Studying Heme Electrochemistry in Heme Proteins and Quinone Binding in Purple Bacterial Reaction Center Using Multi-Conformation Continuum Electrostatics

By

Zhong Zheng

A dissertation submitted to the Graduate Faculty in Physics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2011

This manuscript has been read and accepted for the

Graduate Faculty in Physics in satisfaction of the

dissertation requirement of the degree of Doctor of Philosophy

Marilyn Gunner

Date

Chief of Examining Committee

Steven Greenbaum

Date

Executive Officer

Ronald Koder

Ruth Stark

Marco Ceruso

Harel Weinstein

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Studying Heme Electrochemistry in Heme Proteins and Quinone Binding in Purple Bacterial Reaction Center Using Multi-Conformation Continuum Electrostatics

by

Zhong Zheng

Advisor: Professor Marilyn R. Gunner

Hemes are important redox cofactors. They are found in a variety of proteins and show a diversity of functions. The free energy of heme reduction in different proteins is found to vary over more than 18 kcal/mol. It is a challenge to determine how proteins manage to achieve this enormous range of E_ms with a single type of redox cofactor. Proteins containing 141 unique hemes of *a*-, *b*- and *c*-type, with bis-His, His-Met and aquo-His ligation were calculated using Multi-Conformation Continuum Electrostatics (MCCE). The experimental E_ms range over 800 mV from -350 mV in cytochrome c_3 to 450 mV in cytochrome *c* peroxidase (vs. SHE). The quantitative analysis of the factors that modulate heme electrochemistry includes the interactions of the heme with its ligands, the solvent, the backbone, and sidechains. MCCE calculated E_ms are in good agreement with measured values. The overview of heme proteins with known structures and E_ms shows the lowest and highest potential hemes are c-type, while the b-type hemes are found in the middle E_m range. In solution, bis-His ligation lowers the E_m by ≈205 mV

relative to hemes with His-Met ligands. The bis-His, aquo-His and His-Met ligated *b*-type hemes all cluster about E_ms which are ≈200 mV more positive in protein than in water. In contrast, the low potential bis-His c-type hemes are shifted little from in solution, while the high potential His-Met c-type hemes are raised by ≈300 mV from solution. The analysis shows that no single type of interaction can be identified as the most important in setting heme electrochemistry in proteins. Therefore, different proteins use different aspects of their structures to modulate the in situ heme electrochemistry.

Quinones play important roles in mitochondrial and photosynthetic energy conversion acting as intramembrane, mobile electron and proton carriers between catalytic sites in various electron transfer proteins. They display different affinity, selectivity, functionality and exchange dynamics in different binding sites. The computational analysis of quinone binding sheds light on the requirements for quinone affinity and specificity. The affinities of ten oxidized, neutral benzoquinones (BQs) were measured for the high affinity Q_A site in the detergent solubilized *Rhodobacter sphaeroides* bacterial photosynthetic reaction center. Multi-Conformation Continuum Electrostatics (MCCE) was then used to calculate their relative binding free energies by Grand Canonical Monte Carlo sampling with a rigid protein backbone, flexible ligand and side chain positions and protonation states. Van der Waals and torsion energies, Poisson-Boltzmann continuum electrostatics and accessible surface area dependent ligand-solvent interactions are considered. The affinities are dominated by favorable protein-ligand van der Waals rather than electrostatic interactions. Each quinone appears in a closely clustered set of positions. Methyl and methoxy groups move into the same positions as found for the native quinone.

ACKNOWLEDGEMENTS

There are many people to whom I am indebted and without those guidance, assistance this work would not have been accomplished. It has certainly shaped me as a person and has led me where I am now.

First of all, I'd like to thank my supervisor, Marilyn Gunner, for all the hope she has put on me and the encouragement and support she provides throughout the course of my study.

Much of the data presented herein would not have been possible without the previous work of Junjun Mao and Yifan Song, who developed the program MCCE. I'd also like to thank the numerous people I have had the pleasure to work with during the past six years: Christian Fufezan, Xinyu Zhang, Jun Zhang, Minghui Dong, Muhamed Amin, Xuyu Zhu, Jianxun Lu and Predrag Kukic.

Lastly, I would like to thank my friends and family. My husband, Desheng Ma, who loves me for who I am, has always supported me unconditionally. My parents, Qingxian Zheng and Lilin Chen, have always been proud of me and I am proud of them. To them I dedicate this thesis.

This work is dedicated to my family.

All human wisdom is summed up in two words:

Wait and Hope.

Table of Contents

CHAPTER I. INTRODUCTION **1**

1.1 Heme 3
1.1.1 Heme and heme proteins 4
1.1.2. The Redox Potentials of Hemes 6
1.1.3. Calculations of Heme E_ms 8

1.2 Quinone in Reaction Center 10
1.2.1. The purple photosynthetic bacteria reaction center 10
1.2.2. Quinones 13
1.2.3. The acceptor quinone (Q_A) of the bacteria reaction center 15

CHAPTER II. ANALYSIS OF HEME ELECTROCHEMISTRY **18**

2.1 Methodology: E_m calculations in MCCE 18

2.2 Calculated contribution of different electrostatic energy terms to E_m 24

2.3 Comparison of the electrochemistry of the different classes of heme protein 32
2.3.1 b- vs. c-type 32
2.3.2 bis-His vs His-Met vs aquo-His 36

2.4 Error Analyses and Parameter Sensitivity 38
2.4.1 Eorror Analyses 38
2.4.2 Parameter sensitivity 42

2.5 Summary 46

CHAPTER III. THE CALCULATED AFFINITY OF SUBSTITUTED BENZOQUINONES FOR THE Q_A SITE OF BACTERIAL REACTION CENTERS **48**

3.1 Methodology: Calculations of quinone affinity in MCCE 48
3.1.1 Preparation of protein structure file. 50
3.1.2 Ligand Docking. 50
3.1.3 Calculating the ligand binding energy. 54

3.2 Experimental vs. Calculated relative binding free energies 58

3.3 Relaxation of the quinone in the binding site 64

3.4 Energy decomposition in the calculation of the quinone affinity 69

3.5 BQ orientation in the Q_A site 71

3.6 Modification of the van der Waals terms 76

3.7 Outliers in the calculations 79

3.8 Summary 80

APPENDIX **82**

REFERENCES: **92**

LIST OF TABLES

Table 2-1 Experimental and MCCE calculated E_m range for each ligand and heme type (mV) 27

Table 2-2 Reference reaction field energy (meV) using different charge sets calculated with ε_{in}=4, ε_{out}=80. 39

Table 3-1 Experimental and calculated binding energies 62

Table 3-2 Calculated results using different approaches. 63

LIST OF FIGURES

Figure 1-1 Structures of heme a, b, c, and d. 4

Figure 1-2 The reaction center from *Rhodobacter sphaeroides*. (PDB ID: 1AIJ)..... 12

Figure 1-3 The Q_A binding site of the *Rba. sphaeroides* reaction center. 15

Figure 2-1 Energy terms contributing to the $E_{m,calc}$ vs. $E_{m,expt}$. 24

Figure 2-2 Residue backbone dipoles with strong interaction with the heme. 29

Figure 2-3 MCCE calculated vs. experimental redox potential (mV). 33

Figure 2-4 MCCE Calculated E_ms with different charge distribution and dielectric constants..... 43

Figure 3-1 Overview of K_d calculations. 51

Figure 3-2 Experimental vs. calculated binding free energies (kcal/mol). 64

Figure 3-3 The numbering of the quinone ring..... 65

Figure 3-4 The overlap of DQ and the backbone of Ala M249 and Ala M260 in crystal structure (magenta), GROMACS minimized structure (cyan) and MCCE optimized structure (yellow). 68

Figure 3-5 Residues within 4Å of the Q_A headgroup. 72

Figure 3-6 The overlap of the crystal structure (magenta) with the final MCCE minimized structures with all selected protein and ligand conformers (blue). 74

Chapter I. Introduction

All living beings are made of cells, the smallest structural and functional unit of organisms. The simplest forms of life are isolated cells that breed by dividing in two. Higher organisms, such as human being, have groups of cells carrying out specific functions and are linked by complicated system of cellular communication.

A living cell is composed of a restricted set of elements. Four of them are C, H, N and O, which make up nearly 99% of the cells' weight.[1] Biological molecules are formed by simple combinations of these atoms and certain structures repeatedly recur, such as methyl ($-CH_3$), hydroxyl (-OH), carboxyl (-COOH). Broadly speaking, cells contain just four major families of small organic molecules: the sugars, the fatty acids, the amino acids and the nucleotides.

The studies of the structures and functions of proteins are the focus of this dissertation. The amino acids are the building blocks of proteins. They all contain a central α-carbon atom, a carboxylic acid group, an amino group and a side chain that varies. There are 20 common amino acids in proteins, which occur over and over again in all proteins. To a large extent, cells are made of proteins, which amount to more than half of their dry weight.[1] They control the shape and structure of the cell.

Proteins consist of polypeptide chains formed of amino acids. The associations of the side chains of the amino acids with each other and with water establish non-covalent bonds, which lead the polypeptide chain to fold into one particular and energetically favorable conformation. The chemical properties of a protein, which largely determines

its biological functions, depend on its exposed surface residues. When side chain activities are not sufficient enough, proteins bind substrates for the extension of its functional versatility. There are a variety of substrates and cofactors that are employed in protein. Cytochrome c proteins use hemes for electron transfer while flavoproteins use flavin to perform the same function.[2]

Many proteins bind multiple substrates to accomplish more complicated chemical reactions. The bacterial reaction centers, which will be studied here, convert light energy into electrochemical potential across the membrane, which requires in total 9 cofactors. A pair of bacteriochlorophlls absorbs light energy and release one electron. They don't release another e^- until they have been reoxidized. One bacteriochlorophll and one bacteriopheophytin pass the electron to the primary quinone Q_A. The ubiquinone at Q_B site gathers 2 electrons from Q_A and takes up 2 protons to form QH_2, the product of the reaction. The three other cofactors, an iron, a second bacteriochlorophll and a second bacteriopheophytin, are not functional. Another example of a multi-cofactor protein is cytochrome c oxidase. It's a membrane protein that generates a proton gradient. All the redox-active cofactors are held by subunit I and II. Subunit II contains the di-nuclear Cu_A, which serves as a one-electron donor/acceptor. Meanwhile subunit I binds hemes *a*, a_3, and Cu_B. Heme a_3 and Cu_B form the "bi-nuclear" center where dioxygen reduction to water occurs releasing energy to pump the protons.

This dissertation concentrates on two types of cofactors: hemes and quinones. Hemes are common electron carriers, which are largely found in cytochrome proteins. A major challenge for understanding the relationship between protein structure and function is to recognize the forces that proteins use to modify the behavior of bound ligands. The

simplicity of the heme electrochemistry with its one electron oxidation/reduction reaction, the extreme range of modification of the chemistry in situ and the wealth of data and structures allows in depth computational study of this group of proteins. Earlier calculations on heme proteins analyzed either one protein [3-5], or several proteins covering a small range of E_ms.[6,7] The work reported here is the first using a single method to analyze proteins with hemes covering an E_m spread of 800 mV, almost twice as large as the 450 mV range previously explored.[7] Most bis-His, His-Met, and aquo-His ligated heme containing proteins with known E_ms as of March 2007 were analyzed. Using the standard continuum electrostatics analysis there are no free parameters in the E_m calculation as long as the appropriate E_m of the heme with the ligand in solution ($E_{m,sol}$) is known.[4] The study here will be used to explain how proteins modulate heme electrochemistry.

Quinones are very important lipid soluble electron carriers. Different types of quinones were selected by nature for different proteins. The aim here is to compute binding affinities to shed light on the range of native binding sites that are observed to accommodate quinones. The initial target is the primary quinone (Q_A) binding site of the photosynthetic reaction centers of *Rps. sphaeroides.* Calculations of ligand affinities combine search functions to generate ligand poses in the protein and scoring functions to rank their energies.

1.1 Heme

1.1.1 Heme and heme proteins

Hemes are redox active cofactors found in a variety of proteins. It consists of a large heterocyclic ring system called protoporphyrin IX with an iron in the center. Four nitrogen atoms on the heme porphrin are ligated to the center iron atom.

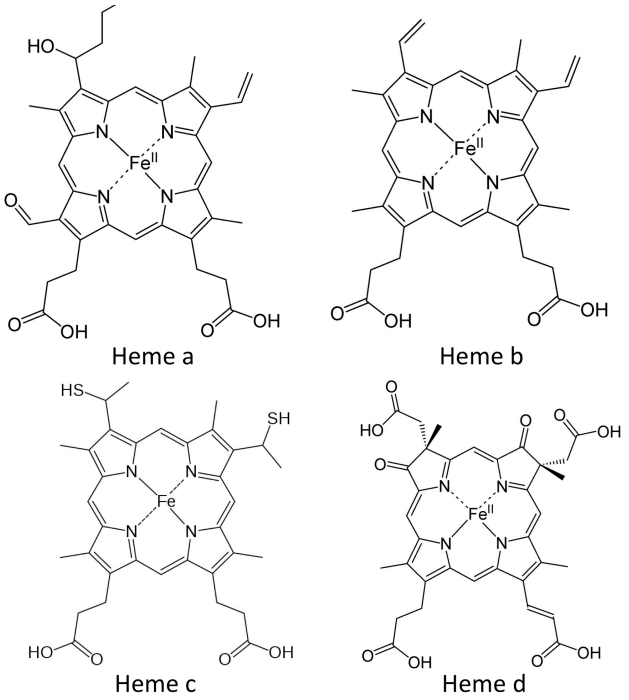

Figure 1-1 Structures of heme a, b, c, and d.

Back in the 1920s, heme proteins were classified on the basis of the position of their lowest energy absorption band in the reduced state, such as cytochromes a(605 nm), b(~565nm), and c(550 nm).[15] Among them, the most common heme tetrapyrrole macrocycle is heme b, also called protoheme. (See Figure 1-1) It has four methyl groups, two vinyl groups and two propionates on the ring. It serves as the basic structure from which heme a and c are biosynthetically derived.[16] Heme c is the most structurally similar to heme b. (See Figure 1-1) The thioether bonds to cysteine covalently link the heme macrocycle replacing the vinyl groups. This type of heme is most often found in

cytochrome c proteins. Cytochrome c proteins typically carry a CXXCH sequence motif from which the two cysteines link to the porphyrin macrocycle and the histidine binds to the iron. Other heme types, such as a, d, o, are also derivatives of heme b and relatively rare in proteins.

Protein scaffolds bind hemes via axial ligands bound to the iron, the polar interactions with the propionates and hydrophobic interactions with the heme macrocycle. The axial ligands are provided by protein via amino acid side chains that contain oxygen, nitrogen or sulfur atoms. Fufezan et al used a non-redundant heme protein database to study heme binding motif.[17] In the non-redundant database, the most common ligand is histidine with 82.6% of the 132 hemes having at least one histidine ligand. More histidine ligands were observed for c-type hemes (95.1%) than b-type hemes (71.8%). Other common ligands include Met, Asn and Asp. The most frequently observed ligation motifs are the bis-histidine and mono-histidine with a free ligation site, which together make up 71% of the total heme non-redundant database.

Besides the axial ligands to the heme that determine the basic heme electrochemistry, the interaction of nearby amino acids in the protein also modulates heme function. The heme binding sites are dominated by hydrophobic residues, such as Leu, Ile, Met, etc.[18] The interactions with the hydrophobic residues are important for heme binding energies. They are essential for proteins like myoglobin, which mainly binds five-coordinated hemes and lacks the covalent bond for the c type heme.[19,20] At least one aromatic residue (Phe, His, Trp and Tyr) near the heme are observed in the majority of heme protein structure. They stablize the heme cofactor by π-stacking or edge-to-face interactions.[21] Polar residues are also seen in the binding site. They make a contribution by forming

hydrogen bonds and regulate the electrostatic protein environment. Heme propionate groups can interact with polar residues. They are often seen salt-bridged to Arg resides. All these polar interactions are critical to modifying the electrochemical properties of the hemes.

A group of versatile and important proteins incorporate hemes as cofactors. These proteins carry out diverse biological functions. For example, *c* type cytochromes are responsible for electron transport in respiratory pathways.[22,23] Cytochromes P450 carry out substrate oxidation.[24] Heme proteins also play roles that do not involve redox reactions, in sensing, storing and transporting small ligands [25], metal ions [26] and gases such as oxygen and carbon monoxide.[27,28] The release of cytochrome *c* from the mitochondria into the cytosol initiates programmed cell death.[29,30] Several six-coordinate heme proteins, such as cyt cd_1, are known to swap ligands upon oxidation/reduction to gate the electron-transfer event.[31] In addition, heme has become a favored cofactor to be bound to de novo designed and bioinspired proteins to generate new functionality.[32-34]

1.1.2. The Redox Potentials of Hemes

In biological electron transfer chains, the cofactor midpoint redox potential controls the direction of favorable transport between individual cofactors.[35,36] Thus, the in situ heme E_m determines the role the protein will play. Reported heme E_ms cover an enormous range of 1000 mV from -550 mV for His-Tyr *b* type heme in HasA[37] to 450 mV for His-Met *c* type heme in cytochrome *c* Peroxidase[38]. This represents a shift of 23.5 kcal/mol for the reduction reaction, equivalent to a protein changing the pK_a of a given residue type by 17.4 pH units.

The redox potential of the heme proteins are fundamentally coupled to the difference in heme binding affinity in the oxidized/reduced states. The 23.5 kcal/mol change in E_m values represents a 10^{18} shift in the ratio of the ferric (Fe III) and ferrous (Fe II) heme association constants. Thus high-potential heme proteins are more stable and tightly bound in their reduced states, while low-potential heme proteins are more stable in the oxidized state. This property can facilitate heme release from the transport ligand. For example, the heme acquisition system A (HasA) has an extremely low reduction potential of -550 mV vs SHE. That indicates a very tight ferric heme affinity (K_d≈10pm) and a very weak ferrous heme affinity (K_d≈10μm).[37] This protein extracts heme from hemoglobin in oxidized state and delivers it to HasR receptor in reduced state.

The ligand coordination determines a starting value for the redox potential of the heme complex. The measured E_m of bis-His microperoxidases (MPs) in aqueous solution is ≈-220 mV[39] vs SHE while the potential is -15 to -70 mV vs SHE[40,41] for the His-Met type. Based on the hard-soft acid-base principle,[42] the Fe(II) is softer than Fe(III) and binds softer ligands, such as Met, more tightly, while the Fe(III) binds harder ligands, such as His, more tightly. This is consistent with the measured redox potentials shown above. The bis-His ligand type binds more tightly to the oxidized heme and less tightly to the reduced heme than the His-Met type, which results in a lower E_m for the bis-His type heme in the absence of other interactions. The five-coordinate hemes bind a water or hydroxide as a sixth ligand. The measured solution E_m for His-water heme is -120 mV and -200 mV for His-hydroxyl,[43] which lies between that of -15 mV for His-Met type to -220 mV for bis-His type. Thus water and hydroxyl are relatively harder ligands than Met but still softer than His ligand.

In addition to the intrinsic factors that affect the heme redox potentials, the protein environment also greatly modulates the heme E_ms. Among the heme proteins calculated in this thesis, the bis-His heme proteins have the largest E_m span of 650 mV from low potential bis-His c heme in soluable cyt. c_3 (-350 mV) to high potential bis-His a heme in membrane protein cyt c oxidase (300 meV).[44] This is a $\approx 10^{12}$ difference in the binding affinities between the oxidized/reduced heme states. The other two heme types included in the calculation are the His-Met and aqua-His. The His-Met type is high potential heme proteins with E_ms ranging from -60 mV in photosynthesis reaction centers to 450 mV in cyt c peroxidase. The aqua-His type is in the middle E_m range with a smaller observed E_m span from -260 mV to 103 mV. Other His ligand, such as His-Tyr, which is not calculated here, has a large E_m range as the bis-His type from -550 mV (HasA) to 287 mV (cyt cd_1).[18]

The redox potential for all these heme types are shifted from the solution values displaying the effects of the protein environment. The bis-His heme proteins favors oxidized hemes lowering the E_ms as well as reduced hemes raising the E_ms. The redox potentials for the His-Met type hemes are mostly increased with some by as much as ≈500 mV. This trend shows that His-Met heme proteins bind reduced heme tighter than the oxidized heme. The aqua heme types exhibit the smallest E_m variation from the solution values. This is probably because the five-coordinate hemes are mostly situated close to the protein surface thus altered less electrochemically by the protein.

1.1.3. Calculations of Heme E_ms

Cytochrome E_ms have been analyzed by various computational techniques to understand how the protein modifies the heme electrochemistry. Studies have used

classical continuum electrostatics (CE),[3,4,6,45] MultiConformation Continuum Electrostatics (MCCE),[7,44] Protein Dipoles Langevin dipoles (PDLD); semi-microscopic PDLD (PDLD/S) analysis;[46,47] Molecular Dynamics (MD),[48] and QM and QM/MM[49] methods. Calculations on the *Rhodopseudomonas viridis* reaction center, which have four hemes with E_ms spanning 450 mV, successfully identified the high- and low-potential hemes.[4,50] Mao investigated a variety of soluble cytochromes with E_ms spanning 450 mV using MCCE, comparing the ways different protein folding motifs modulate the electrochemistry.[7] A variety of studies have shown the importance of the propionic acids, which are peripheral ligands to the heme in determining the E_m and the pH dependence of the E_ms.[3,51,52]

Many of the methods that analyze heme E_ms in proteins are conceptually similar to those that calculate residue pK_as in proteins, analyzing how the protein shifts the proton or electron affinity of a site.[53-58] However proteins are found to shift heme E_ms over a much larger range than amino acid pK_as. Thus protein pK_as are usually measured from pH 2 to 11 where proteins remain stable.[8,53,59] While the observed pK_as for a given residue type range over only a few pH units,[60,61] some highly perturbed species can be found where the protein shifts the pK_as by as much as 4-5 pH units [62,63]. This represents a shift in free energy of ionization of 240-300 meV, much smaller than found routinely for the shifts in heme E_ms.

The earliest analysis of heme protein electrochemistry focused on how the desolvation of the heme in the protein will raise the E_m relative to that found for a heme in water.[64-66] Modern continuum electrostatics techniques then combine continuum electrostatics energies and Monte Carlo sampling of the possible combinations of ionization states in a

protein. This analysis starts with a reference $pK_{a,sol}$ or $E_{m,sol}$ in water for each group of interest calculating only the shift in pK_a or E_m when the group is moved into the protein.[54] The free energy of the reaction is assumed to be shifted by the change in solvation energy and by electrostatic pair-wise interactions with charges and dipoles between the water and protein, which are calculated by the Poisson-Boltzmann (PB) equation[67] or Generalized Born methods.[68,69] Monte Carlo sampling establishes the Boltzmann distribution as a function of pH and E_h, yielding the calculated pK_as and E_ms within the protein. There have long been arguments about the most appropriate value for the protein dielectric constant. Different values have been tried, from as low as 4[8,70,71] to 8[60], 20[6,72] and to as high as 80.[73] There have also been methods using non-uniform dielectric constants.[74,75] All these represent different levels of averaging the response to changes in charge.

MCCE adds extensive rotamer sampling to the sampling of ionization states found in traditional continuum electrostatic calculations. This added degree of freedom raises the local effective dielectric response and improves the match between observed and calculated E_ms or pK_as.[7,8,76] MCCE has proven to be a good tool in calculating pK_as and E_ms of residues and cofactors in proteins.[7,11,51,77-79]

1.2 Quinone in Reaction Center

1.2.1. The purple photosynthetic bacteria reaction center

All reaction centers (RCs) from green plants, algae or photosynthetic bacteria carry out the same function, which is to convert light energy into high energy oxidizing and reducing products by charge separation. These products are then passed on to subsequent electron transport chains. The energy will ultimately drive ATP synthesis and ion

transport.[80] Over the years, the RC has been an important testing ground for the understanding of electron transfer theories and quinone binding. The rich experimental dataset of thermodynamic and kinetic measurements has allowed this protein to be used for developing methods for computational analysis on large transmembrane proteins.[81] Based on the identity of the terminal electron acceptors, the reaction centers are classified into two basic types: Type I (or Fe-S type) and Type II (or pheophytin-quinone type). [82] Type I reaction centers are found in green, brown sulfur bacteria (*Chlorobiaceae* and *Heliobacteriaceae*), PS I bacteria (*Cyanobacteria*) and PSI in plants, while Type II RCs are found in all purple bacteria, green bacteria (*Chloroflexaceae*) and plant PS II. [83]

In the mid-1980s, the structures of the Type II reaction centers of purple photosynthetic bacteria, *Rhodopseudomonas viridis* and *Rhodobacter sphaeroides*, were determined by X-ray crystallography.[84,85](Figure 1-2) This was a critical breakthrough in photosynthesis research. Ever since then, these complexes, which appear to have some common design features to all reaction centers, have been subjected to extensive structural and functional analysis.

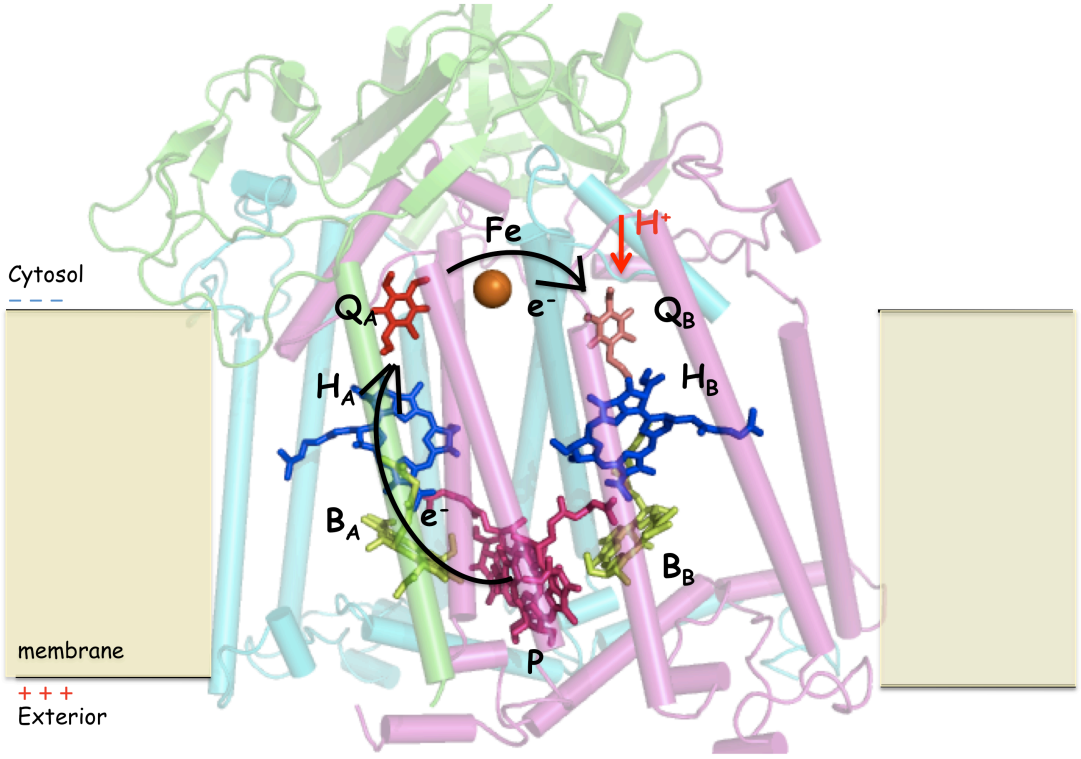

Figure 1-2 The reaction center from *Rhodobacter sphaeroides*. (PDB ID: 1AIJ)

The functional core of the purple RC complex is a non-identical heterodimer (subunits L and M). The axis of the twofold symmetry is oriented perpendicular to the plane of the membrane. (Figure 1-2) A third subunit H caps the LM dimer on the cytosolic side of the membrane, which stabilize the protein structure and is involved in the proton uptake. The LM dimer binds all the active cofactors, including 4 bacteriochlorophylls (BChl), 2 bacteriopheophytins (BPhe), 2 ubiquinones and an iron atom.

