## Title

New conformations in the force-delivering element of a walking molecular motor and a computable analytic expression for the partition function of the first solvation shell



## Author

Sindelar, Charles Vaughn

## Publication Date

2002

Peer reviewed|Thesis/dissertation

*New Conformations in the Force-Delivering Element of a Walking Molecular Motor*

**and**

*A Computable Analytic Expression for the Partition Function of the First Solvation Shell*

by

*Charles Vaughn Sindelar*

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

*Biophysics*

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

.....

Date University Librarian

Degree Conferred:.....

*Without the support of all my friends and family through graduate school,*
*I don't know where I'd be.*

*Thank you, Jason, Jesse, Ken, Lillian, Steph, Vic, and Ward.*
*Thank you, Herb.*
*And thank you most of all, Mom, Dad, and Joel—for always being there when it counted.*

*CVS*
*April, 2002*

**So, I don't do much work anymore...**

**I'm too busy thinkin'!**

*From "The Paperwork Explosion" (audio recording, 1967)*

*By Jim Henson and Raymond Scott*

# Acknowledgements

***Advisors and thesis committee***

As a beginning graduate student, I was very lucky to find myself in a masterfully run laboratory. My advisor, Robert Fletterick, made managing a large, successful lab look easy—it often appeared as if he were simply gliding through his existence as a professor and letting the lab manage itself. Of course, managing a lab is never easy, as quickly becomes apparent to anyone who has actually spent time in one. Robert just happened to have the talent to make it *look* that way. While being highly thoughtful and critical of the science going on in his lab, he remained at the same time wonderfully supportive and inclusive of the people doing the science—no matter how they were struggling. This created an atmosphere where lab members felt encouraged to pursue their own ideas and to communicate freely with each other. Robert also managed to have one of the most effective "open-door policies" I have ever seen. Somehow, despite the fact that he was frequently off on trips, it always seemed like when I needed someone to bounce my latest crazy idea off, there he was in his office. And unless he was already in a meeting, he always had a moment. I cannot emphasize enough how much these traits of his helped me through my times of soul-searching and self doubt, and helped me to grow as a person. A graduate student could not ask for more from an advisor. And my acknowledgement of Robert would be remiss if I did not mention his generosity in hosting lab ski trips, lavish parties, and dinner outings in the city. I never knew the academic life could be so good!

My "other" advisor, Roger Cooke, also deserves special credit for his part in this thesis work. In fact, it was a suggestion of Roger's that led to my first truly interesting result in the kinesin field, causing my graduate career to finally take off— after four years of long-shot experiments (of my own choosing) that never seemed to work out quite right. Like Robert, Roger also had a strong open-door policy— one that may even have been too successful for his own good. Roger, in fact, became so popular with his lab that

*scheduling* a meeting with him was usually impossible. Instead, the best thing to do was just go to his office and get in line. It was always worth waiting to speak with Roger. His gift for reducing problems to their basic, physical elements (often with the aid of the infamous "back of the envelope") and his remarkable knack for teaching did much to develop my skills for careful, critical, scientific thought. Roger's generosity in hosting parties and other social events also did a great deal to make UCSF seem like more of a home, than merely a place of science.

Then, of course, there is Ken Dill—my *other* other advisor. Ken took the bold step of indulging my interest in theoretical work, developing a "side" project for me even though it was entirely separate from my principal thesis topic. Although he never showed it, I am sure this must have taken some patience on his part, given my sometimes-sporadic progress as I continually struggled to switch my mind between "kinesin" mode and "hydrophobic theory" mode. Ken shared with my other advisors the "UCSF spirit" of friendly encouragement and support, not to mention brilliant scientific insight, and I am deeply grateful for the opportunity to work with him. Happily, my work in Ken's lab took a very interesting turn, and I look forward to the opportunity to pursue it more fully upon the completion of my Ph.D.

I am also grateful for the opportunities I had to interact with my fourth thesis committee member, Ron Vale. With his very large lab and busy traveling schedule, Ron was not always as easy to track down as the rest of my committee, but when found, he always came up with deep and insightful comments.

***Laboratory collaborators***

Coming into a laboratory with absolutely no wet-lab experience, as I did, is inevitably challenging and often frustrating. I was very lucky in this regard to have a close mentor, Elena Sablin, who guided me through the initial, trying times (which actually lasted quite a while) of "learning" molecular biology (a generous statement), as well as protein purification and crystallization, from scratch. Elena managed to stay patient, and be my

friend, through all of this. Of course, Elena was far more than a teacher of laboratory skills. She also happened to be one of the world's foremost experts in the molecular mechanisms of motor proteins, with an encyclopedic knowledge of sequence and structure information, and brilliant insight into how these pieces all fit together. Sharing a lab bench space with her for six years gave me the chance to pepper her with many more questions than she should have had time to answer (and yet answer them she did). I learned more from her than I can say.

The crystal structure presented in Chapter 1 is in many ways the centerpiece of this thesis. For this structure, I am very, *very* indebted to the work of Mary Jane Budny, who joined me on the crystallization project as my determination was flagging. In truth, the crystals proved extremely hard to grow. Mary Jane was responsible for the *one* crystal that grew large enough to give atomic-resolution diffraction. But for her patience and skill, I might still be on the lab bench trying to grow that crystal.

In the kinesin chapters, I rely heavily on spectroscopic data (much of it still unpublished) collected by my fellow classmate Sarah Rice, together with coworkers Nariman Naber and Marija Matuska in the Cooke lab—and was there ever data—vast, prodigious amounts of data. I feel extremely lucky to have been able to spend my time here at UCSF collaborating with such brilliant, generous, and hard-working friends. One of the great rewards in science is to be able to sit down with new data, or new ideas, and have a jam session with someone to hash out the details. I was lucky enough to have many jam sessions in the Cooke lab. I would especially like to thank Sarah for all her enthusiasm and support, even after I began snooping through her old data, and even when I started offering "reinterpretations" of it. After moving to Stanford and getting involved in new projects there, Sarah maintained a high level of commitment to the projects she left behind at UCSF. Generously offering to drive back up here, as she has done many times, was way beyond the call of duty. Working with her has truly been a joy.

*Theory collaborators*

The hydrophobic theory work I describe in Chapter 3 involved the strong collaborative efforts of a fellow graduate student, Noel Southall, working in Ken Dill's laboratory. Together, Noel and Ken sowed the seeds of the central idea, a matrix-based analytical theory, which is developed in this final chapter. Noel deserves special thanks for making space for me on this project, which began as a part of his thesis studies. In addition to making valuable intellectual contributions to the theory side, Noel also did a lot of the supporting work to validate the analytical ideas developed in Chapter 3—running simulations and writing new code for data analysis. He also provided a balanced perspective to the project, having read and absorbed a tremendous body of literature on the hydrophobic effect (of which I have still barely scratched the surface).

I am also grateful to a postdoctoral fellow in Ken's laboratory, Tom Truskett, for many astute observations and literature referrals—his insights proved especially helpful in the development of the equations presented in Appendix 3.

*UCSF support*

My lab was a very happy place to spend so many years in graduate school, and I have great memories of the friends I worked with here: to Jon, Linda, Angela, Kai, Shirleko, Elena, Russ, Peter, Ben, Eric, Russ, Eugene, Jennifer Turner, Jenny, Sarah, Pam, Richard, Sabine, Jennifer Ekstrom, Maia, Manish, Tiffane, Carolyn, John, Stephanie, Bea, Mary Jane, thank you. The Fletterick lab, to me, has been a place full of laughter, latté, mind-expanding music, and many other splendid antidotes to the inevitable pitfalls of laboratory work (when will they learn to how to make *nice*-smelling versions of DTT and beta-mercapto-ethanol??).

I joined UCSF together with a most excellent band of classmates. We worked so well together in our classes the first year that it almost felt like cheating. I learned much, much more than I ever would have on my own. I owe this to the wonderful, inclusive

spirit (and awesome organizing power) of Erin, Sarah, Bridget, Mike, Eric, Jack, Manish, Russ, Dave, and Chris. I feel so, so lucky to have classmates that I can also call good friends. It would not have been the same without you all.

The biophysics program at UCSF is amazing in the way that people (both students and professors) continually foster a fun, supportive, and intellectual climate. There just isn't space or time to acknowledge everyone, but I must name one more person: Tack Kuntz. I am happy to add myself to the long, long list of people that Tack has graced with advice and guidance of the very best kind.

I feel like there is a special atmosphere that pervades the whole university that includes a bit of artistic flair. The annual Gordon Tomkins lectures, with their combination of music and broad, thoughtful lecture topics are a perfect example of this. If that wasn't great enough, the 10$^{th}$ floor of the UCSF Health Sciences Building (where the Fletterick lab is currently located) seems to have its own peculiar energy. Maybe it is the vivacious presence of Julie Ransom (more on her shortly), or maybe just the high concentration of structural researchers all crammed together in a tiny amount of lab space. Whatever the reason, the 10$^{th}$ floor has always been a place where you could find people to talk to, share frustrations or glee, or go grab a cup at Pasqua's (before it became the "evil" Starbuck's). While people did sometimes complain about the lack of lab space, I have often wondered whether some of the camaraderie might have been a happy side effect of the crowding. In any case, I sincerely hope the "10$^{th}$ floor spirit" persists after the imminent campus move to Mission Bay.

I am very grateful for the chances I had to interact with Peter Kollman—first as a rotation student in his lab, then as an honorary member of numerous lab ski retreat "conferences," and as a student in his most enthusiastically led quantum chemistry class. Peter was a wonderful role model as a scientist and a person, and my memories of him will always be an inspiration.

***The Julie section***

Julie Ransom deserves a very special acknowledgement for her continuing role as the heart and soul of the UCSF biophysics graduate program. Through endless recruitment seasons, pay fiascos, and even a house that caught on fire, Julie has unfailingly supported us with her unique blend of compassion and ultimate efficiency. For the devotion and spirit she gives to her job, no thanks could ever be enough. Still, thank you Julie...

***Before UCSF***

I would like to acknowledge three professors I met at MIT for their roles in inspiring me towards, and getting me started in, biophysics— a field I had never even heard of as an undergraduate. Thank you, Eric Lander, Jamie Williamson, and Bruce Tidor. In Bruce's lab I had a fabulous introduction to many of the topics I continue to study to this day. The encouragement and inspiration of my friends there, including Zak Hendsch, Eric Simon, Gareth White, Tau-Mu Yi, and Christina Jarque-Uribe, also greatly helped my decision to continue forward to graduate school.

And my gratitude to the original, the one, the only, Brian Farrell for showing me that science could be *cool*.

*New Conformations in the Force-Delivering Element of a Walking Molecular Motor*

**and**

*A Computable Analytic Expression for the Partition Function of the First Solvation Shell*

Charles Vaughn Sindelar

# Abstract

The current model for the kinesin molecular motor protein critically involves a short force-transmission segment called the neck linker, which is thought to drive kinesin's ATP-powered movement along the microtubule track. Here I present a series of crystallographic and spectroscopic experiments that address two crucial issues related to the neck linker: how it is controlled by kinesin's nucleotide-sensing 'relay' mechanism, and the number of specific conformational states it visits during a complete enzymatic cycle.

By solving a new crystal structure of the kinesin construct used in these studies I conclusively demonstrated that the neck linker explores two principal structural states (one of which is disordered) under microtubule-free conditions. Furthermore, by examining site-specific EPR labels on the neck linker, I found that kinesin's nucleotide-sensing mechanism is apparently 'turned off' when the motor is not bound to the microtubule. The crystal structure comparison revealed a likely explanation for why nucleotide sensing does not happen until microtubule binding occurs.

Using new analysis techniques developed in these studies, I then globally reexamined the available probe data for the neck linker. Remarkably, the analysis suggested two new, highly ordered conformations of kinesin's neck linker, critically involved in the motor's motility cycle. These neck linker conformations are likely to interact directly with the microtubule, and form the basis of an elaborate new model for how dimers of kinesin

move along a microtubule protofilament with strikingly asymmetrical "left" and "right" steps.

In the final chapter I examine a theoretical topic in biophysics, the calculation of solvation effects on biomolecules. I introduce an analytical formula called SWYZLE (Shell Waters Yield Z by Lattice Enumeration) that exhaustively treats water configurations in the localized region surrounding the solute, thereby generating the partition function of this subvolume. Preliminary tests with a simple water energy function have demonstrated that SWYZLE can make solvent structure predictions on par with the conventional molecular dynamics or Monte Carlo methods, at reduced computational cost. Furthermore, the subvolume partition function calculated by SWYZLE can be related to large-system thermodynamics, which may allow solvation free energies of biomolecules to be calculated in a new way.

Robert J. Fletterick, Ph.D

Advisor

# TABLE OF CONTENTS

# LIST OF FIGURES

***Introduction to the kinesin studies***



***Chapter 1: Two Conformations in the Human Kinesin Powerstroke Defined by X-ray Crystallography and EPR Spectroscopy***



***Chapter 2: Two New, Highly Ordered Conformations of Kinesin's Putative Force-Delivering Element***

***Chapter 3: A Computable Analytic Expression for the Partition Function of the First Solvation Shell***



***Appendix 2: Summary of EPR data***

# LIST OF TABLES

# Introduction to the kinesin studies

When I started graduate school, in 1995, it was a very exciting time to begin the study of molecular motor proteins. Only two years earlier, the first atomic-resolution structure had been solved of myosin(Rayment, Rypniewski et al. 1993), the essential motor protein that drives muscle movement. The X-ray structure of myosin appeared to confirm the long-standing hypothesis that this protein operated in an appealing, mechanical way—by grabbing an apparently passive partner filament (actin) and delivering a "rowing stroke" (powered by the hydrolysis of one molecule of ATP) in order to slide past it. One year after the myosin structure, the X-ray crystal structure of the mammoth F1-ATPase was published—which resembled (and was later confirmed to be) nothing less than a *rotary* motor, complete with a crankshaft, crafted at the molecular level(Abrahams, Leslie et al. 1994).

Then Elena Sablin and Jon Kull in Robert Fletterick's laboratory at UCSF delivered two atomic-resolution crystal structures(Kull, Sablin et al. 1996; Sablin, Kull et al. 1996) of a third kind of molecular motor, kinesin (Figure 1A). Somewhat similar to myosin, kinesin is typically involved in movement of, or along, filaments and is driven by ATP hydrolysis. However, compared to the rather large, whale-shaped myosin (composed of >1100 amino acids, counting essential light-chain domains), kinesin is quite lean, having a functional domain less than half the size of myosin's. This small size made kinesin especially suitable for premier structural techniques like X-ray crystallography. By the time I joined the graduate program at UCSF, the dimer of kinesin was well known to "walk" continuously along its partner filament, the microtubule, as a single molecule(Howard, Hudspeth et al. 1989). I found this feature very appealing, coming as I had to UCSF with an undergraduate background in mechanical engineering. The fact that kinesin had already been crystallized added to the appeal, as it meant that a lot of the biochemical groundwork had already been laid out. Perhaps, I thought, even a wet-lab neophyte like myself could ease into experimental studies of a motor protein without too much trauma and pain.

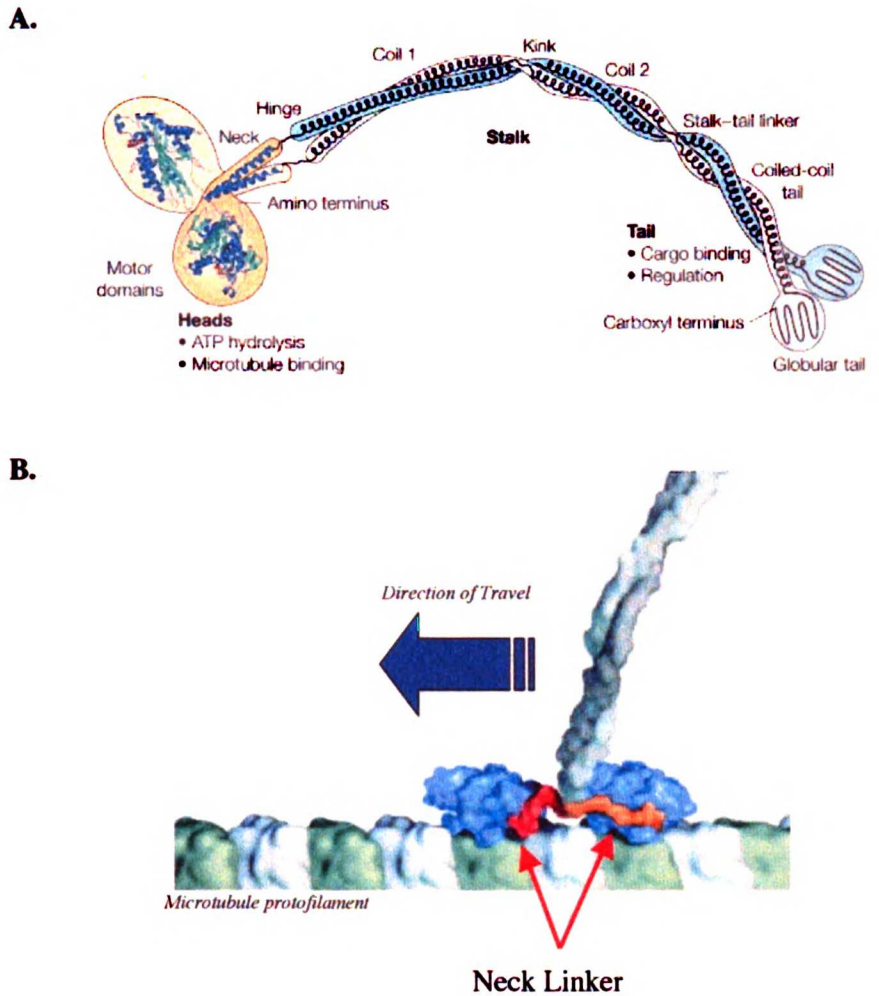



Figure 1. Overall architecture of the kinesin molecular motor protein

A. In conventional kinesin, dimers form a dumbbell shape. A long, coiled-coil "stalk" connects a pair of motor domains, at one end of the dumbbell, to a pair of cargo-binding domains at the opposite end. Figure courtesy of Woehlke and Schliwa(Woehlke and Schliwa 2000). The motor domains by themselves are fully capable of supporting kinesin's "walking" movement, either singly or as artificially-connected dimers (Michio Tomishige, manuscript in preparation). The end of stalk nearest the motor domains is called the "neck," and is not essential for motor function (although certain sequence-specific features of the neck apparently enhance the "run length" of single-molecule walks). B. Proposed geometry of a walking dimer of kinesin. The C-terminal segment of the motor domain, connecting to the neck segment of the stalk dimerization region, is called the neck linker (colored red and orange in the two dimer subunits). The neck linker changes conformation depending on which nucleotide is bound in the motor domain, a feature that is crucial to kinesin's motility. Figure taken from the review by Vale and Milligan(Vale and Milligan 2000)

My decision to enter the molecular motor field was cemented by the very strong collaborations that existed (and persist) between professors both here at UCSF and outside the university. My soon-to-be professor, Robert Fletterick, was part of a close-knit group that included Roger Cooke, Ron Vale, and Bob Mendelson at UCSF, as well as Jim Spudich at Stanford and Stefan Highsmith at the University of the Pacific. Funded by a Program Project NIH grant, the labs of these professors met regularly and formed an active, vivacious community united towards the common purpose of studying motor proteins.

