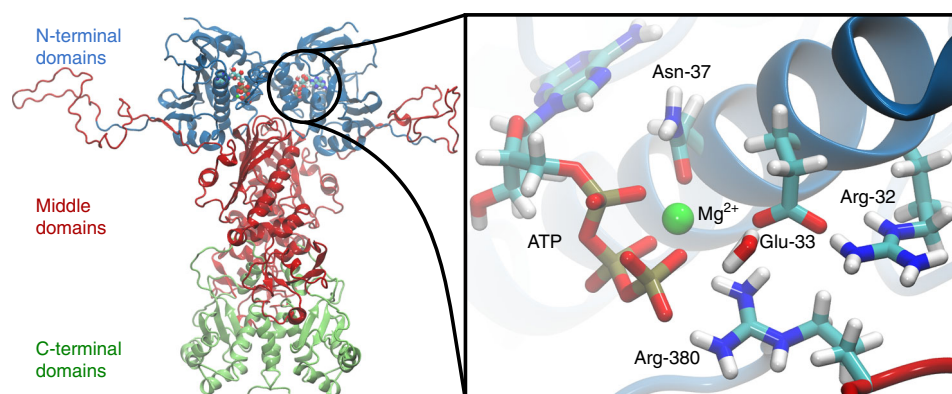
Hsp90 and its catalytic mechanism using hybrid quantum/classical (QM/MM) free-energy calculations in combination with large-scale atomistic molecular dynamics (MD) simulations. Our mechanistic models are further probed by site-directed mutagenesis experiments in combination with Förster-resonance energy transfer and nuclear magnetic resonance (NMR) spectroscopy, as well as by small-angle X-ray scattering (SAXS) measurements.

## Results

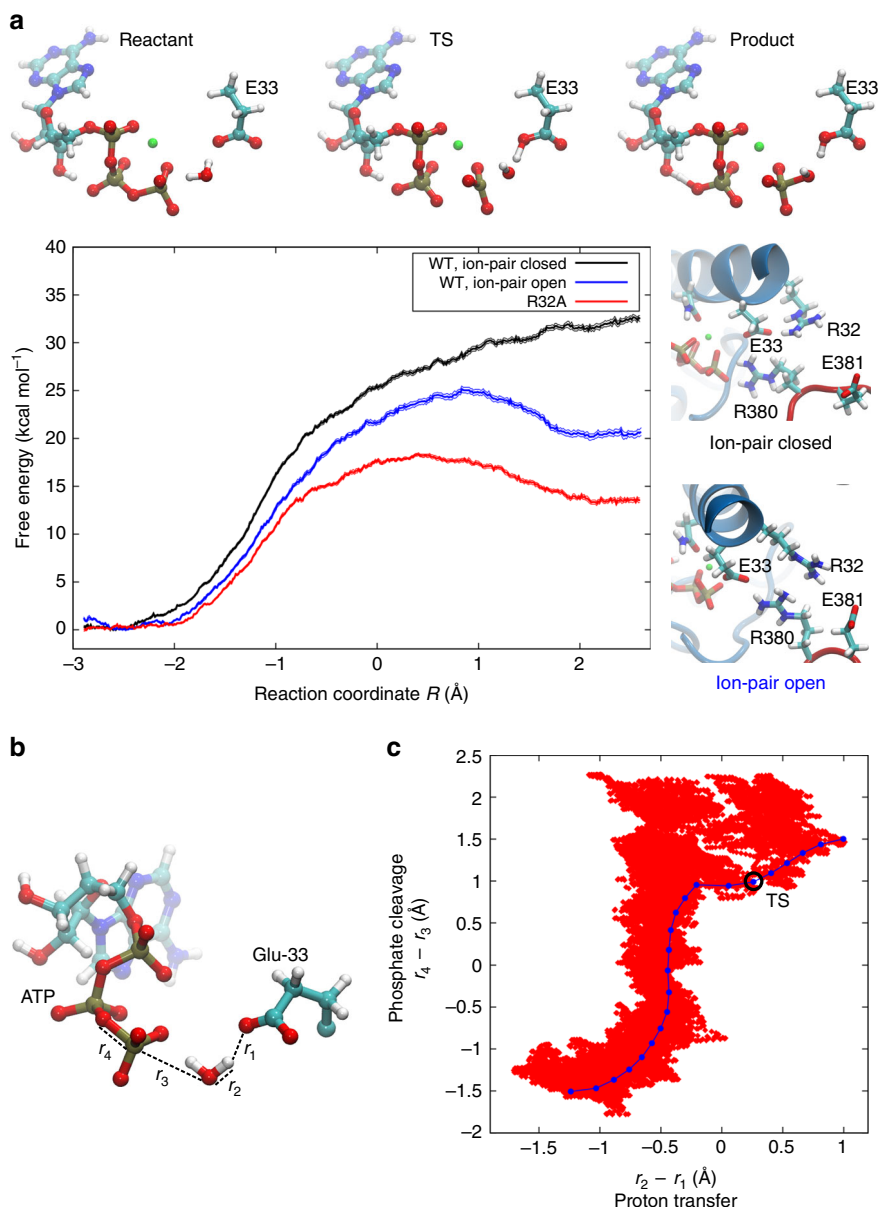
**Catalytic mechanism of ATP hydrolysis in Hsp90.** To study the bond-formation/bond-breaking energetics during catalysis within the active site of Hsp90, we created hybrid QM/MM models, where the electronic structure of the chemically reacting system is computed on-the-fly, based on quantum mechanical density functional theory (DFT), and polarized by a classical force field model of the remaining protein framework. These QM/MM calculations were used to predict a free-energy surface for the ATP-hydrolysis reaction, by dynamically sampling conformational changes within the protein during the chemical transformation (Supplementary Fig. 1).

Our QM/MM free-energy calculations suggest that the ATP-hydrolysis reaction is initiated by proton transfer between a water molecule, coordinating the ATP  $\gamma$ -phosphate, and Glu-33 (Fig. 2a, b). This water molecule is not present in the resolved X-ray structures of Hsp90<sup>9</sup>, but it originates from the bulk water by spontaneously coordinating between the  $\gamma$ -phosphate and Glu-33 during the atomistic MD simulations. After the water deprotonation reaction, the transient OH<sup>−</sup> species immediately attacks the  $\gamma$ -phosphate, resulting in the cleavage of the  $\beta/\gamma$ -phosphate bond. The computed reaction free-energy profiles predict that the proton transfer and phosphate cleavage processes take place semi-concertedly rather than by associative or dissociative reaction steps (Fig. 2c)<sup>23</sup>. The calculations also suggest that the transition state comprises a penta-coordinated  $\gamma$ -phosphate, with elongated axial bonds, resembling the transition states found in other ATPases (Fig. 2a)<sup>24–26</sup>.

Interestingly, despite the chemically feasible reaction intermediates, the QM/MM calculations performed on a model of Hsp90 extracted from the experimental X-ray structure, suggest that the free-energy barrier for the hydrolysis reaction is  $>30$  kcal mol<sup>−1</sup> (Fig. 2a black trace, Supplementary Fig. 2, and Supplementary Table 1), which implies that the reaction rate is kinetically inaccessible (see Methods). Similar reaction energetics are obtained using different DFT approximations in the quantum chemical calculations, suggesting that the results are robust (Supplementary Table 2). Endergonic free-energy profiles have



**Fig. 1** Structure of the yeast Hsp90 dimer (PDB ID: 2CG9)<sup>9</sup>. The N-terminal domains of Hsp90 are shown in blue, the middle domains in red, and the C-terminal domains in green. The inset shows a structure of the active site with a bound ATP molecule obtained from an MD simulation, where Asn-37 undergoes a rotation to form a stronger coordination to the magnesium.