Though the structure of the purple bacterial reaction center is pretty symmetrical, the electron pathway occurs exclusively via A cofactors.[83] The primary donor (P^*) absorbs light energy and forms an excited state. It passes one electron via BChl (B_A) to the BPhe

(H_A). All these occur on a time scale of few picoseconds. The BPhe (H_A) further transfers the electron to the primary quinone, Q_A, which takes ≈0.2 ns. From Q_A^-, the electron crosses the symmetry axis and reaches the secondary quinone, Q_B. This is a much slower process of 10-100 μs.

The positively charged P^+ captures a second electron from the nearby cytochrome so returns to the ground state. Following the same pathway, the Q_B is doubly reduced to QH_2 with the uptake of 2 protons from the solution. It then leaves the RC and diffuse into the membrane. The Q_B site binds a new oxidized quinone (Q) from the membrane pool. Thus the net proton uptake by RCs is the result of quinol production, which requires $2H^+$ per $2e^-$. The essential principle illustrated here is that the reaction centers uses light energy to transfer electrons from weak electron donor (P^*) to strong electron donor (quinone in reduced state). The excitation energy stored in the electron will generate the electrochemical gradient over the membrane, which will be used later in ATP synthesis.

1.2.2. Quinones

Quinones are the most important lipid-soluble electron and proton carriers in the membranes of mitochondria, chloroplasts and oxygenic bacteria.[86,87] They transfer electrons between proteins carrying out the proton-coupled electron transfer reactions that generate transmembrane proton gradients.[88,89] Quinone redox chemistry occurs in specific binding sites in these proteins, which select for different quinones[90] and that modulate in situ quinone chemistry.[91-93] The biological quinone redox reactions shift the quinone between stable quinone and hydroquinone (quinol) species, requiring two electron and two proton transfers. The tuning of affinity, specificity and electrochemistry in quinone

sites is of prime importance since the intermediate, singly reduced semiquinone is a source of reactive oxygen species (ROS).[94-97]

As quinones function as intermediaries in the bioenergetic transfer chain there are quinone binding sites in many proteins. The Krebs cycle products yield QH_2 via $NADH_2$ or succinate oxidation in NADH-DH (complex I) and succinate-DH (complex II).[98] In aerobic respiration, the cytochrome bc_1 complex gives electrons from QH_2 to cyt c[93,99], which in turn provides the electrons to reduce O_2 to water in cytochrome c oxidase. Different organisms can use different quinones. Animals and plants use only ubiquinone (UQ) in their mitochondrial membrane electron transfer proteins, while plants and algae use plastoquinone and phylloquinone in their chloroplast photosynthetic reactions. Bacteria use a mixture of quinones including UQ, menaquinone and rhodoquinone for photosynthetic and other transmembrane, bioenergetic reaction. Quinone binding sites are also the targets for anti-malarial drugs[100]; nafuredin, an antibiotic[101] and herbicides[102] and can be the source of harmful reactive oxygen species.

Quinones are seen in a slowly growing list of membrane protein crystal structures, including in photosynthetic reactions centers of bacteria (RCs) (Q_A, Q_B)[103], PSI and PSII, cytochrome bc_1 and b_6f oxidoreductases (Q_o, Q_i)[104,105], quinol-fumarate reductase (Q_P,Q_D)[106], succinate dehydrogenase[97] and the enzyme DsbB.[107] However, inspection of the quinone binding sites does not provide a well-defined binding motif.[108] Binding sites have from 1 to 6 aromatic residues (Phe, Tyr and Trp) within 4Å of the quinone ring and tail. Some are parallel extensions of the quinone ring or over the ring in tilted or parallel orientations. The two quinone oxygens are hydrogen bond acceptors when the quinone is oxidized (Q) or donors to the neutral reduced cofactor (QH_2). Hydrogen bond partners

include the ionizable Glu, Asp, Lys; the polar His, Ser, Tyr, and backbone NH. The number of hydrogen bond partners varies from 0 as for one of the two quinone binding sites in *Escherichia coli* fumarate reductase to 5 as in RCs. Water molecules are seen in about two thirds of the quinone binding sites.

1.2.3. The acceptor quinone (Q_A) of the bacteria reaction center

In *Rba. Sphaeroides* bacterial reaction center, the Q_A is a ubiquinone-10 (UQ-10). (Figure 1-3) The crystal structure shows that the two carbonyl oxygens of the quinone are hydrogen bonded: O_1 to the peptide NH of Ala M260 and O_4 to the $N_\delta H$ of His M219. The bond lengths are 2.84 and 2.79 Å respectively. Residues that come to close contacts with the Q_A are shown in figure 1-3. They are mostly hydrophobic with a few polar residues.

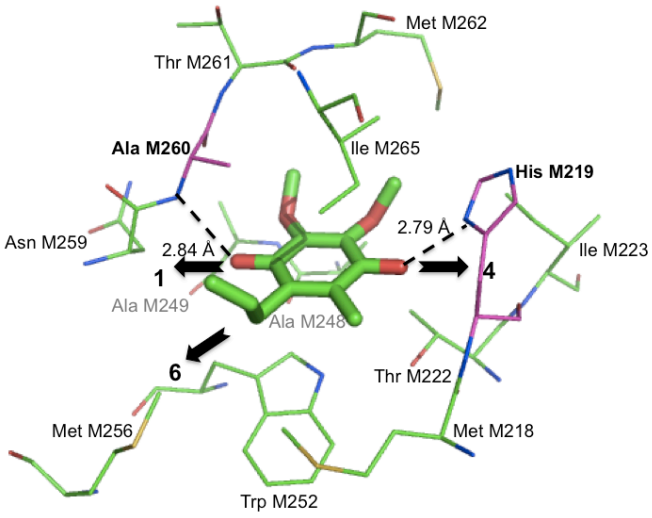

Figure 1-3 The Q_A binding site of the *Rba. sphaeroides* reaction center.

The bacterial reaction centers of photosynthetic bacteria (RCs) have proved to be a useful and robust system for experimental studies of quinone electrochemistry and binding.[109-111] The electrochemical midpoints, E_ms, of Q_A[91] and Q_B[112,113] of *Rb.*

sphaeroides RC have been measured, even down to cryogenic temperature.[114,115] Computational analysis has followed. The Q to $Q^{\bullet-}$ E_m of the native UQ in the Q_A and Q_B sites have been satisfactorily calculated in wild type[9,11,71,116,117] and mutant RCs[118]. Likewise, the affinity of many compounds for the Q_A site of *Rb. sphaeroides* RCs have been measured, providing a qualitative picture of the importance of the quinone ring substituents and the tail structure.[90,119-125] However, there has been little computational analysis of quinone affinity. This is largely because calculations of binding affinity[126-129] are far more challenging than for E_ms and pK_as.[54,55,130]

All methods introduce simplifications in protein and ligand degrees of freedom and in the complexity of the scoring functions in sake of computational speed.[131,132] The fastest techniques use a rigid receptor and ligand, allowing only translation and rotation.[133] More refined methods provide the ligand with flexibility and the protein with optional limited flexibility (e.g. DOCK[134], AutoDock[135], FlexE[136], GOLD[137]). Inevitably conformational freedom greatly expands the size of the problem. The relative protein-ligand interaction energies are evaluated with functions that range from simple statistical functions as found in QSAR[138], to MM-PBSA or MM-GBSA[139,140] methods which use standard Molecular Dynamics (**MD**) bonded and non-bonded parameters and continuum electrostatics (**CE**) analysis[141] to more complete physics-based approaches with significant computational cost, such as alchemical free energy calculations[142] or DFT analysis.[143]

MCCE, a new MM-PBSA type method, which carries many novel features will be used here. It was developed for analysis of E_m and pK_a values and has been used extensively in the analysis of the quinone electrochemistry in RCs.[8-11] MCCE has also

been applied to analyze ion binding,[12] which represents an extension of the program to determine small molecule affinity. The binding of ten fully oxidized neutral benzoquinone (BQ) substitutes at the Q_A site of the *Rhodobacter sphaeroides* photosynthetic reaction center will be calculated. The considered interactions between the protein and the ligand combine molecular mechanics non-electrostatic interactions and Poisson-Boltzmann continuum electrostatics interactions. The binding affinity is determined with Grand Canonical Monte Carlo sampling, allowing flexible side chains[13] and multiple binding modes[14] within a rigid backbone. Unlike almost all simulations, which need to pre-assign the ionization states of the protein and ligand, all protonation and redox states for each residue and cofactor can change in the MC sampling that determines the ligand affinity.

Chapter II. Analysis of heme electrochemistry

2.1 Methodology: E_m calculations in MCCE

Protein Selection: All structures were obtained from the Protein Data Bank.[144] Forty-two different heme-containing proteins (63 pdb files, 96 structures, 141 hemes) with experimental E_ms spanning 800 mV were examined. These proteins range in size from the 76 residue cyt. c_{553} to the 1212 residue bc_1 complex. Most structures are solved by X-ray crystallographer with a resolution of 2.5Å or better. When there are multiple copies of a protein in one PDB file, each independent copy was analyzed separately, keeping all results. Three NMR structures were considered here (cytochrome b_5 1AW3, cytochrome b_5 1B5A, cytochrome b_{562} 1QPU), all of which contain a single structure in the PDB file. With the exception of a single heme *a* in cytochrome *c* oxidase, there are only *b*- and *c*-type hemes included, with ligand types of bis-His, His-Met or aquo-His. All crystal waters were deleted from the structures. Five coordinate hemes, with a single protein ligand, were fitted with an aquo ligand. This can be a water or hydroxyl, with a pK_a that is dependent on the heme oxidation state.[43] A dielectric constant of 80 was applied to the surrounding water. Protein cavities were filled with continuum water. The default value for the protein dielectric constant was 4.

Three membrane proteins are included: bovine bc_1 complex, *Rhodobacter sphaeroides* cyt. *c* oxidase, and the *Rhodopseudomonas viridis* reaction center. The hemes in the reaction center are in an extra-membrane cytochrome c subunit, which was isolated and analyzed independently. These heme E_ms are not affected by removing the other three transmembrane subunits (Gunner unpublished results). A membrane slab of

~35Å was built for cyt. *c* oxidase and the bc_1 complex using the program IPECE.[77,145] IPECE is embedded in MCCE as an optional procedure. Membranes were assigned a dielectric constant of 4. In cyt *c* oxidase, the E_m for the bis-his heme *a* was calculated with all other cofactors fixed in their oxidized states.

***MCCE*:** Multi-conformation continuum electrostatics (MCCE) is a semi-empirical method combining classical continuum electrostatics and molecular mechanics.[8,9,60] It samples the ionization state of each residue and cofactor in the protein as a function of E_h and pH. Version MCCE2.4 was used here. In MCCE, the protein backbone and the heme ring heavy atoms are fixed. MCCE generates rotamers in 60º steps around each rotatable bond for both ionizable and non-ionizable sidechains and the heme propionic acids. Each possible conformation or ionization state for a residue is called a conformer. All pairs of conformers with only modest clashes (<5 kcal/mol) have their positions optimized. All acidic and basic residues have conformers for the neutral and ionized states and multiple hydroxyl positions are found in different conformer. Solution of the finite-difference Poisson-Boltzmann equation using DelPhi[67] yields an energy look-up table, which contains the electrostatic conformer-conformer pairwise interactions and reaction field energy for each conformer. A salt concentration of 0.15 M, a water probe radius of 1.4 Å and a 2.0 Å Stern ion-exclusion radius were used. Protein residues were assigned PARSE charges and radii.[146] Focusing runs give a final resolution better than 2.0 grids/Å.[147] The non-electrostatic van der Waals interactions and torsion energies were calculated with standard AMBER parameters.[148] For a protein with ~110 residues, ~1200 conformers are made, with each conformer representing a choice of atomic position and ionization state for a sidechain, chain termini or cofactor. For every protein microstate, a single

conformer was selected for each residue. Monte Carlo sampling calculates the Boltzmann distribution of all conformer ionization states and positions at 25°C. Each calculation result reported here was the average of two independent Monte Carlo sampling runs. The total energy for one microstate (ΔG^n) is:

$$\Delta G^n = \sum_{i=1}^{M} \delta_{n,i} \left\{ \left[2.3 m_i RT(pH - pK_{sol,i}) + n_i F(E_h - E_{m\,sol,i}) \right] + (\Delta\Delta G_{rxn,i} + \Delta G_{pol,i}) \right\} + \sum_{i=1}^{M} \delta_{n,i} \sum_{j=i+1}^{M} \delta_{n,i}(\Delta G_{ij}) \quad (1)$$

RT is 0.59 kcal/mol. F is the Faraday constant. $\delta_{n,i}$ is 1 for the conformer i that is occupied and 0 for all other conformers in the residue. m_i is 1 for bases, -1 for acids and 0 for all other groups. n_i is the number of electrons. M is the total number of conformers in the protein. $pK_{sol,i}$ and $E_{m\ sol,i}$ are the pK_a and midpoint potential in solution for residue. $\Delta\Delta G_{rxn,i}$ is the desolvation energy, the loss of reaction field energy when conformer i is moved from solution to protein. $\Delta G_{pol,i}$ is the pair-wise electrostatic and nonelectrostatic interaction between the conformer i and the backbone dipoles. ΔG_{ij} is the electrostatic and nonelectrostatic interaction between two conformers. A detailed description of the method can be found in.[7,8,11,60,77]

***Midpoint Reduction Potential (E_m) Calculation*:** MCCE follows the percentage of group ionization for each residue or cofactor as a function of pH or E_h (Eqn. 1). The midpointredox potential is the E_h where half of the hemes are oxidized. The MCCE procedure keeps all protonation states in equilibrium with the heme as it changes redox state during the titration. This allows the proton uptake coupled to electron transfer to be determined from the number of protons bound to the ionizable groups as a function of the

cofactor redox state. The default calculations are carried out at pH 7. The following proteins were analyzed at alternate pH where the E_m was measured: cyt c_3 3CAO, 3CAR at pH 7.6; cyt c_3 1CZJ at pH 8.1; cyt *c* 1M1R at pH 9.0.

The influence of the protein on the E_m can be determined from:

$$E_m = E_{m,sol} - \Delta G_{protein} / nF \quad (2)$$

$\Delta G_{protein}/nF$ is the effect of the protein on the free energy of the redox reaction.[54] This term can be decomposed in MCCE calculation as:

$$\Delta G_{protein} = (\Delta\Delta G_{rxn} + \Delta G_{pol}) + \Delta G_{res,prop} + \Delta G_{res,prot} \quad (3)$$

This incorporates the difference in the loss in reaction field energy (desolvation energy) moving from solvent to protein ($\Delta\Delta G_{rxn}$), the change in interaction with the backbone dipoles (ΔG_{pol}) and contributions from the electrostatic interaction with the propionates ($\Delta G_{res,prop}$) and the rest of the protein ($\Delta G_{res,prot}$). In each case the difference between the interaction of the oxidized and reduced heme with the protein determines the resultant E_m shift. $\Delta\Delta G_{rxn}$ always destabilizes heme oxidation, raising E_ms[64,65,149] because the solvation energy loss is always greater for the more highly charged oxidized heme. Ionized propionates always have favorable interaction ($\Delta G_{res,prop}$) with the heme, lowering E_ms. ΔG_{pol} and $\Delta G_{res,prot}$ depend on the protein context. They are as likely to favor the reduced as the oxidized heme.

Heme-ligand Complex: Heme, its axial ligand sidechains and, for the c-type hemes, the Cys sidechains that are covalently bound to the hemes, are treated together as a single unit with an oxidized or reduced conformer. The backbone atoms for the residues that

function as ligands are considered part of the protein backbone. The two propionates are treated as independently ionizable groups.

The measured E_m of –220 mV (vs. SHE) of bis-His microperoxidases (MPs) in solution was used as the solution E_m ($E_{m,sol}$) for the bis-His hemes.[39] The $E_{m,sol}$ of -15 mV derived from site-directed mutation from bis-His to His-Met axial ligands[40,41] was used for His-Met hemes (see Mao et al for a more complete discussion[7]).

Five-coordinate hemes bind a water or hydroxide as a sixth ligand. Solution pK_a ($pK_{a,sol}$) and $E_{m,sol}$ of this heme complex were also taken from measurements on microperoxidase.[150,151] The hydroxyl $pK_{a,sol}$ is 9.6 in the ferric and 10.9 in the ferrous heme. $E_{m,sol}$ is -120 mV for His-water heme and -200 mV for hydroxyl-His heme.[43,152,153]

The reference reaction field energy ($\Delta G_{rxn,sol}$) of the heme complex was precalculated by DelPhi for each ligand-heme complex by moving the heme with the axial ligands from dielectric constant of 4 to 80 in the reduced and the oxidized states. In calculations with a protein dielectric constant of 8, the heme complex was moved from e=8 to 80 in Delphi to generate $\Delta G_{rxn,sol}$. The Cys peripheral ligands were included in the atomic structure of the c-type heme complex. The propionic acids, which are considered as independently ionizable groups, were removed. The same procedure was carried out to generate the solution reaction field energy for the ionized and neutral acids and bases.[60]

The propionates are given at least two protonated conformers, with a proton on either carboxylic oxygen, and one deprotonated conformer.[7] Each of these can have multiple heavy atom rotamers if these can be accommodated in the protein structure. Protonated conformers have a total charge of zero, while deprotonated ones have -1.00. A

standard solution pK_a of 4.9 for carboxyl group was used for propionates.[154] The reference reaction field energies are -2.5 kcal/mol for the protonated conformers and -18.23 kcal/mol for the deprotonated conformer.

The default calculations use a metal centered charge distribution on the hemes.[4] This has -0.5 charge on each heme ring nitrogen atom and +2 or +3 charge on the iron in the reduced or oxidized state respectively.[4] Ligand side chains have PARSE charges on each atom, with a net charge of zero in both the neutral and ionized heme complex. In addition, a Mulliken charge set was obtained by Gaussian03 calculation with HF/6-31G and B3LYP/6-31G basis sets for the His-Met *c*-type heme[155] and results were compared with those obtained with the metal centered charge. Here charges of +1.20 and +1.34 were assigned to the iron in reduced and oxidized heme. All ligands, His, Met, Cys, have a non-zero net charge in each state. The total heme complex (including two propionates) has -2 and -1 charge. With the heme reduced, the covalently attached propionic acid each has an assigned net charge of -1. However, when the heme is oxidized there is a charge shift diminishing the assigned propionic acid charge to -0.81. A -1.00 charge is assigned to each of the propionates in the calculation regardless of the heme oxidation state. The missing +0.38 charge was evenly added to each atom in the plane of the oxidized heme. Electrostatic potential charge (ESP) and natural bond orbital (NBO) charges for the bis-His *c*-type heme were also obtained using Jaguar[156] (Zhang et al. in preparation*)*. The iron should have a positive atomic charge. Thus with positively charged heme iron of +0.69 and +1.04 in reduced and oxidized, NBO charges appear to be more physically reasonable than a -0.65 and -0.23 charge on iron in ESP.

2.2 Calculated contribution of different electrostatic energy terms to E_m

MCCE interprets the in situ midpoint redox potential as shifted from solution by the protein because of the loss in the self (reaction field) energy ($\Delta\Delta G_{rxn}$), which always favors heme reduction raising E_ms,[64,65] and pair-wise electrostatic interactions with backbone dipoles, propionic acids and charged or polar side chains (ΔG_{pol}, $\Delta G_{res,prop}$ and $\Delta G_{res,prot}$) (Eqn. 3, Table 2-1, Figure 2-1).[7,43,145] The method allows decomposition of these electrostatic terms to determine what features yield the observed electrochemistry in each protein. This new analysis of proteins with an E_m range of 800 mV overturns some conclusions from an earlier MCCE study analyzing a group of higher potential hemes with $E_{m,expt}$ spanning 450 mV from -102 to 350 mV.[7]

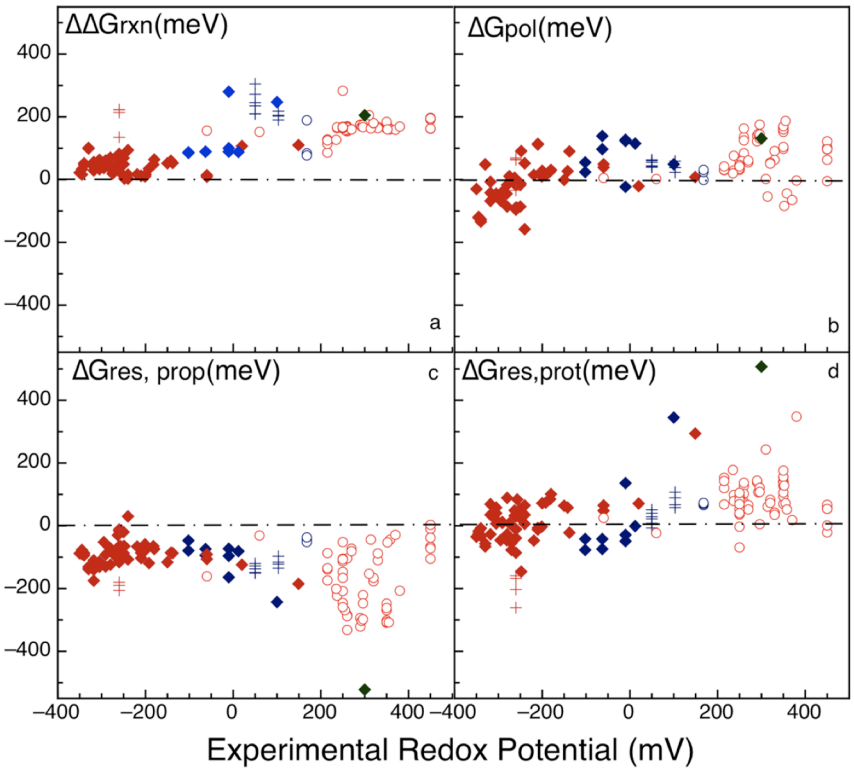

Figure 2-1 Energy terms contributing to the $E_{m,calc}$ vs. $E_{m,expt}$.

(a) $\Delta\Delta G_{rxn}$: loss of reaction field energy (desolvation energy); pairwise interaction with (b) ΔG_{pol}: backbone dipoles; (c) $\Delta G_{res,\ prop}$: propionic acids; (d) $\Delta G_{res,\ prot}$: sidechains. The slope and R^2 values for each figure can be found in table 2-1.

	# proteins	# pdb		$E_{m,expt}$	$E_{m,cal}$	$\Delta\Delta G_{rxn}$	ΔG_{pol}	$\Delta G_{res,prop}$	$\Delta G_{res,prot}$
All	42	63	Range	[-350,450]	[-420,336]	[3,305]	[-158,187]	[-522,30]	[-261,507]
			AVE±STD	0±263	-77±207	113±78	24±74	-128±90	53±107
			Slope	n/a	0.728	0.198	0.173	-0.165	0.199
			R^2	n/a	0.898	0.493	0.373	0.223	0.302
			Slope*	n/a	0.969	n/a	n/a	n/a	n/a
			R^{2*}	n/a	0.931	n/a	n/a	n/a	n/a
bis-His *a*	1	1	Range	300[$]	261	205	131	-522	507
bis-His *b*	3	7	Range	[-102,100]	[-268,100]	[85,280]	[-23,139]	[-243,-47]	[-77,345]
			AVE±STD	-12±69	-85±116	147±91	71±59	-120±70	54±160
			Slope	n/a	0.913	0.361	0.267	-0.490	0.988
			R^2	n/a	0.358	0.107	0.114	0.322	0.251
bis-His *c*	14	16	Range	[-350,149]	[-420,2]	[3,110]	[-158,114]	[-185,30]	[-146,216]
			AVE±STD	-230±104	-260±85	47±30	-13±62	-89±40	17±73
			Slope	n/a	0.674	0.046	0.272	-0.058	0.427
			R^2	n/a	0.568	0.024	0.164	0.020	0.371
His-Met *b*	1	2	Range	168	146	117	18	-44	70
His-Met *c*	24	33	Range	[-60,450]	[0,336]	[86,283]	[-84,187]	[-332,3]	[-69,348]
			AVE±STD	287±101	133±77	169±29	56±75	-168±96	92±77
			Slope	n/a	0.490	0.089	0.239	-0.008	0.173
			R^2	n/a	0.336	0.091	0.083	0.000	0.046
aqua-His *b*	2	10	Range	50,103	20,67	246,204	50,40	-139,-117	24,80
aqua-His *c*	1	1	Range	-260	-186	199	40	-147	-198
b	6	19	Range	[-102,168]	[-268,201]	[77,305]	[-23,139]	[-243,-37]	[-77,345]
			AVE±STD	28±86	-31±127	161±81	59±50	-113±61	55±128
			Slope	n/a	1.344	0.183	-0.143	0.003	0.593
			R^2	n/a	0.687	0.030	0.064	0.000	0.234
c	34	44	Range	[-350,450]	[-420,336]	[3,283]	[-158,187]	[-332,30]	[-261,348]
			AVE±STD	-10±277	-91±211	101±68	17±75	-124±79	46±88
			Slope	n/a	0.714	0.190	0.175	-0.156	0.180
			R^2	n/a	0.923	0.653	0.399	0.239	0.347

bis-His	18	24	Range	[-350,300]	[-420,261]	[3,280]	[-158,139]	[-522,30]	[-146,507]
			AVE±STD	-183±155	-219±141	69±74	3±70	-104±81	35±116
			Slope	n/a	0.755	0.289	0.301	-0.252	0.406
			R^2	n/a	0.749	0.452	0.372	0.283	0.348
His-Met	24	33	Range	[-60,450]	[0,336]	[77,283]	[-84,187]	[-332,3]	[-69,348]
			AVE±STD	282±102	133±76	167±30	54±74	-164±97	91±75
			Slope	n/a	0.425	0.121	0.264	-0.127	0.172
			R^2	n/a	0.279	0.131	0.109	0.010	0.050
aqua-His	3	11	Range	-260,50,103	-186,20,67	199,246,204	40,50,40	-147,-139,-117	-198,24,80
low-potential	13	16	Range	[-350,-102]	[-420,-99]	[-10,223]	[-158,114]	[-206,30]	[-261,101]
			AVE±STD	-251±63	-267±73	50±37	-12±64	-86±38	-3±64
			Slope	n/a	0.877	0.015	0.561	0.176	0.140
			R^2	n/a	0.463	0.000	0.272	0.055	0.012
mid-potential	10	22	Range	[-63,168]	[-230,201]	[9,305]	[-23,139]	[-243,-31]	[-74,345]
			AVE±STD	35±78	-28±118	146±78	41±51	-120±59	75±121
			Slope	n/a	1.213	0.258	-0.127	0.108	0.490
			R^2	n/a	0.508	0.045	0.039	0.022	0.127
high-potential	23	31	Range	[215,450]	[0,336]	[86,283]	[-84,187]	[-522,3]	[-69,507]
			AVE±STD	314±56	148±85	178±50	63±76	-188±116	116±109
			Slope	n/a	0.737	0.210	0.246	0.092	0.228
			R^2	n/a	0.362	0.118	0.048	0.002	0.026

Table 2-1 Experimental and MCCE calculated E_m range for each ligand and heme type (mV)

MCCE calculated energy terms as discussed in methods (meV). Slope*, R^2*: Slope and R^2 after adding a +160 mV offset for His-Met c type hemes. $ $E_{m,expt}$ for bis-His a varies from 300 mV, 340 mV to 430 mV in different species. The value for *Rb. sphaeroides* is used in the figures and data analysis.

ΔG_{rxn}: Moving into the protein from the water always destabilizes the more highly charged oxidized heme more than the reduced, raising the E_m.[64,65] Deeply buried hemes suffer a bigger loss than more exposed ones. The variation of heme exposure to solvent

has been suggested to be a primary determinant of the in site E_m.[157-159] However, previous CE[160] and MCCE[7] studies found a small range of $\Delta\Delta G_{rxn}$ of only ~130 meV in hemes with an $E_{m,expt}$ range of ~ 450 mV and thus suggested that variation in solvation energy is not the dominant factor in determining heme E_ms. Over the wider E_m range covered here, $\Delta\Delta G_{rxn}$ varies from 3 meV in small soluble tetra-heme cytochrome *c* protein ($E_{m,expt}$ -248 mV)[161] to 305 mV in the sperm whale myoglobin ($E_{m,expt}$ 50 mV)[162]. This 302 meV range, while significant, is still insufficient to account for the whole range of $E_{m,expt}$ of 800 mV. In addition, the correlation between $\Delta\Delta G_{rxn}$ and $E_{m,expt}$ shows a slope of only ~0.2 with an R^2 of 0.493 (Table 2-1).