***Initial findings of the kinesin X-ray structures***

Elena and Jon's crystallography work provided motor protein researchers with a treasure trove of structural information, priming kinesin for a new round of detailed interrogation with the more sophisticated probe techniques. The kinesin structures gave strong clues as to the whereabouts of kinesin's microtubule binding domain, soon to be confirmed by experiments where the crystal structures were fit into cryo-electron microscope images of the motor-microtubule complex(Woehlke, Ruby et al. 1997). Also, certain elements in the kinesin structures, with implied roles in nucleotide sensing, were found to strongly resemble analogous elements in the structure of myosin(Kull, Sablin et al. 1996). The similarity was quite surprising in light of the utter lack of sequence similarity between the two motors, and has since led to the hypothesis that kinesin and myosin are highly divergent descendants of a common ancestor(Kull, Vale et al. 1998). Kinesin's (and myosin's) nucleotide-sensing elements were also found to resemble, more remotely, those found in the G-protein family involved in signal transduction(Sablin, Kull et al. 1996).

Taken as a whole, these features clearly suggested a general functional format for these molecular motors. The nucleotide-sensing elements somehow detected and relayed conformational change from the ATP binding pocket to select regions of the protein structure, including the microtubule-binding domain. Even after the first two crystal structures of kinesin had been solved, however, it was still not clear what nucleotide-

induced conformational changes would look like in these motors, or how they could contribute to a motility mechanism.

### *A motility-driving element: the neck linker*

A subsequent round of discoveries, motivated by features of the crystal structures, soon began to shed light on the question of nucleotide-driven conformational change in kinesin. First came mutagenesis experiments that identified what appeared to be a class-specific, modular element, the "neck linker," that could be swapped between various kinesin motor domains. Stunningly, swapping the neck linker between different kinesin family members was found to be able to *reverse* kinesin's walking direction along the microtubule, implicating the neck linker directly in the motility mechanism(Case, Pierce et al. 1997; Henningsen and Schliwa 1997). Soon thereafter, site-specific probe measurements by my classmate Sarah Rice (working with her advisor, Ron Vale, and Roger Cooke) revealed that the neck linker actually made a nucleotide-dependent conformational change during kinesin's motility cycle(Rice, Lin et al. 1999). The suite of structure information obtained by Sarah and her collaborators led quite directly to a simple, intuitive model that explained the walking mechanism of a kinesin dimer. The essential idea of this model was that the neck linker behaved as a "pulling" element, yanking on the dimer partner to displace it towards the site of the next walking step.

Subsequent experiments have provided a plethora of new data, including a variety of crystal structures, which continue to refine our view of kinesin's mechanism. It has become quite clear, for example, that kinesin uses its so-called nucleotide "switch elements" not only to regulate microtubule-binding affinity, but at the same time to control the state of the neck linker, in a concerted way that allows the motor to "walk" (see, for example (Vale, Case et al. 2000; Vale and Milligan 2000)).

### *Questions to be addressed in this work*

Many atomic-resolution features of kinesin's mechanism have remained elusive. For example, it has not been established how many discrete conformations the neck linker actually visits during kinesin's walking cycle. Thus, it has been difficult to say whether kinesin delivers a true "powerstroke" like its cousin myosin is thought to do (Block 1996). A second problem lies in understanding conformational changes in kinesin's nucleotide switch elements. These switch elements can be inferred to follow the same general principles as the ones found in the G-proteins and in myosin. Still, while kinesin's switch elements have been crystallized many times, the structures have refused to yield direct evidence for how the nucleotide sensing actually takes place(Muller, Marx et al. 1999; Kikkawa, Sablin et al. 2001) (see Appendix 1 and Chapter 1). These two related problems—the nucleotide switch mechanism, and the role of the neck linker—became the focus of my thesis work on kinesin.

A chief problem in studying the atomic details of kinesin's nucleotide-controlled "switch" mechanism is that kinesin forms a complex with a gigantic macromolecular partner, the microtubule, during much of its functional cycle. The kinesin-microtubule complex is beyond the reach of conventional high-resolution structural techniques like X-ray crystallography or NMR spectroscopy. Microtubule binding almost certainly alters kinesin's "switch" machinery, making inaccessible many atomic-resolution features of kinesin's nucleotide sensing apparatus. A great deal of deduction has therefore been required in order to infer kinesin's true mechanistic features from the available, microtubule-free crystal structures in combination with the biochemical data. Nevertheless, I hope this thesis will make it clear that many new, interesting, and even startling features of this motor's mechanism are now at hand.

# Chapter One:
# Two conformations in the human kinesin powerstroke defined by X-ray crystallography and EPR spectroscopy

## Summary

Crystal structures of the molecular motor kinesin show conformational variability in a structural element called the neck linker. Conformational change in the neck linker, initiated by ATP exchange, is thought to drive kinesin's movement along the microtubule track. We use site-specific EPR measurements to show that, when microtubules are absent, the neck linker exists in equilibrium between two structural states (disordered, and "docked"). This equilibrium is *not* controlled by kinesin's active-site nucleotide; however, we find that sulfate can specifically bind near the nucleotide site, and can stabilize the docked neck linker conformation as we confirmed by solving a new crystal structure. Comparing the docked and undocked crystal structures of our construct reveals how microtubule binding may activate kinesin's nucleotide-sensing mechanism, allowing neck linker transitions to power motility.

## Introduction

Kinesin is an essential, ubiquitous molecular motor protein in organisms that contain microtubules. Powered by ATP, kinesin performs a variety of essential cellular functions, transporting vesicles in axons and participating in restructuring during cell division(Hirokawa 1998; Goldstein 2001). Dimers of so-called "conventional" kinesin (hereafter called "kinesin") proceed along the protofilament tracks of microtubules in a stepwise fashion, with alternating binding and release steps of the monomer catalytic domains(Block 1998; Woehlke and Schliwa 2000; Howard 2001). Many essential details of this walking mechanism, related to the motor's ability to convert the chemical energy of ATP hydrolysis through a series of conformational changes to a translational mechanical force and work, are not yet elucidated.

Recently, a highly conserved structural element called the "neck linker" has been implicated in kinesin's mechanism of movement and force production(Rice, Lin et al. 1999; Case, Rice et al. 2000; Vale, Case et al. 2000; Rosenfeld, Jefferson et al. 2001). This short (~15 amino acid) segment, located at the very C-terminus of conventional

kinesin's catalytic core, connects this domain to the coiled-coil stalk that leads to cargo-binding domains and links kinesin to its dimer partner. Experiments have led to a model in which the neck linker delivers a "power stroke," changing its conformation to drive the stalk forward over the microtubule-bound catalytic core, like a leg drives a walking person forward so the next foot can step.

The conformations that the neck linker explores during kinesin's catalytic cycle remain poorly understood. In the crystal structure of human kinesin the neck linker is disordered(Kull, Sablin et al. 1996), but in crystallized rat kinesin and others the neck linker extends along the side of the catalytic core domain, stabilized in a rigid conformation(Kozielski, Sack et al. 1997; Sack, Muller et al. 1997; Kikkawa, Sablin et al. 2001; Song, Marx et al. 2001). The binding of ATP at kinesin's catalytic site has been suggested to trigger this "docked" neck-linker conformation(Rice, Lin et al. 1999; Vale and Milligan 2000), which would offset the stalk and partner head in the direction of travel when kinesin is complexed to the microtubule. In this way, ATP binding would transform the neck linker away from some other conformation, possibly disordered, that led backwards to where the partner head attached in the previous step. However, in an apparent contradiction of the proposed mechanism, the docked neck-linker structures can be formed with ADP in the active site, and do not require the binding of ATP (or chemical analog)(Muller, Marx et al. 1999; Song, Marx et al. 2001). The crystallized complex of the kinesin family member KIF1A together with the ATP analog AMPPCP was shown to have a "docked" conformation (although much of the neck linker was not visible), but this conformation could also be produced with ADP(Kikkawa, Sablin et al. 2001).

To explain why crystallized conformations of the neck linker are variable in the presence of ADP, it has been proposed that both docked and undocked conformations of kinesin's neck linker coexist in this nucleotide state (Kikkawa, Sablin et al. 2001) when microtubules are absent. If both neck-linker states are in close equilibrium under such conditions, however, it becomes necessary to explain why ATP exchange was not found to push this equilibrium towards the docked form in probe studies of the free motor(Rice,

Lin et al. 1999). In other words, does nucleotide exchange in kinesin, free in solution, initiate the same conformational change in the neck linker as it does in microtubule-bound kinesin? If so, why does the conformation of the neck linker appear not to be controlled by nucleotide in available crystal structures? Because these questions have not yet been answered, it remains ambiguous which crystallized form of kinesin (if any) displays the neck linker conformation of the ATP-bound, microtubule-complexed motor(Schief and Howard 2001).

Here we use EPR spectroscopy to show that without microtubules, the neck linker exists in equilibrium between an apparently disordered conformation, and one that is docked to the motor core. Strikingly, a crystallization reagent (sulfate) perturbs microtubule-free kinesin's neck linker towards the docked conformation. Furthermore, a new x-ray crystal structure of our kinesin construct under high-sulfate conditions confirms that neck-linker docking occurs. EPR measurements indicate that this docked conformation of the neck linker is similar (or identical) to one that occurs in ATP-bound, microtubule-bound kinesin, in agreement with previous conclusions(Rice, Lin et al. 1999; Vale and Milligan 2000; Kikkawa, Sablin et al. 2001).

Our EPR measurements show that nucleotide exchange has no effect on neck-linker docking under microtubule-free conditions, in contrast to kinesin-microtubule complexes where the motor's nucleotide state determines the extent of the docking(Rice, Lin et al. 1999). Comparing the docked and undocked crystal forms now available for our construct suggests a reason for this difference. Under microtubule-free conditions, the observed disorder and plasticity in a key loop (L11) appears to introduce uncoupling in a relay element that includes the so-called switch II, such that nucleotide-controlled docking of the neck linker is disabled. We propose that microtubule interactions with L11 could rigidify the connection between switch II elements, explaining how microtubule binding engages the nucleotide-driven docking mechanism.

## Results

### *In microtubule-free kinesin, neck-linker docking is insensitive to the nucleotide state*

To test the influence of various conditions on the structure of kinesin's neck linker, we monitored the structure of the kinesin neck linker on a monomeric human construct. In our experiments, structure was reported by a single EPR probe attached at any of three introduced cysteine sites (see Figure 3) along the neck linker of a kinesin construct with its natural cysteines eliminated (these were the same constructs used by Rice et al.(Rice, Lin et al. 1999)). The length of the constructs, 349 amino acids, was the same as for the previously crystallized human construct.

In microtubule-free, ADP-complexed kinesin, neck-linker docking was not strongly favored, as indicated by absence of a strong low-field shoulder in the spectrum in Figure 1A. Furthermore, in microtubule-free kinesin the neck linker is apparently insensitive to whether ADP or non-hydrolyzable ATP analogs are present. In striking contrast to observations of the microtubule-bound motor, under microtubule-free conditions the EPR spectra of kinesin in different nucleotide states are superimposable, as indicated by the overlaid spectra for ADP and AMPPNP in Figure 1B. The EPR signal agreement for ADP and ATP analogs extended over the entire range of temperatures examined, from 2° C to 35° C, and over all three neck-linker probe sites examined. The nucleotide analogs ADP•aluminum fluoride and ADP•beryllium fluoride also produced docking indistinguishable from the ADP conditions (results not shown). While nucleotide exchange could therefore not be confirmed by our experiments, more sensitive monitoring techniques have shown that AMPPNP, ADP•aluminum fluoride and ADP•beryllium fluoride will bind to our construct under the conditions reported here(Rosenfeld, Correia et al. 1996; Rosenfeld, Rener et al. 1996; Xing, Wriggers et al. 2000).

A deconvolution procedure was used to quantify the populations of docked and undocked neck linker, by fitting EPR signals to linear combinations of experimentally determined

"pure" signal components (see Figure 7 and Methods). These calculations demonstrate that at least 20% of the neck linker is docked at room temperature (20°C), and that the fraction increases to > 50% at the lowest temperature examined (2°C). A linear trend is seen in the corresponding van't Hoff plot where the docked and undocked fractional populations are converted to a free energy (Figure 2A). The large slope of the plots indicates an exceptionally strong enthalpy for the docking transition (estimated as –70 kJ/mol for AMPPNP-bound and –52 kJ/mol for ADP-bound kinesin, based on the linear fit), which is consistent with EPR observations of neck-linker docking equilibrium in microtubule-bound kinesin(Rice, Lin et al. 1999).

Also evident in Figure 2A is the equivalence of docking in the ADP and AMPPNP-bound, microtubule-free motor. Taking the various sources of error into account (see Methods), the free energy and enthalpy of neck-linker docking is identical, for kinesin with either ADP or AMPPNP bound, in the microtubule-free state. This observation contrasts with the nucleotide-dependent docking seen by analogous measurements of microtubule-bound kinesin in Figure 2A, where the binding of AMPPNP enhances neck-linker docking by 4kJ/mol relative to ADP binding. Thus, microtubule binding is apparently required in order for the bound nucleotide to exert control over the neck linker conformation.

***High-sulfate conditions used for crystallization enhance neck linker docking***

While the neck linker had no detectable response in the above nucleotide-exchange experiments, it was sensitive to a reagent found in the crystallization condition for monomeric rat kinesin, lithium sulfate. As seen in Figure 1C, lithium sulfate generates a low-field shoulder on the EPR spectrum indicating neck linker docking on the protein core domain. The signal strongly resembles that produced by the neck linker of microtubule-complexed, AMPPNP-bound kinesin, as demonstrated by the overlay in Figure 1C. By contrast, crystallization conditions used for the originally reported human crystal structure, with PEG4K (~10% weight/volume) as the precipitant, did not significantly modify the neck-linker signal. The lithium sulfate docking effect was seen

at all three neck linker positions monitored by EPR (results for positions C328 and C330 not shown).

Quantifying the EPR signal components for the lithium-sulfate experiments, as shown in Figure 2B, revealed a high enthalpy for the docking transition, as found for the neck linker in other conditions explored by EPR. The estimated value by the linear fit to the van't Hoff plot, -40kJ/mol, was somewhat lower than seen for microtubule-free conditions (Figure 2A). The docking free energy was also determined as a function of lithium sulfate concentration as shown in Figure 2C. The stabilizing effect on the free energy of neck linker docking was linear with concentration over the entire range examined, 0M to 1M lithium sulfate. This linear dependence, even at 0.1M sulfate, confirms that crystal nucleation (which is highly cooperative) does not produce the structural change.

Repeating the EPR measurements in the presence of either 1M ammonium sulfate or 1M lithium chloride separated the effects of lithium and sulfate. The ammonium sulfate condition increased neck linker docking identically as 1M lithium sulfate, while lithium chloride had no effect (data not shown). Thus, the enhancement of neck linker docking was not due to simple charge screening. Phosphate buffer (1M) was found to enhance neck linker docking to a similar degree as the sulfate-containing condition (Figure 2C). However, a number of other added salts and buffers including 1M NaSCN, 1M $(CH_4)_3NCl$, and 1M $NaNO_3$ failed measurably perturb the docking equilibrium (Figure 2C). Significantly, crystallization conditions used for the originally reported human kinesin crystal structure (PEG 4000 in combination with acetate buffer pH 4.6) caused the protein solution to become cloudy, indicating aggregation. Even in this case, however, the neck linker EPR signal remained identical to that found in standard buffer conditions (data not shown). Thus, the neck linker docking produced by sulfate and phosphate is a highly specific effect, and is not readily explained by either aggregation or ionic screening.