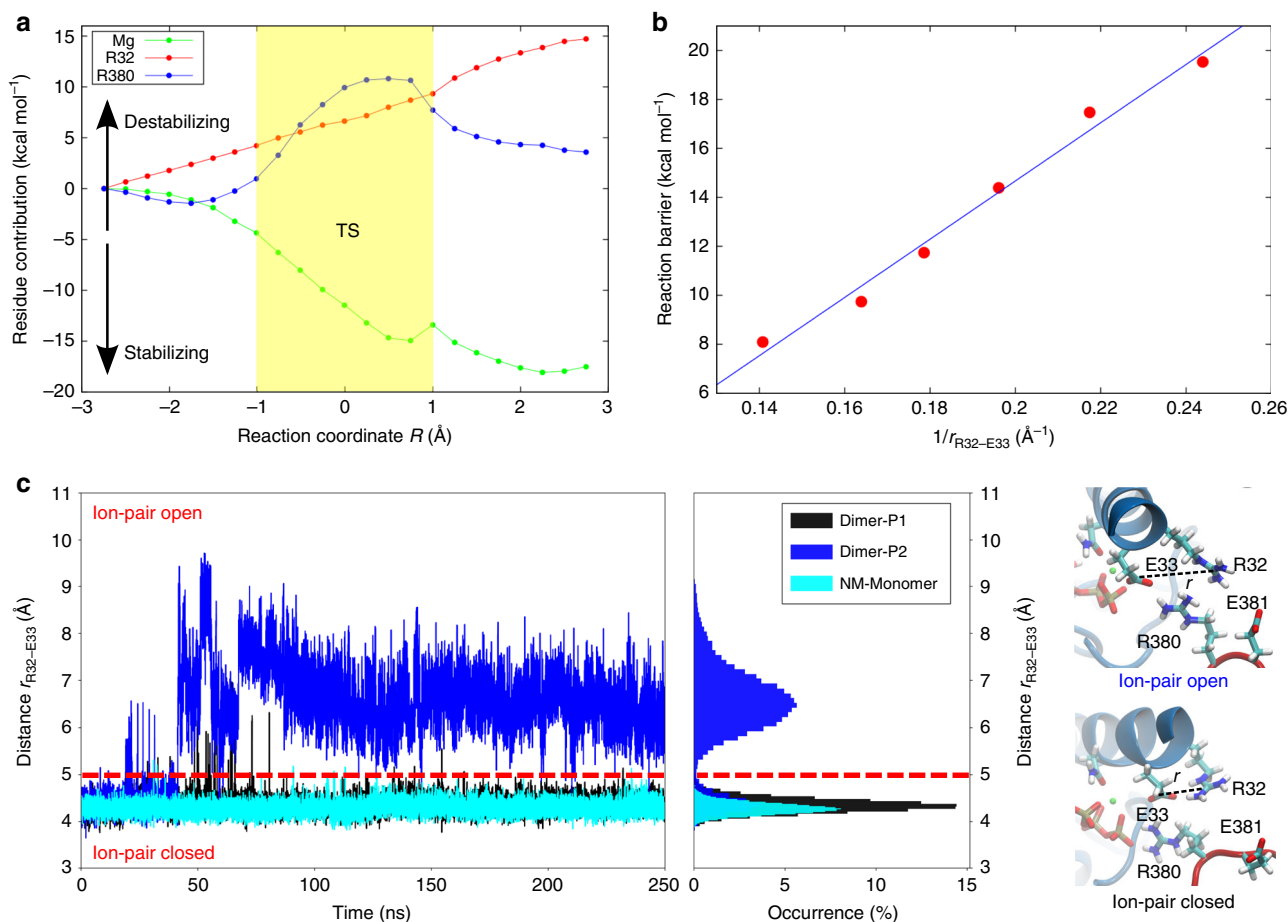


**Fig. 2 Energetics and dynamics of ATP hydrolysis in Hsp90.** **a** Reactant, transition state (TS), and product structures showing ATP,  $\text{Mg}^{2+}$  (in green), and the sidechain of Glu-33 (E33), extracted from QM/MM calculations of the ATP-hydrolysis reaction. QM/MM free-energy profiles were calculated with the Arg-32/Glu-33 ion pair closed (black,  $r_{\text{R32-E33}} < 5$  Å) and open (blue,  $r_{\text{R32-E33}} > 5$  Å, see Fig. 3), as well as for the R32A mutant (red). **b** The reaction coordinate used for QM/MM calculations is  $R = r_4 - r_3 + r_2 - r_1$ , a linear combination of distances between Glu-33, the attacking water molecule, and the  $\gamma$ -phosphate of ATP.  $r_2 - r_1$  is the difference between the bond-breaking ( $r_2$ ) and bond forming ( $r_1$ ) distances for the proton transfer from the water molecule to Glu-33.  $r_4 - r_3$  is the difference between the bond-breaking ( $r_4$ ) and bond forming ( $r_3$ ) distances for the phosphate cleavage<sup>27,28</sup>. The reaction coordinate was optimized from reactants ( $R = -2.9$  Å) to products ( $R = 2.6$  Å). **c** Semi-concerted ATP-hydrolysis mechanism from QM/MM free-energy calculations (red dots) and during reaction path optimization (blue dots), showing the sampled reaction coordinates with the Arg-32/Glu-33 ion pair closed. The transition state (TS) of the reaction path optimization is marked with a black circle.

also been obtained in other QM/MM studies on phosphate cleavage processes<sup>27,28</sup>, whereas entropic effects are likely to stabilize the product state upon  $\text{P}_i$  dissociation up to ca. 17 kcal mol<sup>-1</sup> (Supplementary Table 3), which could be transduced into driving conformational changes in Hsp90. Moreover, the free-energy profile does not show a clear minimum for the product state, indicating that the crystal structure may not represent a hydrolytically fully active state.

To elucidate the molecular basis for this high barrier, we decomposed the energetic contributions of individual residues by switching off their interactions with the active site region during the calculations (Fig. 3a and Supplementary Fig. 3). These

calculations suggest that the high reaction barrier associated with the proton transfer reaction originates from the low proton affinity of Glu-33, which in turn, results from the stabilizing interaction with the nearby Arg-32, favouring the deprotonated form of the glutamate. Calculations based on DFT models show that the reaction barrier has a coulombic 1/distance-dependence on the Arg-32/Glu-33 sidechain distance, suggesting that the reaction is electrostatically tuned (Fig. 3b and Supplementary Fig. 4). The analysis further shows that the  $\text{Mg}^{2+}$  ion pulls away electron density from the  $\gamma$ -phosphate, lowering the reaction barrier, whereas Arg-380 stabilizes the  $\text{OH}^-$  group attacking the  $\gamma$ -phosphate, as well as the resulting phosphate ( $\text{HPO}_4^{2-}$ )



**Fig. 3 Analysis of the catalytic effects in Hsp90.** **a** Energy decomposition of the catalytic effects. The figure shows the energy difference relative to the wild-type calculation, obtained by switching off individual residue contributions, with the transition state (TS) region marked in yellow (see also Supplementary Fig. 3). **b** Reaction barrier for ATP hydrolysis at different Arg-32/Glu-33 distances (see also Supplementary Fig. 4). The energy barrier is ca. 20 kcal mol<sup>-1</sup> in the DFT models and thus slightly lower as compared with the QM/MM models with explicit treatment of the protein surroundings. **c** Arg-32/Glu-33 distance from MD simulations of the full-length Hsp90 dimer (in blue and black) and for the monomeric NM-domain model of Hsp90 (in cyan). All Arg-32/Glu-33 distances are measured between the C $\zeta$  (Arg-32) and C $\delta$  (Glu-33).

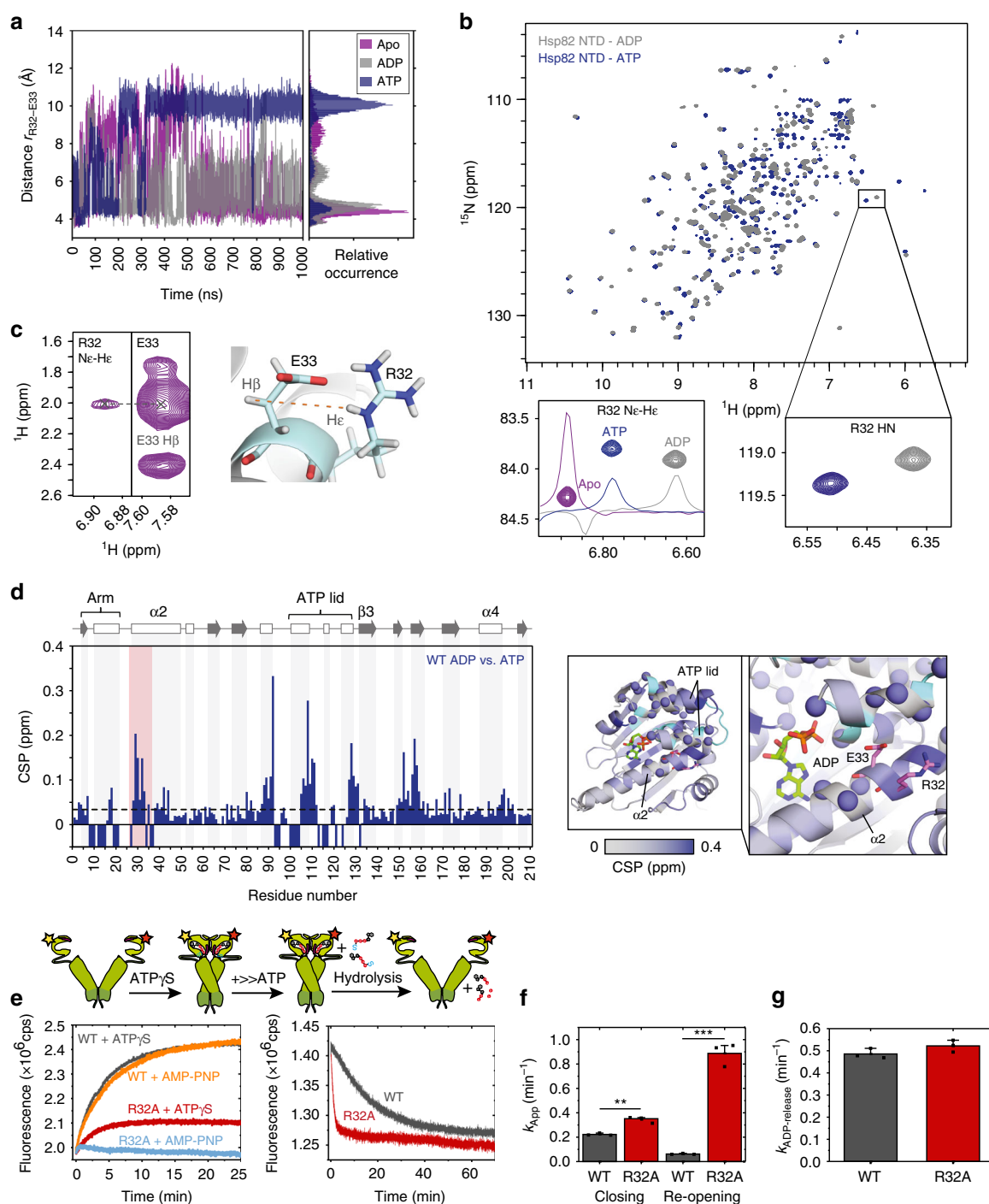
product (Fig. 3a, negative curvature for the R380 profile at  $R < -2.5$  Å and  $R > 1$  Å). Arg-380 could thus be central for stabilizing the transient hydroxide species, but is not directly involved in the water deprotonation reaction. These findings are consistent with the kinetically slow but viable R380A mutant<sup>29</sup>.