ΔG_{pol}: Each amino acid has an associated backbone amide group, which has a dipole moment larger than that of water. Earlier electrostatic surveys of protein electrostatics found that on average the protein backbone dipoles are oriented in all proteins to make the interior more positive.[163] In previous MCCE analysis of heme E_ms, interactions with the protein backbone were always found to destabilize heme oxidation, raising the redox potential.[7] Here, as the potential range is expanded to low potential hemes and new high potential proteins are added, there are hemes found where the backbone dipoles favor oxidation. ΔG_{pol} ranges from -158 meV in *Thermosynechococcus elongatus* cytochrome c_{550} ($E_{m,expt}$ -240 mV[164]) to 187 meV in *Rhodobacter sphaeroides* cytochrome c_2 ($E_{m,expt}$ 355 mV[165]). However, overall there is little correlations between ΔG_{pol} and $E_{m,expt}$ (Table 2-1).

Small differences in protein structure can lead to significant differences in ΔG_{pol}. For example, *B. taurus* cytochrome b_5 ($E_{m,expt}$ -10 mV[166]) and *T. elongatus* cyt c_{550} ($E_{m,expt}$ -240 mV[164]) both have a Pro adjacent to the His axial ligand, which makes the

most significant contribution to ΔG_{pol} (Figure 2-2). In cytochrome b_5 the Pro raises the E_m by ≈50 mV while in cytochrome c_{550} the analogous residue lowers it by a similar amount. This difference results from the orientation of the Pro carbonyl, which points towards the heme in cyt. c_{550} and away in b_5. The backbone is the one structural element that is not optimized in MCCE conformational sampling. The comparison of the results calculated with different atomic structures provides the only way to access the variability of ΔG_{pol} for a given protein.

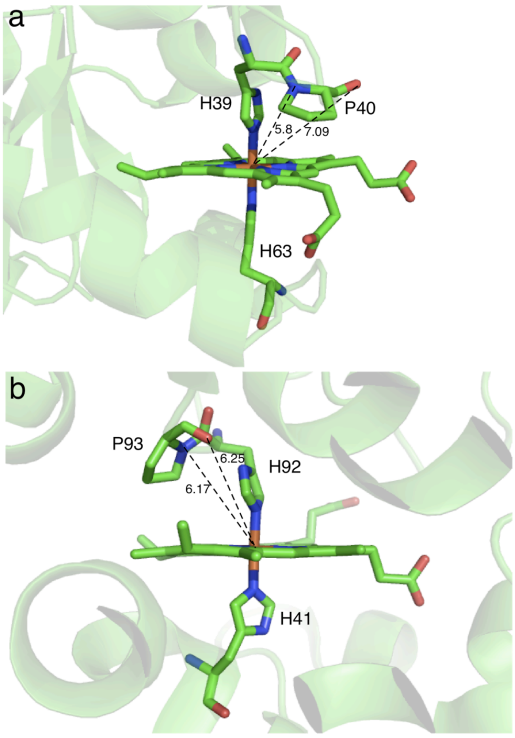

Figure 2-2 Residue backbone dipoles with strong interaction with the heme.

(a) bovine b_5 1CYO with a total ΔG_{pol} of 123 meV. (b) Cyt. c_{550} 1MZ4 with a total ΔG_{pol} of -158 meV. Distances between the iron and the carbonyl group, amine are labeled in Å. The two axial Histidine ligands and one Proline adjacent to the ligand are shown.

$\Delta G_{res,prop}$: All hemes have two propionic acids as peripheral ligands. When ionized, they lower the midpoint potential, favoring heme oxidation. Strong unfavorable

interactions between the two propionates can keep them from both being fully ionized. The more deeply buried acids lose reaction field energy, which destabilizes ionization and also increase the strength of the repulsion between the two propionates.

Oxidation of the hemes always shifts the pK_a of the acid down, leading to proton release if the pK_a is near the solution pH.[7,167, 51] The net charge on the two propionic acids is <–1 (>50% ionized) in 114 out of 141 (81%) hemes in the reduced state, with 40% of the propionic acids more than 90% ionized. When the hemes are oxidized now 129 hemes (91%) have their propionic charge between -1 and -2 and 65% are >90% ionized. The associated proton release on heme oxidation makes the propionic acids important contributors to the pH dependence of many heme E_ms.

The $\Delta G_{res,prop}$ ranges from ≈0 to –522 meV. This is calculated at pH 7 at the E_h where the heme is 50% ionized in Monte Carlo sampling (i.e. $E_h = E_{m,calc}$) (Eqn 1). The lowest value is -522 meV for a deeply buried propionic acid on Heme *a* in cytochrome *c* oxidase. The deeply buried acid remains charged due to its interactions with two Arg (R481, R481) less than 3Å away. Another nearby Arg (R52), a Mg^{+2} and a Ca^{+2} ions also help stabilize both ionized heme acids. Of course the positive potential from the nearby cations will raise the potential at the heme. With the exception of cytchrome c oxidase, the most favorable heme-propionate acid interaction is –332 mV in cyt *c* ($E_{m,expt}$ 260[168]). In the low potential cyt c_{550} ($E_{m,expt}$ -240 mV[164]) and high potential cyt c_4 ($E_{m,expt}$ 450 mV[169]), the propionates are fully protonated at pH 7, so the polar, neutral acids have very small unfavorable interaction with the oxidized heme of 30 and 3 meV respectively.

The heme redox potential correlates poorly with the fractional propionate ionization or with the electrostatic interaction between these attached acids and the rest of the heme (Table 2-1, Figure 2-1c). Interestingly, there is a weak negative correlation between $\Delta G_{res,prop}$ and $E_{m,expt}$. This arises because deeply buried, ionized propionic acids, which themselves lower the E_m, must in turn be surrounded by cations which will tend to raise the heme E_m.

Earlier studies by Mao calculated that, in reduced cytochromes, the acids have an interaction with the hemes ranging from -40 to -170 meV, while with an oxidized heme, it is more favorable, ranging from -60 to -270 meV.[7] The interactions reported here on the same proteins are more favorable. The differences are due to improvements in the MCCE methodology. The earlier calculations used an ad hoc SOFT function,[60] which weakened strong electrostatic interactions. This was necessary because of the added low dielectric material introduced by the multi-conformation method which has all conformers present when calculating pairwise interactions.[8,60] In contrast, the current method uses a more accurate, physical correction for errors in the boundary conditions. The pairwise electrostatics interactions between the single conformers from the original structure are considered correct. The same interaction is also calculated in the multiconformation structure and correction is defined to recover the correct value. The same factor applied to all the conformer-conformer electrostatics pairwise interactions between the two residues.

$\Delta G_{res,prot}$: The ionizable and polar amino acids in the protein also modify the heme redox potential via electrostatic pairwise interactions. Here, $\Delta G_{res,prot}$ varies from -261 meV in the di-heme peroxidase ($E_{m,expt}$ -260 mV[38]) to 507 meV in cyt. *c* oxidase, much

larger than the range from -60 to 165 meV found by Mao et al[7]. As seen in Table 2-1 and Figure 2-1d, there is no correlation between the $\Delta G_{res,prot}$ and redox potential.

2.3 Comparison of the electrochemistry of the different classes of heme protein

2.3.1 b- vs. c-type

Hemes are attached to the protein by purely non-bonded interactions in the a- or b-type hemes or via two Cys linkages in the c-type hemes. In addition, the protein provides one or two axial ligands to the heme. The *b*-type heme, the most common heme tetrapyrrole macrocycle, is the structure from which heme *a* and *c* are derived[17,18]. The dataset here includes one a-type heme, 118 c-type hemes and 22 b-type hemes. Only bis-His, His-Met and aquo-His hemes with measured E_ms were examined here producing a dataset with more c-type than b-type hemes. The b-type hemes are more likely to be found with other ligands.[17]

The b- and c- type hemes are found clustered in different E_ms ranges. The *b*-type hemes have their E_ms in the middle of the E_m range measured for heme proteins, while *c*-type hemes are mainly found either above 250 mV or below -100 mV. Thus, the *c*-type hemes span an 800 mV $E_{m,expt}$ range, from -350 mV in *Desulfovibrio desulfuricans* Norway cytochrome c_3[170] to 450 mV in *Nitrosomonas europaea* peroxidase[38], while *b*-type hemes span less than 300 mV, ranging from -102 mV in *Rat microsonal* cytochrome b_5[171] to 168 mV in *Escherichia coli* cytochrome b_{562}.[172] It can be seen that each of the electrostatic terms are larger for the bis-His b type hemes than for the more exposed bis-His c-type hemes (Table 2-1). The covalent attachment of the heme to the protein via two Cys linkages has been assumed to modify the E_m only by the small shifts in the reference

reaction field energy. This is supported by by the small shift in E_m measured when the c-type heme in cytochrome c_{552} is mutated to a b-type heme by replacing both Cys ligands by Ala.[173] Thus, the difference between the $E_{m,calc}$ for the bis-His b- and c-type hemes is calculated to be solely due to electrostatic terms.

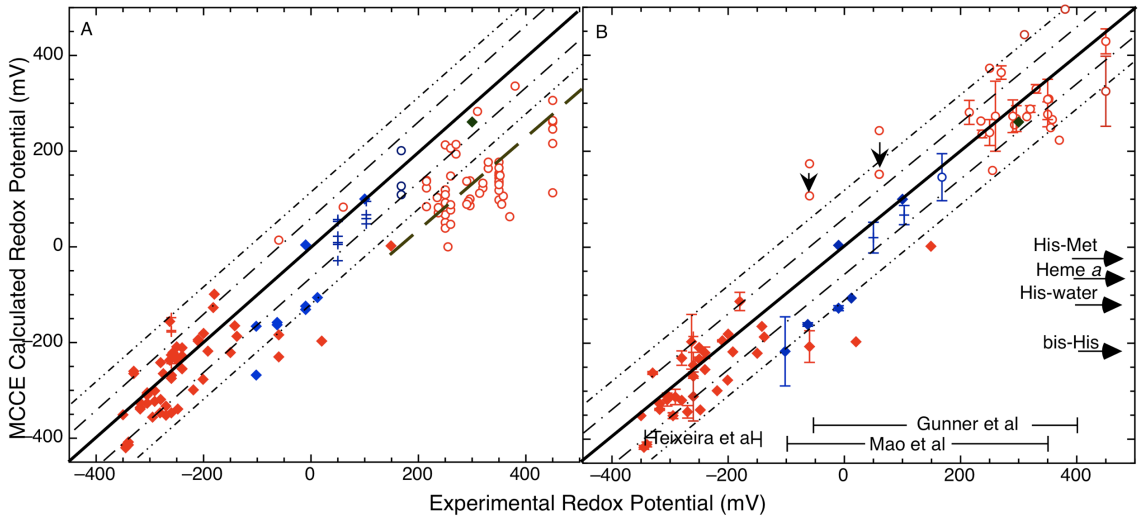

Figure 2-3 MCCE calculated vs. experimental redox potential (mV).

(A) All $E_{m,calc}$ vs. $E_{m,expt}$ (mV). (B) Averaging $E_{m,calc}$ for all proteins with multiple structures available (mV). Error bars show standard deviation of calculated E_ms. A +160 mV offset is added to all calculated His-Met *c* hemes. The two black arrows pointing downwards indicate the shifts in $E_{m,cal}$s when the propionic acids were forced to be fully ionized in *Rps. viridis* reaction centers and *B. pasteurii* cyt c_{553}. ◆: bis-His *a* type. ◆: bis-His *b* type. ◆: bis-His *c* type. ○: His-Met *b* type. ○: His-Met *c* type. +: aqua-His *b*. +: aqua-His *c*. Central black solid line is where experimental and calculated E_ms are equal. The dashed line in (A) is the 160 mV offset for His-Met *c*-type. The black dot dashed and double dot dashed lines are the ±60 and ±120 mV error lines. The arrows pointing to the y-axis show the $E_{m,sol}$s used in this calculation, which are -220 mV for bis-His, -120 mV for His-water, -60 mV for bis-His *a* and -15 mV for His-Met. The range of previous heme E_ms calculations is also shown.[4,6,7]

Independent of the axial ligation, the E_ms for the b-type hemes are ≈200 mV more positive than the $E_{m,sol}$ for the respective ligand type (Figure 2-3, Table 2-1). The heme ligand types studied here have $E_{m,sol}$ of -220 mV, -120 mV and -15 mV and for bis-His, aquo-His and His-Met ligands respectively. The average $E_{m,expt}$ is -12±69 mV for bis-His hemes (a shift of 208 mV from $E_{m,sol}$), 168 mV for the single His-Met heme (183 mV shift) and 76±27 mV for aquo-His hemes (196 mV shift). A positive E_m shift, with its destabilization of the oxidized state of the heme relative to solution, indicates the protein binds the reduced heme more tightly than the oxidized form. A 200 mV E_m shift represents a 2000 fold difference in affinity, representing a 4.5 kcal/mol difference in binding affinity.[11,174] In contrast, for *c*-type hemes, the bis-His ligated hemes have an average $E_{m,expt}$ of -230±104mV, shifting on average only 10 mV from $E_{m,sol}$. Thus, on average, the oxidized and reduced states are bound with similar affinity. The single aquo-His *c*-heme has an $E_{m,expt}$ of -260 mV, lower than the $E_{m,sol}$ of -120 mV. While, the His-Met hemes have an average E_m more than 500 mV higher. Their average $E_{m,expt}$ of 287±101 mV represents a shift of 302 mV from $E_{m,sol}$.

The biological selection of *b*- and *c*-type hemes for different functions is still a matter of debate.[18,174-176] With two covalent thioether attachments, the *c*-type hemes will bind essentially irreversibly to the protein. The covalent attachment appears to allow an increase in the ratio of heme to protein. This seems especially important for the group of low potential multi-heme proteins. For example in the c_3 proteins, four *c*-type hemes are successfully accommodated into a compact but elegant soluble protein of only ~110 residues. Overall the low potential *c* cytochromes have from 23 to 135 amino acids/heme. It is hard to imagine that without the covalent attachment there would be sufficient non-

bonded interactions between each heme and the protein. However, the small covering of protein does not interact much with the heme, yielding the smallest values for each of the MCCE electrostatic energy terms (Table 2-1, Figure 2-1). Thus, the low potential c-type cytochromes have relatively small shifts in E_m from $E_{m,sol}$.

The high potential *c*-type hemes, which employ His-Met ligation, may use a covalently bound heme for a different reason. The His-Met ligands have a more positive $E_{m,sol}$. However, on average the in situ $E_{m,expt}$s are 300 mV more positive than the His-Met $E_{m,sol}$ of –15 mV. This indicates the oxidized heme binds 10^6 times more weakly than the reduced cofactor. In general the oxidized hemes tend to be more weakly bound, potentially destabilizing the protein.[177] As suggested by Gibney and coworkers[174] the covalent attachment in the high potential c-type proteins may be needed to ensure reversible redox reactions with a very positive E_m shift without heme loss. These proteins with their greatly shifted E_ms tend to have large values for all MCCE electrostatic energy terms.

The *c*-type hemes do generally show somewhat better R^2 values than *b*-type hemes for the comparison of the decomposed electrostatic energy terms with the experimental E_ms. However this is largely due to the fact that *c*-type hemes cover a significantly larger range of E_ms. Analyzing an E_m range from -138 to 149 mV in *c*-type hemes similar to that found in the *b*-type, the slopes for each energy term in the c-type hemes do not change much. However, the R^2 values drop significantly. They are now 0.225 and 0.4 for $\Delta\Delta G_{rxn}$ and ΔG_{pol}, a little better for *b*-type hemes, and 0 and 0.027 for $\Delta G_{res,prop}$ and $\Delta G_{res,prot}$, which are even worse.

2.3.2 bis-His vs His-Met vs aquo-His

The sixty-four His-Met type hemes have an $E_{m,expt}$ span of 510 mV, from -60 mV (in tetra-heme *Rhodopseudomonas viridis* reaction center[178]) to 450 mV (in di-heme *Nitrosomonas europaea* peroxidase[38]). In the sixty-three bis-His type hemes $E_{m,expt}$ varies from -350 mV (in tetra-heme *Desulfovibrio desulfuricans Norway*[170]) to 300 mV (in *Rhodobacter sphaeroides* cytochrome *c* oxidase[179]). The fourteen aquo-His ligated hemes have a smaller $E_{m,expt}$ range from -260 mV (*Nitrosomonas europaea* peroxidase[38]) to 103 mV (*monoeric clam* hemoglobin[162]). There are only two proteins with aquo-His *b* hemes, and one protein each with aquo-His *c* or His-Met *b* hemes. Overall, although hemes in proteins have a total E_m range of 800 mV here, proteins with a particular heme and ligand type are likely to be found in a range of 200-500 mV. In the calculations, the different axial ligand types are assumed to add an non-electrostatic offset via the differences in $E_{m,sol}$.

For the c-type hemes the low and high potential hemes can be compared. Their average E_ms vary by ~500 mV. This 500 mV range of $E_{m,expt}$, equivalent to an 11.3 kcal/Mol shift in the reaction equilibrium constant is well calculated by the MCCE method (Figure 2-3).The high potential hemes use His-Met rather than bis-His ligands, raising $E_{m,sol}$ by 205 mV; the high potential hemes have ≈120 mV less solvation energy, 150 mV more positive interaction with the backbone dipoles and the sidechains. The more buried c-type hemes also have larger interactions with their propionic acids lowering the potential by ≈80 mV. However, as will be described below MCCE does not account for all of the destabilization of the oxidized His-Met hemes, resulting in $E_{m,calc}$ which are systematically lower than $E_{m,expt}$ (Figure 2-3A).

There are several hemes, which are shifted more significantly by the protein than the norm for the given heme and ligand type and these pose the best test of the MCCE calculation method. While the majority of bis-His *c* type hemes are at low potential, ranging from -350 to -60 mV, one in the 400 residue reaction center cytochrome c-subunit and one in the 500 residue heme binding subunit of the quinohemoprotein amine dehydrogenase (QH-AmDH) have a much higher $E_{m,expt}$ of 20 and 149 mV respectively. These are calculated to have E_ms near 0 mV, well separated from the low potential bis-His b-type hemes. Traditionally, hemes are predicted to move to high potential because of the loss of reaction field energy, which destabilizes the oxidized heme.[64,65,180] This is in agreement with our observation here. The electrostatic analysis of these two high potential bis-His c-type hemes shows that they are better buried in the protein than the average bis-His c-type heme with $\Delta\Delta G_{rxn}$ of ≈60 meV, ≈80 meV more positive than the average value for this group. However this is not enough to raise the heme E_m in QH-AmDH by ~400 mV from the averaged E_m of its type. Therefore other forces must be enhancing heme reduction. The pair-wise electrostatic interactions with the protein sidechains are also much stronger in QH-AmDH. Four Arginines (R42, R102, R108, R114) near the heme make a total contribution of ~400 meV raising the $E_{m,calc}$. However, this is largely canceled by a ~130 meV interaction with the nearby heme propionic acids.

His-Met *c*-type hemes usually have high $E_{m,expt}$, in the range of 250 to 450 mV. There are two with unusual low redox potentials, at 60 meV in cyt. c_{553} and -60 mV in the *Rb. viridis* reaction center. Solvent heme exposure has been suggested to be one of the most important factors responsible for the uncommon low cyt. c_{553} E_m.[159] However, the calculations here show that $\Delta\Delta G_{rxn}$ of 152 meV, is only ~20 meV lower than the average

of His-Met *c* type. Comparing these hemes with the average for the members of their class (Table 2-1) shows that ΔG_{pol} and $\Delta G_{res,prot}$ are the terms which yield the lowered $E_{m,calc}$.

2.4 Error Analyses and Parameter Sensitivity

2.4.1 Eorror Analyses

MCCE calculation results are in reasonable agreement with the experimental data. The slope comparing all experimental and calculated values is 0.73 with an intercept at -77 mV and a reasonable R^2 of 0.90. Thus, the calculations are able to reproduce an E_m range of 756 mV, from -420 mV in cytochrome c_3 to 336 mV in reaction center. By using the experimental values for $E_{m,sol}$ for each ligand type there are no free parameters in the analysis. Overall this reproduces ≈700 mV of the 800 mV found experimentally (Figure 2-3A). This is the first attempt to use a single method to calculate such a large range of heme E_ms. All atomic force field parameters used here are the same as used for MCCE calculations of pK_as in proteins.

Despite the good agreement between calculation and experiment (Figure 2-3), the His-Met *c* type E_ms, are systematically lower than the experimental values. If the $E_{m,calc}$ for each of these 24 proteins is shifted up by 160 mV, the slope matching the experimental and calculated values significantly increase to 0.969 with an R^2 of 0.931 (Table 2-1, Figure 2-3B). With the correction, 65% of the hemes have an absolute error smaller than 60 mV, equivalent to an error of 1 pH unit (1.36 kcal/mol) in a calculation of a group pK_a and 92% are within 120 mV. This can be compared with the most recent version of MCCE analysis of the pK_as of 320 amino acids in 38 proteins where 79% of the residues have errors smaller than 1 pH unit and 97% errors smaller than 2 pH

units.(ref) However, the calculation of heme electrochemistry may represent a more difficult problem since the heme is generally well buried in the protein. In contrast the large majority of amino acids used in pK_a benchmark studies are on the protein surface where they are perturbed little by the protein.[181]

There are several possible sources of the systematic shift of the His-Met hemes. $E_{m,sol}$ influence all hemes of the same type in the same way. The $E_{m,sol}$ used here are the same as applied in earlier MCCE analysis of heme electrochemistry.[7,43] $E_{m,sol}$ for the bis-His and aquo-His hemes are taken from measurements on microperoxidases. The $E_{m,sol}$ of -15 mV for His-Met hemes is derived from the E_m shift found by mutating one His ligand to a Met in cyt *c*[40,182], in cyt c_{551}[41] and in cyt c_3[183]. The $E_{m,sol}$ for His-Met microperoxidase is -70 mV.[39] Using this value would lower the $E_{m,calc}$ by 55 mV, increasing the needed offset.

Heme Type	Ligand Type	Charge Set	$G_{rxn,sol,ox}$	$G_{rxn,sol,red}$	$\Delta G_{rxn,sol}$
c	bis-His	Metal Centered	**-511**	**-260**	**-251**
		ESP[a]	-439	-204	-235
		NBO[b]	-574	-435	-139
	His-Met	Metal Centered	**-468**	**-136**	**-332**
		Quantum[c]	-478	-126	-352
b	bis-His	Metal Centered	**-510**	**-232**	**-278**
	His-Met	Metal Centered	**-473**	**-124**	**-349**
Mao et al[d]		Metal Centered	-430	-20	-410

Table 2-2 Reference reaction field energy (meV) using different charge sets calculated with ε_{in}=4, ε_{out}=80.

Charges derived from Jaguar calculations using [a] ESP: electrostatic surface potential method. [b] NBO: natural bond orbital theory. [c] Charges obtained with Gaussian from Autenrieth et al. [d] Reference reaction field energy used in Mao et al. for both bis-His and His-Met ligands. Numbers in bold are used as default in calculation.

Another term that is used for all calculations on a given type of heme is the reference reaction field energy $\Delta G_{rxn,sol}$ (Table 2-2). This provides the maximal stabilization of the heme ligand complex when it is immersed in water. A larger difference between the reaction field energy of oxidized and reduced heme allows greater positive shifts of the E_m as the heme is buried. The default $E_{m,calc}$ groups the heme and its axial ligands, with standard PARSE charges, together as a heme complex using the metal centered charges. However, in earlier calculation from this laboratory the ligands carried no charge.[7] Despite the identical shape and net charge on the heme complex the distribution of atomic charges produces a large change in the solution reaction field energy, especially for the neutral reduced heme. In addition, the amino acid axial ligand charges produce a significant difference between the $\Delta G_{rxn,sol}$ for the bis-His and His-Met hemes. Making the solution reaction field energy less negative will shift the $E_{m,calc}$ to more negative values, because there can be a smaller penalty for oxidation of a buried heme. However, this difference again tends to lower the bis-His hemes more than the His-Met hemes so does not explain the result that the His-Met hemes are systematically too low.

As $\Delta G_{rxn,sol}$ is quite sensitive to the atomic charge distribution and radii used, it would be helpful to have an experimental system to use as a model. Formally the $\Delta G_{rxn,sol}$ would be measured by the free energy of transfer of the oxidized and reduced heme ligand complexes from water into a solvent with a dielectric constant of 4.[54] Alternately,

the difference in the transfer energy of oxidized and reduced hemes can be determined from the E_m shift of the heme measured in the two solvents.[71] E_m shifts have been measured for heme in a dendritic microenvironment in water (ε~80), MeCN (ε~40) and CH_2Cl_2 (ε~10).[184] This system shows increasing heme E_ms when the dendrite shell increased or when the solvent dielectric constant decreased as expected from a continuum electrostatic analysis. The smallest model system is viewed as most similar to the microperoxidases used here as the models for $E_{m,sol.}$ The measured E_m is -210 mV in CH_2Cl_2, 80 mV higher than in water. The calculated change in $\Delta G_{rxn,sol}$ moving from ε=4 to ε=80 was compared with the transfer from ε=4 to ε=10. With the charge distribution on the heme-ligand complex used here, the $\Delta\Delta G_{rxn,sol}$ is 96 meV for *b*-type hemes and 80 meV for *c*-type hemes, in agreement with the dendritic model system measurements. In contrast, the charge distribution used previously with no charges on the ligands gives a much larger $\Delta\Delta G_{rxn,sol}$ of 200 mV. Thus, the current charge distribution yields more reasonable solution reaction field energies. However, the experimental model system does not provide any information about possible differences between bis-His and His-Met hemes.

Another possible cause of the $E_{m,calc}$ for His-Met c-type hemes being lower than $E_{m,expt}$ is there that there is an error in the calculated ionization state of the propionates. The His-Met c-type proteins have the largest calculated $\Delta G_{res,prop}$ which lowers $E_{m,calc}$ by 80 mV, on average, more than they do in the bis-His *c*-type hemes. The calculated net charge on the two acids is often between -1 and -2. If the calculated pK_as were too low and the acids too ionized this will lower the E_ms.

2.4.2 Parameter sensitivity

The sensitivity to the propionic acid ionization can be seen in the comparison of the results of calculations on each structure in an ensemble of 40 NMR structures of reduced *horse heart* cytochrome *c* (pdb 2GIW, $E_{m,expt}$ 260 mV[168]). The calculated E_ms range from -14 mV to 255 mV with an average of 107 mV. This averaged value is close to the $E_{m,calc}$ obtained starting with a crystal structure. Comparing the NMR structures with the lowest and the highest calculated E_m, shows that the structure with the $E_{m,calc}$ of -14 mV has one propionate fully and the other 72% deprotonated at the midpoint of the titration. Meanwhile, the structure with an $E_{m,calc}$ of 255 meV only has one propionate ionized. This difference contributes 80 mV to the difference in $E_{m,calc}$. There is also a ≈200 meV difference in the interaction with the backbone dipoles, with the backbone of the Met80 ligand contributing 70 meV to the difference. The MCCE method can sample different propionic acid conformations, reducing the differences between different starting structural models.[60] However, differences between backbone conformations are not sampled in the calculations. This degree of difference found for an ensemble of NMR structures is not unusual in MCCE. However, the averaged pK_a[43,60] or E_ms[7] are generally similar to those found for x-ray crystal structures. The results from each structure in the ensemble often provide interesting information about how changes in conformation can affect the result electrochemistry.

The shifts in $E_{m,calc}$ with ionization of the propionic acids can also be seen in the lowest potential His-Met c-type hemes, both in *Rps. viridis* reaction centers ($E_{m,extp}$ -60 mV[178]) and in cyt c_{553} of *B. pasteurii* ($E_{m,expt}$ 60 mV[185]). In both cases the propionic acids are partially protonated. In cyt c_{553} the A ring propionic acid is trapped at the protein

surface and so only 20% ionized when the heme is fully reduced. The shifts in $E_{m,calc}$ obtained when both propionic acids on these hemes are fixed in their ionized form are shown in Fig. 1B. The $E_{m,calc}$ is now ~90 mV lower than before and reducing the absolute error to ~90 mV.