*Crystallization of the Human Construct in High-Sulfate Conditions*

The structural effects of sulfate on the human monomeric construct K349 were examined by solving the 2.7Å X-ray structure of a crystal grown in the presence of 1.8M lithium sulfate (Table 1). Satisfying the prediction of EPR experiments, the neck linker (red, in Figure 3A) of the new crystal structure was docked. Furthermore, the neck linker docking was accompanied by a more global conformational change of the catalytic motor domain, as has been seen in other kinesin X-ray structures that display docked neck linkers. Docking in the new crystal structure (the other K349 structure will be called "undocked") is accompanied by a concerted, >2Å movement by a group of ~40 connected residues, the switch II cluster (residues 255-295, shown in green).

While the neck linker and the switch II cluster comprise the regions of greatest conformational change, Figure 3B reveals that smaller shifts occur extensively across many of kinesin's surface subdomains, leaving only part of the central ß-sheet unchanged (<0.5Å) relative to the least-squares alignment of the K349 structures. The subdomain shifts are reflected in a larger-than-expected root-mean-square (RMS) deviation of K349's α-carbon backbone between the two structures, 0.9Å, even after omitting the switch II cluster and the neck linker from the comparison. A movement to accommodate the new conformation of the neck linker is made by the three-strand ß-sheet that points upward in Figure 3 to form the tip of kinesin's arrowhead. This element bends toward the neck linker to make a ß-sheet interaction with it in the docked structure, resulting in a minor distortion of the kinesin arrowhead. In addition, kinesin's N-terminal segment (residues 3-8) rearranges to form additional ß-sheet interactions to support the docked neck linker on its opposite, solvent-facing side. Other segments displaying minor shifts, such as L5 and the cluster L1-ß1a-ß1b, do not appear to correlate directly with neck-linker docking or switch II cluster movement. These shifts are accompanied by multiple changes in the surface salt-bridge network, none involving highly conserved residues.

The conformation of the new K349 crystal structure is markedly similar to that found in both monomeric and dimeric(Kozielski, Sack et al. 1997; Sack, Muller et al. 1997)

crystal structures of rat kinesin, both of which also display docked neck linkers. This agreement is almost certainly triggered by the presence of >1M sulfate in all three crystallization conditions (see Discussion). However, despite a 90% sequence similarity between human and rat kinesin sequences and similar crystallization conditions, the RMS deviation of the least-squares $C_\alpha$-carbon alignment between the docked human and rat monomer structures is 0.8Å, reflecting significant conformational variability. As shown in Figure 3C, shifts in surface loops cause the RMS increase, and the pattern of involvement is similar to that seen in the K349 structure pair (Figure 3B). These results show that the precise placement of surface loops such as L5 and L1-ß1a-ß1b depends on subtle variations in protein sequence or crystallization conditions, suggesting these regions have inherent flexibility. Surface loop flexibility has also been inferred by other cross-species comparisons of kinesin crystal structures(Sablin, Kull et al. 1996; Sack, Kull et al. 1999) and is seen in structures of KIF1A(Kikkawa, Sablin et al. 2001).

***Conformation-dependent Hydrogen Bonds and Salt Bridges***

The switch II cluster breaks four surface salt bridges when it relocates in docked K349, but the residues involved are poorly conserved. A total of 10 other surface salt bridge pairs break or form in K349's conformational transition, again with no residues showing a strong conservation pattern. Two absolutely conserved salt bridge pairs do change their status, however, in the switch from an undocked to a docked structure. A network involving the conserved triad Arg203 (from switch I), together with Glu236 and Asn255 (both part of switch II) appears in the docked K349 structure, as was also reported in the structure of monomeric rat kinesin(Sack, Muller et al. 1997). The network is disrupted by the ~2.5Å retreat of the switch II cluster, which breaks the Asn255-Arg203 hydrogen bond and apparently destabilizes the Arg203-Glu236 salt bridge which is lost as well.

The other absolutely conserved salt bridge pair that makes a change is involved in an ion-binding site, shown in Figure 5. Together with Lys187 and His191, the absolutely conserved Arg190 forms a positively charged interaction site that contains a specifically bound sulfate ion. An acetate anion occupies this same site in the original, undocked

K349 structure. Presumably in response to the greater charge of the sulfate, the guanidinium group of the arginine flips by 180° to move closer to the anion in the docked K349 structure, in the process losing a salt bridge to the conserved, mostly buried Glu136.

## Discussion

The EPR experiments presented here demonstrate that, without microtubules, the kinesin neck linker exists in equilibrium between two states, docked and disordered. As we have shown, this equilibrium is modified by the presence of molar quantities of sulfate. Such conditions stabilize the neck linker's docked state and allowed us to produce a new stable conformer of the monomeric human kinesin construct, K349. Our crystal structure contains the first complete, high-resolution characterization of a docked neck linker in human kinesin, and is very similar to the reported structures from the rat(Kozielski, Sack et al. 1997; Sack, Muller et al. 1997; Muller, Marx et al. 1999)—and contrasts with the published K349 structure, whose neck linker is disordered. The two conformations now available for our construct supply detailed structural models for both neck-linker states implied by the EPR experiments. In combination with the information gained by EPR analysis, these structures force a reinterpretation of kinesin's nucleotide-sensing apparatus, as we now show.

### *Nucleotide exchange does not perturb neck linker docking*

Our nucleotide exchange studies with EPR show that AMPPNP does not have a significant effect on neck linker docking under microtubule-free conditions— in the microtubule-bound motor, by contrast, bound AMPPNP promotes neck-linker docking relative to ADP(Rice, Lin et al. 1999). Both of our K349 crystal structures contain ADP at the nucleotide-binding site, proving that the conformational change is not a result of nucleotide exchange. Moreover, sulfate does not occupy the γ-phosphate site of the nucleotide active site, even though our EPR results show that this ion drives the conformational changes observed in the neck linker of the crystallized K349 structures.

These structures, therefore, do not provide direct evidence for how ATP might propagate conformational change through the switch regions to the neck linker. Furthermore, the microtubule-free EPR results imply that even were AMPPNP to be seen in a K349 crystal structure, it would not induce productive docking of the neck linker. If KIF1A kinesin, which has been crystallized in both docked and undocked conformations(Kikkawa, Sablin et al. 2001), obeys the same mechanism then the nucleotide exchange in these structures (from ADP to AMPPCP) most likely did not alter the energetics of neck-linker docking either. Consistent with this view, the *docked* form of KIF1A could be crystallized with either ADP or AMPPCP.

Our EPR measurements indicate that both conformational states represented by the K349 crystal structures are readily accessible under microtubule-free conditions. As seen in Figure 2, the free energy difference between conformations is near zero at ~4°C, the temperature at which kinesin is commonly crystallized. This closeness in energy suggests that, barring other factors, crystal structures of the conventional kinesins might show either docked or undocked neck-linker conformations. This prediction seems to be borne out by the dual conformers of KIF1A•ADP, and also the structures of *neurospora* kinesin(Song, Marx et al. 2001) (where the neck linker is docked) and *eg5*(Turner, Anderson et al. 2001) (where the neck linker is undocked).

***Sulfate (or phosphate) can cause neck-linker docking in K349***

The presence of sulfate in the crystallization condition perturbs the neck linker of K349 by stabilizing docking (as in Figures 1, 2). EPR measurements show that 75% (or more) of the neck linker population is docked at the highest sulfate concentration tested, 1.5M, at 4°C (results not shown). The sulfate effect, therefore, favored crystals of docked K349 under these conditions. Sulfate also likely contributed to the appearance of docked neck-linker conformations in the reported X-ray structures of both monomer and dimer rat kinesin (crystallization conditions of both included >1M sulfate, unlike the conditions used to crystallize KIF1A, *eg5* and *neurospora* kinesin).

While sulfate's mechanism for enhancing neck-linker docking remains unclear, EPR experiments ruled out several possibilities, including aggregation, nonspecific ionic screening, and crystallization contacts (see Results). And, while sulfate can strengthen the hydrophobic effect, other ions known to strongly enhance ($N(CH_3)_3+$) or weaken (SCN- and $NO_3$-) hydrophobic interactions(Baldwin 1996) had no measurable effect on neck linker docking (Figure 2B). Another possibility is that sulfate interacts nonspecifically with the K349 structure in a way that cannot be visualized by crystallography. However, this latter scenario demands an explanation for why sulfate and phosphate had a significant docking effect while all the other ions tested did not.

Could one or both of the specifically bound sulfate ions near switch I of the docked K349 structure, seen also in the high-resolution rat monomer structure (Figures 3A, 3C, 5), effect an allosteric change that enhances neck docking? While the conformational changes seen in K349 do not suggest an obvious mechanism, features of the binding site support a biochemically important role. The left-hand sulfate in Figure 5 is poised between the γ-phosphate site of the nucleotide pocket and the solvent exterior, and is in close proximity to three positively charged residues, including the absolutely conserved Arg190. As noted in the Results, this arginine breaks a salt bridge with conserved Glu136 in the sulfate-bound structure. Mutation of Arg190 to alanine is accompanied by a ~3-fold velocity decrease in microtubule motility assays (R. Vale, personal communication). Also, phosphate and sulfate are interchangeable in our EPR experiments (Figure 2B), suggesting that sulfate may actually behave like a phosphate mimic.

It is therefore conceivable that the ion site may be a sensor, functioning as a transitory site for freed phosphate following ATP hydrolysis, stabilizing neck-linker docking even after the γ-phosphate site is vacated. Since kinesin may dissociate from the microtubule while in the ADP•$P_i$ state(Schief and Howard 2001), such a feature could be important for the motility mechanism, but this line of investigation needs more work. In particular, the ~1M amounts of phosphate that are required to activate the mechanism here are

unphysiological; we note, however, that *microtubule*-bound kinesin binds phosphate with sub-millimolar affinity(Rosenfeld, Rener et al. 1996).

***The position of the switch II cluster is variable***

In a model proposed to explain nucleotide-driven structural changes in kinesin, the switch II cluster acts as a rigid relay element, communicating changes at the nucleotide-binding site to the remote site of neck linker docking(Vale and Milligan 2000; Kikkawa, Sablin et al. 2001). In the K349 structure pair described here, the switch II cluster makes the predicted, rigid-body (self-aligned RMSD of the cluster's alpha carbons is 0.7Å) shift in response to neck-linker docking, consistent with a role in such a mechanism. Relative to the global K349 structure alignment, the cluster's center-of-mass moves by 2.9Å while the long helical axis (residues 258-271) rotates by 7.5° (using the program GEM(Browner, Fauman et al. 1992)). In the transition between docked and undocked states of KIF1A, the switch II cluster's movement is even greater, with a center-of-mass movement of 5.1Å and a rotation of 20°(Kikkawa, Sablin et al. 2001).

Alignments among the docked structure and undocked kinesin structures (all of which have docked neck linkers), however, reveal significant alignment differences even in what are presumably the same "upstroke" states of the cluster. Even when comparing between rat structures only, the cluster's long helical axis can rotate by as much as 5°. As a result of this variability, geometry permits the switch I-switch II hydrogen bond between Arg203 and Asn255 (discussed above) in only two of the four rat and human docked structures, suggesting this interaction is not strong enough to specify the switch II cluster's upstroke position.

Variability in the switch II cluster's positioning is made possible by two architectural features of kinesin. The cluster rests on a large, uniform bed of hydrophobic sidechains from kinesin's core ß-sheet domain, so the resulting interface is devoid of connecting hydrogen bonds; it is also characterized by poor packing interactions(Sablin, Kull et al. 1996). Direct evidence of poor packing at the switch II interface is seen in the transformation from undocked to docked K349 structures: the sidechain of Phe 82 from

the central beta sheet rotates by 90° to find a slightly larger cavity in the docked motor state, whereupon the ring density disappears (apparently due to rotational mobility). More importantly from the standpoint of a relay mechanism, both the N- and C-terminus of the switch II cluster feature flexible hinges(Kikkawa, Sablin et al. 2001). As illustrated in Figure 4, the N-terminal hinge has particular functional significance, because it connects the cluster to the conserved nucleotide-sensing segment of switch II. Flexibility in this hinge, L11, which is extended and partly disordered, permits the protein backbone in the sensing segment to move no more than 0.7Å in the global structural alignment between the two conformational states of K349. Just on the other side of L11, the switch II cluster moves by > 2Å in the same alignment. This relative movement is accommodated by rearrangements in the visible residues at the base of L11, whose conformation is highly variable in crystal structures(Sablin, Kull et al. 1996; Kikkawa, Sablin et al. 2001; Yun, Zhang et al. 2001).

The flexibility introduced between switch II elements by loop L11 is a key difference between the switch II architecture of kinesin and that of myosin, as illustrated in Figure 4B. In myosin, a rigid-backbone geometry replaces the floppy L11 in the connection between the nucleotide sensing segment and the relay helix (equivalent to the switch II cluster), implying a sterically enforced communication pathway between ATP and the relay elements of switch II(Vale and Milligan 2000). In kinesin, on the other hand, it is not clear that ATP-induced change in switch II's nucleotide-sensing segment could productively reposition the switch II cluster enough to influence neck-linker docking(Wriggers and Schulten 1998; Xing, Wriggers et al. 2000). Indeed, an uncoupling of these switch II components could plausibly explain why AMPPNP does not dock the neck linker, at least in the absence of microtubules.

Microtubule binding, however, could modify uncoupling in switch II, because L11 interacts directly with the microtubule(Sosa, Dias et al. 1997; Woehlke, Ruby et al. 1997). In fact the tip of L11, although disordered in many crystal structures, contains two highly conserved amino acids (Leu248 and Glu250) whose purpose is not known. We therefore propose (Figure 6) that loop L11 is roughly analogous to the clutch in a manual-

transmission automobile. Free in solution, L11 is flexible and disordered, and the "clutch" is disengaged. By cementing the structure of L11 (perhaps into one of the ordered conformations seen in crystal structures(Sablin, Kull et al. 1996; Kikkawa, Sablin et al. 2001; Yun, Zhang et al. 2001)) and supporting a rigid connecting geometry between the two switch elements, microtubules could "engage" the clutch. Once switch II was engaged, ATP binding could reposition the rigidified switch II cluster and directly modulate docking of the neck linker, providing a displacement in kinesin's force-generation cycle.

## Conclusions

Previous work has clearly demonstrated that kinesin's neck linker becomes immobilized in the microtubule-bound, AMPPNP complex. However, the proposal that this neck-linker conformation is the same as seen in crystal structures of "docked" kinesin has been supported largely by circumstantial evidence, because atomic-resolution images of the kinesin-microtubule complex are not available. Here we have detected a docked neck-linker conformation in *free* kinesin, and the EPR signature of this state strongly resembles that seen in microtubule-bound, AMPPNP-complexed kinesin. Furthermore, we have tracked the docked conformation through conditions that stabilize it sufficiently to be crystallized. These experiments provide another link between kinesin's difficult-to-characterize microtubule-bound states and the available crystal structures. Our EPR-monitored nucleotide-exchange experiments, moreover, clarify the microtubule's role in kinesin's docking mechanism. These data strongly suggest that, without a microtubule, there *is* no nucleotide-dependent docking mechanism—a feature that must be taken into account by any proposed structural model of these motors. These observations provide a framework for future structural studies of kinesin's motility mechanism.

# Methods

### *Protein Expression and Purification*

Monomeric kinesin constructs containing the first 349 amino acids of wild-type human kinesin (K349) used here for crystallization experiments have been described(Kull, Sablin et al. 1996). K349 was expressed and purified as described previously, with the addition of a final Mono-S column-binding step to enhance purity. In this final purifying step, K349 was bound to the column under the lowest-strength ionic conditions that maintained solubility (~30mM KCl), and eluted with a gradual salt gradient.

### *Labeling with Spin Probes*

A 349-amino-acid 'cys-light' kinesin construct (containing only 3 nonreactive cysteines) with a single introduced cysteine (C328, C330, or C333) was incubated in EPR labeling buffer (25 mM PIPES (pH 7), 50 mM NaCl, 2 mM $MgCl_2$, 1 mM EGTA, 10 $\mu$M ATP) with 4-maleimido-2, 2,6,6-tetramethyl-1-piperidinoxy (MSL; Sigma Chem. Co.) overnight at 4° C. Repeated concentration and dilution of the protein in a 3000 MWCO centricon (Millipore, Bedford, MA) removed the excess spin label and introduced the final measurement buffer (EPR labeling buffer minus ATP), with a final protein concentration of 50-250$\mu$M. The labeling stoichiometry was determined by measuring protein concentration (Bradford assay with BSA standards) and probe concentration by comparing labeled protein to known concentrations of spin label(Naber, Cooke et al. 1997). Experimental conditions with added $Li_2(SO_4)_2$; etc. retained all reagent components and concentrations of the measurement buffer.

### *Nucleotide Exchange and Sample Preparation*

Samples for the ADP-bound experiments contained 2mM added ADP; for ADP•$AlF_x$.or ADP•$BeF_x$ complexes, 2mM AlCl or BeCl, together with 10mM NaF were also added. Samples for the AMPPNP-bound conditions contained 5mM added AMPPNP. Kinesin samples in altered nucleotide conditions were incubated for > 1hour and still showed

quantifiably the same EPR signal. All measurements included at least a 5-min. incubation time for nucleotide conditions to equilibrate.

### *EPR Spectroscopy*

EPR measurements were performed with an ER/200D EPR spectrometer from IBM Instruments, Inc. (Danbury, CT). First derivative, X-band spectra were recorded in a TE011 microwave cavity using 50 sec, 100 Gauss wide magnetic field sweeps. The instrument settings were as follows: microwave power, 25 mW; gain, $1.0*10^4$ - $1.0*10^6$; center field, 3455-3460 Gauss; time constant, 200 ms; frequency, 9.3 GHz; modulation, 1 Gauss at a frequency of 100 kHz. Each spectrum used in data analysis is an average of 3-5 sweeps from an individual experimental preparation.