**Protein conformational changes modulate catalytic activity.** To probe the conformational dynamics linked with the catalytic steps, we performed atomistic MD simulations for 250 ns of both the full-length Hsp90 dimer and a monomeric NM-domain model, as well as 1  $\mu$ s MD simulations for models comprising only the NTD in different ligand states (Supplementary Figs. 5 and 6). Interestingly, we observe that Arg-32 undergoes a conformational change in the full-length Hsp90. In these simulations, Arg-32 flips away from Glu-33 and forms transient contacts with a network of charged/polar residues in the M-domain on 50–100 ns timescales (Fig. 3c and Supplementary Fig. 7). Opening of the Arg-32/Glu-33 ion pair increases the pK<sub>a</sub> value of Glu-33 by ca. 10 pK<sub>a</sub> units, indicating a strong influence of Arg-32 on the pK<sub>a</sub> of Glu-33 (Supplementary Fig. 8). Although we do not expect to achieve exhaustive sampling of these transient states, the electrostatic models, nevertheless, also support that the proton affinity of Glu-33 strongly increases upon the ion-pair dissociation, consistently with the QM/MM results (Fig. 2a). In stark contrast

to simulations of the full-length Hsp90, the Arg-32/Glu-33 ion pair remains closed during the MD simulations of our monomeric NM-domain model, built from a high-resolution NTD-dimerized crystal structure (see Methods), suggesting that the dimerization and global conformational state of Hsp90 affects the ion-pair dynamics. However, in simulations of the monomeric NTD, we find that replacing ADP by ATP triggers opening of the Arg-32/Glu-33 ion pair, mimicking the conformational dynamics of this region observed in the full-length Hsp90 (Fig. 4a). Analysis of the Arg-32/Glu-33 sidechain distance in a selection of Hsp90 crystal structures shows that this ion pair is closed in most of the structures, except in an ATP-bound structure of the NTD, supporting the results from our MD simulations (Supplementary Table 4). In simulations of the full-length Hsp90, we observe ion-pair opening only in one of the two Hsp90 protomers (Fig. 3c), but a systematic exploration of the phase space would be required to determine whether the domains truly operate asymmetrically<sup>11</sup>. However, to probe the reproducibility of our simulations, we performed 100 ns duplicate MD simulations for the NM-domain and the NTD models. These simulations support similar ion-pair dynamics as observed in the original simulations (Supplementary Fig. 9).

To probe how the ATP-hydrolysis reaction is affected by opening of the Arg-32/Glu-33 ion pair, we re-calculated the





**Fig. 4 Biophysical characterization of Hsp90.** **a** Arg-32/Glu-33 distance from classical atomistic MD simulations, showing that the ion pair is closed in the apo-form (purple) of the Hsp90-NTD and its conformation is sensitive to nucleotide binding (ADP in grey/ATP in blue). **b**  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra of the Hsp90-NTD bound to ADP (grey) and ATP (blue), with close-ups of the resonances corresponding to the Arg-32 amide group (bottom right) and its sidechain He-Ne (bottom left), for which the  $^1\text{H}$  traces are also included (apo shown in purple). **c** NOE cross-peak between Arg-32 He and Glu-33 H $\beta$  in the apo-NTD, confirming the spatial proximity of the sidechains. **d** Left: chemical shift perturbations (CSPs) of the Hsp90-NTD bound to ADP and ATP. Negative bars represent residues that could not be assigned due to conformational dynamics. Secondary structure elements are shown on the top and altered residues next to Arg-32 are highlighted in red. Right: distribution of CSP on the NTD for ATP/ADP exchange. Amides of residues that are not visible due to conformational dynamics are represented by blue spheres, whereas unassigned residues are shown in cyan. **e** FRET experiments show that closing (left) in Hsp90-R32A (in red and light blue) takes place to a smaller extent in comparison with the wild type (WT, in black and orange). Right: re-opening of the closed dimer takes place faster in Hsp90-R32A (red) than in WT Hsp90 (black). The schematic figure on the top illustrates the experimental procedure. **f** Rates for closing and re-opening of WT Hsp90 and Hsp90-R32A obtained from mono-exponential fitting of the FRET curves shown in **e**. Error bars represent SD from three independent measurements ( $n = 3$ ), shown as black dots. Statistical significance was assessed using a two-sided two-sample  $t$ -test and a level of significance of 0.01 (\*\*) and 0.001 (\*\*\*). The fit yields a higher closure rate for R32A relative to WT Hsp90. **g** ADP-release assays show that WT Hsp90 and Hsp90-R32A have similar global ATP turnover. Error bars represent SD from three independent measurements ( $n = 3$ ), shown as black dots. Source data are provided as a source data file.

reaction energetics for this conformation. Although the ATP-hydrolysis reaction still takes place by a semi-concerted proton transfer and phosphate cleavage process in these models (Supplementary Fig. 10), the free-energy barrier is reduced to ca. 25 kcal mol<sup>-1</sup> (Fig. 2a), consistent with electrostatic tuning of Glu-33. According to transition state theory, this barrier corresponds to a reaction rate in the hours timescales (see Methods). Moreover, substitution of the arginine by an alanine, R32A, mimicking a fully shielded and extended open ion-pair conformation, shows a free-energy barrier of ca. 18 kcal mol<sup>-1</sup> (Fig. 2a), corresponding to a reaction rate in the seconds timescales according to transition state theory ( $k = 1.3 \text{ s}^{-1}$ , see Methods), and is thus comparable to experimentally observed turnover rates of Hsp90 in the ~0.1–1 min<sup>-1</sup> timescales<sup>10</sup>. MD simulations of the full-length R32A suggest that the substitution leads to re-arrangement of ion pairs within the N–M interface (Supplementary Fig. 11). The R32A variant shows indeed more ion pairs between the two M-domains in comparison with the WT Hsp90, whereas less ion pairs and a weaker interaction energy is observed between the two NTDs of R32A in the MD simulations (Supplementary Fig. 12). Our findings thus suggest that the conformational changes associated with the contact between the NTD and M-domain strongly modulate the ATP-hydrolysis energetics (see below).

**Probing the conformational state of Arg-32.** To gain insight into the conformational changes of Arg-32 in the NTD of Hsp90, we employed NMR spectroscopy in combination with MD simulations in different nucleotide-bound states. Based on the observation that replacing ADP by ATP may trigger opening of the Arg-32/Glu-33 ion pair (Fig. 4a), we measured <sup>1</sup>H-<sup>15</sup>N heteronuclear single quantum coherence (HSQC) spectra of the NTD of yeast Hsp90 in the apo-form and bound to ADP and ATP. The Nε–Hε proton provides a good reporter for the side-chain of Arg-32, which appears in the apo-NTD as a single intense NMR signal, as expected for a sidechain with restricted mobility (Fig. 4b, bottom left panel). The presence of a Nuclear Overhauser effect (NOE) cross-peak between Arg-32 Hε and Glu-33 Hβ confirms the spatial proximity of the two sidechains in the apo-form (Fig. 4c), consistent with our MD simulations of the NTD in the apo state (Fig. 4a). Structural rigidity of the Arg-32 sidechain is supported by {<sup>1</sup>H}-<sup>15</sup>N steady-state heteronuclear NOE experiments, which are sensitive to fast (ns–ps) dynamics<sup>30</sup>. However, a relatively high value of  $0.58 \pm 0.04$  indicates that the Arg-32 sidechain is not flexible, consistent with its interaction with Glu-33.

Interestingly, we find that nucleotide exchange induces significant chemical shift perturbations (CSPs) near the Arg-32 region (Fig. 4b, d) and in the ATP-lid, which is partially similar to previous results by Zhang et al.<sup>31</sup> for human Hsp90, although the yeast protein studied here has several features that significantly differ from the human isoform. For example, in contrast to yeast Hsp90 (Fig. 4b, c), the neighbouring residues around the Arg-46 (Arg-32) region could not be assigned in the human Hsp90<sup>31</sup>, possibly due to signal broadening originating from dynamics on the micro- to millisecond timescales.

CSP around the Arg-32 residue could reflect a higher population of an open ion-pair conformation in the ATP-bound NTD, as predicted by the MD simulations (Fig. 4a). This notion is further supported by the sidechain heteronuclear NOE values, which decrease from  $0.58 \pm 0.09$  for the ADP-bound NTD, to  $0.48 \pm 0.09$  for the ATP-bound NTD. In addition, the increased line-width of the Hε signal observed in the ADP- and ATP-bound states (Fig. 4b) is consistent with dynamics at μs–ms timescales. Breaking the ion pair in the apo state of the NTD-R32A variant