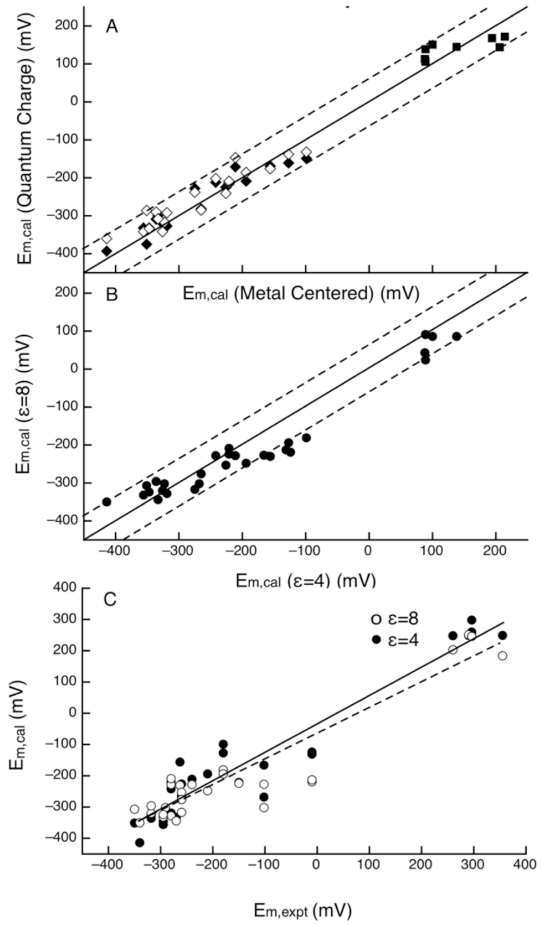

Figure 2-4 MCCE Calculated E_ms with different charge distribution and dielectric constants.

(A) (■) Calculated E_ms of His-Met hemes using quantum charge set from Autenrieth et al.[155] vs. metal centered charge set. Calculated E_ms of bis-His hemes using (◆) ESP and (◇) NBO charge sets vs. metal centered charge set. (B) (●) Calculated E_ms of selected proteins in ε=8 vs. ε=4. Two dashed lines are ±60 mV error lines. (C) Calculated E_ms using both ε=8 and ε=4 vs. experimental E_ms. Here, a +160 mV offset is added to all

calculated His-Met c hemes for better comparison with the experimental data. The dashed line and dark line are the best-fit linear regression for calculated E_ms with ε=8 (slope 0.84, R^2 0.89) and ε=4 (slope 0.92, R^2 0.92) respectively.

***Charge set*:** The sensitivity of the calculations to the heme-ligand complex charge distribution was tested (Figure 2-4A). The same reference reaction field energy calculated with the metal centered charge set was used for each ligand type.

Three charge sets for the bis-His hemes include the metal centered default set and ESP and NBO charges, which were derived from Jaguar calculation (Zhang et al. in preparation). The difference in calculated E_ms between the metal centered charge set and the two quantum charge sets are less than 65 mV, with both providing similar agreement with the measured data. The loss in reaction field energy ($\Delta\Delta G_{rxn}$) and backbone interactions (ΔG_{pol}) are rather similar with the different heme partial charges. The ionization states of the propionates shift very little. No significant conformation or ionization state changes in the protein were observed.

For His-Met hemes the metal centered charge set was compared with those obtained by previously published Gaussian calculation.[155] Small changes, which are no more than 62 mV in the calculated E_ms, were seen. The energy terms in MCCE calculation were quite constant between the two charge sets. Small variations in residue ionization states were seen. Nevertheless, its affect on the final redox midpoint potentials is small.

The influence of charge set on $E_{m,calc}$ used one value of $\Delta G_{rxn.sol}$. $\Delta G_{rxn,sol}$ is dependent on the charge distribution on the heme-ligand complex (Table 2-2). However,

except for the NBO charge set, the difference in $\Delta G_{rxn,sol}$ between charge sets for each ligand type is less than 20 meV. The NBO charge gives a reference reaction field energy ~100 meV lower than the other two, which would lower the $E_{m,calc}$ of all hemes by that much. This would tend to lower the absolute $E_{m,calc}$ but not the relative E_m for a given heme type.

Dielectric constant (ε): The value for the dielectric constant of the protein remains one of the most contentious issues for the continuum electrostatics methodology.[59,65,186-188] MCCE calculation was carried out using both a dielectric constant of 4 and 8 on a selected group of proteins (See Figure 2-4) with the default metal centered charges. $\Delta G_{rxn,sol}$ was calculated with the same ε_{in} as the protein. Of the 29 hemes selected in the calculation, five have their $E_{m,calc}$ shifted by more than 65 mV. Bovine cytochrome b_5 (pdb: 1CYO) changes by 95 mV. Here $\Delta G_{res,prot}$ change by ~13 meV. The heme loses ~50 meV less reaction field energy in ε=8 than in ε=4 but it interacts 60 meV more favorably with the backbone dipoles. As a result, the heme oxidation is favored by ~110 meV more in ε=8 than in ε=4 lowering the $E_{m,calc}$.

The change in dielectric constant shows the balance between reaction field energy and pairwise interactions. Thus, calculations with ε=4 always have a more significant loss in reaction field energy than with ε=8, raising the $E_{m,calc}$. The individual pairwise interactions will have a larger absolute value in the lower dielectric constant, which can lower or raise the $E_{m,calc}$ depending on the sign of a charged group or orientation of a dipole. In the larger, high potential His-Met c-type hemes the calculated $\Delta\Delta G_{rxn}$s with ε=4 are more unfavorable by 42 to 89 meV. While in the small, low potential-bis-His c-type heme proteins, $E_{m,cal}$ shifts only by -2 to 36 meV. The backbone interaction which

tend to raise the $E_{m,calc}$ in the higher potential hemes do so by a smaller amount with the higher dielectric constant. In the small proteins where this term is small in ε=4, relatively similar backbone interactions are seen with ε=8. In figure 2-4C, the calculated E_ms both using ε=8 and ε=4 were compared with measured data. Best-fit linear regression lines show that the results with ε=4 have a slope of 0.92 (R^2=0.92), while with ε=8 the slope is 0.84 (R^2=0.89). Thus, overall calculations with ε=8 underestimate the protein contribution to the heme E_ms.

2.5 Summary

It is a long-standing challenge to demonstrate how proteins modify the in situ properties of bound substrates and cofactors. The enormous range of heme E_ms in proteins allows us to test the hypothesis that the E_m differences in individual proteins arise predominately from difference in the electrostatic interactions of the cofactor with the protein and the surrounding solvent. Here with only one free parameter, $E_{m,sol}$ taken from experiments,[39-41] the calculations are shown to account for ≈73% of the measured change in E_ms. Thus, a good fraction of the observed range of heme E_ms in proteins can be assigned to perturbation of the reaction by changing the heme ligand and by the non-bonded electrostatic interactions with the protein and solvent, which are captured in the MCCE calculations. The free energy changes that arise from proton release upon heme oxidation are included in this analysis. The errors are not uniformly distributed. Rather the his-met c-type hemes are calculated to be too easily oxidized in their protein environment. A number of possible errors and uncertainties are considered here without satisfactorily identifying the source of the problem. Nevertheless, raising this group of E_ms with a 160 mV offset produces calculated results with a slope of 0.97 with an R^2 of

0.93 comparing calculated and experimental E_ms. With this correction 65% of the E_ms have errors of <1.4 kcal/mol and only 7% having errors >2.8 kcal/mol. The 800 mV span studied here represents the largest range of E_ms calculated using a single method. The results provide a better understanding of heme electrochemistry in proteins and may provide rules to help in engineering heme proteins with desired properties.

Given the success of the analysis in calculating the E_ms, it becomes possible to ask what factors are most important for modulating the in situ electrochemistry. The interactions between heme and protein are divided into the loss of heme solvation energy ($\Delta\Delta G_{rxn}$), and pairwise interactions of the heme with the covalently attached propionic acids ($\Delta G_{res,prop}$) and the sidechains ($\Delta G_{res,prot}$) and backbone of the protein (ΔG_{pol}). Each of these energy terms is quite different in the different proteins. However, each alone is poorly correlated with the E_m. It has long been suggested that the loss of solvation energy, which raises all E_ms would be the dominant factory in tuning the in situ electrochemistry.[157-159] While, the $\Delta\Delta G_{rxn}$, which is the continuum electrostatics measure of the solvation energy has the best correlation with the measured E_m, the R^2 is only 0.5 and the slope is 0.2. This indicates that no more than 20% of the E_m range can be assigned to changes in solvation energy. Proteins can also tune their E_ms by local differences in the backbone structure, by the distribution of nearby sidechains, by the ionization state and solvent exposure of their propionic acids and by the exposure of the heme itself to water. Overall this analysis shows that different proteins employ a different mixture of forces to regulate its heme E_m.

Chapter III. The Calculated Affinity of Substituted Benzoquinones for the Q_A site of Bacterial Reaction Centers

3.1 Methodology: Calculations of quinone affinity in MCCE

In MCCE analysis, the term 'conformer' specifies a specific position, protonation and redox state of a side chain or ligand. These are preselected during an initial sampling and optimization cycle and residues can have 1 to 100's of conformers to choose from. The protein backbone is held fixed. Monte Carlo sampling then determines the Boltzmann distribution of conformers. Thus, all acidic and basic amino acids throughout the protein can be either protonated or deprotonated unless otherwise stated, while the side chains and ligands sample all, pre-selected positions. The conformer distribution is determined given the Poisson-Boltzmann electrostatic interactions and ligand solvation energies, full AMBER Lennard-Jones and torsion energies[148] and an SAS based non-electrostatic ligand-solvent interaction energy. The ligand affinity is derived from Grand Canonical Monte Carlo (GCMC) Sampling where the ligand can be bound to the protein or free in solution.[189] A similar GCMC analysis has been applied previously to calculate the chloride affinity to three proteins using MCCE.[12]

The generation of the pre-selected side chain conformers for final Monte Carlo sampling uses a multi-scale approach. This aspect of optimization of side chain and ligand positions is novel in MCCE and is described in detail below. MCCE QUICK conformer placing is the default for the protein as a whole.[190] This retains the input non-hydrogen (C,O,N,S) positions, while allowing additional, essentially isosteric positions for hydroxyls, His tautomers, and the interchange of the O and N in the Asn and Gln side chains. MCCE analysis then focuses more sampling of side chain positions near the

quinone as described below. The program screens, optimizes and prunes these initial position conformers with an analytical energy function that includes AMBER torsion and Lennard-Jones energies as described in detail previously.[190] The electrostatic energies in the prescreening step are obtained using Coulomb's law with a dielectric constant of 6. In the GCMC ligand binding analysis, DelPhi[67] is applied to solve the Poisson-Boltzmann equation with a protein dielectric of 4 and 80 for water at pH 7 and with 150 mM salt using Parse charges and radii.[191] AMBER[148] non-electrostatic parameters are applied for the Lennard-Jones and torsion energies. The quinone affinities are measured with solubilized protein, with the detergent kept below its critical micelle concentration so no membrane is included in the calculations.

During each Monte Carlo titration of ligand binding, the position of the four His (M266, M219, L190 and L230) and one Glu (M234) that are ligands to the RC iron atom are fixed in the crystal structure position and maintained with the His neutral, Glu ionized and Fe^{2+} state. This fixes the position and tautomer for His M219, which also makes a hydrogen bond to the quinone. All the bacteriochlorophylls, bacteriopheophytins and the ubiquinone at the Q_B site are fixed in their neutral ground state. There is an acidic cluster of Glu L212 and Asp L213 near the Q_B site. The acidic pair interacts strongly and the total net charge of the cluster remains -1.[11,71] Here Glu L212 is set to be neutral. In addition, Glu L104, which makes a hydrogen bond to the ring V keto carbonyl of bacteriopheophytin, is fixed in its protonated state. The ionization states of all other protonatable residues are determined by MCCE Monte Carlo sampling.

3.1.1 Preparation of protein structure file.

Identical quinone docking and affinity calculations are carried out on 1AIJ[192] and 2UWU[193] structures taken from the Protein Data Bank (PDB).[194] They have a resolution of 2.20 and 2.04 Å respectively. All crystallographic water molecules, lipid and detergent molecules are removed.

Each of the ten BQs to be docked in the protein is optimized and its atomic electrostatic potential (ESP) partial charge distribution calculated using DFT analysis in the Gaussian98 program[195] with the UB3LYP/6-31G basis sets. This charge set was used in MCCE for the ligand within the protein and for reference calculations of the solvation energy in water.

One cycle of minimization is carried out for each ligand with the GROMACS molecular dynamics program.[196] These calculations use a reduced atomic model, where there are no hydrogens, so the quinone proton partial charge from the Gaussian98 analysis are redistributed to the closest carbon atom. The ab initio RC cofactor parameters derived by Ceccarelli, Procacci and Marchi[197] were converted to GROMACS format and used here. The bonded parameters for the other BQs needed for GROMACS were derived from the parameters for the native UQ.

3.1.2 Ligand Docking.

The calculations start with the native quinone in the crystal structure stripped of substituents, leaving only the flat, six membered ring with the two para-carbonyl groups. MCCE adds the substituent methyl and methoxy groups to the ring, building the ten 1,4 BQs by applying information in topology files that define bond angles and length, and

proton positions. Each BQ is docked in the protein in all possible poses taking into account the symmetry of the substituents. For example, while maintaining the positions of the two carbonyls, the methyl-1,4-BQ has four starting orientations with its methyl group docked into the four quadrants, while there is only one starting position needed for the 4-fold symmetrical DQ. This procedure initiates all subsequent calculations.

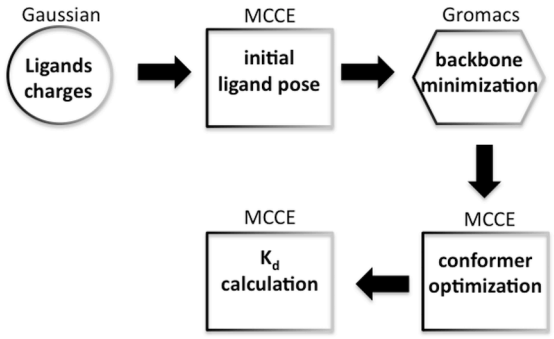

Figure 3-1 Overview of K_d calculations.

Gaussian generates ESP charges for quinone in a given redox and protonation state. MCCE finds lowest energy substituent orientation starting from native quinone ring. This conformer is added to GROMACS energy minimization of backbone and side chains. The GROMACS derived backbone will be used for all subsequent MCCE runs with this ligand. MCCE generates potential side chain and ligand conformers through cycles of translation, out-of-plane and then in-plane rotation. Good poses are further optimized with smaller motions. MCCE K_ds are obtained by competing bound and free ligands in Grand Canonical Monte Carlo sampling as a function of quinone concentration. This uses the full energy function and all protein and ligand conformers selected in any round of conformer optimization merged onto the same backbone.

Each quinone undergoes one cycle of GROMACS optimization to relax the backbone (Figure 3-1). GROMACS molecular dynamics can consider only one starting conformation for each residue and cofactor. As is unlikely that the cofactor can change

substituent orientation during the short simulations used here, the most occupied MCCE quinone replacement orientation is added to the RC crystal structure as the GROMACS starting structure.[196] In addition, the side chain protonation states are fixed given the MCCE results. Ten cycles of energy minimization, each consisting of 500 0.002 ps steps are carried out on the whole protein for each quinone. Energy minimization uses the method of steepest descent. All runs converge in less than 500 steps. The RMSD for the changes in backbone generated by GROMACS in all structures is less than 0.11 Å for the whole protein and less than 0.14 Å in the region of the protein within15Å of the binding site. MCCE calculations with the different resultant protein backbone structures yield negligible differences in ionization state of the protein residues in the final calculations.

Multiple MCCE runs are used to generate many side chain rotamers and quinone positions for calculation of the relative affinity of each ligand. Each conformer generation cycle starts with the final GROMACS side chain and backbone positions for that ligand and the lowest energy ligand position found in the previous cycle of MCCE optimization (Figure 3-1). Side chain conformers are generated and optimized. Ligand conformers are constructed by a cycle of translation, out-of-plane rotation and then in-plane rotation. These are carried out independently to avoid an explosion of conformers. First ligand translation is carried out in 2 to 5 0.2~0.5 Å steps in each ±x,y,z directions. Four step translations create a 9^3 grid with ≈730 conformers. The final, lowest energy position is rotated out-of plane in 30° steps around three axes through the opposite carbons of the quinone ring plane (C1-C4, C3-C6 and C2-C5), yielding 12^3 conformers for each BQ orientation in the site. Finally, the best position is subjected to in-plane rotation in 30° steps along each of the six axes perpendicular to the quinone plane through every ring

carbon atom. Nearby side chain conformers are created and sampled with the quinone conformers in each cycle of optimization. Each type of ligand rotation is coupled to the swing function which allows each rotatable bond to move by 3.0°~ 5.0° for finer sampling. For each transformation stage 12 rotamers are created for each methoxy substituent's two rotatable bonds. In each cycle tens of thousands of conformers are made, screened with an analytical function leading to 10-30 positions that are subjected to Monte Carlo sampling with the complete energy function using Poisson-Boltzmann electrostatics and AMBER non-electrostatics parameters. All side chain and ligand conformers that are chosen to be bound to the protein in Monte Carlo sampling are saved and pooled.

The search for conformers is first conducted in an isolated 15Å protein sphere including the 160 residues around the quinone headgroup. For the ≈12 residues within 4Å of the quinone headgroup, 12 rotamers are made for each rotatable bond. QUICK rotamer making is used for the other residues.[190] For each quinone the ligand is translated and rotated as described above, side chain rotamers are made and their goodness assessed by MC sampling as described above. Then the same cycle of optimization is carried out in the full 840 residue protein. The calculation again starts with the protein residue positions taken from the GROMACS minimized structure. The ligand conformation is in a low energy position taken from the previous optimization cycle built in the 15Å sphere. The region where 12 side chain rotamers are considered for each bond is increased to ≈7Å from the quinone, now encompassing ≈35 residues. With the exception of the first pre-GROMACS calculations all occupied conformers from each stage of optimization are pooled and added to the subsequent cycles of Delphi calculation and Monte Carlo

sampling. This builds up an extensive set of conformers to be included in the final Monte Carlo sampling that gives the quinone K_d. For each quinone the cycle of translation, in plane rotation, out of plane rotation and swing angular displacement is carried out 3 times in the 15Å sphere surrounding the quinone and then twice in the whole protein. The initial complete protein calculation starts with 10-30 thousand conformers. Analytical energy pre-screening functions reduce this to ≈2300 conformers, of which ≈450 are for the quinone and nearby residues. This increases to ≈2700 conformers after the first cycle and ≈3300 conformers following the second cycle of conformer optimization with ≈1500 conformers for the quinone and surroundings. Additional cycles of optimization are not found to increase the affinity showing the process has converged. Currently the entire procedure takes 3~5 days on a single 3.06 GHz processor.

3.1.3 Calculating the ligand binding energy.

All protein and ligand conformers selected during the MCCE conformer generation cycles are preserved for the final analysis of the binding energy for that ligand. The bound ligand conformers compete against a conformer in solution, which has torsion, van der Waals self-energy and solvation energy of an isolated quinone in water and no interactions with the protein. The van der Waals and torsion energy for the solution conformer are taken from the most occupied conformer in the protein so that these non-electrostatic self-energy terms do not affect the affinity. The quinone titration is modeled by a series of Monte Carlo runs. An energy that is proportional to the effective chemical potential is added to the solution conformer to represent the effective concentration of the quinone.[12] When the calculated occupancy of the bound and free ligand is equal, the energy difference between the apo and holo states is zero. The quinone solution chemical

potential that gives 50% bound ligand is compared for the various quinones to obtain the relative K_d.

The binding free energy is the change in the system free energy when a ligand in water binds to the RC Q_A site. Using the Boltzmann distribution of ligand and side chain conformers found in the ligand titration, the binding energy can be decomposed in a mean field analysis as:[128]

$$\Delta G_{bind} = \Delta G_{solv} + \Delta G_{P\text{-}L} + \Delta G_{entropy} + \Delta G_{strain}$$

$$= (\Delta G_{solv,ele} + \Delta G_{solv,np}) + (\Delta G_{prot,ele} + \Delta G_{prot,vdW}) + \Delta G_{entropy} + \Delta G_{strain} \quad (1)$$

ΔG_{solv} accounts for the loss of ligand-solvent interactions and $\Delta G_{P\text{-}L}$ the gain of new ligand-protein interactions. These terms may be favorable or unfavorable depending on the ligand, the solvent and the binding site. $\Delta G_{entropy}$ is a generally unfavorable term that accounts for the loss of degrees of freedom of the ligand on binding. ΔG_{strain}, accounts for the strain that builds up in the protein or ligand on binding.

ΔG_{solv}. The interaction of water and ligand is decomposed into three terms. The electrostatic component ($\Delta G_{solv,ele}$) accounts for the favorable interaction of water with the dipolar ligand. It is calculated by MCCE using DelPhi,[67] in which the solvent is modeled implicitly as a medium with a dielectric constant of 80 and the solute with an interior dielectric constant of 4. Experimentally it is expected that some water will occupy a site that is empty of quinone. No water enters the empty binding site and the dielectric constant remains 4 in the calculation. However, any errors that this yields should be the same for all quinones.

The non-polar solvation term consists of the unfavorable free energy required to form a cavity for the ligand in water (ΔG_{cav}), which is the basis of the hydrophobic effect[198] and the favorable solute-solvent van der Waals interaction (ΔG_{vdW}).[199,200] Both terms are estimated by an empirical linear relationship between the total non-polar solvation free energy ($\Delta G_{solv,np}$) and the solvent accessible area (A) so they cannot be readily separated. Here:

$$\Delta G_{solv} = \Delta G_{solv,ele} + \Delta G_{solv,np}$$

$$= \Delta G_{solv,ele} + (\gamma_1 A + B) = \Delta G_{solv,ele} + (\Delta G_{cav} + \Delta G_{vdW}) \quad (2)$$

where γ_1 is -5 cal·mol^{-1}·Å^{-2} and B is zero (Calc1 in Table 3-1).[201,202]

ΔG_{P-L}. The protein-ligand contribution to the binding energy is also decomposed into electrostatic and non-electrostatic terms:

$$\Delta G_{P-L} = \Delta G_{prot,ele} + \Delta G_{prot,vdW} \quad (3)$$

Both contributions are treated explicitly in the full MCCE analysis. The electrostatic pair-wise interactions with the protein ($\Delta G_{prot,ele}$) are calculated by DelPhi.[67] The pair-wise van der Waals interactions ($\Delta G_{prot,vdW}$) are calculated using standard AMBER parameters.[148] Torsion energies are bundled into the latter term. Pair-wise terms are further decomposed into side chain and backbone interactions in the mean field analysis of the calculated quinone affinity.[190]

In addition, binding is analyzed using an implicit van der Waals protein-ligand interaction that is proportional to the ligand surface area (Calc2 in Table 3-1). Eqn 1 can be reorganized as:

$$\Delta G_{bind} = (\Delta G_{solv,np} + \Delta G_{prot,vdW}) + (\Delta G_{solv,ele} + \Delta G_{prot,ele}) + \Delta G_{entropy} + \Delta G_{strain} \quad (4a)$$

$$= \gamma_2 A + (\Delta G_{solv,ele} + \Delta G_{prot,elec}) + \Delta G_{entropy} + \Delta G_{strain} \quad (4b)$$

The coefficient (γ_2) derived from the best fit line for the solvent accessible area (A) term is -55 cal·mol^{-1}·Å^{-2}.[203] This adds the favorable protein-ligand van der Waals interaction ($\Delta G_{prot,vdW}$) to the implicit water-ligand non-polar interactions (Eqn. 2). The electrostatic part of the binding energy ($\Delta G_{solv,ele}$+$\Delta G_{prot.ele}$), the change in entropic energy and the strain energy are the same in equations 4a and 4b.

$\Delta G_{entropy}$. Changes in entropy can make significant contributions to the overall free energy of ligand binding.[128,204,205] For extended ligands, simulations estimate that as much as 25 kcal/mol can be lost so the entropy loss can oppose binding nearly as strongly as the enthalpy favors it. However, the BQs considered here are flat six-member rings with at most two rotatable methoxy groups. The loss of rotational/translational entropy on being confined in the binding site is assumed to be the same for all ligands. Neglecting this term will contribute to the absolute calculated binding energies being more favorable than those measured. The configurational entropy loss is assumed to be proportional to the number of rotatable bonds affected by binding.[206-208] A range of values from 0.3,[209] 0.45,[210] to 0.7[211] kcal/mol per rotatable bond have been used in ligand binding simulations. Assuming the methoxy trans and ±gauche orientations yield three states of equal energy, when they are fixed in one location the entropy loss is:[211]

$$T\Delta S_{rot} = RT\ln 3 \approx 0.7 \text{ kcal/mol/methoxy} \quad (5)$$

A correction of 0.7 kcal/mol/methoxy is used here for the 2,6-dimethoxy BQ. A smaller energy penalty of 0.4 kcal/mol/methoxy is used for the Q_0 and methyl-Q_0 because the

adjacent methoxy groups restrict each other, so that one is in plane and the other out of plane even in solution.[90]

MCCE also makes an entropy correction when there are different numbers of conformers for distinguishable states.[190] Each quinone ligand has ≈20 different bound conformers competing with a single conformer representing the unbound state. The disparity in the number of conformers artificially favors the bound state. The entropy correction is calculated dynamically during Monte Carlo sampling to renormalize the number of low energy bound and free conformers.[190] The correction is generally less than 2 kcal/mol.

ΔG_{strain}. Upon binding, the protein and/or ligand may be distorted.[128] This costs energy, favoring the ligand in water. The strain energy is included in MCCE calculation when different conformations for the protein and ligand are adopted in the bound and unbound states. Thus, the Grand Canonical Monte Carlo sampling naturally includes this contribution to the free energy change.

3.2 Experimental vs. Calculated relative binding free energies

The affinity of ten neutral benzoquinones (BQ) for the Q_A site of the *Rhodobacter sphaeroides* photosynthetic reaction center was measured with protein solubilized in detergent suspension below the critical micelle concentration. The relative binding free energies were then calculated using MCCE. The native quinone found in the crystal structure is Ubiquinone-10 (UQ_{10}), a 2,3-dimethoxy-5-methyl-6-isoprenyl-BQ with a 10 unit, 50 carbon, isoprene tail. Methyl-Q_0, a UQ analogue with a methyl group in place of

the isoprene tail, is chosen as the reference compound. Its K_d is measured to be 0.63 μM. All the values discussed below are relative to this quinone unless noted.

MCCE was developed to allow extensive rotamer sampling during calculations of residue and ligand pK_as and E_ms in proteins.[190] It has thus focused on processes that change the charge of the protein so electrostatic interactions are dominant. However, as side chain motions accompany protonation or redox changes, state energies also include Lennard-Jones and torsion terms. More recent calculations have used the Monte Carlo sampling of multiple positions and charge states to study proton coupled chloride binding.[12] In the work reported here the calculations of the affinities of quinones represent an extension of the program into the analysis of a redox active ligand. These binding free energy calculations compare affinities determined using implicit and explicit Lennard-Jones intra-protein interactions in two independent RC crystal structures. The computed results are in good agreement with the experimental affinity measurements, permitting an evaluation of the factors that generate the different affinities of the different compounds.

Measured affinities: The measured affinities of the quinones for the Q_A binding site span 5.04 kcal/mol, from -0.14 kcal/mol for DQ to 4.90 kcal/mol for the unsubstituted BQ (Table 3-1). The standard deviation of the measurements is ≈0.25 kcal/mol. The overall trend is that the smaller quinones bind more weakly than the larger ones. The first methyl adds 0.3 kcal/mol to the affinity. Comparing the affinity of 2,3 dimethyl-BQ with the 2,5 and 2,6 isomers shows the second methyl adds ≈1.0 kcal/mol irrespective of its position. The third and fourth methyl raise the affinity by an additional ≈2 kcal/mol and ≈1.6 kcal/mol. Thus, the experimental results suggest that the first

substituent to BQ adds less to the affinity than do the latter ones. Changing 2,6 dimethyl to 2,6-dimethoxy increases the affinity by 0.9 kcal/mol, while the methyl-Q_0 essentially binds as tightly as DQ.