### *EPR Signal Decomposition*

The amount of docked and undocked neck linker was quantified by approximating measured EPR signals as linear combinations of pure signal components ("deconvolution"), using the least-squares fitting function LINEST from Microsoft Excel. A sample of high-temperature (35°C), microtubule-free kinesin was used to approximate a "pure undocked" reference spectrum (Figure 7A), while very low-temperature (2.5° C), microtubule-complexed kinesin was used to obtain an approximate "pure docked" reference spectrum (Figure 7B). The 2.5°C sample had a distinctly lower splitting between high- and low- field peaks (65 Gauss) than was seen in a definitely frozen sample (72 Gauss), proving it had not been overcooled.

In addition to "pure docked" and "pure undocked" EPR signal components a third, intermediate component was identified (Figure 7C). Because this intermediate signal was never the majority population under the explored conditions, it had to be derived by manipulating low- and high- temperature spectra of AMPPNP-bound, microtubule-complexed kinesin. Two low-temperature spectra of kinesin in this state were subtracted to remove the high-mobility ("undocked") component by visual inspection, then this process was repeated with two high-temperature spectra. The two resulting difference

spectra were themselves subtracted to eliminate the low-mobility ("docked") component, leaving the remainder spectrum seen in Figure 7C. Adding in this third component to the spectral deconvolution radically improved the fits, which were poor ($\chi$-squared frequently exceeded 15%) using only the first two components in Figure 7. The three components in Figure 7A-7C were successfully used to analyze spectra obtained from all three neck-linker positions (328, 330, 333), in most cases giving $\chi$-squared values below 2%.

The intermediate component in Figure 7C, which was detected in nearly every EPR measurement, could represent an intermediate ("undocked") in the neck linker's "folding pathway" between docked and undocked states. Alternatively, however, it could signify another probe arrangement available with the *docked* neck linker. Therefore, two parallel analyses were performed, reflecting either choice of assignment for the intermediate signal. The conclusions reached in either case were very similar, with just one caveat. If assigned as a docked state, the intermediate signal uniformly shifts the reported free energies in Figure 2 (which are calculated assuming an undocked conformation for the intermediate signal) downward by ~3 +/- 1 kJ/mol in every case. Therefore, while we are not entirely certain about the absolute magnitude of the docking free energy, the *relative* stability of the neck linker when comparing various conditions is well established.

Another potential inaccuracy in the analysis was that the "pure" reference spectrum for "undocked" neck linker was taken as that of ADP-bound kinesin at 35°C (this data point is omitted from Figure 2A). This signal will contain a component caused by residual docking, even though it is interpreted as 100% undocked. This assignment leads to an underestimation of the docked component at the highest temperatures, because in these conditions the marginal amount of docking present in experimental spectra will be comparable to the docking erroneously present in the "pure undocked" component spectrum (from 35°C). This error is responsible for the upward curvature at the high-temperature end of the van't Hoff plots in Figure 2A. However, the plot becomes linear at lower temperatures, where the amount of docking present overwhelms the error in the "pure undocked" spectrum. The precision of our measurements was estimated to be +/-

0.2 kJ/mol for the free energy values, and +/- 5kJ/mol for the enthalpy, based on the standard error reported for the linear fits to the van't Hoff plots.

### *Crystallography of K349*

Purified kinesin (in 25mM HEPES pH 6.8, ~150mM KCl, 1mM DTT, 2mM $MgCl_2$) was concentrated in a 3000 MWCO centricon (Millipore, Bedford, MA) to ~10 mg/mL and mixed with mother liquor (1.5-2M $Li_2SO_4$, 25mM HEPES pH 7.5, 0-50mM KCl). Crystals appeared in 10-20$\mu$L hanging drops after ~1 week. A crystal was harvested and transferred to a cryosolvent mixture in incremental steps (final cryosolvent condition was 30% sucrose combined with all mother liquor components, the first transfer was to a 5% sucrose condition followed by 10%, etc.). X-ray diffraction data were collected at the Advanced Light Source (Lawrence Berkeley National Laboratory) beamline 5.02 ($\lambda$ = 1.1Å), and indexed and scaled by the program *HKL2000*(Otwinowski and Minor 1997). A molecular replacement solution was found by the program EPMR(Kissinger, Gehlhaar et al. 1999) using the coordinates of monomeric rat kinesin (Protein Data Bank(Berman, Westbrook et al. 2000) entry 2kin) as a search model. Subsequent refinement was accomplished using CNS(Brunger, Adams et al. 1998); an initial rigid-body refinement step was followed multiple rounds of manual rebuilding (using the program QUANTA) into composite omit maps, alternating with simulated annealing and minimization. The final model contains a total of 334 visible residues, with the first two, and the last, and a twelve-residue loop segment from L11 (239-251) disordered. Two sulfate anions were identified by electron densities exceeding a 3-sigma cutoff in $2F_o$-$F_c$ omit maps, which also clearly showed density for ADP. The average B-factor for the structure, 25.8, is somewhat higher than typical for cryo-frozen crystals. The structure was analyzed by WHATIF(Vriend 1990) to check for refinement errors.

***Table 1. Statistics for crystallographic data collection and refinement***

| | |
|---|---|
| Space Group | $P2_12_12_1$ |
| Unit Cell | a = 73.8 Å, b = 74.1 Å, c = 91.5 |
| λ | 1.1 Å |
| Measurements | 247,497 |
| Unique reflections | 14,266 |
| Completeness | 99.8% (100.0%) |
| Resolution | 2.7Å |
| $R_{symm}$ | 7.9% (27.0%) |
| $\langle I \rangle / \langle \sigma(I) \rangle$ | 23.0 (6.9) |
| $R_{cryst}$ ($R_{free}$) | 21.1% (25.9%) |
| R.M.S. deviation from ideality (bonds / angles) | 0.007 Å / 1.3° |
| Molecules per asymmetric unit | 1 |

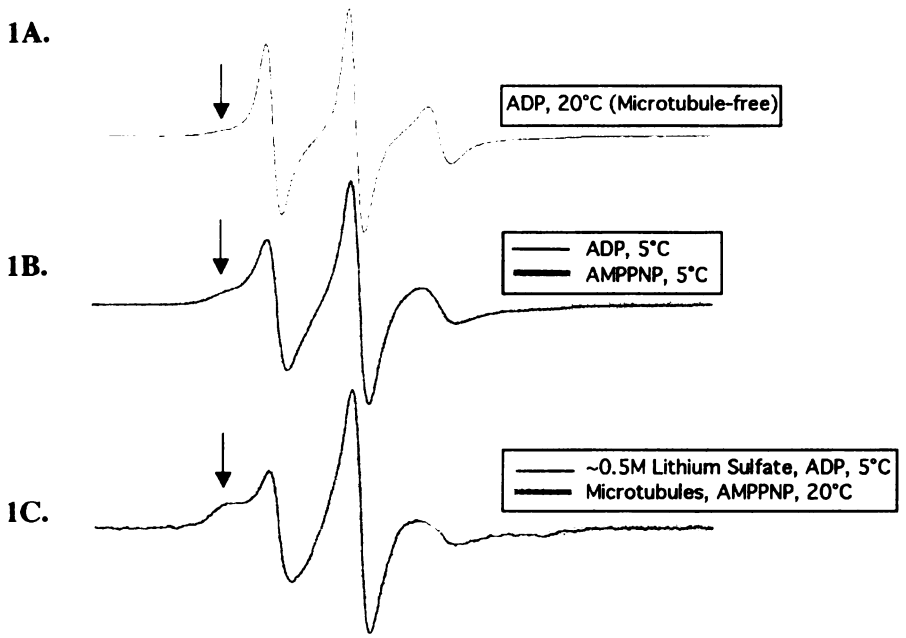



Figure 1. EPR spectra from site-labeled probes on kinesin's neck linker, no microtubules present, show that neck-linker docking is independent of nucleotide. However, sulfate enhances docking, generating a similar signal to what is seen in the kinesin•microtubule•AMPPNP complex. All displayed spectra are from the C333MSL-labeled construct, but results were duplicated for two other probe label positions, C328 and C330 (data not shown).

**A.** Kinesin, no microtubules present, ADP, 20°C. This signal is dominated by a high-mobility component with narrower field splitting between peaks than other components seen in measurements below.

**B.** Kinesin, no microtubules present, ADP or AMPPNP, 5°C shows a low-field shoulder in the EPR signal (arrow) indicating the presence of a low-mobility component corresponding to a docked conformation of the neck linker. The signals from the two nucleotide states are superimposable, indicating the neck linkers are docked in similar proportions. The height of the shoulder reflects the relative proportion of the probe population that has entered the low-mobility state; probe populations were quantified by a fitting procedure (see the Methods and Figure 7), leading to the plots of Figure 2. The 5°C ADP and AMPPNP signals both show 45% of the probes in the highly immobilized state, while the 20°C spectrum in (**A**) shows 20% of the probes immobilized.

**C.** Overlay of kinesin-microtubule complex plus AMPPNP, 20°C (gray) with kinesin plus ~0.5M lithium sulfate, 5°C (black). Adding sulfate to the conditions in (**B**) enhances neck-linker docking, increasing it to the same levels as seen in the microtubule-bound, AMPPNP-complexed spectrum (both show 65% of probes in the lowest-mobility state). The superposition of signals in these conditions indicates the EPR probe environment is similar, consistent with the proposal that the docked neck-linker conformation seen in crystal structures is the same as in microtubule-bound, AMPPNP-complexed kinesin. Microtubule•AMPPNP complex was prepared as described(Rice, Lin et al. 1999).

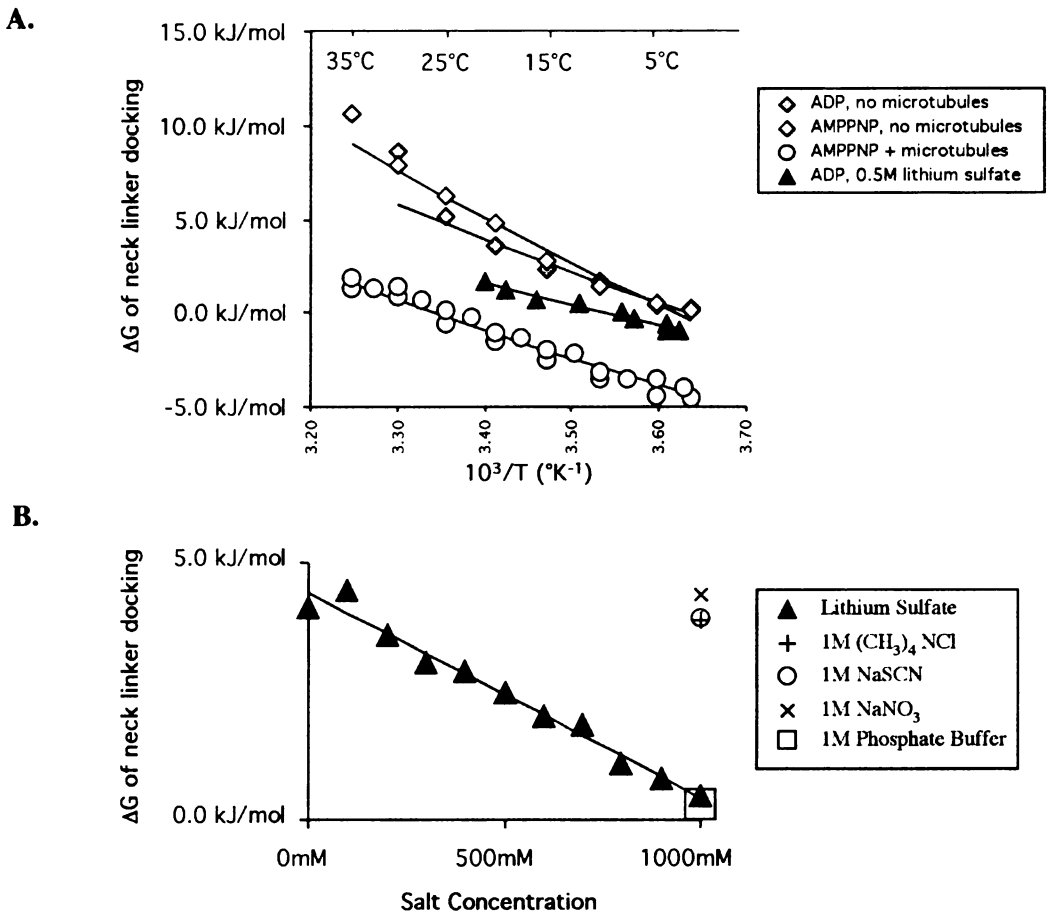



Figure 2. Quantifying the effects of various conditions on kinesin's neck linker, as monitored by EPR. The free energy of docking is reported from label position C333, as determined by population analysis of the reported EPR probe signal. The conservative upper limit of the free energy is presented. A second interpretation of the data is possible (discussed in the Methods) in which points are uniformly more favorable by approximately 3 +/-1 kJ/mol, while other features of the graphs remain qualitatively similar.

A. In standard buffer conditions and in the absence of microtubules, docking of the neck linker is slightly disfavored, but does not depend on which nucleotide is bound. Adding microtubules *and* AMPPNP enhances neck-necker docking by ~4 kJ/mol; adding sulfate (no microtubules) has a similar, although weaker, effect. Solid lines represent linear fits to the corresponding van't Hoff plot (where the data are replotted with the y axis, -RTlnK, divided by T to give -RlnK). The enthalpies determined by the linear fits are in the range -40 - -70 kJ/mol.

B. The docking effect of sulfate is linear as a function of concentration, and phosphate buffer at 1M has the identical effect as 1M sulfate. However, other ionic reagents including $(CH_3)_4$ NCl, NaSCN, and 1M $NaNO_3$ (shown) did not produce a detectable change in the docking. Buffer conditions were otherwise identical in these experiments (see Methods). Experiments were performed at ~20°C..

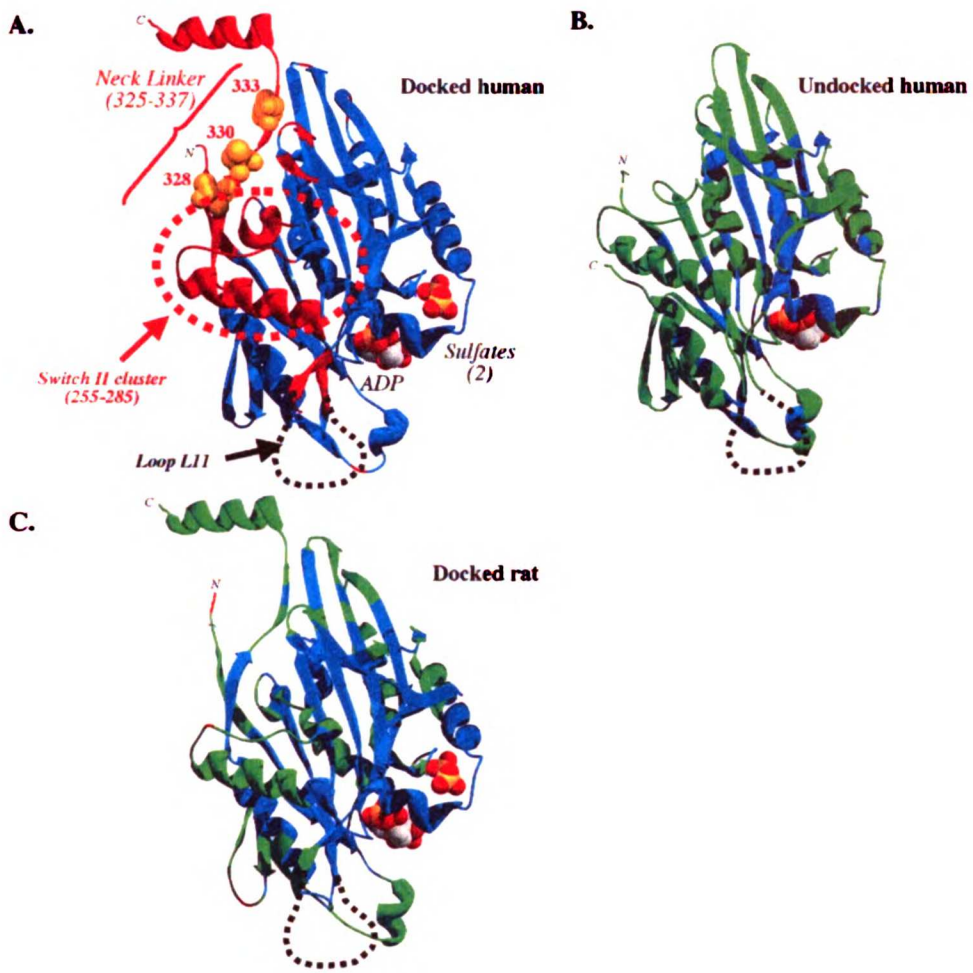



Figure 3. Conformational change in two crystal structures of K349.

A. Regions of greatest conformational change in the crystal structure of the monomeric human kinesin construct K349, with respect to the published form. Residues that have moved more than 2Å relative to a least-squares alignment of the two structures are colored in red, as are two stretches of ordered chain (the neck linker, ordered residues in the new structure that were disordered in the original K349 structure. The ADP active-site nucleotide is represented as a space-filled model, as are two coordinated sulfate anions bound at on the rear face of the molecule from this viewing angle. The three positions on the neck linker used for site-labeled EPR probes are indicated by gold space-filling models. The tip of the "arrowhead" (referred to in the text) is up.