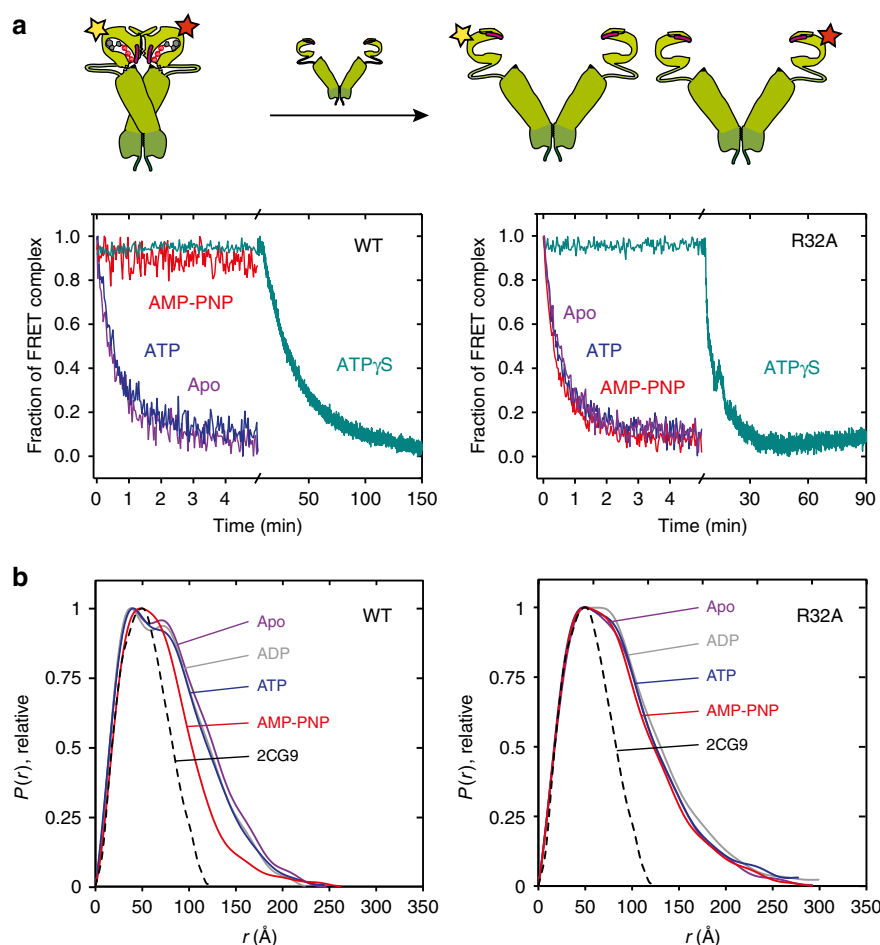
causes significant shifts not only in the environment of the mutation point but also at the C-terminal helix α4 and in the ATP lid, emphasizing its role on the structure and conformation of the NTD (Supplementary Fig. 13).

**Effects of the R32A mutation on Hsp90 conformation and function.** To probe further how the computationally identified R32A substitution influences the global conformational structure and dynamics of Hsp90, we employed a FRET system with the NTDs labelled by fluorescent donor and acceptor dyes that induce resonance energy transfer upon closing of the Hsp90 dimer<sup>6</sup>. Although our current MD simulations and NMR data capture short- to medium-range conformational changes in the protein structure, FRET and SAXS provide powerful techniques to capture global effects linked with the R32A substitution, providing complementary data about the Hsp90 structure and function.

Our FRET experiments on the full-length Hsp90 show that the R32A substitution strongly inhibits the closing of the dimer (Fig. 4e). Interestingly, the non-hydrolysable ATP-analogue, adenylyl-imidodiphosphate (AMP-PNP), and the slowly hydrolysable ATPγS, induce a similar closing behaviour in the WT Hsp90, whereas the R32A variant forms only a partially closed FRET dimer with ATPγS. At the same time, re-opening of the R32A dimer takes place faster as compared with the WT Hsp90 (Fig. 4e). Although R32A-Hsp90 does not close to the same extent as the WT Hsp90 dimer, exponential fitting of the FRET data gives a higher closure rate for R32A-Hsp90 as compared with the WT (Fig. 4f). A possible explanation for this observation is that the partially closed R32A-Hsp90 dimer is formed faster than the fully closed WT Hsp90 dimer. This hypothesis is supported by a ca. 10-fold faster re-opening of the R32A-Hsp90 dimer relative to the WT Hsp90 dimer (Fig. 4e, f). The ratio of the apparent rates (Supplementary Table 5) is consistent with a lower population of the closed conformation in R32A and thus indirectly supports a similar or even lowered intrinsic ATP-hydrolysis barrier in the R32A variant relative to WT Hsp90.

In chase experiments (Fig. 5a), where we used unlabelled yeast Hsp90 (Hsp82) to disrupt the preformed FRET dimer<sup>6</sup>, the WT Hsp90 shows no subunit exchange from its closed compact conformation in the presence of AMP-PNP, in line with previous observations<sup>6,13</sup>. In stark contrast, the R32A variant has a significant subunit exchange, consistent with the perturbed closing equilibrium of the mutant. These findings suggest that the closed R32A dimer is not stabilized as strongly as the closed dimer of WT Hsp90. The perturbed closing also affects the co-chaperone binding of Sba1, which is important for the fine-tuning of the chaperone cycle and stabilizes the closed conformation (Supplementary Fig. 14)<sup>32</sup>. To this end, Sba1 shows a lower binding affinity for the R32A variant as compared with the WT Hsp90.

To obtain further structural information of the R32A variant, we performed in-line size-exclusion chromatography coupled to SAXS experiments. The analysis of the scattering profiles with the Guinier approximation and the pair-wise distance distribution functions ( $P(r)$ ) shows that the apo-form of both the WT and R32A variant of the full-length Hsp90, populate an ensemble of a fully open conformation with a radius of gyration ( $R_g$ ) of ca. 62–66 Å (Fig. 5b, Supplementary Fig. 15, and Supplementary Tables 6 and 7). Nucleotide binding to the WT enzyme induces an overall decrease in the  $R_g$  to 59.4 Å and 57.6 Å for the ADP- and ATP-bound states, respectively, leading to a more compact conformation, consistent with previous observations<sup>33</sup>. However, in stark contrast to the WT enzyme, ligand binding triggers no significant structural changes in the R32A variant, suggesting that



**Fig. 5** Characterization of the Hsp90 dimer by FRET and SAXS experiments. **a** Chase experiments with the closed wild type (WT, left) and R32A (right) Hsp90 dimers in the apo-form and bound to different nucleotides (ATP, AMP-PNP, ATP $\gamma$ S). Chase experiments were initiated by addition of unlabelled Hsp90. **b** Left: pair distance distributions ( $P(r)$ ) obtained from the SAXS profiles of apo wild-type (WT) Hsp90 (purple) and bound to different nucleotides (ADP, ATP, AMP-PNP). The  $P(r)$  distribution calculated from the X-ray structure of Hsp90 in the closed conformation (PDB ID: 2CG9)<sup>9</sup> is shown for comparison (in black). Right: pair distance distributions of the R32A variant in different nucleotide states. AMP-PNP favours a compact conformation of wild-type (WT) Hsp90, whereas the R32A variant shows no major conformational rearrangements.

the coupling between the NTD and M-domain is drastically perturbed, thus affecting the overall conformational state of the Hsp90 dimer. The involvement of Arg-32 in the closing process is particularly pronounced upon binding of the non-hydrolysable AMP-PNP, which induces a compact Hsp90 conformation in the WT enzyme ( $R_g = 53.1$  Å), reducing the inter-atomic distance distribution, but leaves the global conformation of R32A largely unaffected ( $R_g = 62.3$  Å; Fig. 5b and Supplementary Tables 6 and 7). These observations support the notion that the free energy from the nucleotide binding site is not effectively transduced into the conformational state in the R32A variant.

To probe the influence of the R32A substitution *in vivo*, we used a plasmid shuffling approach to introduce the Hsp90 variant as the sole source of Hsp90 in the yeast *Saccharomyces cerevisiae*, where Hsp90 is essential for cell survival. Interestingly, the shuffling experiments show that the R32A strain is not viable (Supplementary Fig. 16), possibly due to the impaired closing dynamics, which is crucial for the maturation and regulation of client proteins, suggesting that the R32A variant cannot fulfil its tasks in the cell. A possible explanation for this observation is that R32A does not adopt the fully closed conformation, which is necessary for the activation of client proteins<sup>7,9</sup>.

As our calculations predict a lower catalytic barrier for the ATP-hydrolysis in R32A-Hsp90, we tested the interaction of

R32A-Hsp90 with ATP. The R32A variant binds ATP to a similar extent as the WT Hsp90 (Supplementary Fig. 17) and is able to hydrolyse ATP, consistent with its lowered catalytic barrier (Fig. 2a), but remarkably, the ADP-release kinetics in the R32A variant ( $0.55 \pm 0.05$  min<sup>-1</sup>) is practically identical to that of the WT Hsp90 ( $0.50 \pm 0.05$  min<sup>-1</sup>) (Fig. 4g). The NM-fragment of R32A-Hsp90 does not, however, show ATPase activity, which is also lacking from the NM construct of the WT protein (Supplementary Fig. 18). Moreover, the coupling between the NTD and the M-domain in the full-length Hsp90 is indirectly supported by our experiments using the E381Q variant, which show an increased closing rate and ATPase activity relative to the WT enzyme (Supplementary Fig. 19). We speculate that the E381Q substitution could, however, also affect the conformation of R380, which has been linked to formation of the “closed 2” state<sup>13,29</sup>.