A. 1AIJ	Expt	Calc[1]	Calc[2]	SAS	TΔS	Electrostatics				vdW[1]				vdW[2]
						De-solv	bkbn	side chain	total	SAS	bkbn	side chain	total	total
unsub	**4.90**	**9.70**	**6.06**	-134	-0.80	-0.96	0.44	0.00	-0.52	0.67	6.11	3.12	9.90	7.37
methyl	**4.60**	**6.62**	**4.04**	-107	-0.80	-0.62	0.11	-0.54	-1.06	0.54	6.58	-0.16	6.96	5.90
2,3-dimethyl	**3.45**	**2.49**	**3.41**	-89	-0.80	-0.79	0.10	0.00	-0.70	0.45	4.86	-2.83	2.48	4.91
2,5-dimethyl	**3.64**	**4.26**	**3.08**	-86	-0.80	-0.78	0.34	-0.41	-0.85	0.43	4.87	-0.50	4.80	4.73
2,6-dimethyl	**3.63**	**3.96**	**3.42**	-79	-0.80	-1.04	0.88	0.00	-0.16	0.40	2.26	0.94	3.60	4.37
tri-methyl	**1.52**	**2.04**	**1.87**	-61	-0.80	-0.94	0.37	-0.14	-0.70	0.31	3.22	-2.10	1.43	3.37
tetra-methyl (DQ)	**-0.14**	**1.74**	**0.51**	-41	-0.80	-1.20	0.25	0.00	-0.95	0.21	5.97	-4.49	1.68	2.26
2,6-dimethoxy	**2.73**	**-0.99**	**1.31**	-36	0.60	-0.10	0.20	-1.36	-1.26	0.18	1.11	-3.09	-1.80	1.98
2,3-dimethoxy, 5-methyl (Q_0)	**1.59**	**2.60**	**0.71**	-14	0.00	0.05	0.17	-0.27	-0.06	0.07	2.18	0.18	2.43	0.77
2,3-dimethoxy, 5,6-dimethyl (methyl-Q_0)*	**0.00**	**0.00**	**0.00**	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Absolute energies methyl Q_0						2.67	-1.49	-1.77	-0.59	-1.93	-14.01	-14.82	-30.76	

B. 2UWU	Expt	Calc[1]	Calc[2]	SAS Å^2	TΔS	Electrostatics				vdW[1]				vdW[2]
						De-solv	bkbn	side chain	total	SAS	bkbn	side chain	total	total
Unsub	**4.90**	**7.15**	**6.83**	-134	-0.80	-1.17	0.61	0.82	0.26	0.67	4.71	1.21	6.59	7.37
Methyl	**4.60**	**4.32**	**4.39**	-107	-0.80	-0.60	0.02	-0.13	-0.71	0.54	3.88	0.81	5.23	5.90
2,3-dimethyl	**3.45**	**1.53**	**3.57**	-89	-0.80	-0.80	-0.01	0.27	-0.54	0.45	2.34	-1.11	1.68	4.91
2,5-dimethyl	**3.64**	**1.86**	**3.25**	-86	-0.80	-0.74	0.20	-0.14	-0.68	0.43	1.72	0.28	2.43	4.73
2,6-dimethyl	**3.63**	**3.28**	**3.35**	-79	-0.80	-1.12	0.49	0.41	-0.22	0.40	1.42	1.30	3.11	4.37
tri-methyl	**1.52**	**1.89**	**1.28**	-61	-0.80	-0.91	-0.10	-0.27	-1.29	0.31	3.82	-0.77	3.35	3.37
tetra-methyl (DQ)	**-0.14**	**-0.21**	**0.83**	-41	-0.80	-1.16	0.26	0.27	-0.63	0.21	0.19	0.54	0.94	2.26
2,6-dimethoxy	**2.73**	**-2.84**	**1.39**	-36	0.60	-0.05	0.09	-1.22	-1.18	0.18	-0.26	-3.17	-3.25	1.98
2,3-dimethoxy, 5-methyl (Q_0)	**1.59**	**3.45**	**0.65**	-14	0.00	-0.06	0.48	-0.54	-0.12	0.07	-2.28	5.23	3.01	0.77
2,3-dimethoxy, 5,6-dimethyl (methyl-Q_0)*	**0.00**	**0.00**	**0.00**	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Absolute energies methyl Q_0						2.68	-1.23	-1.77	-0.33	-1.93	-12.29	-15.22	-29.44	

C. 1AIJ	Expt	Calc[1]	Calc[2]	SAS	TΔS	Electrostatics				vdW[1]				vdW[2]
						De-solv	bkbn	side chain	total	SAS	bkbn	side chain	total	total
Unsub	**-3.54**	**-21.76**	**-14.98**	252	0.00	1.71	-1.05	-1.77	-1.11	-1.26	-7.89	-11.70	-20.86	-13.88
Methyl	**-3.84**	**-24.85**	**-17.00**	279	0.00	2.04	-1.38	-2.31	-1.65	-1.40	-7.42	-14.98	-23.80	-15.35
2,3-dimethyl	**-4.98**	**-28.98**	**-17.63**	297	0.00	1.87	-1.39	-1.77	-1.29	-1.49	-9.14	-17.66	-28.29	-16.34
2,5-dimethyl	**-4.79**	**-27.20**	**-17.96**	300	0.00	1.88	-1.14	-2.18	-1.44	-1.50	-9.14	-15.33	-25.97	-16.52
2,6-dimethyl	**-4.80**	**-27.50**	**-17.62**	307	0.00	1.63	-0.61	-1.77	-0.75	-1.53	-11.74	-13.88	-27.16	-16.88
tri-methyl	**-6.91**	**-29.43**	**-19.17**	325	0.00	1.73	-1.12	-1.90	-1.29	-1.63	-10.78	-16.93	-29.34	-17.88
tetra-methyl (DQ)	**-8.57**	**-29.73**	**-20.53**	345	0.00	1.47	-1.24	-1.77	-1.54	-1.73	-8.04	-19.31	-29.08	-19.00
2,6-dimethoxy	**-5.70**	**-32.45**	**-19.73**	350	1.40	2.57	-1.29	-3.13	-1.85	-1.75	-12.89	-17.92	-32.56	-19.27
2,3-dimethoxy, 5-methyl (Q_0)	**-6.84**	**-28.86**	**-20.32**	372	0.80	2.72	-1.32	-2.04	-0.65	-1.86	-11.82	-14.64	-28.33	-20.48
2,3-dimethoxy, 5,6-dimethyl (methyl-Q_0)*	**-8.43**	**-31.46**	**-21.04**	386	0.80	2.67	-1.49	-1.77	-0.59	-1.93	-14.01	-14.82	-30.76	-21.25

Table 3-1 Experimental and calculated binding energies

A,B: Relative experimental and calculated binding energies using methyl-Q_0 as the reference compound. Results of multiple calculations vary by <0.4 kcal/mol, while the experimental results have errors of ≈0.25 kcal/mol. A: Calculation in crystal structure 1AIJ. B: 2UWU. C: absolute experimental and calculated binding energies in 1AIJ. $Calc^1$ with vdW^1 represents full MCCE binding calculations with AMBER protein-ligand van der Waals interactions and -5 cal/mol/Å^2 surface area dependent ligand-solvent non-electrostatic interactions (Eqn. 2). Electrostatic desolv: the loss of the DelPhi calculated solvation energy when the ligand is moved into the protein. $Calc^2$ with vdW^2 uses a -55 cal/mol/Å^2 SAS dependent implicit van der Waals energy for both ligand to protein and ligand to solvent interaction (Eqn. 4). TΔS: entropy correction due to expected loss of methoxy rotation degrees of freedom. All energy values are in kcal/mol. *: methyl Q_0 is chosen as the reference compound. The $Calc^1$ energies are derived from the quinone solution chemical potential needed to have equal amounts of quinone bound and free in Monte Carlo sampling. The energy breakdown is obtained using equation 1. This uses a mean field analysis given the conformer distribution generated by the Monte Carlo sampling under conditions where half of the proteins have quinone bound.

Calculated affinities: The analysis presented here focuses on the relative rather than the absolute affinity of the BQs. The binding free energy calculations were compared in two crystal structures, 1AIJ and 2UWU (Table 3-1).[192,193] On average the binding energies relative to Methyl-Q_0, are 1.1 kcal/mol less favorable in 1AIJ than in 2UWU. This value is taken from the best-fit line. In 1AIJ the average error of the relative affinity is 0.72 kcal/mol, while it is -0.6 kcal/mol in 2UWU. With either structure the standard deviation of the error is 2.3 kcal/mol. The slope of the best-fit comparison between experiment and calculations is 1.28 for 1AIJ and 0.92 for 2UWU (Table 3-2). The results in the two crystal structures are highly correlated, with a slope of 1.0, R^2 of 0.83, and with no individual K_d differences >±2.5 kcal/mol. Values in the text will refer to 1AIJ

unless noted and will focus on the explicit van der Waals simulation (Calc1 in Table 3-1), since it provides the most structural information. The calculations are reasonably successful in reproducing the experimental values (Figure 3-2, Table 3-1). Seven out of nine calculated relative binding energies have errors with absolute value ≤ 2.0 kcal/mol. All side chain and ligand conformations selected by Monte Carlo sampling during multiple cycles of conformer generation are included in the final affinity calculation. The binding energy is found to increase relatively smoothly as additional conformers are added. However, when cycles of conformer additions beyond those described in the Methods sections were tested the calculated affinity varies by less than 0.4 kcal/mol, showing the process has converged.

	Expt	$Calc^{single}$	$Calc^{1}$	$Calc^{scale}$	$Calc^{2}$
range	5.04	10.7	10.7	6.7	5.5
slope	n/a	0.016	1.28	0.81	0.97
R^2	n/a	0	0.54	0.78	0.84
avg. abs. error	n/a	3.46	1.76	1.14	0.64
RMSD	n/a	4.2	2.29	1.4	0.77

Table 3-2 Calculated results using different approaches.

Expt: measured values; $Calc^{single}$: Single GROMACS minimized backbone, side chain and quinone position is evaluated with a QUICK calculation where only side chain isosteric and protonation conformers are sampled; $Calc^{1}$ and $Calc^{2}$ are defined as in Table 3-1. $Calc^{scale}$: a scaling factor of 0.62 is applied to all protein-ligand van der Waals interactions. Range is the difference in affinity between the tightest and weakest binding quinone in kcal/mol. The other terms characterize the correspondence between the calculated and experimental relative affinities.

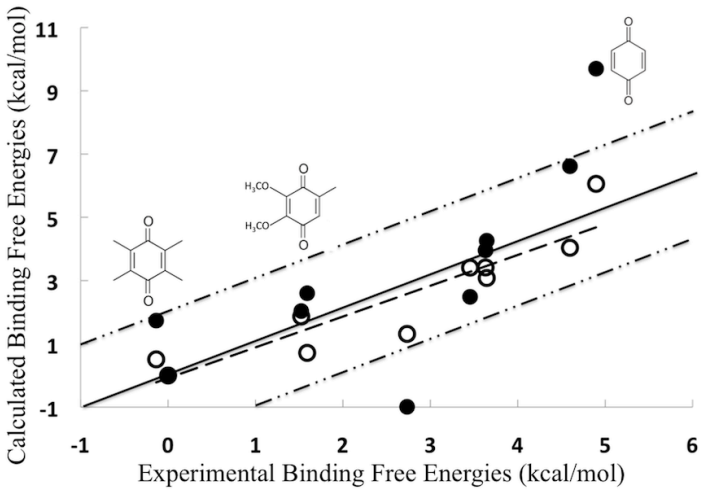

Figure 3-2 Experimental vs. calculated binding free energies (kcal/mol).

(●): Explicit AMBER Lennard- Jones energies. Dashed-dotted lines show errors of ±2 kcal/mol. Eight out of ten quinones are within this range. The best fit slope is 1.28 with a RMSD of 2.29. (○): Implicit SAS term for protein-ligand Lennard-Jones interactions. The average |error| is now 0.64±0.44 kcal/mol. Dashed line: best fit line with a slope of 0.97 and R^2 of 0.84.

3.3 Relaxation of the quinone in the binding site

The importance of protein flexibility in modulating site recognition in protein-ligand binding is well established.[132,212] However, a full analysis of all possible ligand protein conformations is computationally very expensive. Additional degrees of freedom also introduce noise in the calculations due to fluctuations in the protein conformation that are unrelated to ligand binding. Here MCCE starts the calculation with ligand positions near the UQ_{10} bound in the structure rather than doing an exhaustive search for the best global ligand position. Following a single GROMACS relaxation of the protein, the protein backbone is fixed, while the side chains and ligands can sample multiple positions. This

is similar to programs such as GLIDE,[213] which allow the protein side chains and ligand to make small movements to relieve the van der Waals clashes.

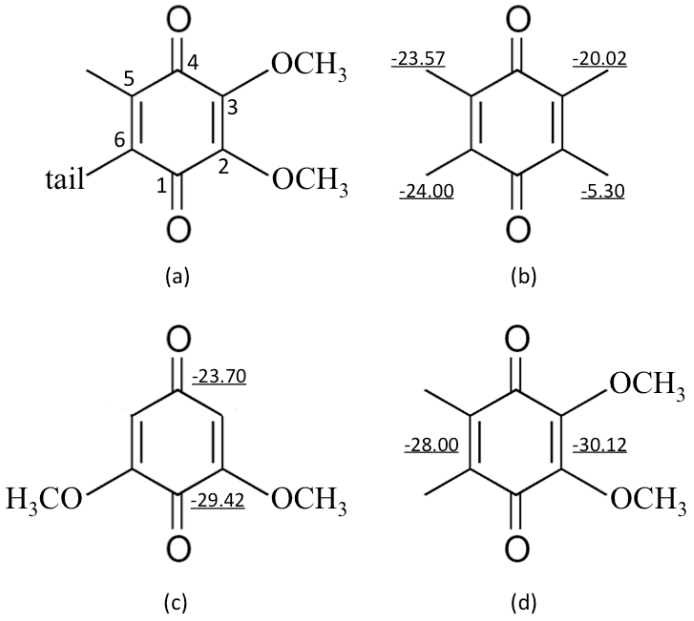

Figure 3-3 The numbering of the quinone ring.

(a) The headgroup of the native quinone. The carbons on the ring are labeled from 1 to 6. The native ubiquinone has two carbonyl groups, hydrogen bonded to Ala M260 (position 1) and His M219 (position 4), methoxy groups at ring carbons 2 (near Thr M261) and 3 (near Ile M265) and, a methyl group at carbon 5 (towards Met M218) and the tail at carbon 6 (near Met M256). These orientations are used throughout the paper. (b-d) The energy in kcal/mol of each orientation is in the protein is underlined. (b) Energies for methyl groups in methyl-BQ docked into the four quadrants onto the crystal structure ring. Position 6 is the most favorable. (c) The 2,6 di-methoxy BQ is docked with the Methoxyls in the 2,6 or the 3, 5 positions. The first is at lower energy. (d) The Methyl-Q_0 is docked with the methoxys in the 2,3 or the 5,6 positions. The former is favored.

The analysis of the quinone affinity starts with the generation of conformers for the ligand and the protein. The quinones are first bound with the substituent added to the crystal structure Q_A ring. The quinone has two carbonyl groups, hydrogen bonded to Ala

M260 (position 1) and His M219 (position 4), methoxy groups at ring carbons 3 (pointed towards Ile M265) and 2 (near Thr M261), a methyl group at carbon 5 (towards Met M218) and the tail at carbon 6 (near Met M256) (Figure 3-3). These numbers will be associated with the 6 orientations in the protein throughout the discussion. Asymmetric substituents are allowed to sample all 4 orientations in the proteins keeping the carbonyls fixed, yielding from one (BQ and DQ) to four (methyl-BQ, tri-methyl-BQ and Q_0) starting positions. A QUICK MCCE calculation of the energies of the distinguishable substituent orientations give differences in affinity for the best and worst orientations in the original crystal structure ranging from 11 (2,6-dimethyl BQ) to 28 kcal/mol (either dimethoxy BQ).

The crystal structure, with the ligand substituent pose at the lowest energy in an initial MCCE run, is minimized with GROMACS to relax the backbone. The affinity of the various BQs in their GROMACS optimized structures, were determined with a QUICK run with the quinone position fixed. This QUICK run binding energy is in some ways similar to a MM-PBSA analysis of a single protein docked ligand conformation with pre-assigned ionization states. However, the MCCE calculation still allows isosteric conformer positions and acidic and basic residue protonation states to be optimized in the calculation that determines the affinity.[190] A number of the smaller quinones show less than 1 kcal/mol difference in their affinity compared with that found for this quinone orientation in the original crystal structure. However, the backbone relaxation increases the affinity of DQ and 2,6-dimethoxy-BQ by ≈6 kcal/mol. For Q_0 and methyl-Q_0 the single pose recovered from the GROMACS minimized structure binds 3~4 kcal/mol more weakly than the best quinone pose in the crystal structure. In both cases, the

backbone van der Waals interactions are ≈5 kcal/mol more favorable in the minimized structure. However the 3-methoxy group, which is fixed in the QUICK run clashes with Thr M222 raising the energy by ≈7.5 kcal/mol. This will be relaxed when MCCE side chain optimization is added. The RMSD between experiment and the MCCE QUICK calculation following GROMACS backbone relaxation is 4.2 kcal/mol and the slope and R^2 are close to zero (Table 3-2).

DQ is the tightest binding quinone studied here with an affinity similar to methyl-Q_0, the analogue of the native UQ. However, it has the smallest affinity, 13 kcal/mol weaker than methyl-Q_0, when the methyl substituents are simply built onto the quinone ring in the crystal structure. The MCCE QUICK run shows GROMACS optimization increases the DQ affinity by ≈6.7 kcal/mol by improving the van der Waals interactions, with negligible changes in the electrostatic energy. This is mostly due to changes in position of the Ala M249 and M260, which increases the affinity by ≈7.3 kcal/mol. The interaction of the DQ with the backbone of these two residues becomes ≈7.9 kcal/mol more favorable, while there are small unfavorable interactions with the side chain groups. The C_α hydrogen of M249 starts 1.6 Å from the DQ 2-methyl resulting in a 15.4 kcal/mol van der Waals clash (Figure 3-4). GROMACS relaxation increases this distance to 1.9 Å, reducing the energy to 4.8 kcal/mol. However, this pulls M260 closer, increasing the unfavorable interaction from 1.7 kcal/mol to 4.4 kcal/mol. These compromises represent an example of the strain energy built up on binding. The RMSD changes for the two Ala backbones are each ≈0.45 Å. The significant GROMACS relaxation of 2,6-dimethoxy BQ results from the release of a similar clash between a methoxy and the backbone of Ala M260 (Figure 3-4).

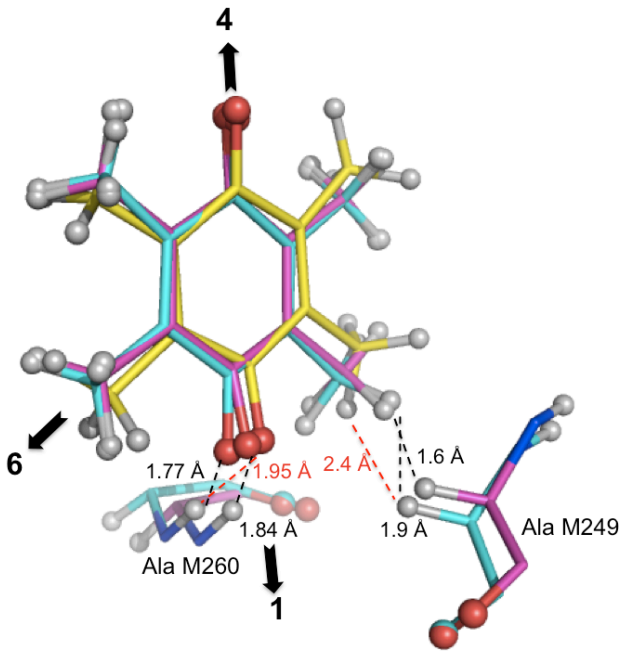

Figure 3-4 The overlap of DQ and the backbone of Ala M249 and Ala M260 in crystal structure (magenta), GROMACS minimized structure (cyan) and MCCE optimized structure (yellow).

The backbone amides are identical in GROMACS and MCCE structures so only the cyan is seen. Dashed lines label the distance between the quinone and the backbones of the two Ala residues. Black: crystal structure and GROMACS structure. Red: MCCE structure. The black arrows label the direction of the quadrants. Position 6 is occupied by the quinone tail in the native UQ_{10}.

Each GROMACS structure, optimized for each quinone, is now subjected to MCCE sampling of the side chain and ligand positions. This further increases the binding affinity by 0.3 to 7 kcal/mol. BQ, with the smallest van der Waals interactions, has the smallest and DQ the largest change. Small movements of DQ away from the GROMACS position considered above relieve clashes with the Ala M249 and Ala M260 backbone by ≈6 kcal/mol each. However the van der Waals interactions with the protein side chains, in particular Ile M265, become ≈4 kcal/mol less favorable. In total, the MCCE ensemble of

accepted ligand and side chain position increases the DQ binding affinity by ≈7 kcal/mol. Thus, for the group of BQs coupled ligand, side chain and backbone relaxation is needed. While no dramatic conformational changes are found, small changes in the positions of the side chains of Ile M265, Met M262, Met M218, Thr M222 and Ile M223 lead to notable increase in the binding affinity.

3.4 Energy decomposition in the calculation of the quinone affinity

MCCE considers both explicit interactions, which depend on the atomic positions of the protein and ligand, and implicit energy terms, which depend on the atomic solvent-accessible surface area (SAS) or on the dielectric constant assigned to the protein and solvent. The implicit desolvation energy, and explicit electrostatic pair-wise interactions between the ligand and protein are obtained using DelPhi[67] to solve the Poisson Boltzmann equation. Explicit non-electrostatic Lennard-Jones and torsion energies are calculated with the AMBER force field.[148] An implicit non-polar solvation energy term combines the hydrophobic energy, favoring quinone binding and favorable Lennard-Jones interaction between the quinone and the solvent. A value of -5 cal·mol^{-1}·Å^{-2}, used here, is taken from an earlier analysis of the vacuum to water transfer free energy of 67 molecules, including analogs to the polar amino acid side chains and protein amide backbone, using Delphi electrostatic energy with PARSE charges as does MCCE.[201,202]

Electrostatic energies: The BQs whose affinities are calculated here are all neutral. As each has two carbonyls they are quadrapoles. The absolute total electrostatic contributions to binding range from -0.59 to -1.85 kcal/mol (Table 3-1). The biggest contributors are the ferrous non-heme iron that lies between Q_A and Q_B and His M219, which individually stabilize binding by up to 2.1 kcal/mol. The neutral His is both a

ligand to the iron and a hydrogen bond donor to the quinone O_4. No other individual residue or backbone dipole has electrostatic interactions of more than 0.7 kcal/mol with any BQ. The loss in solvation energy, which opposes binding, is also not large, contributing from 1.5 to 2.7 kcal/mol. While the electrostatic interactions contribute a small amount to the affinity, they add 1.3 kcal/mol to the range of relative affinities of the different quinones. (See Table 3-1A) If the electrostatic interactions were ignored the average absolute error would increase from 1.75 to 2.0 kcal/mol.

One unique feature of the MCCE program is to be able to couple protein ionization with ligand binding. However, the quinone ligands studied here are neutral and situated at a relatively hydrophobic binding site. Thus calculated ionization states of the residues change little upon binding.

Non-electrostatic energies: The protein-ligand van der Waals interactions and the non-polar part of the solvation energy are the most significant contributors to the binding affinity, both in magnitude and in their differences between different quinones (Table 3-1). These strongly distance sensitive interactions are the most dependent on the optimization of the quinone position in the binding site. After relaxation the favorable van der Waals interactions range from -20.86 (BQ) to -32.56 kcal/mol (2,6 dimethoxy BQ).

As the size of the ligand increases, these interactions generally become more favorable. The most important contributions are from Trp M252 and Ile M265, ≈3.5 Å away from the Q_A ring plane in the original crystal structure. The interactions with the Trp M252 side chain range from -2.0 (BQ) to -5.3 kcal/mol (methyl-Q_0). Mutation of this

conserved Trp leads to an RC that will not bind Q_A.[214,215] The interactions with Ile M265 can be as favorable as -3 kcal/mol. Thr M222 and His M219 can also contribute ≈-2 kcal/mol each.

Two energies are summed in the implicit non-polar interaction of the ligand with the solvent. One is the unfavorable energy to create a cavity for the ligand in solution, which is the source of the hydrophobic effect.[216] The second is the favorable van der Waals interaction between the solvent and the ligand. Each of these is treated as being proportional to the surface area of the ligand (Eqn 2).[199] A value of -5 cal·mol^{-1}·Å^{-2} is used,[201,202] which favors moving the ligand out of the water. This provides a small contribution to the affinity ranging from 0.67 for unsubstituted BQ to 0.07 kcal/mol for Q_0 relative to methyl-Q_0 (Table 3-1).

3.5 BQ orientation in the Q_A site

Ligand conformers: MCCE generates multiple conformations for the ligands at the RC Q_A site. In the final analysis of affinity, there are usually ≈20 conformers available for each ligand. Monte Carlo sampling assigns occupancy for each one, ranking the quinone orientations in the binding site. All ten BQs have several closely clustered conformers occupied in the Boltzmann distribution of states. The optimized orientations provide a view of the local interactions around the quinone. When the original crystal structure with the substituted BQs in place of the native quinone is compared with the optimized protein structure with all selected ligand conformers, no large conformational changes are found. The binding site is quite flat, restricting rotation around the carbonyl axis so only positions in the native quinone plane are found (Figure 3-5B). The smallest quinone studied is the unsubstituted BQ, which is 4.9 Å wide including hydrogens, while

2,6 dimethoxy is the widest stretching ≈9 Å. The smaller BQ and methyl-BQ occupy a number of positions near the native quinone. As the quinone gets larger its position is better defined. Comparing only the six carbons in the quinone ring and two carbonyl oxygens, the RMSD between the crystal structure Q_A ring and the most favorable conformer of each ligand ranges from 0.17 Å in Q_0 to 2.6 Å in unsubstituted BQ. This clustering of the quinone in the binding site is not unexpected given that all of these BQs can reconstitute UQ function, which requires them to be in position to be reduced within 10 ns of the excitation of the reaction center.[217,218]

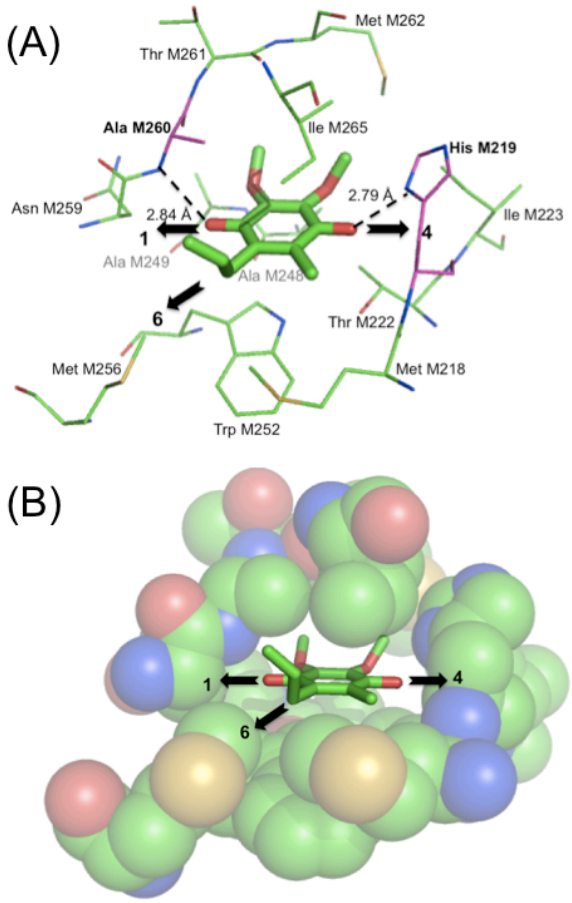

Figure 3-5 Residues within 4Å of the Q_A headgroup.

A and B are both views close to the ring plane but from slightly different angles. The same residues are shown in the two figures. The two H-bond partners to the ubiquinone are Ala M260 and His M219, which are highlighted in magenta in A. The black arrows label the quadrant designation. Figure B shows the binding site is too flat to allow significant rotation of the quinone out of plane.