B. The originally reported crystal structure of human kinesin K349, with residues colored green if they have moved more than 0.5Å relative to a least-squares alignment with the new high-sulfate structure.

C. The crystal structure of monomeric rat kinesin, colored in comparison to the new K349 crystal structure according to the same scheme as in (B). Red ribbon segments indicate two single-residue insertions in the rat sequence relative to human, and an ordered N-terminal residue not visible in the K349 structure. Two sulfate ions (space-filled) are found at the same positions as in the docked K349 structure in (A). This figure and all following molecular drawings were generated by the Swiss PDB Viewer(Guex and Peitsch 1997) and rendered using the MegaPOV-ray software package (www.povray.org)

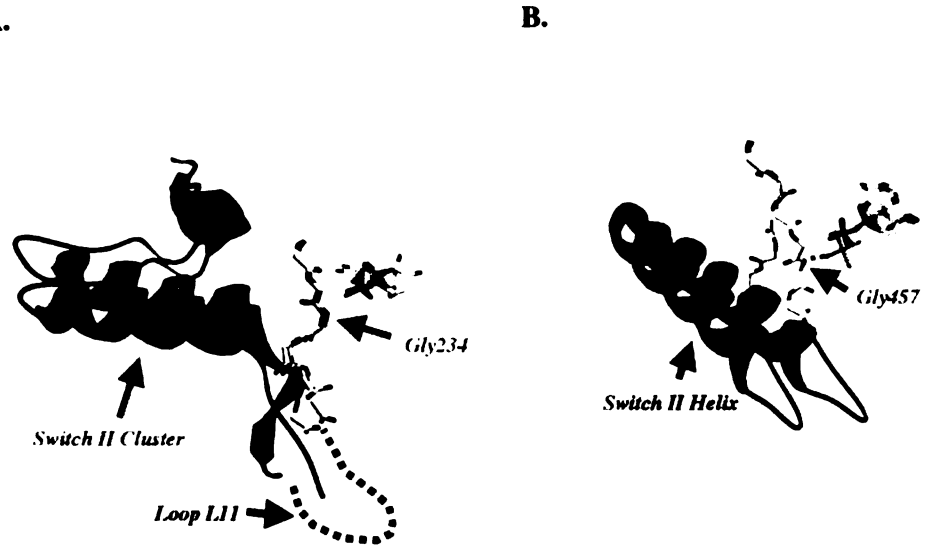



Figure 4. Comparison of switch II transitions in kinesin and myosin

A. Uncoupling in kinesin's switch II. In microtubule-free human kinesin, the switch II glycine (Gly234) is connected to the moveable switch II cluster (residues 255-295) by the disordered loop L11 (dashed line). This allows the switch II cluster to shift downward and to the right in the docked-neck linker crystal structure (green) relative to the undocked structure (blue), in the global structural alignment. The conserved switch II nucleotide-sensing segment (DLAGSE, 231-236, drawn in a backbone-only ball-and-stick representation) moves less than 0.7Å in this alignment. Note that our definition of the switch II cluster includes only residues that move together primarily as a rigid body, and so does not include the nucleotide-sensing segment or L11— this differs slightly from the original definition proposed by Kikkawa et al(Kikkawa, Sablin et al. 2001).

B. A view of triphosphate-induced conformational change in myosin illustrates a through-bond communication between the switch II glycine (Gly457, equivalent to Gly234 of human kinesin) and the remainder of switch II that follows, including the switch II, or "relay" helix . Shown are the ADP•aluminum-fluoride and the beryllium-fluoride complex of Dictyostelium discoideum myosin(Fisher, Smith et al. 1995), with the switch II helix colored as a green or blue ribbon, respectively. The planar aluminum-fluoride moiety, appearing as a gray cross, forms a hydrogen bond with the amide nitrogen of Gly457. In the beryllium fluoride complex, Gly457 moves 4.5Å away from the nucleotide site, accompanied by the entire C-terminal region of switch II including the relay helix.

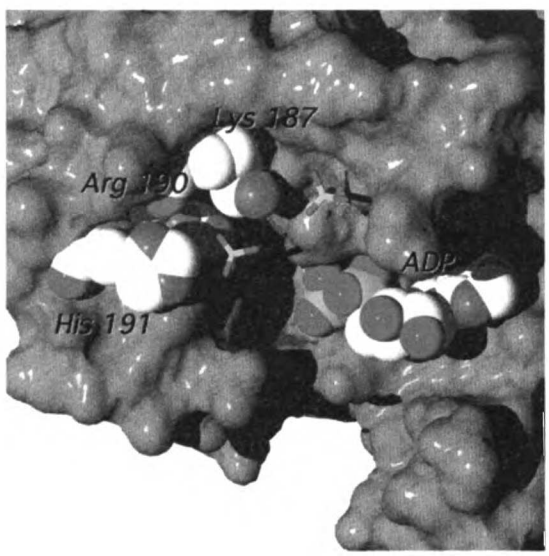



Figure 5. A conserved anion binding site in K349 that could be poised to capture the orthophosphate product of hydrolysis.

This view is rotated approximately 90° about the vertical axis (counterclockwise, when viewed from top) from the view in Figures 3 and 4. One sulfate, replaced by an acetate anion in the published structure of human kinesin, is coordinated by Lys187 and His 191, and is in close vicinity to the 100% conserved, partially buried Arg190. A second coordinated sulfate ion appears nearby. From this viewing angle, kinesin's microtubule-binding interface is on the opposite side.

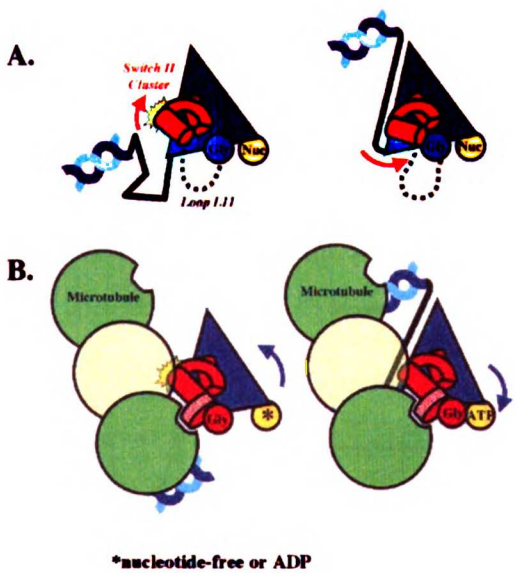



Figure 6. A model for how microtubules introduce tight coupling between ATP binding and the neck linker.

A. Free in solution, the switch II cluster (green) is in equilibrium between conformations that obstruct neck linker docking (left) and ones that permit it. Flexibility in loop L11 (dashed line) permits this movement in all nucleotide states.

B. Binding to the microtubule protofilament (pale green circles) could rigidify the conformation of L11 so that when ATP binds and attracts the switch II glycine (as occurs in G-proteins and myosin), the switch II cluster moves relative to the rest of kinesin's catalytic core. This movement of the cluster would recruit the neck linker to dock, translating the coiled-coil (blue helix) in the direction of travel, towards the microtubule plus end (up, in this view). In accordance with the proposal of Kikkawa et al(Kikkawa, Sablin et al. 2001), we portray the switch II cluster fixed on the microtubule surface, so that the cluster "movement" actually causes the rest of the catalytic domain (blue triangle) to rotate relative to the microtubule.

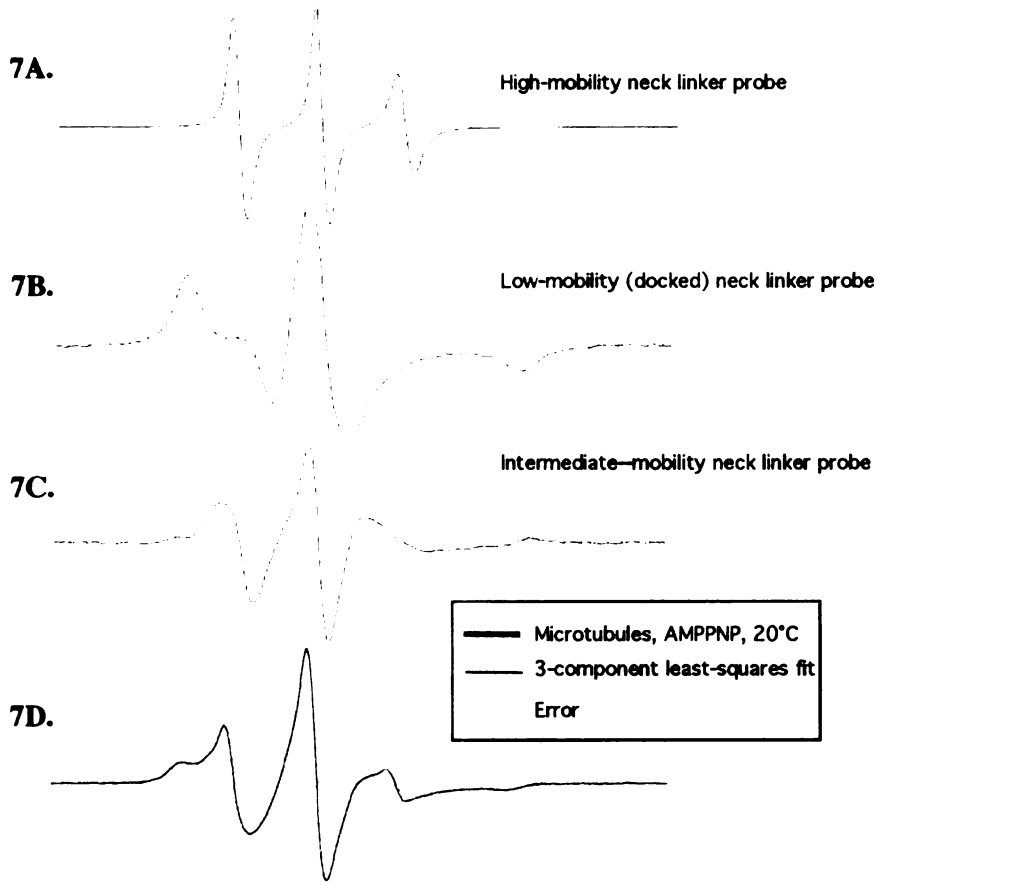



Figure 7. EPR spectra used for signal decomposition. In the probe population analysis, EPR signals from the kinesin neck linker were approximated as weighted sums of these three signals, which we identify as low, intermediate, and high-mobility states of the neck linker probe. All displayed spectra are from the C333MSL-labeled construct.

A. Low-mobility signal obtained from the neck linker of microtubule-bound kinesin, 2.5°C. This signal corresponds to a docked conformation of the neck linker.

B. High-mobility signal obtained from microtubule-free kinesin at high temperature (35°C). This signal reflects a mobile, disordered state of the neck linker.

C. Intermediate-mobility signal discovered by weighted differences of kinesin neck-linker spectra from microtubule-complexed, AMPPNP-bound kinesin at various temperatures (see Methods).

D. Decomposition of the microtubule-bound, AMPPNP spectrum in Figure 1C using the 3 spectra from (A) - (C) as a basis set. The LINEST function of Microsoft Excel was used to find the linear combination of basis set spectra with the guaranteed optimal least-squares fit. The resulting coefficients for the basis set spectra determined the relative population of spin probe in each conformational substate. The same basis set was used in all fits of EPR spectra reported here.

# Chapter Two:
# Two New, Highly Ordered Conformations of Kinesin's Putative Force-Delivering Element

## Summary

The current model for the kinesin motor protein is based on conformational changes of a short force-transmission segment called the neck linker, which has been proposed to switch between ordered and disordered states while the catalytic head is bound to its microtubule track: a disordered conformation in the ADP-complexed or nucleotide-free states, and docked one along the motor core in the ATP state. Here we demonstrate, using a revised analysis of site-specific EPR probe data, that significant ordered populations of the neck linker exist in all motor nucleotide states. The ordered populations were clearly present in monomeric kinesin, but were dramatically higher in the dimer and reached nearly *identical* levels (~75%) for both the *ADP-bound* and *AMPPNP-bound*, microtubule-complexed motor. Therefore, for the dimer at least, any nucleotide-dependent transition by the neck linker is likely a conversion between well-ordered states. This conclusion is reinforced by published cryo-EM images of the monomer construct showing well-ordered neck linker conformations pointing backwards (towards the microtubule minus end) in the ADP-bound microtubule complex, and pointing forwards in the AMPPNP microtubule complex. Furthermore, the cryo-EM data show not one, but *two* backwards-bound neck linker conformations, and these likely interact with tubulin- not with the motor core domain. Our results are explained by an asymmetric "ice climber" model, in which kinesin clings tightly to the microtubule surface and advances with alternating left- and right-handed power strokes of the neck linker.


## Introduction

The current model of kinesin's motility is based on the neck linker's switching between two states(Rice, Lin et al. 1999), disordered and docked along the motor core (as in crystal structures). The purpose of this chapter is to scrutinize the available data of the neck linker: EPR (much of it unpublished), FRET, and cryo-EM, and compare it with the predictions of the current neck-linker driven model for kinesin's motility cycle.

Up to this point, EPR measurements have supplied some of the more suggestive evidence that only *one*, "docked," conformation of the neck linker exists in the microtubule-bound motor, independent of the different nucleotide states examined. Ordered neck-linker conformations, evidenced by a distinctive low-mobility EPR signal component, could be detected in all microtubule-bound motor states, notably in the presence of ADP, AMPPNP, or in the nucleotide-free condition (see Figure 2 below, and Appendix 2). Furthermore, this ordered EPR component appeared to be very similar after nucleotide exchange, maintaining its characteristic features (signal shape and splitting width). Remarkably, this agreement between nucleotide states was found across all three neck linker positions examined. Since EPR probes are considered to be sensitive monitors of conformational change, the most parsimonious explanation of the data was therefore that the neck linker visited the *same* docked conformation in all nucleotide states.

Only one difference in kinesin's neck linker between nucleotide states has yet been conclusively established. The fraction of neck linkers found in an ordered state *decreases* in the microtubule-bound, ADP-complexed motor relative to the AMPPNP-complexed form(Rice, Lin et al. 1999) (see Figures 5 and 7, below). Across the temperature range examined, AMPPNP docked the neck linker with ~1 kcal/mol greater affinity than was seen in the other two nucleotide states tested, ADP or nucleotide-free. This feature allows a model to be proposed (Figure 1) in which neck linker docking provides a 1 kcal/mol "power stroke," so that a microtubule-bound kinesin monomer domain could "throw" the second head of the dimer forward, positioning it for the next step.

This power stroke, however, would be feeble compared to the free energy available from ATP (~12 kcal/mol at cellular levels of nucleotide substrate). From a straightforward mechanical standpoint, it is not obvious how this type of mechanism could provide kinesin with its impressive performance specifications. In particular, kinesin's exceptionally high stall force (~6 pN, for 5 kcal/mol of work per 8-nm step) and high reaction barrier to backwards stepping (~3kcal/mol, H. Higuchi, personal communication) may prove problematic for a motor with only 1 kcal/mol available to distinguish between forwards and backwards steps. Whether subtle aspects of

thermodynamics and brownian movement might permit such "weak power stroke" models to work remains unclear, and is the subject of ongoing investigation.

With these details in mind, we set out to reanalyze the available data in search of additional neck-linker features that might provide an alternative mechanism for motor stepping, using the new EPR analysis tools developed in the last chapter. Surprisingly, we discovered that current models, which involve at most one ordered neck-linker conformation, are very difficult to reconcile with the spectroscopy data. In fact, the data suggest that nucleotide-free kinesin (as well as ADP-bound kinesin) can adopt not one, but *two* ordered neck linker conformations in the course of the motility cycle—neither of which resemble the crystallized neck-linker conformation.

Revising kinesin's motility cycle to incorporate these new neck-linker states can explain two puzzling structural observations that appear to be at odds with previously proposed models of kinesin's cycle. First, the new neck-linker conformations explain why cryo-EM images of dimeric kinesin complexed to microtubules seem to show the second, unbound head pointing towards the microtubule plus end—regardless of the motor nucleotide state(Arnal and Wade 1998; Hirose, Henningsen et al. 2000; Hoenger, Thormahlen et al. 2000). In fact, our model predicts that the connectivity of the dimeric heads in these maps should be reassigned so that the ADP –bound (and nucleotide-free) motors actually direct their unbound heads *backward*, towards the microtubule minus end.

Second, the model can account for recent experiments which have demonstrated that kinesin, if it truly alternates between catalytic heads as it walks, must display a marked asymmetry between "left" and "right" steps(Hua, Chung et al. 2002). The two new ordered neck-linker conformations that we infer from the experimental data explain how kinesin may operate in such an "asymmetric, hand-over-hand" mode. Several experiments are currently possible to test the expanded role for the neck linker that we propose.

# Results

***EPR may not distinguish between different ordered neck-linker conformations***

EPR signals are generally regarded as sensitive indicators of conformational change. In the case of the three neck-linker probes used with kinesin, however, the position-dependent effects on EPR line shape proved to be quite subtle. Remarkably, the same basis set of three signal components (see last chapter), developed to analyze spectra from position 333 on the neck linker, were found to perform equally well in the analysis of data from *both* other probe positions, 328 and 330 (Figure 2). In fact, the $\chi^2$ error of the fit was actually *better* for position 330 data than for the position 333 data from which the components for analysis were derived. This level of agreement seems to suggest that these probes do not interact very specifically with their protein environment, so their signal behavior is mainly dominated by the general restraints of a cysteine linkage to the protein backbone. The placement of all three probes on the uncomplimented, solvent-facing edge of a ß-strand (see Figure 3 of the last chapter) gives them a similar environment and allows significant freedom of movement, probably contributing to the effect.