Interestingly, despite the perturbed closing dynamics, we observe that the R32A variant shows similar ADP-release kinetics as the WT Hsp90. To find an explanation for the seemingly contradicting experimental results, we constructed a simplified kinetic model of the Hsp90 cycle (Supplementary Fig. 20 and Supplementary Table 8). Although the intrinsic ATPase rate is expected to increase as a result of the R32A substitution, our kinetic models suggest that the overall ADP-release rate could



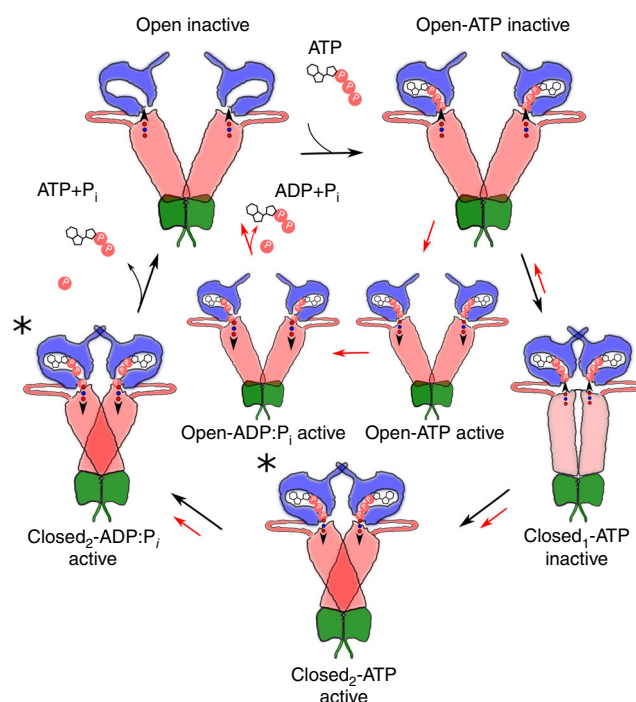
remain unchanged if the population of the closed conformation is reduced at the same time, as also suggested by our FRET data (Fig. 4e, f). The kinetic models suggest that even a 100-fold rate increase in the catalytic step, would require only a two- to threefold reduction in formation of the fully closed conformation to account for the observed ADP-release data. In other words, although the intrinsic ATPase reaction is strongly modulated by the conformational state of Hsp90, the less efficient population of the closed conformation could account for the unchanged ADP-release rate. The R32A variant is, however, likely to undergo a more complex reaction cycle than that modelled here, but elementary rates for the transitions remain unknown. Nevertheless, taken together, our combined findings suggest that the Arg-32 site has drastic effects on both the local and global conformational dynamics of Hsp90, which in turn is directly coupled to the reaction barrier for the ATP-hydrolysis process.

## Discussion

We have shown here that the catalysis leading to the hydrolysis of a phospho-anhydride bond within the active site of Hsp90 is modulated by both local and global conformational changes in the protein. Based on molecular simulations, we identified that Arg-32 within the active site functions as a key coupling element that mediates the communication across the different domains of Hsp90. The ATPase reaction barrier is electrostatically tuned by conformational transitions in Arg-32 and, due to microscopic reversibility, perturbation of the catalytic machinery also results in an impaired global closing behaviour that is central for the chaperone cycle. Our data indicate that the conformational switching takes place via a conserved network of ion pairs between the N- and M-domains of the enzyme involving, e.g., E381 (Supplementary Figs. 7 and 19), with mechanistic similarities to, e.g., respiratory complex I, a mitochondrial redox-driven proton pump where quinone reduction triggers proton pumping, up to 200 Å from the active site, by conformational changes in a network of conserved ion pairs<sup>34–37</sup>.

It is puzzling that although the R32A variant is not viable *in vivo* and unable to properly form a closed compact state, the enzyme still hydrolyses ATP with an unchanged ADP-release rate. The catalytic reaction is not considered rate-limiting for the WT Hsp90 machinery, but is normally limited by the closing kinetics<sup>6,38</sup>. However, our calculations showed that the ATPase reaction has a significant barrier also in the WT Hsp90, especially prior to conformational changes within Arg-32 that are captured by both our MD simulations and NMR experiments. In contrast to the WT enzyme, our R32A variant has a significantly perturbed closing equilibrium, but it is still able to hydrolyse ATP with an overall unchanged cycle turnover. These findings indicate that Arg-32 functions as a switching point, the perturbation of which decouples the catalysis from global conformational changes. To this end, our kinetic models support that the ADP-release rate could indeed remain unchanged if the closing kinetics is perturbed at the same time (Supplementary Fig. 20), as also supported by our FRET experiments (Fig. 4e, f).

The R32A variant could, however, also run another, fully decoupled cycle in parallel to the regular Hsp90 cycle, where the ATP-hydrolysis reaction is independent of the formation of a closed compact state (Fig. 6). Binding of the co-chaperone Sba1 to R32A supports that the closed conformation is populated at least to some extent in the mutant (Supplementary Fig. 14) and kinetic competition between these cycles is thus consistent with the non-viable, but catalytically active R32A variant. This suggests that R32A-Hsp90 could hydrolyse ATP in the (partly) open state, suggesting that this variant might also have an increased ATPase



**Fig. 6 Putative Hsp90 reaction cycle.** ATP-binding to the open, catalytically inactive state, induces formation of a compact ATP-bound state (closed<sub>1</sub>-ATP inactive). Conformational changes (red/blue dots followed by arrow) within the NTD and M-domain lead to a catalytically active state (active) that triggers formation of a fully compact active form of Hsp90 (closed<sub>2</sub>-ATP active), with chaperone activity (marked with an asterisk). The hydrolysis reaction leads to splitting of the phospho-anhydride bond and formation of the closed<sub>2</sub>-ADP:P<sub>i</sub> state, which further dissociates to the open state. The R32A decoupling variant lowers the barrier for switching between inactive and active catalytic states that also lowers the ATP-hydrolysis barrier, but increases the free energy of forming the compact state (red arrows). The decoupling mutant could also run an ATP-hydrolysis cycle in parallel to the regular cycle that is fully decoupled from the formation of the closed form (red arrows). The NTD is shown in blue, the middle domain in red, and the CTD in green.

activity in an NM-fragment. We were, however, unable to detect ATP catalysis in the NM-fragment in either the WT or R32A variant (Supplementary Fig. 18), suggesting that the NM construct might sample different conformations than the full-length Hsp90 dimer, as dimerization initiated at the CTD is most likely required for correctly organizing the active site.

In conclusion, by using an integrated computational and experimental approach, we showed here how the catalytic activity couples to conformational changes in the molecular chaperone Hsp90. We identified a central ion pair that switches between the domain harbouring the active site and the M-domain of the enzyme, mediating long-range signals across the protein framework. Conformational changes in this ion pair favour the ATP-hydrolysis reaction by electrostatic tuning. The computationally derived R32A variant was experimentally expressed and its properties were probed by NMR, SAXS, and FRET experiments. The R32A mutation impedes the formation of a compact Hsp90 dimer conformation that, in turn, is a pre-requisite for the chaperone activity. Substitution of the identified site decouples catalysis from global conformational changes, leading to a non-viable mutant that hydrolyses ATP without activating the chaperone cycle as indicated by our *in vivo* experiments. These

results open up possibilities to unravel the exact molecular principles on how the energy from the ATPase/closed state is transduced into the chaperone activity in Hsp90. Our combined data thus highlight new features of the remarkable coupling between catalysis and biological activity in Hsp90.

## Methods

**DFT and QM/MM models of the active site in Hsp90.** Monomeric QM and QM/MM models were constructed based on the crystal structure of Hsp90 from *Danio rerio* (PDB ID: 4IVG)<sup>39</sup>. The calculations were performed at the B3LYP-D3/CHARMM36 (QM/MM) and B3LYP-D3 (QM) level using def2-SVP/def2-TZVP (Mg) basis sets<sup>40–43</sup>. The QM region comprised ATP, Mg<sup>2+</sup>, Arg-32, Glu-33, Asn-37, Ser-99, and Arg-380, as well as six water molecules, in addition to the backbone of residues 118–124 that was included only in the QM models (see Supplementary Table 9 for residue numbering in different species). Link atoms were introduced between the C $\beta$  and Ca atoms in the QM/MM models, whereas terminal carbon atoms were kept fixed during structure optimization in the QM models. In the QM models, the protein surroundings were treated as a polarizable medium with a dielectric constant of  $\epsilon = 4$ <sup>44</sup> and reaction pathways were optimized using a chain-of-state method<sup>45</sup>. Vibrational and entropic contributions were estimated at the B3LYP-D3/def2-SVP/def2-TZVP (Mg)/ $\epsilon = 4$  level by calculating the Hessian numerically<sup>46</sup>. The catalytic tuning effect of Arg-32 was studied by scanning the distance between the C $\zeta$  of Arg-32 and C $\delta$  of Glu-33 (see Supplementary Fig. 4). Single-point calculations with def2-TZVP basis set were performed on the reactant, transition state, and product structures extracted from the reaction path optimizations. The QM/MM system with ca. 17,000 atoms was cut out from the full protein, using a sphere with a 30 Å radius around the ATP site, and embedded in a water sphere with 150 mM NaCl. Reaction pathways were optimized with the MM region fixed. Energy decomposition was carried out by removing individual residues from the QM/MM system during single-point calculations, without re-optimizing the protein structure. Based on the reaction profiles, 27 ps restrained QM/MM MD simulations were performed with harmonic restraints on the commonly employed reaction coordinate describing the relevant bond-formation and bond-breaking processes<sup>27,28</sup>,  $R = r_4 - r_3 + r_2 - r_1 = [-2.9 \text{ Å}, 2.6 \text{ Å}]$  (see Fig. 2b), using force constants of 100 or 500 kcal mol<sup>-1</sup> Å<sup>-2</sup>. All atoms beyond 15 Å from the QM region were kept fixed in the umbrella sampling (US) simulations. The statistical overlap of the sampled reaction coordinate is shown in Supplementary Fig. 1. Free-energy profiles were computed using the US/weighted histogram analysis method (WHAM) with a convergence criterion of 0.00001 kcal mol<sup>-1</sup><sup>47</sup>. Bootstrap error analysis was performed within WHAM with ten Monte Carlo (MC) trial steps to estimate the statistical uncertainty of the free energy. Rate constants were calculated using the Eyring equation at  $T = 310 \text{ K}$  using a standard pre-exponential factor,

$$k = \kappa \frac{k_B T}{h} e^{-\frac{\Delta G^\ddagger}{RT}} \quad (1)$$

where  $\kappa$  is the reflection coefficient that was set equal to 1,  $\Delta G^\ddagger$  is the free-energy barrier,  $R$  is the gas constant,  $k_B$  is Boltzmann's constant, and  $h$  is Planck's constant. All DFT calculations were performed using TURBOMOLE<sup>48</sup>, which was coupled together with CHARMM in the QM/MM models<sup>49,50</sup>.