H-bonds to the ligand: The calculations can analyze the importance of specific interactions between the ligand and binding site. The two quinone carbonyls are hydrogen bond acceptors, O_4 from HN_δ of His M219 and O_1 from the backbone HN of Ala M260. However, a number of experiments have shown that the para-carbonyl arrangement is not required for binding or function.[123,217,219] A single carbonyl is sufficient for quinone binding.[123] The ortho-napthoquinone reconstitutes Q_A activity[160] and ortho-phenanthroline is an often-used competitive inhibitor for quinone binding.[220] Nevertheless, calculations here generally yield selected conformers with the two carbonyls pointing towards the native hydrogen partners (Figure 3-6).

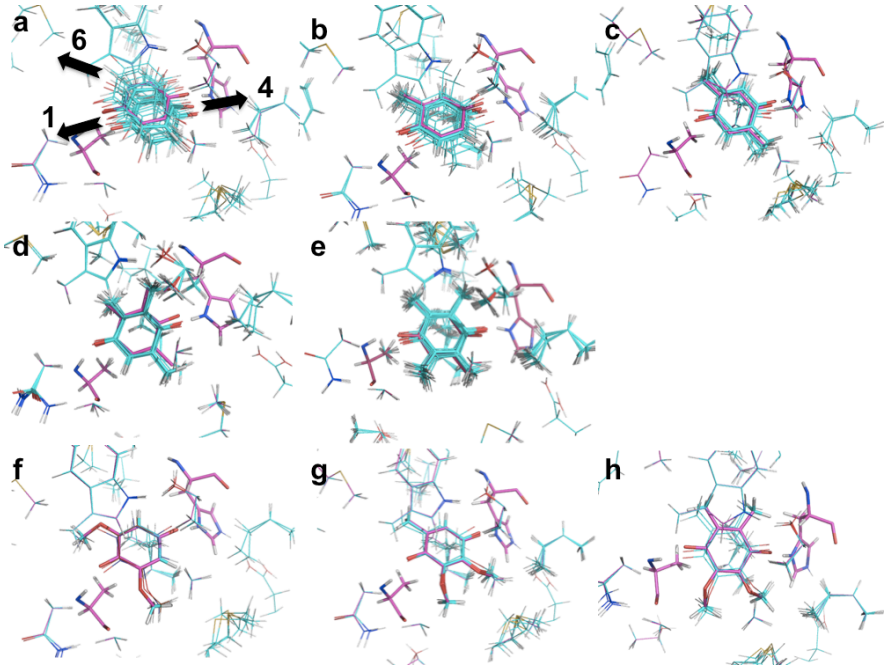

Figure 3-6 The overlap of the crystal structure (magenta) with the final MCCE minimized structures with all selected protein and ligand conformers (blue).

a: unsubstituted BQ, b: methyl-BQ, c: 2,5 di-methyl-BQ, d: tri-methyl-BQ, e: DQ, f: 2,6 di-methoxy BQ, g: Q_0, h: methyl-Q_0. The two H-bond partners to the ubiquinone, Ala M260 and His M219 and the most occupied ligand conformer are thicker sticks. As the ligand gets smaller more ligand poses are observed.

The unsubstituted BQ was used to study the importance of the two H-bonds by rotating the ring moving the carbonyls from 1,4 to 2, 5 or 3, 6 orientations in the binding site (see Fig 3-5A). With the carbonyls in the 2,5 positions, overlaying a methyl and a methoxy in the native quinone, the quinone binds as tightly as in the 1,4 orientation. While the two H-bonds are lost weakening its electrostatic interactions with the protein by 2 and 1.2 kcal/mol respectively, there is now ≈3 kcal/mol more favorable van der Waals interactions with the protein. The affinity in the 3,6 orientation is 2 kcal/mol less. Thus for the unsubstituted neutral BQ, the carbonyl H-bonds are not indispensable for binding, though the rearrangement could change the reactivity.

The preferred methyl group positions: The first methyl increases the measured affinity by only 0.3 kcal/mol while the third increases it by 2.1 kcal/mol (Table 3-1). One challenge of the calculations is to explain these preferences. Methyl-1,4-BQ was built with its methyl group docked in each of the four quadrants (2, 3, 5 and 6) onto the UQ ring in the crystal structure. The relative affinity of each of the 4 orientations was determined by four separate binding calculations with only one conformer available in a QUICK MCCE calculation. The BQ with the methyl group in position 5 or 6 show very similar binding affinities (Figure 3-3). Adding a methyl group at position 3 is 4 kcal/mol less favorable. However the methyl is strongly disfavored at position 2, towards

ThrM261, due to a clash between the methyl hydrogen and the C_α hydrogen of Ala M249 backbone. In the crystal structure there is a methoxy group at this position and the ether oxygen fits where the methyl does not. Thus the methyl, di-methyl and tri-methyl BQs place the first methyl group at position 5 or 6, with position 5 favored. Position 2 is filled only with DQ and requires backbone rearrangement and ligand movement. While this analysis was carried out with limited conformer sampling the same orientation preferences are found when the protein and ligand are allowed multi-conformer sampling. The calculated results do not fully explain the trend in binding affinity when adding methyl groups to the quinone ring. However it does show a strong preference for a methoxy rather than a methyl in this position, which contributes to the specificity of the binding site.

The distribution and orientation of the methoxy groups: There are three di-methoxy BQs studied here. In Q_0 and methyl-Q_0 the two methoxys are adjacent to each other. Monte Carlo sampling places these in protein positions 2, and 3, the same as the native quinone. This orientation binds ≈2 kcal/mol better than with the methoxys in protein positions 5 and 6 (Figure 3-3). The methoxy in the 2 position is required to reconstitute electron transfer from Q_A to Q_B in the RC.[90] The methoxy groups in the 2,6-dimethoxy-BQ each situated adjacent to the same carbonyl prefer protein positions 2 and 6 rather than 3 and 5 by ≈6 kcal/mol due to clashes with Ala M248 and Met M218. In addition, protein position 2 strongly prefers to bind a methoxy group rather than a methyl group sorting the methyl and methoxy groups into the same quadrants as found with the native UQ.

When a methoxy group lies in plane the oxygen lone pairs are conjugated with the π-orbital of the ring system. This increases the ring electron density, decreasing the quinone E_m and lowering the energy of the system, while an out of plane methoxy is electron withdrawing and at higher energy.[90] One consequence is that the methoxy position can tune the quinone redox potential, with an in plane substitution making the quinone harder to reduce.[90,221] For one methoxy group either with or without an adjacent methyl group, the in-plane conformation should be favored.[222] When there are two adjacent methoxy groups, steric hindrance only allows one to be in-plane at a time. Despite the potential importance of the methoxy position, recent investigation of 34 RC structures at better than 2.8Å resolution shows that the methoxy groups are not well anchored in the binding site by either a hydrogen bond to the ether oxygen or by strongly defined contacts.[83] The 3-methoxy group was found at an angle of -77±8° while the 2-methoxy's position at 139±25° lies closer to being in plane, but is more variable.

Calculations were made with no special energy function applied to the methoxy positions. The standard van der Waals interactions will not allow two methoxys to both lie in plane, but did not account for the p electron donation into the ring that favors the in-plane orientation. The result is that all methoxy groups tend to lie out of plane with angles of 135.3±3.5°, 121±12.7° and 139.7±1° in the protein 3, 2 and 6 positions. Orientation 5 towards Met M218 is never occupied with a methoxy group.

3.6 Modification of the van der Waals terms

The van der Waals interactions, given by $A/r^{12} – B/r^6$, are exquisitely sensitive to the precise position of ligand and protein. Thus, small changes in orientation can change the ligand affinity by more than 20 kcal/mol. The modest errors in the calculations with full

MCCE treatment shows that this physics-based approach, developed for examining electrostatic interactions can also handle energies dominated by van der Waals interactions reasonably well. However, two alternative approaches that smooth these non-polar interactions were also investigated. In one the van der Waals energies were scaled, reducing the A and B coefficients by the same amount. In the second the non-polar energies were added solely with an implicit energy that is dependent on the surface area. The electrostatic desolvation energy and protein-ligand interactions continue to be taken from MCCE calculation based on the same multi-conformation protein structure and ligand positions.

The standard MCCE calculations use the full AMBER van der Waals potential. The slope for the comparison of experimental and calculated data is 1.28 implying MCCE overestimates the difference in interaction energy of the different compounds with the binding site (Table 3-2). The average explicit van der Waals energy is ≈ -81 cal·mol^{-1}·Å^{-2}. As described below an average van der Waals interaction of 50 cal·mol^{-1}·Å^{-2} appears to provide a better estimate of the protein-ligand interactions, so both the A and B coefficients were reduced to 62% of their original values. The range of calculated binding affinities is now reduced from 10.7 kcal/mol to 6.7 kcal/mol, closer to the 5.0 kcal/mol found experimentally (Table 3-2). The RMSD is reduced from 2.29 to 1.4 and eight out of ten ligands have smaller errors.

A more extreme change is to treat the non-polar part of the protein-ligand interaction with another implicit, surface area dependent term, which is added to the non-polar solvation energy (Eqn 4). The best fit of the data and simulation comes with a surface area energy of -55 cal·mol^{-1}·Å^{-2}. Honig and coworkers have used a coefficient with a

value of ≈-58 cal·mol^{-1}·Å^{-2} for a similar analysis of the MHC class I protein-peptide binding.[203] Their calculated non-polar contribution to the binding free energy also favors ligand binding. The slope of the plot of experimental versus calculated relative binding free energy is now 0.97 with a R^2 of 0.84. As shown in Table 3-2 all measures of the correspondence between experimental and calculated affinities are improved (Figure 3-2, Table 3-2). The outliers remain the 2,6-dimethoxy BQ, with calculated value 1.4 kcal/mol more favorable and the unsubstituted BQ, which is 1.2 kcal/mol weaker than the measured value. While the analysis with the explicit van der Waals interactions shows that the MCCE sampling is able to generally find relatively good docked positions for each quinone, the excellent agreement with an implicit Lennard-Jones attraction shows that the Q_A binding site can easily accommodate the small quinones studied here.

The use of the fully implicit non-polar model mixes three surface dependent terms: the penalty for forming a cavity in water, the van der Waals interactions between solvent and solute and now the protein-ligand non-polar interactions. Here -5 cal·mol^{-1}·Å^{-2} was used for γ_1 to represent the non-polar solvation energy, a value that favors pushing the quinone into the protein (Eqn 2). Increasing this to -55 cal·mol^{-1}·Å^{-2} (γ_2) indicates an average of -50 cal·mol^{-1}·Å^{-2} for the protein-ligand van der Waals interactions. The water-cyclohexane partition coefficients (log P) have been measured for these ten BQs.[223,224] The best fit line for surface area dependence of the non-polar transfer free energy has a slope of ≈ -21 cal·mol^{-1}·Å^{-2}, which is the difference in the dispersion energy between water and hexane plus the unfavorable cavity term in water. This independently calculated value is smaller than the -55 cal·mol^{-1}·Å^{-2}, which is the best fit for the binding calculations using the implicit protein-ligand van der Waals energy. Thus, the average

non-electrostatic protein-ligand interaction is approximately twice as favorable as the average interaction of the ligand with hexane.

3.7 Outliers in the calculations

The unsubstituted BQ, which binds too weakly and 2,6-dimethoxy, which binds too tightly are the two outliers in the calculations in both crystal structures. The analysis of the small, unsubstituted BQ turns out to be the most challenging. The affinity calculated with explicit Lennard Jones interactions is 2.2 (2UWU) to 4.8 (1AIJ) kcal/mol weaker than measured. It is the smallest BQ included here and is thus the most mobile in the binding site. More conformers are accepted with bigger movements than for any of the other BQ (Figure 3-6) so it may be we do not find the full range of favorable poses. Given its greater mobility in the site, the calculations may also overestimate the entropy loss. As the weakest binding quinone, it also contains more uncertainty in its measured affinity, as it can be difficult to achieve the necessary amounts of mono-disperse quinone in solution.

The 2,6-dimethoxy BQ binds especially tightly when compared with the 2,3-dimethoxy containing quinones with error from -3.7 (1AIJ) to -5.5 (2UWU) kcal/mol. One source of error can be not properly accounting for the cost of pushing the methoxy groups out of plane when they are bound. When the methoxy is forced to take conformations in the ring plane the affinity is weakened by ≈1 kcal/mol reducing the error.

3.8 Summary

Numerous different approaches have been used for calculating binding by docking and scoring,[225-228] each with their own strengths and limitations. Calculations can come to within 0.5 kcal/mol of the experimental values using time-intensive, advanced MD methods.[227] Faster methods using standard molecular-mechanics potential energy function can have a slope as steep as 1.77 with R^2 of 0.57 comparing experiment an calculation.[226] Recent explorations of different scoring functions in MM-GBSA yielded poor to modest results with R^2 values of 0.21-0.36.[225] All these indicate that protein-ligand binding remains a very challenging problem.

In this paper we describes a novel but computationally relatively inexpensive method to calculated binding affinities. MCCE combines pose generation and scoring. Flexibility is allowed for both ligand and protein side chains (but not backbone) using subroutines developed for exploration of side chain rotamer positions. The final calculation of affinity is carried out by Monte Carlo sampling of selected poses for ligand and side chains. Unlike almost all methods, which need to pre-assign the ionization states of the protein and ligand, all protonation and redox states for each residue and cofactor can also change in the Monte Carlo sampling that determines the ligand affinity. This unique feature allows study of the changes in the ionization states of the protein upon binding, which did not prove to be important here. However, it will permit analysis of the coupling of the changes in quinone electrochemistry to the differences in the binding affinity of the quinone and semiquinone.

With implicit protein ligand van der Waals interactions, the numerical results are improved, with a nearly ideal slope of 0.97 and a very good R^2 of 0.84. The success with

implicit van der Waals interactions indicates that the solvent-accessible surface area correlates well with the protein-ligand van der Waals interactions as well as with the non-polar solvation energies. The Q_A binding site can generally accommodate these small compounds. The challenge then is to recover equally good results with a more complete model, which will provide more information about local interactions. A QUICK calculation of quinone binding affinity in the GROMACS optimized structure has an RMSD of 4.2 kcal/mol and an R^2 of ≈0 (Table 3-2). However, adding MCCE optimized conformers and calculating binding by Grand Canonical Monte Carlo sampling yield good results for the 10 quinones. The linear regression line comparing all measured and calculated K_ds has a slope of 1.28 with an R^2 of 0.54 using explicit van der Waals interactions. Thus, a single cycle of GROMACS backbone relaxation and full MCCE optimization of the ligand and side chain position is needed to relax small clashes that result within any rigid structure. The driving force for binding the neutral benzoquinone to the Q_A site is identified via calculation to be the protein-ligand van der Waals interaction, while the electrostatic energies contribute little. The favored distribution of the methyl and methoxy substitute on the benzoquinone ring is the same as the native quinone. In particular, quadrant 2 near ThrM261, binds a methoxy in the native UQ. A methyl will not bind in this position without changes in the backbone structure. Smaller quinones are more mobile at the binding site and allow more binding poses than larger quinones. In summary, calculations provide insight into how the Q_A binding site accommodates different type of quinones.

APPENDIX

	Atom	Metal Centered Charge		NBO		ESP	
		Reduced	Oxidized	Reduced	Oxidized	Reduced	Oxidized
Heme	FE	2.00	3.00	0.69	1.04	-0.65	-0.23
	N A	-0.50	-0.50	-0.52	-0.53	-0.12	-0.19
	N B	-0.50	-0.50	-0.50	-0.53	-0.04	-0.12
	N C	-0.50	-0.50	-0.50	-0.54	-0.04	-0.11
	N D	-0.50	-0.50	-0.51	-0.53	-0.12	-0.20
	CHA	0.00	0.00	-0.27	-0.24	-0.35	-0.36
	CHB	0.00	0.00	-0.20	-0.20	-0.32	-0.32
	CHC	0.00	0.00	-0.24	-0.25	-0.33	-0.34
	CHD	0.00	0.00	-0.20	-0.20	-0.32	-0.33
	C1A	0.00	0.00	0.20	0.14	0.21	0.23
	C2A	0.00	0.00	-0.05	0.03	-0.11	-0.06
	C3A	0.00	0.00	-0.23	-0.21	-0.25	-0.22
	C4A	0.00	0.00	0.14	0.16	0.21	0.23
	CMA	0.00	0.00	0.21	0.23	0.16	0.18
	C1B	0.00	0.00	0.16	0.18	0.20	0.24
	C2B	0.00	0.00	-0.23	-0.21	-0.27	-0.24
	C3B	0.00	0.00	-0.24	-0.21	-0.25	-0.22
	CAB	0.00	0.00	0.21	0.23	0.16	0.18
	CBB	0.00	0.00	0.00	0.00	0.00	0.00
	C4B	0.00	0.00	0.14	0.15	0.17	0.20
	CMB	0.00	0.00	0.21	0.23	0.16	0.18
	C1C	0.00	0.00	0.14	0.22	0.17	0.19
	C2C	0.00	0.00	-0.22	-0.22	-0.25	-0.22
	C3C	0.00	0.00	-0.23	-0.20	-0.27	-0.24
	C4C	0.00	0.00	0.14	0.16	0.20	0.23
	CMC	0.00	0.00	0.21	0.23	0.16	0.18
	CAC	0.00	0.00	0.21	0.23	0.16	0.18

	CBC	0.00	0.00	0.00	0.00	0.00	0.00
	C1D	0.00	0.00	0.15	0.17	0.20	0.24
	C2D	0.00	0.00	-0.22	-0.21	-0.25	-0.23
	C3D	0.00	0.00	-0.03	0.02	-0.12	-0.06
	C4D	0.00	0.00	0.12	0.14	0.21	0.23
	CMD	0.00	0.00	0.21	0.23	0.16	0.18
	HHA	0.00	0.00	0.21	0.23	0.15	0.18
	HHB	0.00	0.00	0.21	0.23	0.15	0.18
	HHC	0.00	0.00	0.21	0.23	0.15	0.18
	HHD	0.00	0.00	0.21	0.23	0.15	0.18
Histidine	CB	0.13	0.13	0.22	0.23	0.18	0.20
	CG	0.16	0.16	-0.09	-0.06	-0.15	-0.11
	NE2	-0.56	-0.56	-0.43	-0.46	0.34	0.26
	CE1	0.16	0.16	0.22	0.24	-0.07	-0.04
	HE1	0.13	0.13	0.22	0.23	0.17	0.18
	ND1	-0.40	-0.40	-0.53	-0.51	-0.29	-0.26
	HD1	0.40	0.40	0.41	0.43	0.35	0.36
	CD2	-0.13	-0.13	-0.04	-0.04	-0.26	-0.25
	HD2	0.13	0.13	0.23	0.23	0.20	0.21

Table App. 1: Metal centered, NBO and ESP charge sets for oxidized and reduced bis-His c-type heme. NBO and ESP charge sets are from Jaguar calculation (Zhang et al. in preparation) . The same charges are assigned to both His ligands.

Residue	Atom	Metal Centered Charge		Quantum Charge	
		Reduced	Oxidized	Reduced	Oxidized
Heme	FE	2.00	3.00	1.20	1.35
	N A	-0.50	-0.50	-0.76	-0.75
	N B	-0.50	-0.50	-0.76	-0.75
	N C	-0.50	-0.50	-0.76	-0.75
	N D	-0.50	-0.50	-0.76	-0.75
	CHA	0.00	0.00	-0.19	-0.15
	CHB	0.00	0.00	-0.19	-0.15
	CHC	0.00	0.00	-0.19	-0.15
	CHD	0.00	0.00	-0.19	-0.15
	C1A	0.00	0.00	0.32	0.33
	C2A	0.00	0.00	0.07	0.08
	C3A	0.00	0.00	0.07	0.08
	C4A	0.00	0.00	0.32	0.33
	CMA	0.00	0.00	-0.10	-0.06
	C1B	0.00	0.00	0.32	0.33
	C2B	0.00	0.00	0.07	0.08
	C3B	0.00	0.00	0.07	0.08
	CAB	0.00	0.00	-0.16	-0.17
	CBB	0.00	0.00	0.00	0.03
	C4B	0.00	0.00	0.32	0.33
	CMB	0.00	0.00	-0.10	-0.06
	C1C	0.00	0.00	0.32	0.33
	C2C	0.00	0.00	0.07	0.08
	C3C	0.00	0.00	0.07	0.08
	C4C	0.00	0.00	0.32	0.33

	CMC	0.00	0.00	-0.10	-0.06
	CAC	0.00	0.00	-0.16	-0.17
	CBC	0.00	0.00	0.00	0.03
	C1D	0.00	0.00	0.32	0.33
	C2D	0.00	0.00	0.07	0.08
	C3D	0.00	0.00	0.07	0.08
	C4D	0.00	0.00	0.32	0.33
	CMD	0.00	0.00	-0.10	-0.06
Histidine	CB	0.13	0.13	-0.07	-0.04
	CG	0.16	0.16	0.32	0.32
	NE2	-0.56	-0.56	-0.49	-0.51
	CE1	0.16	0.16	0.49	0.51
	HE1	0.13	0.13	0.00	0.00
	ND1	-0.40	-0.40	-0.20	-0.16
	HD1	0.40	0.40	0.00	0.00
	CD2	-0.13	-0.13	0.10	0.12
	HD2	0.13	0.13	0.00	0.00
Methionine	CB	0.00	0.00	0.00	0.04
	CG	0.06	0.06	-0.05	-0.07
	CE	0.06	0.06	-0.08	-0.04
	SD	-0.12	-0.12	0.22	0.23
Cysteine	CB	0.00	0.00	0.25	0.25
	SG	0.00	0.00	-0.27	-0.23

Table App. 2: Metal centered and DFT derived charge set for His-Met c-type heme in both oxidized and reduced states. The latter charge set is obtained from Autenrieth et al 2004.

Label	Type	Source	PDB	Resolution[Å]	Heme ID	Heme Type	Exp. Em	Cal. Em	$\Delta\Delta G_{rxn}$	ΔG_{pol}	$\Delta G_{res,prop}$	$\Delta G_{res,prot}$
1	b5	*Rattus N.(wt)*	1AW3*	n/a	96	bis-His *b*	-102	-268	87	24	-79	-77
			1B5A*	n/a	95	bis-His *b*	-102	-166	85	55	-47	-42
2		*B.Taurus(wt)*	1CYO	1.50	201	bis-His *b*	-10	-124	100	123	-73	-49
			1EHB	1.90	A201	bis-His *b*	-10	-131	89	127	-96	-29
3		*B.Taurus(V61H)*	1ES1	2.10	A201	bis-His *b*	12	-106	88	115	-81	-1
4		*Rattus N(V54I/V61I)*	1EUE	1.80	A201	bis-His *b*	-63	-163	89	139	-94	-74
					B201	bis-His *b*	-63	-158	89	97	-74	-42
5	b562	*E.Coli*	1QPU*	n/a	A107	His-Met *b*	168	201	189	-1	-44	66
			256B	1.40	A1	His-Met *b*	168	127	84	24	-52	70
					B1	His-Met *b*	168	109	77	31	-37	73
6	c	*E.Caballus*	1CRC	2.08	A105	His-Met *c*	260	77	169	136	-332	127
					B105	His-Met *c*	260	47	170	131	-287	56
			1HRC	1.90	105	His-Met *c*	260	88	166	125	-284	102
			1WEJ	1.80	F105	His-Met *c*	260	148	154	163	-288	137
			2PCB	2.80	B105	His-Met *c*	260	206	157	84	-104	76
7		*S.Cervisiae*	1YCC	1.23	1	His-Met *c*	290	137	175	140	-294	139
			2YCC	1.90	104	His-Met *c*	290	89	174	122	-321	144
8	c2	*R.Capsulatus*	1C2R	2.50	A120	His-Met *c*	350	98	162	158	-269	70
					B120	His-Met *c*	350	136	163	162	-269	93
9		*R.Centenum*	1JDL	1.70	125	His-Met *c*	293	95	169	146	-303	107
10		*R.Palustris*	1FJ0	1.70	A115	His-Met *c*	350	154	169	163	-305	139
					B115	His-Met *c*	350	144	169	168	-303	139
					C115	His-Met *c*	350	134	169	154	-307	140
					D115	His-Met *c*	350	150	170	166	-307	143
			1I8P	1.95	A115	His-Met *c*	350	130	170	163	-306	127
					B115	His-Met *c*	350	149	166	160	-302	136
					C115	His-Met *c*	350	137	171	152	-309	155
					D115	His-Met *c*	350	165	169	170	-307	156
11		*R.Sphaeroides*	1CXC	1.60	125	His-Met *c*	355	89	165	187	-308	72
12		*R.Viridis*	1CO6	1.60	A108	His-Met *c*	296	100	173	131	-298	121
			1CRY	3.00	108	His-Met *c*	296	86	174	175	-221	5
			1IO3	1.90	A108	His-Met *c*	296	138	167	144	-247	108
13	c3	*D. d. Norway*	1CZJ	2.16	119	bis-His *c*	-318	-326	41	-7	-110	-27
					120	bis-His *c*	-350	-351	22	-30	-88	-35
					121	bis-His *c*	-280	-319	70	-17	-124	-27
					122	bis-His *c*	-150	-221	52	-1	-116	64
14		*D. Gigas*	1GYO	1.20	A111	bis-His *c*	-260	-226	26	1	-65	30
					A112	bis-His *c*	-295	-356	37	-67	-52	-46
					A113	bis-His *c*	-180	-99	64	31	-74	101
					A114	bis-His *c*	-280	-242	44	-86	-72	89
					B111	bis-His *c*	-260	-265	19	11	-81	8
					B112	bis-His *c*	-295	-347	36	-45	-72	-46
					B113	bis-His *c*	-180	-127	48	26	-65	85
					B114	bis-His *c*	-280	-221	71	-90	-65	84
15		*D. vulgaris. M*	1J0O	1.15	A1001	bis-His *c*	-263	-156	53	6	-30	32
					A1002	bis-His *c*	-340	-414	52	-127	-87	-33
					A1003	bis-His *c*	-291	-323	61	-39	-130	9
					A1004	bis-His *c*	-318	-336	56	-68	-174	70
			2CDV	1.80	1	bis-His *c*	-318	-339	29	-42	-140	35
					2	bis-His *c*	-340	-408	47	-134	-94	-8
					3	bis-His *c*	-263	-237	44	2	-105	49
					4	bis-His *c*	-291	-301	58	-52	-92	12
16		*D. vulgaris*	2CTH	1.67	A109	bis-His *c*	-305	-305	36	-45	-114	40
					A110	bis-His *c*	-345	-420	33	-119	-90	-24
					A111	bis-His *c*	-250	-230	49	-12	-65	24
					A112	bis-His *c*	-305	-308	59	-50	-118	19
					B109	bis-His *c*	-305	-310	33	-51	-127	58
					B110	bis-His *c*	-345	-415	16	-122	-77	-17
					B111	bis-His *c*	-250	-236	49	-17	-87	42
					B112	bis-His *c*	-305	-327	63	-47	-131	7
17		*D. Africanus*	3CAO	1.60	A104	bis-His *c*	-210	-197	12	114	-103	-12
					A105	bis-His *c*	-270	-352	26	1	-81	-77
					A106	bis-His *c*	-260	-347	82	-18	-116	-86
					A107	bis-His *c*	-240	-228	16	52	-107	35
			3CAR	1.60	A104	bis-His *c*	-210	-194	12	112	-86	-7
					A105	bis-His *c*	-270	-333	33	4	-95	-55
					A106	bis-His *c*	-260	-275	65	-10	-77	-39
					A107	bis-His *c*	-240	-211	35	14	-119	74
18		*S. Oneidensis*	1M1R	1.02	A801	bis-His *c*	-248	-339	3	91	-65	-146
					A802	bis-His *c*	-219	-299	17	15	-66	-47
					A803	bis-His *c*	-138	-187	53	90	-87	-22
					A804	bis-His *c*	-192	-218	3	52	-107	65
19	c4	*P. Stutzeri*	1ETP	2.20	A199	His-Met *c*	237	71	164	29	-209	98