As can be seen in Figure 2, the same 3-component basis set analyzes neck-linker signals from several different nucleotide states (results for AMPPNP and ADP-bound microtubule complex are shown) with equally low error. Excellent agreement was also found when fitting EPR data from the microtubule-free motor (see the previous chapter, and Appendix 2). This global agreement indicates that the same three signal components seem to be present in every case (the relative occupancy of these components changes with the fraction of bound neck linker). Ordinarily, because of EPR probes' above-mentioned environmental sensitivity, invariance in the EPR signal character might suggest that nucleotide exchange (and/or microtubule binding) did not affect the structure of the bound neck linker. In this case, however, it is clear that signal agreement by itself does not prove structural equivalence: different probe *sites* on the neck linker can give nearly identical signals. These data, therefore, leave open the possibility that the neck

linker might adopt some *other* extended conformation that featured similar probe mobilities at all the sites—which would make the change very difficult to detect via the EPR signal shape. It therefore becomes vital to crosscheck the EPR data with neck-linker information obtained by other spectroscopic methods, in order to truly rule out a nucleotide-induced conformational change.

***FRET data suggest a second docked neck linker conformation***

The three-component EPR basis set developed in the last chapter, as discussed above, permitted a global reanalysis of the three neck-linker probe sites that had been examined in published work on the kinesin monomer K349. Spectra for each of these sites were available, as a function of temperature, in a variety of nucleotide states, both with and without microtubules present (S. Rice, unpublished data; see Appendex 2). As indicated in Figure 3A, microtubule-bound, AMPPNP-bound microtubule-free kinesin (at room temperature) showed the greatest levels of neck linker ordering (in agreement with the published results(Rice, Lin et al. 1999))—with ~65% of the neck linker population in a highly ordered, bound state at 20°C. Surprisingly, however, several other motor conditions whose neck linkers had previously been characterized as disordered were found by the new analysis to have significant bound populations at room temperature. In microtubule-free kinesin, the highly ordered fraction of the neck linker was ~20%-- this case has been discussed in detail in the previous chapter. A ~20% ordered fraction was also found in microtubule-bound kinesin, both in the ADP complex (not shown) or nucleotide-free.

The newly discovered, ordered neck linker populations, while not especially large, proved to have great significance in interpreting neck linker structure. As shown in Figure 3B, FRET measurements(Rice, Lin et al. 1999) show that the neck linker moves emphatically *away* from the tip of the catalytic core, on average, in the microtubule-bound, nucleotide-free state, relative to the microtubule-free motor condition (first two columns of the graph). This distance increase, however, differs from the trend seen in EPR measurements of the same motor conditions. As seen in the Figure, our EPR

analysis shows the *same* degree of neck linker ordering in the two motor states. The FRET data, therefore, suggest a change in the mode of neck-linker docking after kinesin binds to microtubules and releases ADP.

***Reinterpreting Cryo-EM images of nucleotide-free, microtubule-bound kinesin***

In Figure 4B is shown the image for AMPPNP-bound, monomeric kinesin with yellow difference density drawn for a neck-linker-attached gold particle(Rice, Lin et al. 1999). Below the EM density, in Figure 4D, is shown a crystal structure oriented to align with the EM picture (where the alignment was found by docking crystal structures into the EM density maps). This figure illustrates what more precise fitting also indicated(Rice, Lin et al. 1999): that the neck linker is docked to the motor core, almost certainly in the same conformation as seen in crystal structures. On the other hand, gold particles attached to neck-linker sites in MT-bound, ADP-bound and nucleotide-free are seen at two distinct locations in these maps, distinct from the AMPPNP location (maps for the former two nucleotide states showed very similar positioning of neck-linker gold labels). These gold label densities, shown in Figure 4B, were originally interpreted as disordered chain conformations, perhaps freezing artifacts.

Contrary to this conclusion, our revised EPR analysis shows that at the temperatures where the samples were prepared before freezing (5°C), 50% of the neck linkers are bound in a well-ordered conformation. The ordered components in these spectra, in fact, cannot easily be distinguished from the ones for the microtubule-bound, AMPPNP-complexed motor—at all three neck-linker positions (see Figure 2). This specifically bound population of the neck linker, therefore, should have been seen in cryo EM images of the nucleotide-free state... as it apparently was. However, the neck linker is obviously in a different conformation in these images than in the AMPPNP state—and is seen in *two* places (Figure 4B, 4D). Thus, the probe data suggest that two, very well structured neck linker conformations do exist in the microtubule complex (nucleotide-free kinesin or with ADP), and these conformations are distinct from the one seen in the AMPPNP state.

***The putative new neck-linker conformations are likely to point towards the microtubule minus end***

Only one of the two "ADP" neck-linker densities (right-hand red circle in Figure 4B) is even marginally associated with the density of the catalytic motor domain. The other neck-linker density is positioned far from the motor domain, which means that because of the repeating microtubule lattice this gold label occupies *two* symmetrical sites near the motor domain (left-hand red circles in Figure 4B). The originally assigned connectivity(Rice, Lin et al. 1999), indicated by the black lines, places this left-hand neck linker conformation extending forward towards the microtubule plus end. However, cryo-EM experiments cannot conclusively prove that a given connectivity is the right one, because the neck linker is not imaged directly.

In fact, a simple geometric argument demonstrates that this plus end-directed assignment is probably wrong. Figure 4D reveals that the plus end-assigned gold density (upper left red circle) is at least twice as far from the neck linker's connection point on kinesin's catalytic core as is the AMPPNP label density (Figure 4C). The neck linker, however, does not appear to be long enough to support this distance increase, because in the AMPPNP structure it is already fully extended. Also, in order to reach the upper left-hand position, the neck linker would have to wrap tightly around the motor core, implying another bound conformation on the catalytic domain for which there is no evidence.

The lower-left-hand label density is equally far from the neck linker's attachment point on kinesin, but here there is another possibility to explain the distance increase. As shown in Figure 4D the helix (H6), to which the neck linker is C-terminally attached, extends directly towards the lower left-hand label density. The geometry suggests that, if H6 were to change conformations, the neck linker could reposition to reach the lower label density. While rearrangement of H6 has never been suggested before in a conventional kinesin, several observations suggest that it may indeed occur; we will return to this subject in the discussion. In any case, the alternative-- that the neck linker somehow elongates to reach the upper-left density— seems far less probable.

Neither of the gold densities in the nucleotide-free cryo-EM conditions appears to be as well-connected to the motor domain as the gold density in the AMPPNP condition. This raises the possibility that these neck-linker conformations may be bound to tubulin rather than kinesin. Particularly in the case of the left-hand density, it is hard to conceive how an extended neck linker conformation could be immobilized to the extent determined by our EPR experiments *without* some kind of neck linker-microtubule contact. As we show in the following section, we have actually found indirect evidence for a neck linker interaction with the microtubule.

***An anomalous measurement of neck-linker binding in kinesin's microtubule-bound, ADP-complexed state***

In the published EPR measurements of kinesin's ADP-complexed, microtubule-bound state, neck linker docking was reported to be consistently less favorable than found in the AMPPNP complex. Some early unpublished experiments, however, showed a markedly different binding pattern for the neck linker in the ADP state, as shown in Figure 5. Represented in Figure 5A are bound neck linker populations, converted to free energies by the above-described 3-component method, for data collected on two different days (experiments conducted several months apart). The uppermost line in Figure 5A, corresponding to the *second* experiment, represents neck-linker behavior for the ADP state could be reproduced in subsequent tests (S. Rice, unpublished data). These free energies represent a marginally unstable population of docked neck linker, which is ~1kcal/mol less favored then in AMPPNP-complexed, microtubule-bound samples (bottom line in Figure 5A)—consistent with the published report. Data from the *earlier* experiment (middle line in Figure 5A), on the other hand, shows the bound neck linker (ADP state) in a significantly altered, more populated state.

The difference between the two data sets, for the ADP-bound case, is particularly noticeable at the lower temperatures. As shown in Figure 5B, the ADP-bound EPR spectra for the two experiments show a dramatic reversal between mobilized and immobilized peaks. The enthalpy of neck linker binding, reflected in the slopes of the

Figure 5A plots, is also significantly different in the two experiments: -140 kJ/mol for the more strongly bound ADP data set vs. –110 kJ/mol for the less-well bound one. By contrast, the AMPPNP conditions produced neck linker docking that was indistinguishable in the two experiments (both data sets are plotted in Figure 5A, and superimpose as is seen). These data suggest that something changed, between two experimental days, to strongly modify the neck linker docking interaction in the ADP-bound, but *not* AMPPNP-bound, motor.

The more strongly bound ADP data set provided low-temperature spectra, as shown in Figure 5C, that could be directly compared to AMPPNP spectra at the same temperature, having roughly the same relative populations of docked and undocked states. These spectra, therefore, could be scrutinized for very subtle differences in the EPR line shape of the low-mobility component. The comparison is especially sensitive because the high-mobility signal exists at nearly identical levels in the two signals, and is only weakly present. As can be seen, at both 2.5°C and 10°C the immobilized component in the ADP condition has shifted to a narrower field splitting than in the AMPPNP condition, reflecting a slightly enhanced probe mobility in the bound state. While the splitting difference is barely detectable, it is extremely important, because it is the indication of conformational change. With ADP and microtubules, these experiments show, the environment of the neck linker probe has changed, relative to the AMPPNP state.

What experimental conditions could have changed in Figure 5, to cause such a dramatic difference in the ADP-bound signals, but not the AMPPNP-bound ones? The answer most almost certainly lies in the microtubules. The tubulin used here was derived from cow brain, which is known to be heterogeneous. Furthermore, there was a potentially significant change in our laboratories' source of tubulin shortly after the "anomalous" ADP measurement was made. Up to that time, bovine brains from a slaughterhouse were used for tubulin preparation (preps were performed 1-2 times annually to restock the lab supply). However, when the slaughterhouse closed down, a new source of brains had to be found, and the new source turned out to be *veal calves* and not adult cows. Tubulin from these two sources has actually been suggested to have different properties, although

the literature is very old and the issue appears not to have been thoroughly investigated (S. Rice, personal communication).

Therefore, it seems possible, even probable, that differences in *tubulin's* composition could be responsible for the variability we observed in the neck linker of ADP-bound kinesin. Further confirmation of this idea comes from repeat experiments that were done using the same kinesin protein sample that gave the original, "anomalous" neck-linker reading, but using the new tubulin supply (from veal calves). The repeat experiments were found to give the new (less bound) neck linker docking energetics (upper curve in Figure 5A) with ADP-bound motor, and did not recreate the "anomalous" docking (middle curve).

We draw two conclusions from the data in Figure 5. Most importantly, the EPR spectra provide direct evidence for a nucleotide-induced conformational change in the *bound* state of the neck linker. Secondly, experimental variability in both the free energy and enthalpy of neck linker binding in the microtubule-bound, ADP complexed state of kinesin suggests that tubulin heterogeneity may affect the bound neck linker conformation in this motor state (but *not* in the AMPPNP-bound one). This observation meshes nicely with our above synthesis of EPR, FRET and cryo-EM data, which suggests that the neck linker is extended and specifically bound on the microtubule surface when ATP or analogs are absent.

***EPR measurements of the dimer neck linker***

The neck linker of dimeric kinesin was examined by EPR, using the same labels as in monomeric experiments described above. Remarkably, these spectra show very high levels of probe ordering from the lowest temperatures (5°C) examined up to the highest (35°C), as shown in Figure 6. This ordering demonstrates that bound neck linker conformations are much more highly populated in the dimer than the monomer construct, as will be quantified below. Furthermore, signals for the ADP-bound and AMPPNP-bound dimer-microtubule complex superimpose nearly exactly (Figure 7B). This

agreement indicates that nucleotide exchange, under microtubule-bound conditions, does not significantly affect the fraction of neck linker bound.

The dimer EPR spectra were analyzed to extract low- and high-mobility components, using the 3-component method described above. Remarkably, as shown in Figure 6, the purified components derived to fit *monomer* spectra provided a workable fit to the dimer spectra as well. The $\chi^2$'s of the fits (~14% for ADP data, and ~7% for AMPPNP) are much higher than in the monomer. However, inspection of the low-field shoulder (where the agreement is very good) suggests that no *extra* components have appeared, and that instead only minor line-shape shifts are responsible for the fitting error.

Estimating the bound neck-linker fraction by the above fits results in the free energy plots of Figure 4. These plots show that kinesin dimers at 35°C continue to retain levels of neck linker binding seen in the monomer at close to freezing temperatures. Strikingly, the enthalpy of neck linker binding implied by the 3-component analysis is no more than 8 kcal/mol—fully *two-thirds less* than the values found for the monomer, reported above. Bound neck populations and enthalpies will be determined with greater accuracy by rederiving three new basis set EPR components from the dimer data set, but this analysis has not yet been completed.

## Discussion

### *New, specifically bound conformations of the neck linker*

Our EPR experiments show that very well ordered, in other words bound, conformations of kinesin's neck are present in every state of the motor yet tested. Furthermore, our results indicate that, for the microtubule-bound motor, the bound conformation(s) found in the nucleotide-free and ADP-bound states differ from the conformation found in the AMPPNP state. All three kinds of probe measurements described here support a nucleotide-dependent conformation change in the bound neck linker state. As we have shown, EPR detects a slight but significant change in the bound neck-linker environment

between nucleotide-free and AMPPNP-complexed forms of the microtubule-bound motor. FRET measurements show a distance change in nucleotide-free, microtubule-bound kinesin that would be unexpected if the motor had only one highly ordered neck linker conformation available in all its states. Finally, cryo-EM images show an ordered neck-linker population in the nucleotide-free, microtubule-bound motor, in a different orientation than the AMPPNP•microtubule motor state. Collectively, these observations provide highly suggestive evidence for multiple, well-ordered and bound neck linker conformations, controlled by kinesin's nucleotide state—in contrast to the simple model in Figure 1.

In the monomer, the new neck linker conformations we propose are not present in great quantities—although the small (~20%) fraction that occurs has nearly the same degree of ordering, by EPR, as the crystallized neck linker conformation in its docked form. In the *dimer*, however, the neck linker is much more tightly bound (75%) under the conditions (ADP-complexed, microtubule-bound) where we expect it to adopt its alternate, non-ATP-like conformations.

Features specific to the dimer are likely to enhance the stability of bound neck linker conformations. At the end of their neck linkers the dimer constructs have a coiled-coil stalk that interacts with microtubules and influence processivity(Thorn, Ubersax et al. 2000). Even in monomers, there is evidence that residues in the stalk region interact with microtubule. Cryo-EM images of the kinesin family member KIF1A complexed to microtubules(Kikkawa, Sablin et al. 2001) show extra electron density that may reflect an interaction of the stalk with the C-terminal tail of tubulin (known as the "E-hook"). However, dimerization of this stalk could significantly enhance the stalk's ability to interact with the microtubule. Indeed, a heterodimeric kinesin construct with an intact, coiled-coil stalk but only one catalytic head (maybe this should be called a "heteromonomer") has a 4-8 fold slower detachment rate from the microtubule at the end of its hydrolysis cycle(Hancock and Howard 1999). This increased "stickiness" may reflect an enhanced microtubule interaction by the coiled-coil stalk that isn't available to monomeric constructs like K349 with only half a stalk. If so, then constructs with a

coiled coil—such as the dimer or the heteromonomer—might be expected to have enhanced neck linker docking, because binding of the coiled coil would tie down the neck linker's C-terminal end. A second cause for neck linker docking to be enhanced in dimeric constructs could be binding of the second catalytic head, which might stabilize the bound state of the neck linkers for both heads.

Our observation that neck linker docking is the dimer is equally strong in both nucleotide states tends to further disfavor the simple walking model for kinesin illustrated in Figure 1. In dimeric kinesin, our experiments show, none of the nucleotide states seem to show the high levels of disordered neck linker (>= 50%) that the model in Figure 1 predicts to be present in all nucleotide states of the motor. Furthermore, our EPR experiments in dimeric kinesin show that the neck linker binds equally well in both ADP and AMPPNP-bound states of the motor, failing to provide even a weak powerstroke (assuming the neck linker bound the *same* way in both nucleotide states). Even in the monomer, we have encountered conditions where the neck linker binds equally tightly in ADP and AMPPNP states—at temperatures significantly above freezing (see Figure 5). These data, we believe, are most consistent a model in which the neck linker switches between multiple, distinct, bound conformations during kinesin's motility cycle.

***A potential conformational change in helix H6***

Comparing kinesin's X-ray crystal structure to the cryo-EM images of neck linker probes in nucleotide-free and ADP states (see Figure 4 above) suggested that both implied conformations of the neck linker extend "backwards," towards the microtubule minus end. However, as noted, at least one of these backwards-pointing arrangements appears to necessitate a drastic rearrangement of helix H6, the structural element in the catalytic core to which the neck linker is C-terminally attached. This proposal is not unprecedented. Indeed, probe experiments with the kinesin family member NCD (which operates similarly to kinesin but walks in the opposite direction along microtubules) have suggested a microtubule-induced conformational change in H6(Naber, Cooke et al. 1997). A comparison between kinesin and the muscle protein myosin, which has been

shown to share with kinesin crucial aspects of its nucleotide sensing machinery, also suggests that H6 may rearrange. In myosin, the SH1 helix is positioned at the analogous point in space as is H6 in kinesin, relative to the conserved nucleotide-sensing machinery ("Switches" I and II)(Kull, Sablin et al. 1996). Experiments have shown that SH1 loses its helical form in a nucleotide-dependent way, so that in certain conditions it can form an inter-chain sulfhydryl bond ("SH") that the helical structure would not have allowed(Houdusse, Kalabokis et al. 1999). Thus, by analogy helix H6 in kinesin may be implicated in a structural change as well.