**Classical MD simulations of Hsp90.** Classical MD simulations were performed on the full-length dimeric Hsp90 model and the R32A mutant constructed based on the dimeric crystal structure of yeast Hsp90 (PDB ID: 2CG9)<sup>9</sup>. MD simulations were also carried out for the monomeric Hsp90 model with NM-domains constructed based on the X-ray structure of the dimeric NM-fragment of Hsp90 from *D. rerio* (PDB ID: 4IVG)<sup>39</sup> and NTD models of yeast Hsp90 in the apo state and with ADP or ATP (PDB ID: 1AMW)<sup>51</sup>. The protein was embedded in a water-ion environment with 100–150 mM NaCl. The complete simulation setups comprised ca. 72,300 atoms (NM-model), 77,300 atoms (NTD model), and 303,000 atoms (full-length dimer). The constructs were simulated for 250–1000 ns at  $T = 310 \text{ K}$ , using a 2 fs timestep and keeping all covalently bound hydrogens fixed. We employed the CHARMM27/36 force field<sup>43,52</sup> treating the long-range electrostatics by the Particle Mesh Ewald approach. The MD simulations were performed using NAMD<sup>53</sup> and Visual Molecular Dynamics (VMD)<sup>54</sup> was used for visualization and analysis. All Arg-32/Glu-33 distances shown in the paper are measured between the C $\zeta$  of Arg-32 and the C $\delta$  of Glu-33. All simulations are summarized in Supplementary Table 10.

**Calculation of  $pK_a$  values.** Poisson-Boltzmann (PB) continuum electrostatic calculations with MC sampling of  $2^N$  protonation states of the full-length Hsp90 dimer were used for the calculation of  $pK_a$  values. The Adaptive Poisson-Boltzmann Solver (APBS) was used for the PB calculations<sup>55</sup> and Karlsberg+ was employed for the MC sampling<sup>56,57</sup>. The protein was described with atomic partial charges, embedded in an inhomogeneous dielectric continuum with a dielectric constant of  $\epsilon = 4$ . Bulk water was approximated by a homogeneous dielectric continuum with  $\epsilon = 80$ . The molecular surface routine in APBS was used to calculate the boundary interface between the protein and the solvent, using a solvent

probe radius of 1.4 Å and modelling an implicit ionic strength of the solvent of 100 mM KCl. Protonation probabilities were calculated every ns along the 250 ns MD simulation of the full-length Hsp90 dimer.

**Kinetic modelling.** Kinetic models of a simplified chaperone cycle of the WT and R32A variant were simulated using the stochastic Gillespie algorithm in Dizzy<sup>58</sup>. Kinetic parameters are given in Supplementary Table 8.

**Protein purification.** The R32A, E381Q, and D61C (for FRET) single point mutations were introduced by site-directed mutagenesis according to the NEBase Changer protocol in pET28 vectors (Invitrogen, Karlsruhe, Germany) containing the yeast Hsp90 (Hsp82) sequence. Hsp90-NM constructs (aa 1–529) were cloned using a pET28a-SUMO Vector containing an N-terminal 6×His-SUMO-tag. For the NTD constructs (aa 1–210), a pETM-11 vector was used. The plasmids were transformed into the *Escherichia coli* strain BL21 (DE3) cod+ (Stratagene, La Jolla, USA). Full-length proteins (yeast Hsp90, sbal) were expressed for 4 h at 37 °C, NM constructs were expressed at 25 °C overnight. Expression was induced with 1 mM Isopropyl  $\beta$ -D-1-thiogalactopyranoside. After collection, the cells were re-suspended in Ni-NTA buffer A (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 500 mM NaCl, 10 mM imidazole pH 7.5) supplemented with EDTA-free protease inhibitor [SERVA] and DNase1. Cells were lysed using a Cell Disruption System (Constant Systems) at 1.8 kbar. After lysate clarification, the supernatant was loaded on a 5 mL HisTrap HP column (GE Healthcare) and washed with 10 CV Ni-NTA buffer A and 10 CV 5 % Ni-NTA buffer B (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 500 mM NaCl, 300 mM imidazole pH 7.5). The bound proteins were eluted with 100 % Ni-NTA buffer B. Full-length proteins were pooled, diluted to 150 mL with ResQ buffer A (40 mM HEPES, 20 mM KCl, 1 mM EDTA, 1 mM dithiothreitol (DTT) pH 7.5) and loaded onto a ResQ column. NM constructs were supplemented with His<sub>6</sub>-tagged SUMO-protease after Ni-NTA and dialysed against 5 L Ni-NTA buffer A ON at 4 °C. The protein solution was again loaded on a 5 mL HisTrap HP column and the flowthrough was collected. As a final step, proteins were loaded on a Superdex 16/60 75 pg SEC column (GE Healthcare) and eluted with SEC buffer (40 mM HEPES, 150 mM KCl, 5 mM MgCl<sub>2</sub> pH 7.5). For the NMR samples (NTD constructs), cells were grown in M9 minimal media supplemented with <sup>15</sup>NH<sub>4</sub>Cl and <sup>13</sup>C<sub>6</sub>-glucose (Aldrich, Germany), with protein expression and purification conducted following the procedure for the NM constructs. Instead of SUMO-protease, TEV-protease was used. Buffers used were buffer A (50 mM Tris-HCl pH 8, 150 mM NaCl, 5 mM imidazole and 0.02 % NaN<sub>3</sub>), buffer B (50 mM Tris-HCl pH 8, 150 mM NaCl, 500 mM imidazole and 0.02 % NaN<sub>3</sub>), and dialysis buffer (20 mM Tris-HCl pH 8, 300 mM NaCl, 2 mM DTT). For SEC, NMR buffer (see “NMR experiments”) was used. The primer sequences are listed in Supplementary Table 11.

**ADP-release assays.** The ATPase activity was measured spectrophotometrically by following the ADP-release reaction using an enzymatic ATP regenerating system<sup>59</sup>. Full-length (3  $\mu\text{M}$ ) yeast Hsp90 (Hsp82) and 10  $\mu\text{M}$  of the NM constructs were used in 40 mM HEPES (pH 7.5), 150 mM KCl, 5 mM MgCl<sub>2</sub>. Measurements were performed in a Cary 100 UV-Vis photometer (Varian, Inc.) at 30 °C. The reaction was initiated by the addition of 2 mM (full-length Hsp90) and 4 mM (Hsp90-NM) ATP. For the determination of  $K_M$ -values (ATP-binding), different concentrations of ATP from 0.1 mM to 5 mM were used. To subtract the background activity, 50  $\mu\text{M}$  radicicol were added at the end of the measurement. The data were analysed by linear regression using Origin 8.0. ADP-release kinetics were determined using the following equation, where  $m$  is the slope,  $\epsilon_{\text{NADH}}$  is the extinction coefficient of NADH, and  $c_{\text{Hsp90}}$  is the yeast Hsp90 (Hsp82) concentration:

$$k_{\text{ADP-release}} = \frac{-m}{(\epsilon_{\text{NADH}} \cdot c_{\text{Hsp90}})} \quad (2)$$

For the determination of  $K_M$ -values for ATP binding, the ADP-release kinetics were plotted against the respective ATP concentration. The data were fitted using the Michaelis-Menten equation:

$$y = \frac{V_{\text{max}} \cdot x}{K_M + x} \quad (3)$$

**FRET experiments.** FRET experiments were conducted following the protocol of Hessling et al.<sup>6</sup>. Atto488 (donor) and Atto550 (acceptor) (ATTO-TEC GmbH) labelled yeast Hsp90-D61C (200 nM) was used. Measurements were performed in 40 mM HEPES (pH 7.5), 150 mM KCl, 5 mM MgCl<sub>2</sub>, 2 mM DTT in a Fluoromax 4 fluorescence spectrophotometer (Horiba Jobin Yvon) at 30 °C. Closing of Hsp90 was induced by addition of 2 mM AMP-PNP or 2 mM ATPyS. Chase experiments were performed with a tenfold excess (4  $\mu\text{M}$ ) of unlabelled WT yeast Hsp90 (Hsp82) to disrupt the FRET complex. Samples were incubated for 60 min at 30 °C prior to the addition of the unlabelled species. For determination of re-opening rates, closing was induced with 2 mM ATPyS. After the equilibrium was reached, a tenfold excess (20 mM) of ATP was added, to induce re-opening of the Hsp82 dimer. The data were analysed with Origin 8.0 and fitted using a mono-exponential

equation:

$$y = A_1 \left( \frac{-x_1}{t_1} \right) + y_0 \quad (4)$$

$$k_{\text{App}} = \frac{1}{t_1} \quad (5)$$

**Yeast viability assay of Hsp82 variants.** Plasmid shuffling experiments were conducted following the protocol of Nathan et al.<sup>60</sup>, using the  $\Delta\text{PCLDa}$  *S. cerevisiae* strain from S. Lindquist's laboratory deficient in genomic Hsp82 and Hsc82 containing a plasmid coding for WT Hsp82. The pKAT6 plasmid is constitutively expressed under the control of the glyceraldehyde-3-phosphate dehydrogenase gene promoter (GPD promoter) and carries a URA selection marker for the selection of cells that have lost the WT Hsp82 plasmid in the medium supplemented with 5-Fluoroorotic Acid (5-FOA) (Thermo Fisher Scientific). The cells were transformed with either the empty vector p413 (negative control), the p413 vector, coding for the WT Hsp82 (positive control), and the p413 vector, coding for the Hsp82 R32A variant. Hsp82 is essential for yeast survival and the loss of the pKAT6 plasmid, due to 5-FOA-induced selection, inhibits yeast growth. Transformation of p413 vectors containing Hsp82 variants might restore yeast growth, depending on the characteristics of the Hsp82 variants.