					A200	His-Met *c*	330	177	168	91	-110	42
					B199	His-Met *c*	330	164	164	63	-88	35
					B200	His-Met *c*	237	82	167	76	-223	79
20		*A. Ferrooxidans*	1H1O	2.13	A1184	His-Met *c*	350	118	180	98	-248	106
					A1185	His-Met *c*	450	113	162	-5	-10	-21
					B1185	His-Met *c*	350	177	184	104	-262	176
					B1186	His-Met *c*	450	216	163	63	3	1
21	c549	*Synechocystis sp*	1E29	1.21	A136	bis-His *c*	-250	-209	69	-86	-20	53
22		*C.A.Maxima*	1F1C	2.30	A200	bis-His *c*	-260	-268	75	-96	-12	-11
					B400	bis-His *c*	-260	-269	71	-89	-16	-11
23	c550	*T elongatus*	1MZ4	1.80	A135	bis-His *c*	-240	-255	94	-158	30	6
24		*P versutus*	2BGV	1.90	X1121	His-Met *c*	255	0	158	59	-269	76
25	c551	*P.Aeruginosa*	351C	1.60	0	His-Met *c*	270	214	160	55	-73	66
			451C	1.60	0	His-Met *c*	270	194	160	61	-81	52
26	c552	*P.Nautica*	1CNO	2.20	A200	His-Met *c*	250	109	163	46	-147	53
					B200	His-Met *c*	250	51	168	42	-246	95
					C200	His-Met *c*	250	79	166	31	-105	36
					D200	His-Met *c*	250	39	166	43	-208	82
					E200	His-Met *c*	250	67	167	47	-52	-69
					F200	His-Met *c*	250	76	168	35	-166	69
					G200	His-Met *c*	250	92	168	35	-107	40
					H200	His-Met *c*	250	119	167	42	-259	105
27		*H.Thermophilus*	1YNR	2.00	A81	His-Met *c*	215	123	128	31	-174	153
					B81	His-Met *c*	215	84	86	30	-139	120
					C81	His-Met *c*	215	139	114	31	-135	143
					D81	His-Met *c*	215	137	125	42	-87	68
28	c553	*B. Pasteurii*	1C75	0.95	A93	His-Met *c*	60	83	152	2	-31	-23
29	c6	*M. braunii*	1CTJ	1.10	91	His-Met *c*	358	106	160	-44	-44	52
30		*chloroplast*	1CYI	1.90	200	His-Met *c*	370	63	159	-65	-29	18
31		*C.A.Maxima*	1F1F	2.70	A200	His-Met *c*	314	112	163	-54	-44	61
32		*C. glomerata*	1LS9	1.30	A92	His-Met *c*	352	149	160	-84	-53	128
33	c7	*D. acetoxidans*	1HH5	1.90	A69	bis-His *c*	-200	-277	-10	10	-59	-3
					A70	bis-His *c*	-140	-165	60	27	-86	58
					A71	bis-His *c*	-200	-181	16	28	-77	72
34	Q.A.Dehydrogenase	*P denitrificans*	1JJU	2.05	A991	His-Met *c*	235	103	110	8	-145	155
					A992	bis-His *c*	149	2	127	20	-142	216
35	Hemoglobin	*monoeric clam*	1EBT	1.90	144	mono-his b	103	95	202	39	-116	107
			1BOB	2.30	144	mono-His b	103	59	218	59	-135	67
			1FLP	1.50	143	mono-His b	103	48	190	23	-97	58
			1MOH	1.90	143	mono-His b	103	67	205	38	-122	89
36	Myoglobin	*sperm whale*	1A6G	1.15	154	mono-His b	50	57	272	55	-151	51
			1A6K	1.10	154	mono-His b	50	5	210	39	-120	29
			1A6M	1.00	154	mono-His b	50	22	245	60	-149	40
			1HJT	2.50	160	mono-His b	50	53	305	63	-149	-7
			1JP6	2.30	A200	mono-His b	50	-29	209	42	-137	22
			1JP9	1.70	A200	mono-His b	50	9	235	44	-128	8
37	Cyt. *c* Peroxidase	*P.aeruginosa*	1EB7	2.40	A401	bis-His c	-330	-265	99	49	-138	-57
					A402	his-met c	320	133	179	54	-177	83
					B401	bis-His c	-330	-260	101	48	-131	-66
					B402	his-met c	320	123	180	51	-166	73
38		*N europaea*	1IQC	1.80	A401	mono-His c	-260	-175	223	69	-180	-168
					A402	His-Met *c*	450	264	197	102	-67	54
					B401	mono-His c	-260	-244	214	63	-206	-203
					B402	His-Met *c*	450	246	191	122	-105	61
					C401	mono-His c	-260	-178	222	64	-190	-159
					C402	His-Met *c*	450	261	195	96	-72	67
					D401	mono-His c	-260	-148	135	-34	-11	-261
					D402	His-Met *c*	450	306	195	97	-38	54
39	c"	*M. methylotrophus*	1GU2	1.19	A125	bis-His *c*	-60	-230	9	39	-106	48
					B125	bis-His *c*	-60	-184	14	49	-91	65
40	bc1	*B.Taurus*	1BE3	3.00	C1	bis-His *b*	-10	4	280	-23	-164	136
					C2	bis-His *b*	100	100	247	49	-243	345
					D3	His-Met *c*	250	213	283	61	-113	-3
41	Cyto c Oxi	*Rb. Sphaeroides*	2GSM	2.00	A2001	bis-His *a*	300	261	205	131	-522	507
42	RC	*R.Viridis*	1DXR	2.00	C401	His-Met *c*	-60	14	156	5	-161	25
					C402	His-Met *c*	310	283	205	2	-153	243
					C403	His-Met *c*	380	336	169	-3	-207	348
					C404	bis-His *c*	20	-197	107	-21	-124	71

Table App. 3: Experimental, calculated Ems and decomposed energy terms for each protein. *: NMR structures.

BACKBONE STRUCTURE	Crystal structure		GROMACS	
	Crystal Worst pose	Crystal Best Pose	GROMACS Optimized pose	MCCE Final multi-pose
Number of quinone poses evaluated	1-4	1-4	1	≈20
unsub	*-21.46**		-21.76	-21.76
methyl	-6.70	-23.70	-25.70	-24.85
2,3-dimethyl	-9.49	-26.89	-26.24	-28.98
2,5-dimethyl	-8.43	-27.00	-26.35	-27.20
2,6-dimethyl	-10.65	-22.03	-25.03	-27.50
tri-methyl	-12.13	-26.63	-27.33	-29.43
tetra-methyl (DQ)	*-16.00**		-22.70	-29.73
2,6-dimethoxy	>0	-26.95	-32.35	-32.45
2,3-dimethoxy,5-methyl (Q_0)	>0	-29.06	-24.56	-28.86
2,3-dimethoxy,5,6-dimethyl (Methyl-Q_0)	>0	-29.13	-25.96	-31.46

Table App. 3. Calculated binding free energy at different stages of optimization. Crystal structure: [1]Quinone substituents are build onto the native UQ position in all distinguishable orientations and the energy evaluated with a MCCE QUICK calculation where only isosteric conformers are sampled. The energy of the highest (worst pose) and lowest (best pose) energy distinguishable pose is given. *BQ and DQ all 4 orientations are indistinguishable. GROMACS: the GROMACS minimized backbone, side chain and quinone position is evaluated with a QUICK calculation. MCCE: takes the GROMACS backbone and does full MCCE optimization, sampling multiple positions to obtain the binding energy. All energies are in kcal/mol.

	Expt	Calc[1]	Calc[scale]	Calc[2]
unsub	4.9	9.7	6.98	6.06
methyl	4.6	6.62	4.75	4.04
2,3-dimethyl	3.45	2.49	2.29	3.41
2,5-dimethyl	3.64	4.26	3.41	3.08
2,6-dimethyl	3.63	3.96	3.38	3.42
tri-methyl	1.52	2.04	2.12	1.87
tetra-methyl (DQ)	-0.14	1.74	1.41	0.51
2,6-dimethoxy	2.73	-0.99	0.29	1.31
2,3-dimethoxy, 5-methyl (Q_0)	1.59	2.6	3.38	0.71
2,3-dimethoxy, 5,6-dimethyl (methyl-Q_0)	0	0	0	0

Table App.4 Experimental and calculated relative binding energies. Calc[1], Calc[scale] and Calc[2] are defined as in Table 1 and Table 2. Calc[1] represents full MCCE binding calculations with AMBER protein-ligand van der Waals interactions. Calc[2] uses a -55 cal/mol/Å^2 SAS dependent implicit van der Waals energy for both ligand to protein and ligand to solvent interaction (Eqn. 4). Calc[scale] applies a scaling factor of 0.62 to all protein-ligand van der Waals interactions.[2]

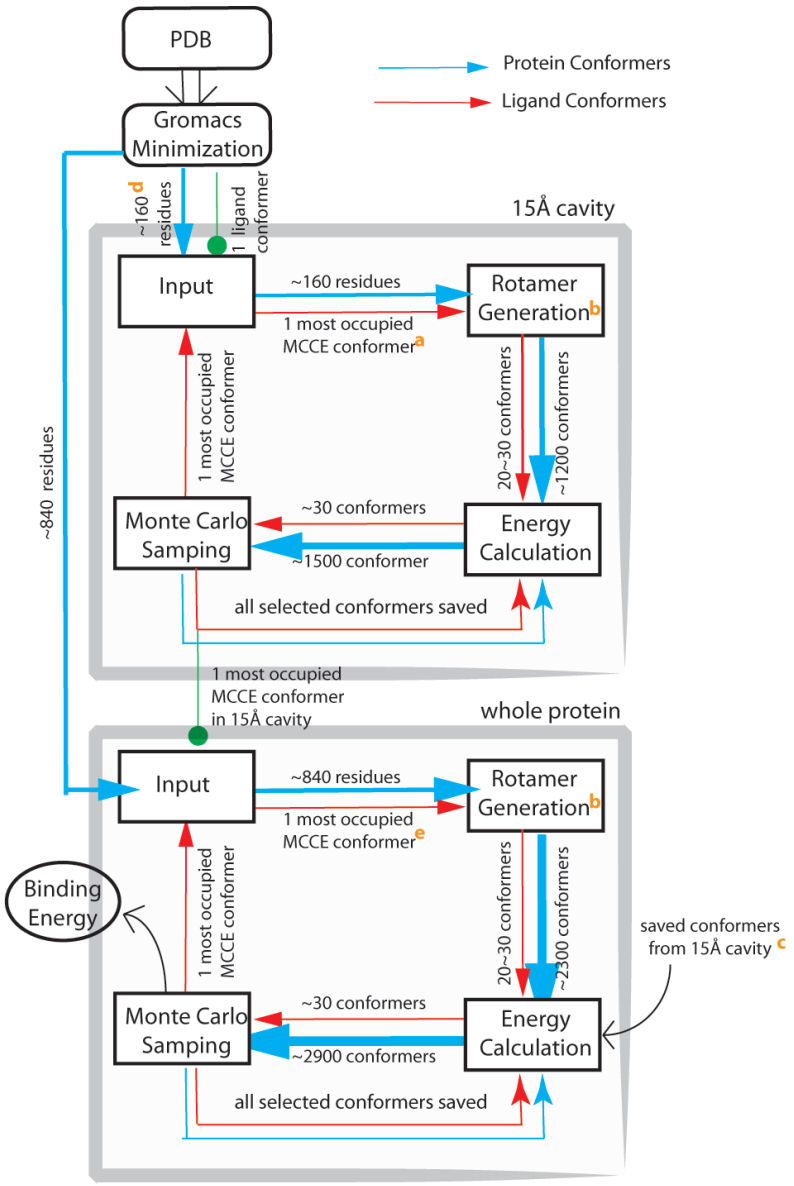

Figure App. 1 The flowchart of the ligand pose generation in MCCE. a: In the initial calculation in the 15Å sphere, the single ligand conformer from the GROMACS minimization is used as the initial, seed conformation. b: A cycle explores ligand translation, in-plane or out-of-plane rotation. c: The saved, Monte Carlo selected conformers from a 15Å sphere around the native quinone are merged with the run on the whole protein. d: The Gromacs optimization is carried out on the whole protein. e[3]: The most occupied ligand conformer from 15Å sphere was only used in the first cycle. In all subsequent steps in MCCE optimization, the most occupied ligand conformer from previous cycle is used. The green ball: In the initial run in 15Å, the single ligand conformer is taken from the Gromacs minimization. In the initial run in the whole

protein, the single ligand conformer is the most occupied ligand pose taken from the last 15Å run.

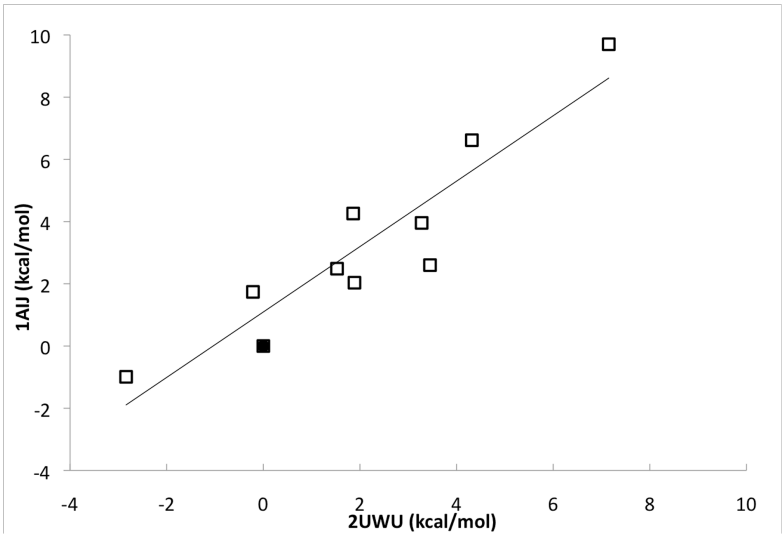

Figure App. 2 The calculated relative binding free energies in 1AIJ vs. 2UWU. The line shown is the best fit line with a slope of 1.05, R^2 of 0.86. The filled square is the reference compound methyl-Q_0.

REFERENCES:

1. Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. Molecular Biology of the Cell. New York, NY: Garland Publishing; 1989.
2. Stuehr DJ, Tejero J, Haque MM. Structural and mechanistic aspects of flavoproteins: electron transfer through the nitric oxide synthase flavoprotein domain. FEBS J 2009;276(15):3959-3974.
3. Ullmann GM. The coupling of protonation and reduction in proteins with multiple redox centers: Theory, computational method, and application in cytochrome c_3. JPhysChem B 2000;104:6923-6301.
4. Gunner MR, Honig B. Electrostatic control of midpoint potentials in the cytochrome subunit of the *Rhodopseudomonas viridis* reaction center. Proc Natl Acad Sci USA 1991;88:9151-9155.
5. Verissimo AF, Sousa FL, Baptista AM, Teixeira M, Pereira MM. Thermodynamic Redox Behavior of the Heme Centers of cbb(3) Heme-Copper Oxygen Reductase from Bradyrhizobium japonicum. Biochemistry 2007;46(46):13245-13253.
6. Teixeira VH, Soares CM, Baptista AM. Studies of the reduction and protonation behavior of tetraheme cytochromes using atomic detail. J Biol Inorg Chem 2002;7:200-216.
7. Mao J, Hauser K, Gunner MR. How cytochromes with different folds control heme redox potentials. Biochemistry 2003;42(33):9829-9840.
8. Alexov EG, Gunner MR. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. Biophys J 1997;72:2075-2093.
9. Alexov EG, Gunner MR. Calculated protein and proton motions coupled to electron transfer: electron transfer from Q_A^- to Q_B in bacterial photosynthetic reaction centers. Biochemistry 1999;38(26):8253-8270.
10. Lancaster CRD, Michel H, Honig B, Gunner MR. Calculated coupling of electron and proton transfer in the photosynthetic reaction center of *Rhodopseudomonas viridis*. Biophys J 1996;70:2469-2492.

11. Zhu Z, Gunner MR. Energetics of quinone-dependent electron and proton transfers in *Rhodobacter sphaeroides* photosynthetic reaction centers. Biochemistry 2005;44:82-96.
12. Song Y, Gunner MR. Using multiconformation continuum electrostatics to compare chloride binding motifs in alpha-amylase, human serum albumin, and Omp32. J Mol Biol 2009;387(4):840-856.
13. Zhao Y, Sanner MF. Protein-ligand docking with multiple flexible side chains. J Comput Aided Mol Des 2008;22(9):673-679.
14. Guan Y, Shi R, Li X, Zhao M, Li Y. Multiple binding modes for dicationic Hoechst 33258 to DNA. J Phys Chem B 2007;111(25):7336-7344.
15. Keilin D. On Cytochrome, a respiratory pigment, common to animals, yeast and higher plants. Proc R Soc B 1925;98.
16. O'Brian MR, Thony-Meyer L. Biochemistry, regulation and genomics of haem biosynthesis in prokaryotes. Adv Microb Physiol 2002;46:257-318.
17. Fufezan C, Zhang J, Gunner MR. Ligand preference and orientation in b- and c-type heme-binding proteins. Proteins 2008.
18. Reedy CJ, Gibney BR. Heme protein assemblies. Chem Rev 2004;104:617-649.
19. Breslow E, Koehler R, Girotti AW. Properties of protoporphyrin-apomyoglobin complexes and related compounds. J Biol Chem 1967;242(18):4149-4156.
20. Breslow E, Koehler R. Combination of Protoporphyrin Ix with Sperm Whale Apomyoglobin. J Biol Chem 1965;240:2266-2268.
21. Liu D, Williamson DA, Kennedy ML, Williams TD, Morton MM, Benson DR. Aromatic Side Chain−Porphyrin Interactions in Designed Hemoproteins. Journal of the American Chemical Society 1999;121(50):11798-11812.
22. Chapman SK, Daff S, Munro AW. Heme: The most versatile redox centre in biology? Metal Sites in Proteins and Models 1997;88:39-70.
23. Pettigrew GW, Moore GR. Cytochromes c: Biological aspects. Rich A, editor. berlin: Springer-Verlag; 1987.
24. Guengerich FP, Miller GP, Hanna IH, Martin MV, Leger S, Black C, Chauret N, Silva JM, Trimble LA, Yergey JA, Nicoll-Griffith DA. Diversity in the oxidation

of substrates by cytochrome P450 2D6: lack of an obligatory role of aspartate 301-substrate electrostatic bonding. Biochemistry 2002;41(36):11025-11034.

25. Antonin E, Brunori M. Hemoglobin and Myoglobin in their Reactions with Ligands: North-Holland: Amsterdam; 1971.
26. Harrison P, Huehns ER. Proteins of iron metabolism. Nature 1979;279(5713):476-477.
27. Rodgers KR. Heme-based sensors in biological systems. Curr Opin Chem Biol 1999;3(2):158-167.
28. Chan MK. Recent advances in heme-protein sensors. Curr Opin Struct Biol 2001;5(2):216-222.
29. Liu X, Kim CN, Yang J, Jemmerson R, Wang X. Induction of apoptotic program in cell-free extracts: requirement for dATP and cytochrome c. Cell 1996;86(1):147-157.
30. Wilson JD, Giesselman BR, Mitra S, Foster TH. Lysosome-damage-induced scattering changes coincide with release of cytochrome c. Opt Lett 2007;32(17):2517-2519.
31. Williams PA, Fulop V, Garman EF, Saunders NF, Ferguson SJ, Hajdu J. Haem-ligand switching during catalysis in crystals of a nitrogen-cycle enzyme. Nature 1997;389(6649):406-412.
32. Discher BM, Koder RL, Moser CC, Dutton PL. Hydrophilic to amphiphilic design in redox protein maquettes. Curr Opin Chem Biol 2003;7(6):741-748.
33. Hecht MH, Das A, Go A, Bradley LH, Wei Y. De novo proteins from designed combinatorial libraries. Protein Sci 2004;13(7):1711-1723.
34. Haehnel W. Chemical synthesis of TASP arrays and their application in protein design. Mol Divers 2004;8(3):219-229.
35. Dutton PL. Redox potentiometry: determination of midpoint potentials of oxidation-reduction components of biological electron-transfer systems. Methods Enzymol 1978;54:411-435.
36. Wilson GS. Determination of oxidation-reduction potentials. Methods Enzymol 1978;54:396-410.

37. Izadi N, Henry Y, Haladjian J, Goldberg ME, Wandersman C, Delepierre M, Lecroisey A. Purification and characterization of an extracellular heme-binding protein, HasA, involved in heme iron acquisition. Biochemistry 1997;36(23):7050-7057.
38. Arciero DM, Hooper AB. A di-heme cytochrome c peroxidase from Nitrosomonas europaea catalytically active in both the oxidized and half-reduced states. J Biol Chem 1994;269(16):11878-11886.
39. Wilson GS. Electrochemical studies of porphyrin redox reactions as cytochromes models. Bioelectrochem and Bioenerg 1974;1:172-179.
40. Raphael AL, Gray HB. Axial ligand replacement in horse heart cytochrome c by semisynthesis. Proteins: Struct Funct Genet 1989;6:338-340.
41. Miller GT, Zhang B, Hardman JK, Timkovich R. Converting a c-type to a b-type cytochrome: Met61 to His61 mutant of Pseudomonas cytochrome c-551. Biochemistry 2000;39:9010-9017.
42. Pearson RG. Hard and Soft Acids and Bases. Journal of the American Chemical Society 1963;85(22):3533-3539.
43. Song Y, Mao J, Gunner MR. Electrostatic environment of hemes in proteins: pK_as of hydroxyl ligands. Biochemistry 2006;45:7949-7958.
44. Zheng Z, Gunner MR. Analysis of the electrochemistry of hemes with E(m)s spanning 800 mV. Proteins 2008;75(3):719-734.
45. Ishikita H, Loll B, Biesiadka J, Saenger W, Knapp EW. Redox potentials of chlorophylls in photosystem II reaction center. Biochemistry 2005;44:4118-4124.
46. Churg AK, Weiss RM, Warshel A, Takano T. On the action of cytochrome c: correlating geometry changes upon oxidation with activation energies of electron transfer. J Phys Chem 1983;87:1683-1694.
47. Sham YY, Chu ZT, Warshel A. Consistent calculations of pK_a's of ionizable residues in proteins: Semi-microscopic and microscopic approaches. J Phys Chem 1997;101:4458-4472.
48. Oliveira AS, Teixeira VH, Baptista AM, Soares CM. Reorganization and conformational changes in the reduction of tetraheme cytochromes. Biophys J 2005;89:3919-3930.

49. Bhattacharyya S, Stankovich MT, Truhlar DG, Gao J. Combined quantum mechanical and molecular mechanical simulations of one- and two-electron reduction potentials of flavin cofactor in water, medium-chain acyl-CoA dehydrogenase, and cholesterol oxidase. J Phys Chem A 2007;111(26):5729-5742.
50. Voigt P, Knapp EW. Tuning heme redox potentials in the cytochrome c subunit of photosynthetic reaction centers. J Biol Chem 2003;278:51993-52001.
51. Hauser K, Mao J, Gunner MR. pH dependence of heme electrochemistry in cytochromes investigated by multiconformation continuum electrostatic calculations. Biopolymers 2004;74:51-54.
52. Rogers NK, Moore GR, Sternberg MJE. Electrostatic interactions in globular proteins: Calculation of the pH dependence of the redox potential of cytochrome c_{551}. J Mol Biol 1985;182:613-616.
53. Gunner MR, Alexov E. A pragmatic approach to structure based calculation of coupled proton and electron transfer in proteins. Biochim Biophys Acta 2000;1458:63-87.
54. Gunner MR, Mao J, Song Y, Kim J. Factors influencing energetics of electron and proton transfers in proteins. What can be learned from calculations. Biochim Biophys Acta 2006;1757:942-968.
55. Bashford D. Macroscopic electrostatic models for protonation states in proteins. Front Biosci 2004;9:1082-1099.
56. Beroza P, Case DA. Calculations of proton-binding thermodynamics in proteins. Method Enzymol 1998;295:170-189.
57. Ullmann GM, Knapp EW. Electrostatic models for computing protonation and redox equilibria in proteins. Eur Biophys J 1999;28(7):533-551.
58. Neves-Petersen MT, Petersen SB. Protein electrostatics: a review of the equations and methods used to model electrostatic equations in biomolecules--applications in biotechnology. Biotechnol Annu Rev 2003;9:315-395.
59. Warshel A, Papazyan A. Electrostatic effects in macromolecules: Fundamental concepts and practical modeling. Curr Opin Struct Biol 1998;8:211-217.

60. Georgescu RE, Alexov EG, Gunner MR. Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins. Biophys J 2002;83:1731-1748.
61. Toseland CP, McSparron H, Davies MN, Flower DR. PPD v1.0--an integrated, web-accessible database of experimentally determined protein pKa values. Nucleic Acids Res 2006;34(Database issue):D199-203.
62. Song Y, Gunner MR. Comparison of bacteriorhodopsin and halorhodopsin by MCCE reveals a possible new chloride binding site in halorhodopsin. Biophys J 2004;86(1):611a-611a.
63. Paddock ML, Rongey SH, Feher G, Okamura MY. Pathway of proton transfer in bacterial reaction centers: Replacement of glutamic acid 212 in the L subunit by glutamine inhibits quinone (secondary acceptor) turnover. Proc Natl Acad Sci USA 1989;86:6602-6606.
64. Kassner RJ. Effects of nonpolar environments on the redox potentials of heme complexes. Proc Natl Acad Sci USA 1972;69(8):2263-2267.
65. Warshel A, Russell ST. Calculations of electrostatic interactions in biological systems and in solutions. Q Rev Biophys 1984;17:283-422.
66. Kadish KM. Electrochemistry of metalloporphyrins in nonaqueous media. Prog Inorg Chem 1986;34:435-605.
67. Nicholls A, Honig B. A rapid finite difference algorithm utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. J Comp Chem 1991;12:435-445.
68. Bashford D, Case DA. Generalized Born models of macromolecular solvation effects. Annu Rev Phys Chem 2000;51:129-152.
69. Onufriev A, Bashford D, Case DA. Modification of the generalized Born model suitable for macromolecules. Journal of Physical Chemistry B 2000;104(15):3712-3720.
70. Velin Z. Spassov HL, Klaus Gerwert and Donald Bashford. pKa Calculations Suggest of an Excess Proton in a Hydrogen-bonded Water Network in Bacteriorhodopsin. J Mol Biol 2001(312):203-219.

71. Rabenstein B, Ullmann GM, Knapp E-W. Energetics of electron-transfer and protonation reactions of the quinones in the photosynthetic reaction center of *Rhodopseudonas viridis*. Biochemistry 1998;37:2488-2495.

72. Antosiewicz J, McCammon JA, Gilson MK. The determinants of pK_as in proteins. Biochemistry 1996;35:7819-7833.

73. Sandberg L, Edholm O. A fast and simple method to calculate protonation states in proteins. Proteins: Struct Funct Genet 1999;36:474-483.

74. Sharp K, Jean-Charles A, Honig B. A local dielectric constant model for solvation free energies which accounts for solute polarizability. J Phys Chem 1992;96:3822-3828.

75. Teixeira VH, Cunha CA, Machuqueiro M, Oliveira ASF, Victor BL, Soares CM, Baptista AM. On the use of different dielectric constants for computing individual and pairwise terms in Poisson-Boltzmann studies of protein ionization equilibrium. J Phys Chem B 2005(109):14691-14706.

76. Gunner MR, Mao JJ, Hauser K, Kim J. Adding conformational flexibility to continuum electrostatics: Assaying the stability and importance of charges buried in proteins. Abstracts of Papers of the American Chemical Society 2003;225:U750-U750.

77. Song Y, Mao J, Gunner MR. Calculation of proton transfers in bacteriorhodopsin bR and M intermediates. Biochemistry 2003;42(33):9875-9888.

78. Haas AH, Lancaster CR. Calculated coupling of transmembrane electron and proton transfer in dihemic quinol:fumarate reductase. Biophys J 2004; 87:4298-4315.

79. Tang CL, Alexov E, Pyle AM, Honig B. Calculation of pKas in RNA: on the structural origins and functional roles of protonated nucleotides. J Mol Biol 2007;366(5):1475-1496.

80. Crofts AR, Wraight CA. The electrrochemical domain of photosynthesis. Biochim Biophys Acta 1983;726:149-185.

81. Gunner MR, Alexov E, Torres E, Lipovaca S. The importance of the protein in controlling the electrochemistry of heme metalloproteins: Methods of calculation and analysis. J Biol Inorg Chem 1997;2:126-134.