Recent structural studies of the kinesin family member KIF1A(Kikkawa, Sablin et al. 2001) may provide a clue as to how the conformation of H6 could be controlled by nucleotide exchange. When the crystal structure of KIF1A was modeled into cryo-EM maps of the motor-microtubule-complex, the motor was found to shift its orientation on the microtubule surface depending on its nucleotide state. In the ADP state of the motor, this rotation appeared to introduce a steric clash between H6 and the microtubule surface, leading the authors of the study to propose that the motor made an additional sliding movement (in addition to the rotation) along the microtubule surface to accommodate the clash. An alternative possibility, however, is that KIF1A relieved the clash by rearranging H6 instead. If so, then KIF1A could exert nucleotide-dependent control of H6 by a simple "bumping" mechanism. Based on our own crystal structures of K349, we have proposed that K349 makes a similar, nucleotide-dependent rotation on the microtubule (see Figure 6 from the last chapter)— as is supported by cryo-EM imaging data, described below. This suggests that a "bump-controlled" change of H6, by rotation of the core domain on microtubules, might be a fairly general mechanism for kinesins.

***Reinterpreting cryo-EM images of the kinesin dimer complexed to microtubules***

We propose that neck linkers of nucleotide-free (and ADP-bound) kinesin are specifically bound to the microtubule and extend backwards towards the minus end, and that in kinesin dimers these backwards-bound neck linker conformations are highly stable. If true, our hypothesis predicts that, in dimers, the two catalytic heads might be found a

considerable distance apart on the microtubule surface when imaged in these nucleotide states. Indeed, unbinding experiments, in which single kinesin dimers were pulled off the microtubule at high speed, detected doubly-bound catalytic heads in the AMPPNP-bound condition(Kawaguchi and Ishiwata 2001). Backwards-bound neck linker conformations may help kinesin dimers to "straddle" adjacent binding sites on the microtubule, as is implied to happen in the unbinding experiments.

The conditions under which kinesin dimers are imaged by cryo-EM, however, almost certainly reduce or eliminate the ability of both heads to bind(Hoenger, Thormahlen et al. 2000). In order to get good image quality, it is necessary to saturate the all the microtubule binding sites. Under these conditions, however, the dimer:binding site stoichiometry approaches 1:1, meaning that dimers most frequently cannot bind their second heads because adjacent tubulin sites are already occupied.

In these ultra-saturated cryo-EM conditions, therefore, a backwards-docked neck linker from the bound head of a kinesin dimer would only be able to *direct* the partner head towards a rearwards site. The partner head, however, would be unable to bind. As a result, these tethered but unbound partner heads might butt up against a rearward-bound dimer neighbor, as illustrated in Figure 10B. This arrangement of heads, in cryo-EM images, would consequently give a misleading impression: they would show the bound head in close contact with a tethered head to which there was no real connection. Indeed, cryo-EM data(Hirose, Henningsen et al. 2000) (shown schematically in the figure) are entirely consistent this result, for nucleotide-free motor conditions where we predict backwards neck linker docking to occur.

Postulating two possible backwards-docked neck linker conformations, furthermore, leads to the prediction that cryo-EM images might actually show a pair of *alternative,* "butting" (but not necessarily connected) free heads. Figure 10C shows that this second prediction is met, in images of the fast-walking *neurospora* kinesin. For this kinesin, it was found that different experimental samples, in nucleotide-free conditions, could yield images with the bound head contacting the "partner" head either on either the upper-left

or the upper-right sides. Changes in the microtubule lattice or in the heterogeneous tubulin preparations are likely differences in the *neurospora* experiments. Such variations could easily affect which of the two backwards-bound neck-linker conformations were favored in a given experiment, explaining the observed variability. Thus, cryo-EM images of the *neurospora* dimer appear to be entirely consistent with our theory of not one, but two backwards-bound neck linker conformations.

### *Two backwards-docked neck linker conformations can produce an "asymmetric, hand-over-hand" model of kinesin motility*

A most curious observation has recently been reported by (Hua, Chung et al. 2002) in single-molecule experiments of walking kinesin. In these experiments, functional copies of dimeric kinesin were rotationally immobilized on a streptavidin-coated glass coverslip, using a pair of biotinylated cysteines introduced near the motor-proximal end of kinesin's coiled-coil dimer interface. They then introduced microtubules together with very low concentrations of ATP, so that microtubules would bind single kinesin molecules but step only rarely, at intervals of seconds or minutes. In these experiments, it was found that after kinesin attached to a microtubule, the microtubule became rotationally stabilized (i.e., did not pivot around the kinesin attachment point) during the prolonged "waiting periods" that separated the very brief, millisecond-order ATP-induced stepping events. Control experiments, on the other hand, indicated that microtubules *could* swivel freely when they were nonspecifically attached to the coverslip. Therefore, these experiments prove that there is some kind of restraint that prevents the coiled coil from rotating relative to the microtubule during these waiting periods.

Perhaps the most straightforward explanation for this rotational immobilization would be that both heads were tightly bound at all times during the waiting period, so restricting the orientation of the coiled coil. However, there is a problem with this idea. After a single step, the identities of the forward and backward heads are swapped, which would imply that the restraint on the coiled coil is reversed (Figure 11A-B). This simple, "doubly-bound" model therefore predicts that, after each step, the coiled-coil should relax

to a new orientation that is flipped 180° relative to the last step—an event that the Hua et al. experiments definitely rule out.

It is an established fact that both heads of a kinesin dimer can bind simultaneously to the microtubule(Kawaguchi and Ishiwata 2001). Therefore, it is necessary to reconcile the observations of Hua et al. with the occurance of doubly attached catalyic heads. In order to do this, a *second* connecting geometry must exist that can connect the two heads of a dimer when they are both microtubule-bound (Figure 11C). This second arrangement allows the light blue head (in the figure) to switch from being bound behind the dark blue to being bound in front of it, without rotating the coiled coil very much. As pointed out by Hua et al., *each* of the two connecting geometries must also be rotationally stable and unlikely to interconvert, because otherwise 180° flipping of the coiled coil would have been observed during the "waiting periods" that were scrutinized experimentally. These requirements, that two discrete connecting geometries exist and that they do not interconvert, constitute a severe restriction (which might be called a "symmetry constraint") on allowable walking models for the kinesin motor.

The two backwards-docked neck linker conformations we have inferred in our experiments provide a satisfying way to fulfill the symmetry constraint. Inspecting the proposed arrangement of motor heads, neck linkers, and coiled coil in Figure 10 reveals why this is so. We propose that the neck linker (red) and the coiled coil both become specifically bound on the microtubule surface, and can do so in two different ways (Figure 11B-C). In this way, the orientation of the coiled coil will be fixed during "waiting periods" in kinesin's stepping cycle, and will not easily interconvert between the two alternatives. This mechanism will work even if the rearward head is not tightly bound—and the experiments of Ishiwata et al. suggest the second head may be bound only half the time, under the nucleotide-free conditions most resembling the "waiting period" we discuss.

The ideas discussed above are sufficient to sketch out a fairly detailed model for kinesin's stepping mechanism, addressing in particular the coordination between catalytic heads.

Before we introduce our model, however, we address a few details related to catalytic events within individual catalytic heads of kinesin.

***Evidence for nucleotide-controlled rotation in our K349 construct***

A striking observation made of the KIF1A kinesin was that its microtubule-binding subdomain, the "switch II cluster," rotated by ~20° relative to the core catalytic domain between two crystallized conformations(Kikkawa, Sablin et al. 2001). This motor was observed to make a similar, and opposite, rotation on the microtubule surface in cryo-EM images of two nucleotide states, suggesting that its switch II cluster was fixed while the *core domain* rotated. In crystal structures for our K349 construct described in the last chapter, the switch II cluster was observed to rotate by 7.5°-- a significant amount, but less than half of KIF1A's. This suggests that if K349's switch II cluster was fixed on the microtubule surface like in the proposed KIF1A mechanism, a nucleotide-induced 7.5° rotation by the core catalytic domain might be hard to detect by cryo-EM imaging.

In the case of K349, however, cryo-EM data are available(Rice, Lin et al. 1999) with an additional feature that may enhance the ability to distinguish rotations of the core domain: a gold label fixed on the tip of the catalytic domain. Remarkably, when these image data are compared for the nucleotide-free and AMPPNP-bound states of the motor, a difference is seen, as shown in Figure 8. While the gold label's density in the AMPPNP state has a localized, relatively spherical shape, in the nucleotide-free state the label density becomes elongated. Furthermore, the elongation of the label density extends in the clockwise direction, relative to the AMPPNP state—the same direction that the KIF1A orientation changes when it makes the analogous conversion (from the AMPPNP-bound state to the ADP state). The magnitude of the apparent arc seen in the gold label density even agrees with the 7.5° rotation predicted by crystal structures (see figure). These images, therefore, suggest that the K349 catalytic domain wobbles (this wobbling should not be confused with the coiled-coil "wobble problem" discussed above) in the nucleotide-free state, likely exchanging between two orientations relative to the switch II cluster.

We offer the following structural interpretation for these results. In the previous chapter, we presented evidence that the ADP state of free K349 is in equilibrium between two conformations similar to the crystallized ones, featuring a 7.5° rotation of the switch II cluster (see Figure 6A from chapter). We propose that when kinesin binds microtubules and releases ADP, this rotational mobility between the cluster and the core domain remains potentially available—but can be restrained by a single, new, significant interaction. The restraint, we suggest, is the "backwards" docking of the neck linker that we have inferred from the probe data of K349.

Our proposal explains kinesin's implied wobbling in the nucleotide-free cryo-EM images, in the following way. In the AMPPNP complex, the switch II cluster is rigidly confined to the orientation seen in the "docked" crystal structure, by interactions between switch II and the nucleotide's γ-phosphate that have previously been proposed. This positioning of switch II also enables the neck linker to dock forwards on the catalytic domain, towards the microtubule plus end (schematic in Figure 8A). In the nucleotide-free state, switch II is liberated and can twist by 7.5°. This free twisting allows the catalytic core to reorient on the microtubule, a movement that creates a clash between the switch II cluster and the neck linker (see Figure 6 from the last chapter) and forces undocking. As discussed above, the undocking may even be enhanced by a steric clash between H6 and the microtubule surface (see Figure 4 and accompanying discussion).

In the newly available motor orientation of the nucleotide-free state (schematic in Figure 8B, left-hand side), the neck linker becomes repositioned near a "backwards" binding site on the microtubule, generating a new interaction. In this way, when the neck linker docks backwards, it restrains the core from rotating back to the orientation found with AMPPNP. This restraint would not be absolute in the case of the nucleotide-free K349 monomer, however, because even near the freezing temperatures of the cryo-EM experiments our EPR results show that the neck linker is only bound half the time (see Figure 4 above). Thus, when the neck linker is unbound, the motor core can resume

wobbling on its switch II cluster "platform" (schematic in Figure 8B, right-hand side), explaining the "arc" of probe density in the nucleotide-free state.

### *Microtubule-bound rotation of the catalytic domain could drive a backwards- to forwards-bound neck linker transition*

The above proposal, relating motor domain rotation on the microtubule to neck linker binding states, is little more than a guess. However, it is at the heart of a simple and physically appealing model in which a monomeric kinesin head could generate a large, 8 nm powerstroke, as illustrated in Figure 10. The model works as follows. When kinesin initially binds to the microtubule and releases ADP (Figure 9A), the neck linker is specifically bound on the microtubule, which pulls the catalytic domain counterclockwise on its flexible base (the switch II cluster). Then, when ATP binds, the catalytic core twists *clockwise* on the switch II cluster due to strong attraction between switch II's nucleotide sensor and the active-site γ-phosphate (as in Figure 6B of the last chapter). The immediate effect of this twisting would be to dislodge the neck linker from its microtubule-binding site (Figure 9B), which might also lead a substep as discussed below. Subsequently, the neck linker would "zipper" up the side of the motor domain, as has previously been proposed(Rice, Lin et al. 1999), to complete the "power stroke" (Figure 9C).

"Twisting off" the backwards-bound neck linker conformation by ATP-induced rotation of the motor domain implies a certain versatility that may actually be required in kinesin's mechanism. If there are *two* backwards-directed neck linker conformations in the nucleotide-free state, as we have suggested, an ATP-induced transition in kinesin would have to be versatile enough to control both of them. The twist-off neck linker mechanism provides a way to do this. The two alternative microtubule-bound neck linker conformations need not satisfy any special requirements, except that they both only be accessible in the nucleotide-free, *clockwise*-positioned orientation of the motor.

### *A detailed model for processive, dimeric movement by kinesin along microtubules*

In figure 12 we present a model for kinesin motility that incorporates the elements discussed above. To illustrate the asymmetry between "left-" and "right-" foot forward steps that we have proposed, these two stepping pathways are both shown, with equivalent panels vertically aligned. Kinesin's movement in this model resembles that of an ice climber, where the coiled-coil is the "body" and the two motor domains are left and right "pickaxes."

According to this model, kinesin begins a hydrolysis cycle in either of two "waiting" states (step 1, top or bottom). The forward, strongly bound head of the dimer is nucleotide-free, while the rear head probably contains ADP (Hackney 1994). The neck linker of the forward head (red) is docked "backwards" on the microtubule at either one of its two alternative binding sites. The neck linker of the rear head probably docks along the motor core domain in moderate amounts similar, to the free K349 monomer (~20%).

Following ATP binding in the forward head (step 2), we propose that the catalytic domain makes a 7.5° rotation and dislodges the neck linker from either of the backward binding sites. Initially, this may lead to a "hydrophobic collapse" of the short neck linker chain, which has 5 hydrophobic residues and a cysteine in the short span of 15 amino acids. If so, the neck linker transition between steps 1 and 2 may provide an explanation for 3- and 5-nm substeps that have been observed experimentally(Coppin, Finer et al. 1996). A significant reaction barrier to backwards stepping may also lie between steps 1 and 2, if the ATP-induced rotation of the catalytic core is tightly coupled and the new orientation does not easily permit rearwards of the neck linker towards the previous microtubule binding site.

The transition initiated by ATP binding continues in step 3, with forwards "zippering" of the bound head. In our model, this represents the second and final substep in which the coiled coil is significantly translated. The remaining steps serve to establish and solidify the attachment of the partner head, and prepare both motors for another cycle of hydrolysis. We note that, under load, the forward zippering in step 3 might not all happen at once, which could lead to gradual pulling phase during this second substep.

Such an effect may have been seen recently in very sensitive measurements of single kinesin molecules using dark field microscopy (H. Higuchi, personal communication).

In step 4, the neck linker of the forward head finds its binding site on the microtubule. The choice of either a "left" or a "right" site in this process is determined by the coiled-coil geometry, which is in turn determined by the arrangement in the previous step. This neck linker binding, like the "zippering" in step 3, could be very fast because the interactions are likely to occur sequentially.

We note that, while step 4 is depicted as a transient in the mechanism, a recent experiment may have captured a piece of it. The very elegant experiments of (Sosa, Peterman et al. 2001) revealed a most unusual microtubule-binding mode by the K349 monomer. In the ADP-complexed state only, the catalytic domain of this construct was found to rock back and forth with high amplitude (a full 180°) while remaining bound to the microtubule. Remarkably, this binding mode appeared to be specific—no observable diffusional motion along the microtubule was found during the rocking. While it was suggested that a loosening of the binding interface might have allowed this type of motion, it is hard to conceive how a protein could cling to an interface while undergoing such a dramatic movement. An alternative explanation suggested, by our model, is evident in the upper, ADP-bound head of step 4. We propose that the *neck linker only* of K349 specifically attaches to the microtubule surface under these conditions, while ADP (which puts kinesin in a weak-binding mode) prevents the full association of the catalytic core domain. Such a conformation would only be weakly associated with the microtubule, which is consistent with Sosa et al.'s finding that the ADP binding mode only occurred at very low ionic strengths.

In the last step before the cycle repeats itself (step 5), the forward head binds the microtubule and releases ADP. This step is important because it is a *strong*-binding event—so once the forward head has found its site, a much stronger force is required to pull the motor backward than would be required, say, to unravel a zippering neck linker. Once step 5 is complete, "all" that remains is for the rearward head to complete its

hydrolysis step and unbind from the microtubule (leading back to step 1). As has been pointed out, however,(Schief and Howard 2001) hydrolysis and unbinding of the rear head is a tricky step. The sequence of events is very important. If the rear head releases phosphate while still bound, then it becomes faced with a choice: either unbind from the microtubule, keeping its ADP—OR—release ADP and reenter a strong binding state. The latter choice would not lead to an effective step, however, because it would tend to pull the forward head *back,* conceivably leading to its unbinding and, consequently, a futile hydrolysis step. However, this is exactly what the *forward* head *does* in step 5—it binds microtubules and releases ADP. One solution to this conundrum would be for kinesin to release from microtubules in the post-hydrolysis ADP•Pi state(Schief and Howard 2001).