**Fluorescence anisotropy.** Fluorescence anisotropy experiments were used to probe the binding of sbal1 to WT yeast Hsp90 and the R32A variant. Measurements were conducted in a JASCO-8500 fluorescence spectrophotometer with polarisers (Jasco, Groß-Umstadt, Germany) at 30 °C in 40 mM HEPES (pH 7.5), 150 mM KCl, 5 mM MgCl<sub>2</sub>, containing 2 mM AMP-PNP. Excitation and emission wavelengths were set to 490 and 530 nm, respectively, using 200 nM of Atto488-labelled sbal1 and 500 nM Hsp90. After pre-incubation of Hsp90 with labelled sbal1, unlabelled sbal1 was added in varying concentrations to compete out the preformed labelled sbal1-Hsp90 complex. The  $K_D$ -value was determined by fitting the anisotropy signal,  $r$ , to the sbal1 concentration,  $c_{\text{sbal1}}$ ,

$$r = F_P - (c_{\text{Hsp90}} + c_{\text{sbal1}} + K_D) - \left( \frac{F_P - F_{\text{PL}}}{2c_{\text{Hsp90}}} \right) \sqrt{(c_{\text{Hsp90}} + c_{\text{sbal1}} + K_D)^2 - 4c_{\text{Hsp90}}c_{\text{sbal1}}} \quad (6)$$

where  $F_P$  is the anisotropy of unbound labelled sbal1,  $F_{\text{PL}}$  is the anisotropy of the labelled sbal1-Hsp90 complex, and  $c_{\text{Hsp90}}$  is the concentration of yeast Hsp90.

**SAXS experiments.** SAXS data were collected at beamline BM29 at the European Synchrotron Radiation Facility (Grenoble, France). Fifty microlitres of sample in 25 mM Hepes pH 7.5, 150 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM TCEP, 0.02 % NaN<sub>3</sub> was injected to a Superdex 200 5/100 GL column (GE Healthcare) connected online to the SAXS capillary. One SAXS frame per second was recorded at a flow rate of 0.15 ml min<sup>-1</sup>. Nucleotide-bound forms were prepared by adding the compounds at a final concentration of 2.5 mM followed by incubation at room temperature. Due to the fast  $k_{\text{off}}$  of the nucleotides, the buffers employed in the chromatographic runs contained 2.5 mM ADP, 2.5 mM ATP, and 1 mM AMP-PNP. SEC-SAXS chromatograms were analysed using the Chromix software<sup>61</sup>. Briefly, >100 buffer frames with constant average intensity were selected. Sample frames were selected from the chromatogram based on (1) chromatographic peak shape and (2) constant  $R_g$  across the selected region. The subtracted averaged scattering profiles were analysed using the Primus software package<sup>62</sup> to extract the  $R_g$  values and the  $P(r)$  distributions were obtained using the Gnom programme<sup>63</sup>. Theoretical scattering profiles were computed from X-ray coordinates using Crysol<sup>64</sup>.

**NMR experiments.** NMR spectra were recorded using Bruker 500, 600, and 950 MHz spectrometer (Bruker, Billerica, USA) at 25 °C, using an NMR buffer with 20 mM sodium phosphate, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 5% D<sub>2</sub>O, and 0.2% NaN<sub>3</sub> pH 6.5, and a protein concentration between 500 and 600  $\mu\text{M}$ . ADP was added at a final concentration of 5 mM, whereas ATP was added at 2.5 mM together with an ATP regenerating system<sup>65</sup>. These saturating conditions were used to avoid interference of different nucleotide affinities in the WT and mutant constructs. Chemical shift assignments were based on previous work for the Hsp90-NTD<sup>33,66</sup> and were extended for the R32A mutant and the nucleotide-bound forms using a combination of triple-resonance HNCA, HNCOCOA, and HNCACB experiments using non-uniform sampling<sup>67</sup>. Spectra were processed using NMRPipe<sup>68</sup> and analysed using the CCPNmr software<sup>69</sup>. CSPs were calculated for backbone amide peaks of 2D <sup>1</sup>H,<sup>15</sup>N-HSQC correlation experiments using the equation,

$$\Delta\delta_{\text{NH}}(\text{ppm}) = \sqrt{\Delta\delta_{\text{H}}^2 + (\alpha \cdot \Delta\delta_{\text{N}})^2} \quad (7)$$

where  $\alpha$  is a scaling factor, calculated from the ratio between the <sup>1</sup>H and <sup>15</sup>N chemical shift ranges ( $\alpha = 0.1689$ ). To observe the arginine guanidine sidechain NMR signals in <sup>1</sup>H,<sup>15</sup>N-HSQC spectra, the <sup>15</sup>N carrier frequency was centred at 105 p.p.m. with a spectral width set to 70 p.p.m. <sup>1</sup>H-<sup>15</sup>N steady-state heteronuclear NOE experiments were performed using modified sequences described by Farrow et al.<sup>30</sup> by collecting two datasets, with and without <sup>1</sup>H saturation,

respectively. Heteronuclear NOE values were obtained from the intensity ratio of R32 H $\epsilon$ -N $\epsilon$  peaks between the saturated and unsaturated spectra, and errors were estimated from spectral baseplane noise root-mean-square deviation (RMSD) as described by Farrow et al.<sup>30</sup>. <sup>15</sup>N-edited NOESY experiments were performed using standard sequences with a mixing time of 120 ms<sup>70</sup>. The NMR assignments of the NTD of the R32A variant have been deposited in the BMRB under the accession number 27858.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The NMR assignments for the N-terminal domain of yeast Hsp90-R32A have been deposited in the BMRB under the accession number 27858. The source data underlying Fig. 4e-g and Supplementary Figs. 14, 17, 18, and 19 are provided as a Source Data file. Other data are available from the corresponding author upon reasonable request.

Received: 26 December 2018; Accepted: 16 February 2020;

Published online: 16 March 2020

## References

- Nathan, D. F., Vos, M. H. & Lindquist, S. *In vivo* functions of the *Saccharomyces cerevisiae* Hsp90 chaperone. *Proc. Natl. Acad. Sci. USA* **94**, 12949–12956 (1997).
- Picard, D. Heat-shock protein 90, a chaperone for folding and regulation. *Cell. Mol. Life Sci.* **59**, 1640–1648 (2002).
- McClellan, A. J. et al. Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. *Cell* **131**, 121–135 (2007).
- Meyer, P. et al. Structural and functional analysis of the middle segment of Hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. *Mol. Cell* **11**, 647–658 (2003).
- Wegele, H., Muschler, P., Bunc, M., Reinstein, J. & Buchner, J. Dissection of the contribution of individual domains to the ATPase mechanism of Hsp90. *J. Biol. Chem.* **278**, 39303–39310 (2003).
- Hessling, M., Richter, K. & Buchner, J. Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90. *Nat. Struct. Mol. Biol.* **16**, 287–293 (2009).
- Southworth, D. R. & Agard, D. A. Client-loading conformation of the Hsp90 molecular chaperone revealed in the Cryo-EM structure of the human Hsp90: Hop complex. *Mol. Cell* **42**, 771–781 (2011).
- Verba, K. A. et al. Atomic structure of Hsp90-Cdc37-Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase. *Science* **352**, 1542–1547 (2016).
- Ali, M. M. U. et al. Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature* **440**, 1013–1017 (2006).
- Panaretou, B. et al. ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone *in vivo*. *EMBO J.* **17**, 4829–4836 (1998).
- Elnatan, D. et al. Symmetry broken and rebroken during the ATP hydrolysis cycle of the mitochondrial Hsp90 TRAP1. *eLife* **6**, e2525 (2017).
- Grenert, J. P., Johnson, B. D. & Toft, D. O. The importance of ATP binding and hydrolysis by Hsp90 in formation and function of protein heterocomplexes. *J. Biol. Chem.* **274**, 17525–17533 (1999).
- Zierer, B. K. et al. Importance of cycle timing for the function of the molecular chaperone Hsp90. *Nat. Struct. Mol. Biol.* **23**, 1020–1028 (2016).
- Jackson, A. P. & Maxwell, A. Identifying the catalytic residue of the ATPase reaction of DNA gyrase. *Proc. Natl. Acad. Sci. USA* **90**, 11232–11236 (1993).
- Okimoto, N. et al. Theoretical studies of the ATP hydrolysis mechanism of myosin. *Biophys. J.* **81**, 2786–2794 (2001).
- Hayashi, S. et al. Molecular mechanism of ATP hydrolysis in F<sub>1</sub>-ATPase revealed by molecular simulations and single-molecule observations. *J. Am. Chem. Soc.* **134**, 8447–8454 (2012).
- Picard, D. et al. Reduced levels of hsp90 compromise steroid receptor action *in vivo*. *Nature* **348**, 166–168 (1990).
- Colombo, G., Morra, G., Meli, M. & Verkhivker, G. Understanding ligand-based modulation of the Hsp90 molecular chaperone dynamics at atomic resolution. *Proc. Natl. Acad. Sci. USA* **105**, 7976–7981 (2008).
- Cunningham, C. N., Krukenberg, K. A. & Agard, D. A. Intra- and intermonomer interactions are required to synergistically facilitate ATP hydrolysis in Hsp90. *J. Biol. Chem.* **283**, 21170–21178 (2008).
- Mickler, M., Hessling, M., Ratzke, C., Buchner, J. & Hugel, T. The large conformational changes of Hsp90 are only weakly coupled to ATP hydrolysis. *Nat. Struct. Mol. Biol.* **16**, 281–286 (2009).
- Taipale, M. et al. Quantitative analysis of HSP90-client interactions reveals principles of substrate recognition. *Cell* **150**, 987–1001 (2012).