82. Heathcote P. Reaction centers: the structure and evolution of biological solar power. Trends Biochem Sci 2002;27:79-86.
83. Wraight CA, Gunner MR. The Acceptor Quinones of Purple Photosynthetic Bacteria - Structure and Spectroscopy: Springer Netherlands; 2008.
84. Allen JP, Feher G, Yeates TO, Komiya H, Rees DC. Structure of the reaction center from *Rhodobacter sphaeroides* R-26: The protein subunits. Proc Natl Acad Sci USA 1987;84:6162-6166.
85. Deisenhofer J, Epp O, Miki R, Michel H. Structure of the protein subunits in the photosynthetic reaction center of *Rhodopseudomonas viridis* at 3 Å resolution. Nature 1985;318:618-624.
86. Nohl H, Kozlov AV, Staniek K, Gille L. The multiple functions of coenzyme Q. Bioinorganic Chem 2001;29:1-13.
87. Cramer WA, Knaff DB. Energy Transduction in Biological Membranes: A Textbook of Bioenergetics. New York: Springer-Verlag; 1991.
88. Mitchell P. The protonmotive Q cycle: a general formulation. FEBS Lett 1975;59(2):137-139.
89. Mitchell P. Protonmotive redox mechanism of the cytochrome bc_1 complex in the respiratory chain: protonmotive ubiquinone cycle. FEBS Lett 1975;56(1):1-6.
90. Wraight CA, Vakkasoglu AS, Poluektov Y, Mattis AJ, Nihan D, Lipshutz BH. The 2-methoxy group of ubiquinone is essential for function of the acceptor quinones in reaction centers from Rba. sphaeroides. Biochim Biophys Acta 2008;1777(7-8):631-636.
91. Woodbury NW, Parson WW, Gunner MR, Prince RC, Dutton PL. Radical-pair energetics and decay mechanisms in reaction center containing anthraquinones or benzoquinones in place of ubiquinone. Biochim Biophys Acta 1986;851:6-22.
92. Gunner MR, Madeo J, Zhu Z. Modification of quinone electrochemistry by the proteins in the biological electron transfer chains: examples from photosynthetic reaction centers. J Bioenerg Biomembr 2008;40(5):509-519.
93. Rich PR. The quinone chemistry of bc complexes. Biochem Biophys Acta 2004;1658:165-171.

94. Guo J, Lemire BD. The ubiquinone-binding site of the Saccharomyces cerevisiae succinate-ubiquinone oxidoreductase is a source of superoxide. J Biol Chem 2003;278(48):47629-47635.
95. Borisov VB. Defects in mitochondrial respiratory complexes III and IV, and human pathologies. Mol Aspects Med 2002;23(5):385-412.
96. Kramer DM, Roberts AG, Muller F, Cape J, Bowman MK. Q-cycle bypass reactions at the Q_o site of the cytochrome bc_1 (and related) complexes. Methods Enzymol 2004;382:21-45.
97. Yankovskaya V, Horsefield R, Tornroth S, Luna-Chavez C, Miyoshi H, Leger C, Byrne B, Cecchini G, Iwata S. Architecture of succinate dehydrogenase and reactive oxygen species generation. Science 2003;299(5607):700-704.
98. Cecchini G, Maklashina E, Yankovskaya V, Iverson TM, Iwata S. Variation in proton donor/acceptor pathways in succinate:quinone oxidoreductases. FEBS Lett 2003;545(1):31-38.
99. Hunte C, Palsdottir H, Trumpower BL. Protonmotive pathways and mechanisms in the cytochrome bc_1 complex. FEBS Lett 2003;545(1):39-46.
100. Siregar JE, Syafruddin D, Matsuoka H, Kita K, Marzuki S. Mutation underlying resistance of Plasmodium berghei to atovaquone in the quinone binding domain 2 (Qo2) of the cytochrome b gene. Parasitol Int 2008;57(2):229-232.
101. Omura S, Miyadera H, Ui H, Shiomi K, Yamaguchi Y, Masuma R, Nagamitsu T, Takano D, Sunazuka T, Harder A, Kolbl H, Namikoshi M, Miyoshi H, Sakamoto K, Kita K. An anthelmintic compound, nafuredin, shows selective inhibition of complex I in helminth mitochondria. Proc Natl Acad Sci U S A 2001;98(1):60-62.
102. Duke SO. Overview of herbicide mechanisms of action. Environ Health Perspect 1990;87:263-271.
103. Stowell MHB, McPhillips TM, Rees DC, Soltis SM, Abresch E, Feher G. Light-induced structural changes in photosynthetic reaction center: implications for mechanism of electron-proton transfer. Science 1997;276:812-816.
104. Gao X, Wen X, Esser L, Quinn B, Yu L, Yu CA, Xia D. Structural basis for the quinone reduction in the bc1 complex: a comparative analysis of crystal structures

of mitochondrial cytochrome bc_1 with bound substrate and inhibitors at the Q_i site. Biochemistry 2003;42(30):9067-9080.

105. Kurisu G, Zhang H, Smith JL, Cramer WA. Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity. Science 2003;302(5647):1009-1014.

106. Iverson TM, Arciero DM, Hooper AB, Rees DC. High-resolution structures of the oxidized and reduced states of cytochrome c554 from Nitrosomonas europaea. J Biol Inorg Chem 2001;6(4):390-397.

107. Zhou Y, Cierpicki T, Jimenez RH, Lukasik SM, Ellena JF, Cafiso DS, Kadokura H, Beckwith J, Bushweller JH. NMR solution structure of the integral membrane enzyme DsbB: functional insights into DsbB-catalyzed disulfide bond formation. Mol Cell 2008;31(6):896-908.

108. Fisher N, Rich PR. A motif for quinone binding sites in respiratory and photosynthetic systems. J MolBiol 2000;296:1153-1162.

109. Dutton PL, Gunner MR, Prince RC. Systemmatic modification of electron transfer kinetics in a biological protein: Replacement of the primary ubiquinone of photochemical reaction centers with other quinones. In: Helene C, Charlier M, Montenay-Garestier T, Laustriat G, editors. Trends in Photobiology. New York: Plenum Press; 1982. p 561-570.

110. Okamura MY, Paddock ML, Graige MS, Feher G. Proton and electron transfer in bacterial reaction centers. Biochim Biophys Acta 2000;1458:148-163.

111. Wraight CA. Proton and electron transfer in the acceptor quinone complex of photosynthetic reaction centers from Rhodobacter sphaeroides. Front Biosci 2004;9:309-337.

112. Mancino LJ, Dean DP, Blankenship RE. Kinetics and thermodynamics of the $P870^+Q_A^-$ ‡ $P870^+Q_B^-$ reaction in isolated reaction centers from the photosynthetic bacterium *Rhodopseudomonas sphaeroides*. Biochim Biophys Acta 1984;764:46-54.

113. Kleinfeld D, Okamura MY, Feher G. Electron transfer in reaction centers of *Rhodopseudomonas sphaeroides*: I. Determination of the charge recombination

pathway of $D^+Q_AQ_B^-$ and free energy and kinetic relations between $Q_A^-Q_B$ and $Q_AQ_B^-$. Biochim Biophys Acta 1984;766:126-140.

114. Xu Q, Gunner MR. Temperature dependence of the free energy, enthalpy and entropy of $P^+Q_A^-$ charge recombination in photosynthetic reaction centers. JPhysChem B 2000;104:8035-8043.

115. Xu Q, Gunner MR. Exploring the energy profile of the Q_A^- to Q_B electron transfer reaction in bacterial photosynthetic reaction centers: pH dependence of the conformational gating step. Biochemistry 2002;41:2694-2701.

116. Ishikita H, Knapp EW. Variation of Ser-L223 hydrogen bonding with the Q_B redox state in reaction centers from *Rhodobacter sphaeroides*. J Am Chem Soc 2004;126:8059-8064.

117. Rabenstein B, Ullmann GM, Knapp EW. Electron transfer between the quinones in the photosynthetic reaction center and its coupling to conformational changes. Biochemistry 2000;39(34):10487-10496.

118. Alexov E, Miksovska J, Baciou L, Schiffer M, Hanson DK, Sebban P, Gunner MR. Modeling the effects of mutations on the free energy of the first electron transfer from Q_A^- to Q_B in photosynthetic reaction centers. Biochemistry 2000;39(20):5940-5952.

119. Stein RR, Castellvi AL, Bogacz JP, Wraight CA. Herbicide-quinone competition in the acceptor complex of photosynthetic reaction centers from *Rhodopseudomonas sphaeroides*: a bacterial model for PS-II-herbicide activity in plants. J Cell Biochem 1984;24:243-259.

120. McComb JC, Stein RR, Wraight CA. Investigations on the influence of headgroup substitution and isoprene side-chain length in the function of primary and secondary quinones of bacterial reaction centers. Biochim Biophys Acta 1990;1015:156-171.

121. Gunner MR, Braun BS, Bruce JM, Dutton PL. The characterization of the Q_A binding site of the reaction center of *Rhodopseudomonas sphaeroides*. In: Michel-Beyerle ME, editor. Antennas and Reaction Centers of Photosynthetic Bacteria. Berlin: Springer-Verlag; 1986. p 298-305.

122. Warncke K, Gunner MR, Braun BS, Gu L, Yu C, Bruce JM, Dutton PL. Influence of hydrocarbon tail structure on quinone binding and electron-transfer performance at the Q_A and Q_B sites of the photosynthetic reaction center protein. Biochemistry 1994;33:7830-7841.
123. Warncke K, Dutton PL. Experimental resolution of the free energies of aqueous solvation contributions to ligand-protein binding: quinone-Q_A site interactions in the photosynthetic reaction center protein. Proc Natl Acad Sci USA 1993;90:2920-2924.
124. Takahashi E, Wells TA, Wraight CA. Protein control of the redox potential of the primary quinone acceptor in reaction centers from *Rhodobacter sphaeroides*. Biochemistry 2001;40(4):1020-1028.
125. Hucke O, Schmid R, Labahn A. Exploring the primary electron aceptor (Q_A)-site of the bacterial reaction center from *Rhodobacter sphaeroides*: binding mode of vitamin K derivatives. Eur J Biochem 2002;269:1096-1108.
126. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: current status and future challenges. Proteins 2006;65(1):15-26.
127. Jain AN. Scoring functions for protein-ligand docking. Curr Protein Pept Sci 2006;7(5):407-420.
128. Gilson MK, Zhou HX. Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct 2007;36:21-42.
129. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct 2003;32:335-373.
130. Garcia-Moreno EB, Fitch CA. Structural interpretation of pH and salt-dependent processes in proteins with computational methods. Methods Enzymology 2004;380:20-51.
131. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 2004;3(11):935-949.
132. Guvench O, MacKerell AD, Jr. Computational evaluation of protein-small molecule binding. Curr Opin Struct Biol 2009;19(1):56-61.

133. Wiehe K, Peterson MW, Pierce B, Mintseris J, Weng Z. Protein-protein docking: overview and performance analysis. Methods Mol Biol 2008;413:283-314.
134. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, Kuntz ID. DOCK 6: combining techniques to model RNA-small molecule complexes. RNA 2009;15(6):1219-1230.
135. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 2009;30(16):2785-2791.
136. Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 2001;308(2):377-395.
137. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins 2003;52(4):609-623.
138. Gussio R, Pattabiraman N, Zaharevitz DW, Kellogg GE, Topol IA, Rice WG, Schaeffer CA, Erickson JW, Burt SK. All-atom models for the non-nucleoside binding site of HIV-1 reverse transcriptase complexed with inhibitors: A 3D QSAR approach. Journal of Medicinal Chemistry 1996;39(8):1645-1650.
139. Stoica I, Sadiq SK, Coveney PV. Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases. J Am Chem Soc 2008;130(8):2639-2648.
140. Rastelli G, Rio AD, Degliesposti G, Sgobba M. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. J Comput Chem 2009;31:797-810.
141. Zhang T, Koshland DE, Jr. Computational method for relative binding energies of enzyme-substrate complexes. Protein Sci 1996;5(2):348-356.
142. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA. Predicting absolute ligand binding free energies to a simple model site. J Mol Biol 2007;371(4):1118-1134.
143. Zhao Y, Truhlar DG. Infinite-basis calculations of binding energies for the hydrogen bonded and stacked tetramers of formic acid and formamide and their use for validation of hybrid DFT and ab initio methods. J Phys Chem A 2005;109(30):6624-6627.

144. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucl Ac Res 2000;28:235-242.
145. Song Y, Michonova-Alexova E, Gunner MR. Calculated proton uptake on anaerobic reduction of cytochrome c oxidase: Is the reaction electroneutral? Biochemistry 2006;45:7959-7975.
146. Sitkoff D, Sharp KA, Honig B. Accurate calculation of hydration free-energies using macroscopic solvent models. J Phys Chem-Us 1994;98(7):1978-1988.
147. Gilson MK, Honig B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. Proteins: Struct Funct Genet 1988;4:7-18.
148. Pearlman DA, Case DA, Caldwell JW, Ross WS, III TEC, DeBolt S, Ferguson D, Seibel G, Kollman P. AMBER a package of computer programs for applying molecular mechanics. normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Computer Physics Communications 1995;91:1-41.
149. Kassner RJ. A theoretical model for the effects of local nonpolar heme environments on the redox potentials in cytochromes. J Am Chem Soc 1973;95(8):2674-2676.
150. Matthew JB, Gurd FRN, Garcia-Moreno B, Flanagan MA, March KL, Shire SJ. pH-dependent processes in proteins. CRC Crit Rev Biochem 1985;18:91-197.
151. Munro OQ, Marques HM. Heme-peptide models for hemoproteins. 1. solution chemistry of N-acetylmicroperoxidase-8. Inorg Chem 1996;35:3752-3767.
152. Vashi PR, Marques HM. The coordination of imidazole and substituted pyridines by the hemeoctapeptide N-acetyl-ferromicroperoxidase-8 (Fe^{II}NAcMP8). J Inorg Biochem 2004;98:1471-1482.
153. Marques HM, Cukrowski I, Vashi PR. Co-ordination of weak field ligands by N-acetylmicroperoxidase-8 (NAcMP8), a ferric haempeptide from cytochrome c, and the influence of the axial lignad on the reduction potential of complexes of NAcMP8. J Chem Soc 2000;2000:1335-1342.
154. Lide DR. CRC Handbook of Chemistry and Physics. Boca Raton: CRC Press; 1990.

155. Autenrieth F, Tajkhorshid E, Baudry J, Luthey-Schulten Z. Classical force field parameters for the heme prosthetic group of cytochrome c. J Comput Chem 2004;25(13):1613-1622.
156. Jaguar, version 7.0: SchrÖdinger, LLC, New York, NY; 2007.
157. Stellwagen E. Haem exposure as the determinate of oxidation-reduction potential of haem proteins. Nature 1978;275(7):73-74.
158. Tezcan FA, Winkler JR, Gray HB. Effects of ligation and folding on reduction potentials of heme proteins. J Am Chem Soc 1998;120:13383-13388.
159. Fantuzzi A, Sadeghi S, Valentti F, Rossi GL, Gilardi G. Tuning the reduction potential of engineered cytochrome c-553. Biochemistry 2002;41:8718-8724.
160. Gunner MR, Nicholls A, Honig B. Electrostatic potentials in *Rhodopseudomonas viridis* reaction center: Implications for the driving force and directionality of electron transfer. J Phys Chem 1996;100:4277-4291.
161. Harada E, Kumagai J, Ozawa K, Imabayashi S, Tsapin AS, Nealson KH, Meyer TE, Cusanovich MA, Akutsu H. A directional electron transfer regulator based on heme-chain architecture in the small tetraheme cytochrome c from Shewanella oneidensis. FEBS Lett 2002;532(3):333-337.
162. Kraus DW, Wittenberg JB. Hemoglobins of the *Lucina pectinata*/bacteria symbiosis. I. Molecular properties, kinetics and equilibria of reactions with ligands. J Biol Chem 1990;265(27):16043-16053.
163. Gunner MR, Saleh MA, Cross E, ud-Doula A, Wise M. Backbone dipoles generate positive potentials in all proteins: origins and implications of the effect. Biophys J 2000;78(3):1126-1144.
164. Roncel M, Boussac A, Zurita JL, Bottin H, Sugiura M, Kirilovsky D, Ortega JM. Redox properties of the photosystem II cytochromes b559 and c550 in the cyanobacterium Thermosynechococcus elongatus. J Biol Inorg Chem 2003;8(1-2):206-216.
165. Drepper F, Dorlet P, Mathis P. Cross-linked electron transfer complex between cytochrome c2 and the photosynthetic reaction center of Rhodobacter sphaeroides. Biochemistry 1997;36(6):1418-1427.

166. Xue L, Wang, Y., Xie, Y., Yao, P., Wang, W., Qian, W., Huang, Z. Effect of mutation at valine 61 on the three-dimensional structure, stability, and redox potential of cytochrome b5. Biochemistry 1999;38:11961-11972.
167. Moore GR. Control of redox properties of cytochrome c by special electrostatic interactions. FEBS Letts 1983;161(2):171-175.
168. Moore GR. Haemoproteins. In: Bendall DS, editor. Protein Electron Transfer. Cambridge: Bios; 1996. p 189-216.
169. Malarte G, Leroy G, Lojou E, Abergel C, Bruschi M, Giudici-Orticoni MT. Insight into molecular stability and physiological properties of the diheme cytochrome CYC41 from the acidophilic bacterium Acidithiobacillus ferrooxidans. Biochemistry 2005;44(17):6471-6481.
170. Gayda JP, Benosman H, Bertrand P, More C, Asso M. EPR determination of interaction redox potentials in a multiheme cytochrome: cytochrome c_3 from *Desulfovibrio desulfuricans* Norway. Eur J Biochem 1988;177:199-206.
171. Rivera M, Seetharaman R, Girdhar D, Wirtz M, Zhang X, Wang X, White S. The reduction potential of cytochrome b5 is modulated by its exposed heme edge. Biochemistry 1998;37:1485-1494.
172. Springs SL, Bass SE, Bowman G, Nodelman I, Schutt CE, McLendon GL. A multigeneration analysis of cytochrome b_{562} redox variants: evolutionary strategies for modulating redox potential revealed using a library approach. Biochemistry 2002;41:4321-4328.
173. Tomlinson EJ, Ferguson SJ. Conversion of a c type cytochrome to a b type that spontaneously forms in vitro from apo protein and heme: implications for c type cytochrome biogenesis and folding. Proc Natl Acad Sci U S A 2000;97(10):5156-5160.
174. Zhuang J, Amoroso JH, Kinloch R, Dawson JH, Baldwin MJ, Gibney BR. Evaluation of electron-withdrawing group effects on heme binding in designed proteins: implications for heme a in cytochrome c oxidase. Inorganic Chemistry 2006;45:4685-4694.
175. Springs SL, Bass, S.E., McLendon, G.L. Cytochrome b562 variants: a library for examing redox potential evolution. Biochemistry 2000;39:6075-6082.

176. Weiland TR, Kundu S, Trent JT, 3rd, Hoy JA, Hargrove MS. Bis-histidyl hexacoordination in hemoglobins facilitates heme reduction kinetics. J Am Chem Soc 2004;126(38):11930-11935.
177. Gray HB, Winkler JR. Electron tunneling through proteins. Q Rev Biophys 2003;36(3):341-372.
178. Dracheva SM, Drachev LA, Konstantinov AA, Semenov AY, Skulachev VP, Arutjunjan AM, Shuvalov VA, Zaberezhnaya SM. Electrogenic steps in the redox reactions catalyzed by photosynthetic reaction-centre complex from *Rhodopseudomonas viridis*. Eur J Biochem 1988;171:253-264.
179. Zhen Y, Schmidt B, Kang UG, Antholine W, Ferguson-Miller S. Mutants of the CuA site in cytochrome c oxidase of Rhodobacter sphaeroides: I. Spectral and functional properties. Biochemistry 2002;41(7):2288-2297.
180. Fujieda N, Mori M, Kano K, Ikeda T. Spectroelectrochemical evaluation of redox potentials of cysteine tryptophylquinone and two hemes c in quinohemoprotein amine dehydrogenase from Paracoccus denitrificans. Biochemistry 2002;41(46):13736-13743.
181. Davies MN, Toseland CP, Moss DS, Flower DR. Benchmarking pK(a) prediction. BMC Biochem 2006;7:18.
182. Wallace CJA, Clark-Lewis I. Functional role of heme ligation in cytochrome c: effects of replacement of methionine 80 with natural and non-natural residues by semisynthesis. J Biol Chem 1992;1992:3852-3861.
183. Mus-Veteau I, Dolla A, Guerlesquin F, Payan F, Czjzek M, Haser R, Bianco P, Haladjian J, Rapp-Giles BJ, Wall JD, et al. Site-directed mutagenesis of tetraheme cytochrome c3. Modification of oxidoreduction potentials after heme axial ligand replacement. J Biol Chem 1992;267(24):16851-16858.
184. Diederich F, Felber B. Supramolecular chemistry of dendrimers with functional cores. Proc Natl Acad Sci U S A 2002;99(8):4778-4781.
185. Stefano Benini MB, Stefano Ciurli, Alexander Dikiy, Matteo Lamborghini. Modulation of Bacillus pasteurii cytochrome c553 reduction potential by structural and solution parameters. JBIC 1998;3:371-382.

186. Gilson MK, Honig BH. The dielectric constant of a folded protein. Biopolymers 1986;25:2097-2119.
187. Harvey S. Treatment of electrostatic effects in macromolecular modeling. Proteins: Struct Funct Genet 1989;5:78-92.
188. Sham YY, Muegge I, Warshel A. The effect of protein relaxation on charge-charge interactions and dielectric constant of proteins. Biophys J 1998;74(4):1744-1753.
189. Liwo A, Czaplewski C, Oldziej S, Scheraga HA. Computational techniques for efficient conformational sampling of proteins. Curr Opin Struct Biol 2008;18(2):134-139.
190. Song Y, Mao J, Gunner MR. MCCE2: improving protein pKa calculations with extensive side chain rotamer sampling. J Comput Chem 2009;30(14):2231-2247.
191. Sitkoff D, BenTal N, Honig B. Calculation of alkane to water solvation free energies using continuum solvent models. J Phys Chem-Us 1996;100(7):2744-2752.
192. Stowell MH, McPhillips TM, Rees DC, Soltis SM, Abresch E, Feher G. Light-induced structural changes in photosynthetic reaction center: implications for mechanism of electron-proton transfer. Science 1997;276(5313):812-816.
193. Koepke J, Krammer EM, Klingen AR, Sebban P, Ullmann GM, Fritzsch G. pH modulates the quinone position in the photosynthetic reaction center from Rhodobacter sphaeroides in the neutral and charge separated states. J Mol Biol 2007;371(2):396-409.
194. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Research 2000;28:235-242.
195. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, J. A. Montgomery J, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB,

Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA. Gaussian 98, Revision A.9: Gaussian, Inc., Pittsburgh PA; 1998.

196. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem 2005;26(16):1701-1718.
197. Ceccarelli M, Procacci P, Marchi M. An ab initio force field for the cofactors of bacterial photosynthesis. J Comput Chem 2003;24(2):129-142.
198. Kauzmann W. Some factors in the interpretation of protein denaturation. Adv Protein Chem 1959;14:1-63.
199. Levy RM, Zhang LY, Gallicchio E, Felts AK. On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. J Am Chem Soc 2003;125(31):9523-9530.
200. Wagoner JA, Baker NA. Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. Proc Natl Acad Sci U S A 2006;103(22):8331-8336.
201. Ben-Naim A, Marcus Y. Solvation thermodynamics of nonionic solutes. J Chem Phys 1984;81(4):2016-2027.
202. Sitkoff D, Sharp KA, Honig B. Accurate calculation of hydration free energies using macroscopic solvent models. J Phys Chem 1994;98:1978-1988.
203. Froloff N, Windemuth A, Honig B. On the calculation of binding free energies using continuum methods: application to MHC class I protein-peptide interactions. Protein Science 1997;6(6):1293-1301.
204. Chang CE, Chen W, Gilson MK. Ligand configurational entropy and protein binding. Proc Natl Acad Sci U S A 2007;104(5):1534-1539.
205. Villa J, Strajbl M, Glennon TM, Sham YY, Chu ZT, Warshel A. How important are entropic contributions to enzyme catalysis? Proc Natl Acad Sci U S A 2000;97(22):11899-11904.

206. Bohm HJ. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des 1994;8(3):243-256.
207. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des 1997;11(5):425-445.
208. Wang R, Lai L, Wang S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J Comput Aided Mol Des 2002;16(1):11-26.
209. Luo R, Gilson MK. Synthetic adenine receptors: Direct calculation of binding affinities. J Amer Chem Soc 2000;122:2934-2937.
210. Nicholls A, Sharp K, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins: Struct Funct Genet 1991;11:281-296.
211. Ajay, Murcko MA. Computational methods to predict binding free energy in ligand-receptor complexes. J Med Chem 1995;38(26):4953-4967.
212. Wong CF, McCammon JA. Protein flexibility and computer-aided drug design. Annu Rev Pharmacol Toxicol 2003;43:31-45.
213. Zavodszky MI, Sanschagrin PC, Korde RS, Kuhn LA. Distilling the essential features of a protein surface for improving protein-ligand docking, scoring, and virtual screening. J Comput Aided Mol Des 2002;16(12):883-902.
214. Stilz HU, Finkele U, Holzapfel W, Lauterwasser C, Zinth W, Oesterhelt D. Influence of M subunit Thr222 and Trp252 on quinone binding and electron transfer in Rhodobacter sphaeroides reaction centres. Eur J Biochem 1994;223:233-242.
215. McAuley KE, Fyfe PK, Ridge JP, Cogdell RJ, Isaacs NW, Jones MR. Ubiquinone binding, ubiquinone exclusion, and detailed cofactor conformation in a mutant bacterial reaction center. Biochemistry 2000;39:15032-15043.
216. Pratt LR, Chandler D. Theory of the hydrophobic effect. J Chem Phys 1977;67:22.

217. Gunner MR, Robertson DE, Dutton PL. Kinetic studies on the reaction center protein from *Rhodopseudomonas sphaeroides*: The temperature and free energy dependence of electron transfer between various quinones in the Q_A site and the oxidized bacteriochlorophyll dimer. J Phys Chem 1986;90:3783-3795.

218. Gunner MR, Dutton PL. Temperature and -ΔG° dependence of the electron transfer from BPh^- to Q_A in reaction center protein from *Rhodobacter sphaeroides* with different quinones as Q_A. J Am Chem Soc 1989;111:3400-3412.

219. Morris AL, Synder SW, Zhang Y, Thurnauer MC, Dutton PL, Robertson DE, Gunner MR. Electron spin polarization model applied to sequential electron transfer in iron-containing photosynthetic bacterial reaction centers with different quinones as Q_A. J Phys Chem 1995;99:3854-3866.

220. Okamura MY, Isaacson RA, Feher G. Primary acceptor in bacterial photosynthesis: obligatory role of ubiquinone in photoactive reaction centers of Rhodopseudomonas spheroides. Proc Natl Acad Sci U S A 1975;72(9):3491-3495.

221. Prince RC, Dutton PL, Bruce JM. Electrochemistry of ubiquinones, menaquinones and plastoquinones in aprotic solvents. FEBS Letts 1983;160:273-276.

222. Nonella M, Boullais C, Mioskowski C, Nabedryk E, Breton J. vibrational spectrum and torsional potential of 2-methoxy-3-methyl-1,4-benzoquinone. J Phys chem 1999;103:6363-6370.

223. Braun BS, Benbow U, Lloyd-Williams P, Bruce JM, Dutton PL. Determination of partition coefficients of quinones by high-performance liquid chromatography. Methods Enzymol 1986;125:119-129.

224. Keske JM. Experimental and theoretical approaches to quinone binding and function at the Q(A) site of photosynthetic reaction centers. Philadelphia: University of Pennslyvania.

225. Yang CY, Sun H, Chen J, Nikolovska-Coleska Z, Wang S. Importance of ligand reorganization free energy in protein-ligand binding-affinity prediction. J Am Chem Soc 2009;131(38):13709-13721.

226. Boas FE, Harbury PB. Design of protein-ligand binding based on the molecular-mechanics energy model. J Mol Biol 2008;380(2):415-424.
227. Jiao D, Golubkov PA, Darden TA, Ren P. Calculation of protein-ligand binding free energy by using a polarizable potential. Proc Natl Acad Sci U S A 2008;105(17):6290-6295.
228. Singh N, Warshel A. Absolute binding free energy calculations: On the accuracy of computational scoring of protein-ligand interactions. Proteins 2010;78:1705-1723.