22. Mollapour, M. et al. Asymmetric Hsp90 N domain SUMOylation recruits Aha1 and ATP-competitive inhibitors. *Mol. Cell* **53**, 317–329 (2014).
23. Kamerlin, S. C. L., Florián, J. & Warshel, A. Associative versus dissociative mechanisms of phosphate monoester hydrolysis: on the interpretation of activation entropies. *ChemPhysChem* **9**, 1767–1773 (2008).
24. Abrahams, J. P., Leslie, A. G. W., Lutter, R. & Walker, J. E. Structure at 2.8 Å resolution of F<sub>1</sub>-ATPase from bovine heart mitochondria. *Nature* **370**, 621–628 (1994).
25. Ban, C. & Yang, W. Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell* **95**, 541–552 (1998).
26. Schmidt, H. & Carter, A. P. Structure and mechanism of the dynein motor ATPase. *Biopolymers* **105**, 557–567 (2016).
27. Cheng, Y., Zhang, Y. & McCammon, J. A. How does the cAMP-dependent protein kinase catalyze the phosphorylation reaction: an ab initio QM/MM study. *J. Am. Chem. Soc.* **127**, 1553–1562 (2005).
28. Lopata, A. et al. Mutations decouple proton transfer from phosphate cleavage in the dUTPase catalytic reaction. *ACS Catal.* **5**, 3225–3237 (2015).
29. Cunningham, C. N., Southworth, D. R., Krukenberg, K. A. & Agard, D. A. The conserved arginine 380 of Hsp90 is not a catalytic residue, but stabilizes the closed conformation required for ATP hydrolysis. *Protein Sci.* **21**, 1162–1171 (2012).
30. Farrow, N. A. et al. Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by <sup>15</sup>N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).
31. Zhang, H. et al. A dynamic view of ATP-coupled functioning cycle of Hsp90 N-terminal domain. *Sci. Rep.* **5**, 9542, 1–9 (2015).
32. Schopf, F. H., Biebl, M. M. & Buchner, J. The HSP90 chaperone machinery. *Nat. Rev. Mol. Cell Biol.* **18**, 345–360 (2017).
33. Lorenz, O. R. et al. Modulation of the Hsp90 chaperone cycle by a stringent client protein. *Mol. Cell* **53**, 941–953 (2014).
34. Sharma, V. et al. Redox-induced activation of the proton pump in the respiratory complex I. *Proc. Natl. Acad. Sci. USA* **112**, 11571–11576 (2015).
35. Di Luca, A., Gamiz-Hernandez, A. P. & Kaila, V. R. I. Symmetry-related proton transfer pathways in respiratory complex I. *Proc. Natl. Acad. Sci. USA* **114**, E6314–E6321 (2017).
36. Warnau, J. et al. Redox-coupled quinone dynamics in the respiratory complex I. *Proc. Natl. Acad. Sci. USA* **115**, E8413–E8420 (2018).
37. Kaila, V. R. I. Long-range proton-coupled electron transfer in biological energy conversion: towards mechanistic understanding of respiratory complex I. *J. R. Soc. Interface* **15**, 20170916 (2018).
38. Richter, K. et al. Conserved conformational changes in the ATPase cycle of human Hsp90. *J. Biol. Chem.* **283**, 17757–17765 (2008).
39. Lavery, L. A. et al. Structural asymmetry in the closed state of mitochondrial Hsp90 (TRAP1) supports a two-step ATP hydrolysis mechanism. *Mol. Cell* **53**, 330–343 (2014).
40. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
41. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).
42. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132**, 154104 (2010).
43. Best, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_2$  dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
44. Klamt, A. & Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. 2*, 799–805 (1993).
45. Plessow, P. Reaction path optimization without NEB springs or interpolation algorithms. *J. Chem. Theory Comput.* **9**, 1305–1310 (2013).
46. Deglmann, P. & Furch, F. Efficient characterization of stationary points on potential energy surfaces. *J. Chem. Phys.* **117**, 9535–9538 (2002).
47. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
48. Ahlrichs, R., Bär, M., Häser, M., Horn, H. & Kölmel, C. Electronic structure calculations on workstation computers: the program system Turbomole. *Chem. Phys. Lett.* **162**, 165–169 (1989).
49. Brooks, B. R. et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
50. Riahi, S. & Rowley, C. N. The CHARMM-TURBOMOLE interface for efficient and accurate QM/MM molecular dynamics, free energies, and excited state properties. *J. Comput. Chem.* **35**, 2076–2086 (2014).
51. Prodromou, C. et al. Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. *Cell* **90**, 65–75 (1997).
52. MacKerell, A. D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
53. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
54. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
55. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041 (2001).
56. Rabenstein, B. & Knapp, E.-W. Calculated pH-dependent population and protonation of carbon-monoxymyoglobin conformers. *Biophys. J.* **80**, 1141–1150 (2001).
57. Kieseritzky, G. & Knapp, E.-W. Optimizing pK<sub>a</sub> computation in proteins with pH adapted conformations. *Proteins* **71**, 1335–1348 (2008).
58. Ramsey, S., Orrell, D. & Bolouri, H. Dizzy: stochastic simulation of large-scale genetic regulatory networks. *J. Bioinform. Comput. Biol.* **3**, 415–436 (2005).
59. Tamura, J. K. & Gellert, M. Characterization of the ATP binding site on *Escherichia coli* DNA gyrase. Affinity labeling of Lys-103 and Lys-110 of the B subunit by pyridoxal 5'-diphospho-5'-adenosine. *J. Biol. Chem.* **265**, 21342–21349 (1990).
60. Nathan, D. F. & Lindquist, S. Mutational analysis of Hsp90 function: interactions with a steroid receptor and a protein kinase. *Mol. Cell. Biol.* **15**, 3917–3925 (1995).
61. Panjkovich, A. & Svergun, D. I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34**, 1944–1946 (2018).
62. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.* **36**, 1277–1282 (2003).
63. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* **25**, 495–503 (1992).
64. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOLE – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* **28**, 768–773 (1995).
65. Karagöz, G. E. et al. N-terminal domain of human Hsp90 triggers binding to the cochaperone p23. *Proc. Natl. Acad. Sci. USA* **108**, 580–585 (2011).
66. Dehner, A. et al. NMR chemical shift perturbation study of the N-terminal domain of Hsp90 upon binding of ADP, AMP-PNP, geldanamycin, and radicicol. *ChemBioChem* **4**, 870–877 (2003).
67. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158 (1999).
68. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
69. Vranken, W. F. et al. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins Struct. Funct. Bioinf.* **59**, 687–696 (2005).
70. Ikura, M., Bax, A., Clore, G. M. & Gronenborn, A. M. Detection of nuclear Overhauser effects between degenerate amide proton resonances by heteronuclear three-dimensional nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* **112**, 9020–9022 (1990).

## Acknowledgements

This work was supported by the German Research Foundation (SFB1035 projects B12 to V.R.I.K. and A03 and Z1 to M.S. and J.B.). Q.L. acknowledges the China Scholarship Council (CSC) for funding. We acknowledge the European Synchrotron Radiation Facility for access to radiation facilities and local contact Gabriele Giachin for support. NMR measurements were performed at the Bavarian NMR Centre (BNMRZ) at TUM, Garching. We are thankful for the computing time provided by SuperMuc at the Leibniz Rechenzentrum (pr84pa, pr53po, and pr83ro). Open access funding provided by Stockholm University.

## Author contributions

V.R.I.K. designed the project. S.L.M., Q.L., A.P.G.H., and V.R.I.K. performed the molecular simulations. J.L. and D.A.R. performed the protein biochemistry, FRET, and fluorescence anisotropy experiments. A.L. performed NMR and SAXS experiments. S.L.M., A.L., J.L., Q.L., D.A.R., A.P.G.H., M.S., J.B., and V.R.I.K. analysed data. V.R.I.K. wrote the manuscript with contributions from all authors.

## Competing interests

The authors declare no competing interests.

**Additional information**

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-15050-0>.

**Correspondence** and requests for materials should be addressed to V.R.I.K.

**Peer review information** *Nature Communications* thanks Timothy Street and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020



# Acknowledgements

I would like to thank Ville for his continuous support and guidance throughout my PhD. I am very grateful for the training I received, and that I got the opportunity to attend conferences and schools all over the world. Here I would also like to thank DFG and the TUM graduate school for providing funding. Thank you, Ville, for all the help with publications, posters, proposals for computing time at SuperMUC, and this thesis. I hope you will never lose your enthusiasm for research.

Of course, I would also like to thank all active and former members of the Kaila group. You have become too many to list by name, but everyone of you contributed to a great atmosphere in the office. Thank you all for sharing your expertise, providing advice in any matter, and giving feedback. Thanks as well for all the interesting scientific and non-scientific discussions, especially during birthday celebrations that were always accompanied by creative birthday cards and presents. I enjoyed all the small and big group activities we had in Garching and Munich, at conferences, and our group trips.

I wish to thank my collaborators from the Buchner, Groll, Itzen, Hess, Lang, and Sattler groups for the fruitful cooperation. It was also always a pleasure to meet the members of the SFB1035 at various retreats and conferences.

I have met some very nice people in Munich and Garching and also at conferences, whom I would like to thank for their support and the great time and activities. I also would like to acknowledge the residents of HHG for the welcoming atmosphere and for all the activities organized there.

Finally, a big thanks goes to my parents and my brothers, as well as my uncles, aunts, and cousins, for their support during this exciting adventure.