***Rationale for kinesin's inferred, elaborate behavior***

There is a reason for kinesin to move with the asymmetrical "left" and "right" steps of Figure 12. The reason is related to motor velocity. Previously, it has been proposed that the free head of a walking kinesin dimer searches, in a fairly random, diffusional way, for the next available microtubule binding site(Rice, Lin et al. 1999). Such a step, however, might be fairly time-consuming. A problem not usually pointed out in these models is that the neck linker of the searching head must be stretched essentially to the limit (and in the proper direction) when forward binding finally occurs. This stretching is certain to extract a strong entropic penalty. The stretched conformation is also likely to take a long time to find, in the absence of other stabilizing interactions.

A microtubule-binding mode for the neck linker offers a solution to these two problems. Favorable interactions with the microtubule could easily stabilize the extended neck linker conformation enough to compensate for the lost entropy. Not only would these interactions speed up the forward head's search for a binding site, they would also favor the final binding interaction. Furthermore, "zippering down" of the neck linker onto the microtubule could happen sequentially, meaning that the neck linker's conformational

search for the extended state would be highly directed. This feature could significantly enhance kinesin's maximum velocity.

However, having only *one* microtubule-binding mode for the neck linker would introduce a problem. This problem is directly related to the "symmetry constraint" discussed in the last section, and was illustrated in Figure 11. After kinesin has taken a single step, the coiled-coil would have to rotate by 180° in order for the *neck linker* of the searching head to reach the same binding site on tubulin as was used by the other neck linker in the equivalent, previous step. However, waiting for a 180° rotation (a "wobble") of the coiled coil after every step could represent a significant kinetic barrier for the walking motor. The resulting slowdown might be particularly noticeable if the cargo (typically vesicles or other relatively large cellular structures) at the end of the coiled coil did not rotate quickly, restricting the coiled coil's geometry. The solution to this "wobble problem" is what we have proposed already: in successive steps, kinesin's neck linker could alternate between two different binding sites on the microtubule. In this way, searching of the forward head would be highly directed, avoiding the potential pitfall of the wobble.

## Conclusions and Proposed Experiments

I personally find the evidence for kinesin's asymmetrical steps, based on the data we present above, to be quite compelling. The alternating, "left" and "right"-handed, *backwards*-bound neck linkers illustrated in Figure 12 seem to me to be an elegant way for a molecular motor to solve the basic engineering problems that it faces. More than this, I have not been able to come up with any other way to explain the large (but admittedly diffuse) web of data that now exists to describe kinesin's mechanism. Perhaps the data are not quite yet at the point where, in the words of Sherlock Holmes, "when you have eliminated the impossible, whatever remains, however improbable, must be the truth"—but to me, the situation is approaching that point. If the right experiments can be designed, it may soon be possible to illuminate many details of the model we have just proposed.

***Selectively knocking out forwards or backwards neck linker docking***

Perhaps the central hypothesis in Figure 12 is that kinesin's neck linker switches between specifically-bound conformation(s) on the microtubule (in the nucleotide-free state), to the crystallized, docked conformation in the AMPPNP-bound state. If this is true, it should be possible to *selectively* disable the forwards, or the backward, neck linker conformations— leaving one or the other unchanged. The way to do this experimentally is to modify the neck linker's interaction partner to *disrupt* its binding partners, while keeping the neck linker sequence itself intact. If this selective structure knockout were possible, then structure indicators (like EPR) should reflect it—showing wild-type levels of order and docking some specific nucleotide states, but much more disorder in others.

Selectively knocking out the neck linker's binding site on the motor *core* is the most straightforward experiment, because crystal structures provide an excellent model for the neck linker structure to be disrupted. Mutating the motor core facing the neck linker to include one or two additional well-placed bulky charged residues would be likely sufficient to eliminate (or at least severely modify) the motor-docked neck linker conformation. We would predict that neck linker docking in these mutants would be destroyed in all microtubule-free motor states, and in the microtubule-bound, ATP-bound state. If our prediction of *backwards-bound* neck linkers is correct, however, such mutants should show virtually no change (relative to wild-type) in neck linker docking levels in microtubule-bound ADP or nucleotide-free states.

The complementary "dock knockout" is to eliminate the *microtubule*-bound neck linker states. This is a more difficult task, because there are no detailed structural models describing the target interaction. However, I believe one likely possibility is that the neck linker interacts with tubulin's C-terminus (sometimes called the "E-hook"). This guess is based on a variety of biochemical data, including cryo-EM imaging of KIF1A that suggests an interaction between the E-hook and kinesin's neck(Kikkawa, Sablin et al. 2001). Thus, one very easy experiment is to remove the microtubule's E-hooks (which

can be done by a subtilisin digestion), and ask how this modification affects kinesin's neck-linker docking in various states. The prediction would be that *only* microtubule-bound, ADP-bound or nucleotide-free motors would show a significant decrease in docking, relative to experiments with untreated microtubules.

In combination, these experiments could provide additional, powerful evidence for the microtubule-bound neck linker conformations we propose, and work on them is currently in progress.

***Detecting asymmetry in kinesin's walking steps?***

In Figure 12 we have proposed the first so-called "asymmetric hand-over-hand" mechanism for kinesin with sufficient detail to be experimentally tested. The asymmetry in this model is entirely localized to the neck linker. Thus, we predict that probes placed on this element will be likely to detect differences between "left" and "right"-handed steps. Single-molecule experiments may be the most straightforward way to look for these differences, but the technical problems involved may be significant. If there is a large angular change in the neck linker's *orientation* between the left and right steps, then one possibility would be to use bifunctional fluorophore probes (for example, linking C330 and C333 of the neck linker, see Figure 3 of Chapter 1). Polarization measurements of single kinesin molecules labeled in this way would then be predicted to alternate between two discrete values, corresponding to the left and the right-handed steps. Such an experiment, however, is probably years away. It remains to be seen whether other, simpler experiments may be capable of providing direct proof of kinesin's walking asymmetry.

***A possible role for Switch I in microtubule unbinding?***

One of the great puzzles to me in kinesin's mechanism is that ADP release strongly modulates kinesin's affinity for the microtubule. Cryo-EM experiments appear to show kinesin binding to the microtubule in a very similar conformation in either the nucleotide-free or the ADP-bound states, with the neck linker in the same, apparently "backwards-

bound," conformation in both(Rice, Lin et al. 1999). Furthermore, both of these nucleotide states are expected have switch II in a "downstroke" position, which would suggest the catalytic domain was docked onto the microtubule in a similar orientation in either case. However, ADP binding *dissociates* kinesin's catalytic domain (but perhaps not the neck linker, see Step 4 of Figure 12 and accompanying discussion) from the microtubules, while the nucleotide-free state remains a tightly bound. How could this difference be explained by kinesin's nucleotide-sensing elements?

My favorite hypothesis to explain this conundrum involves kinesin's "other" nucleotide sensor, Switch I, which has not yet been assigned a strong role in any particular aspect of kinesin's mechanochemical cycle. Perhaps Switch I detects a difference between nucleotide-free and ADP-bound states, while switch II remains more or less unchanged. There is some precedence for this idea, because the Switch I of G-proteins has been found to change conformations depending on whether nucleotide is present. The recently solved structure in our lab of the kinesin analog XKCM1 (Jennifer Turner, personal communication), in a nucleotide-free form, also shows a conformational change of switch I relative to other known kinesin structures (which all have bound nucleotide).

If the switch I conformation in the ADP (but not the nucleotide-free) state clashed with the microtubule, this could dissociate the motor domain from the microtubule. A significant aspect of this hypothesis is that switch I could conceivably remain in the *same* conformation in ADP- *or* ATP-bound states. However, rotation of the motor induced by ATP binding might conceivably *relieve* the switch I clash, even if the switch I conformation did not change. This would allow a fairly simple mechanism to explain why the ATP and nucleotide-free states of kinesin bind microtubules tightly, while the ADP state does not. This idea, of course, is merely a guess— in the absence of much structural data.

***Final thoughts***

It is probable that coordinated interactions involving switch I, switch II, *and* the microtubule work together to govern kinesin's mechanism in ways that are now impossible to predict. Unfortunately, testing of most hypotheses related to kinesin's switch elements and microtubule binding will probably await the high-resolution structural characterization of the kinesin-microtubule complex. Much more work needs to be done before we can even begin to carefully examine how kinesin's structural mechanism relates to overall features like walking velocity, or chemical efficiency. It is fair to say that kinesin will remain a fascinating target for scientific investigation for many years to come.

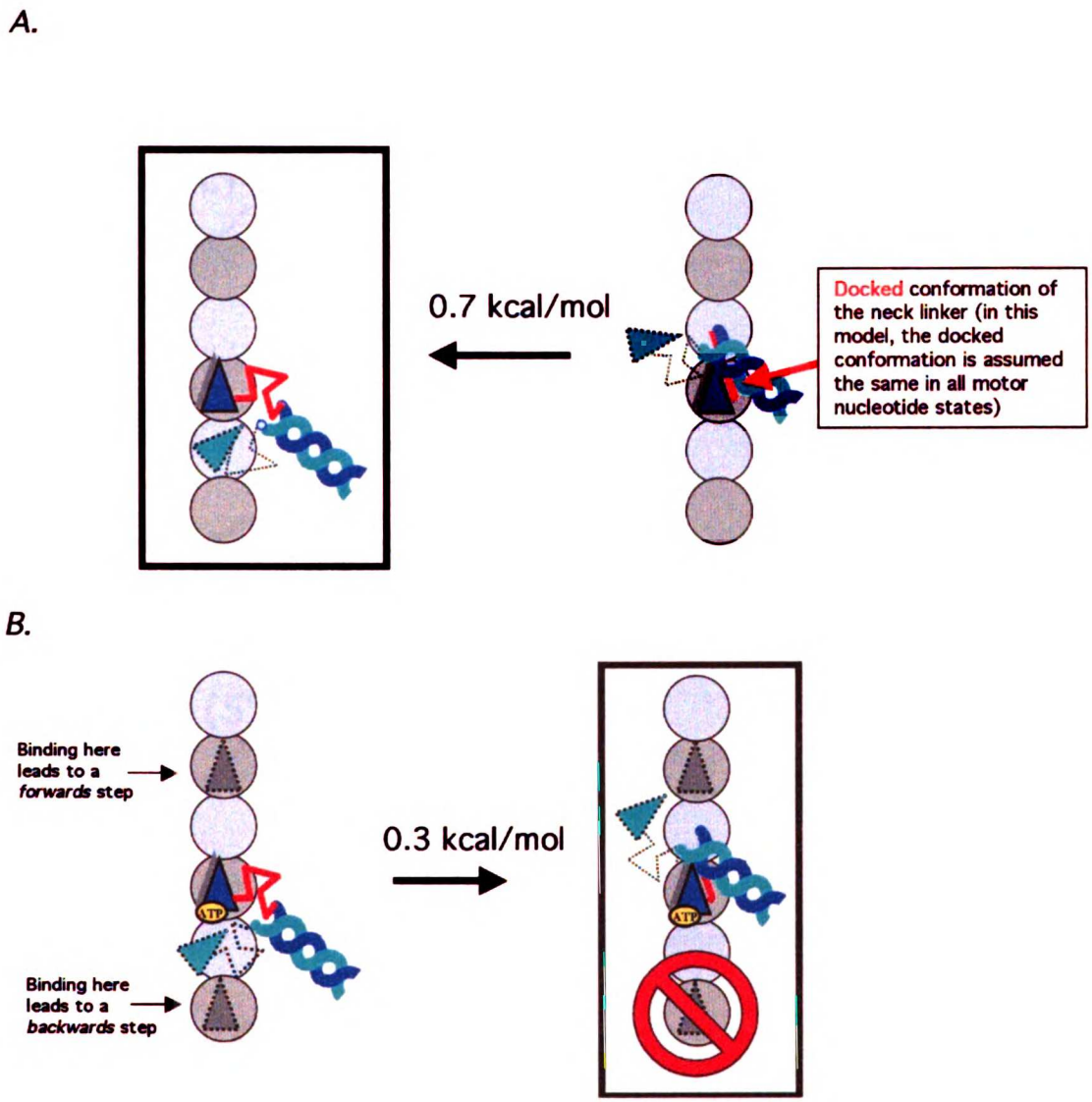



Figure 1. Illustrating how a motility model for kinesin might be fit using only the one currently known docked neck linker conformation.

*A.* In the absence of ATP, the neck linker of the attached monomer head (dark blue) favors a disordered conformation and has little or no specific preference for directing forwards vs. backwards binding of the tethered head (light blue). *B.* Upon ATP binding, a minor preference is introduced for the neck linker to reach its docked state, leading to an enhanced rate of forwards stepping vs. reversed stepping. ATP binding favors neck linker docking by an additional 1 kcal/mol. Displayed free energies for neck linker docking are the ones found by EPR analysis of the K349 monomer (see the last chapter),in experiments performed at 20°C; The "ATP" state was represented by AMPPNP-bound kinesin. Temperature significantly affects these free energies in K349, so that at 35°C docking is disfavored in both ADP and AMPPNP states, by 0.5 and 1.5 kcal/mol, respectively. Superficially applied, this model predicts a backwards stepping rate of ~15% under zero load, which is more than an order of magnitude higher than the experimentally measured rate of < 1% (H. Higuchi, personal communication). High-load behavior of this type of model is not known.

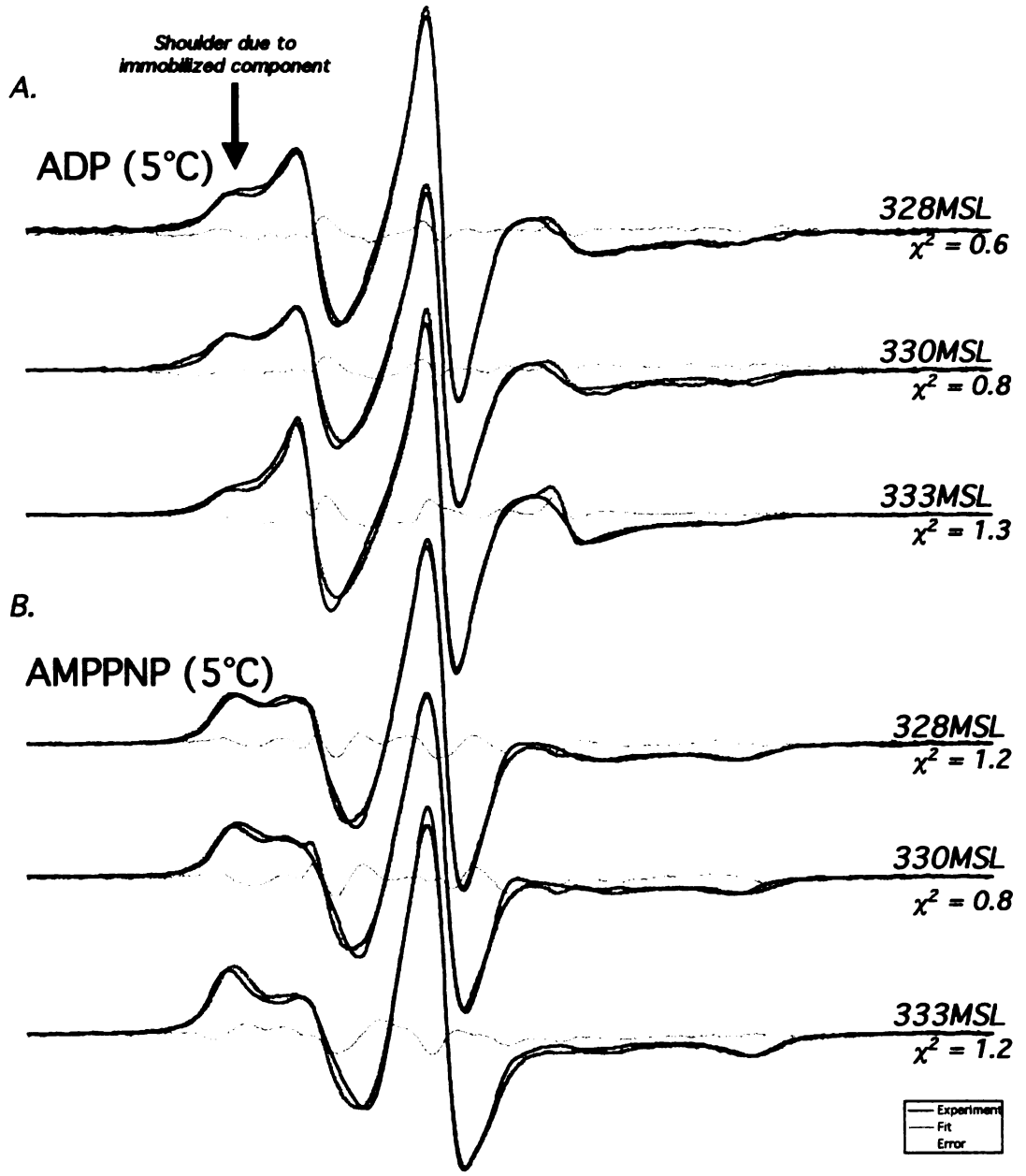



Figure 2. The EPR signals reported by probe labels at different sites on the neck linker are highly similar. All can be well fit ( $\chi^2$ <= 1.3) using the *same* three "basis set" signals (derived from the 333MSL labeled signals, see Figure 7 from the previous chapter). The signal fitting quality extends beyond the two kinesin nucleotide states shown here (AMPPNP and ADP, microtubule-bound motor) to include the microtubule-bound, nucleotide-free state (data not shown) and the microtubule-free states of the motor (see the previous chapter). The similarity of signals for three different probe sites shows that environmental effects at these positions are relatively subtle. This leaves open the possibility that ordered neck-linker conformations could be *different* depending on the nucleotide state of the motor.

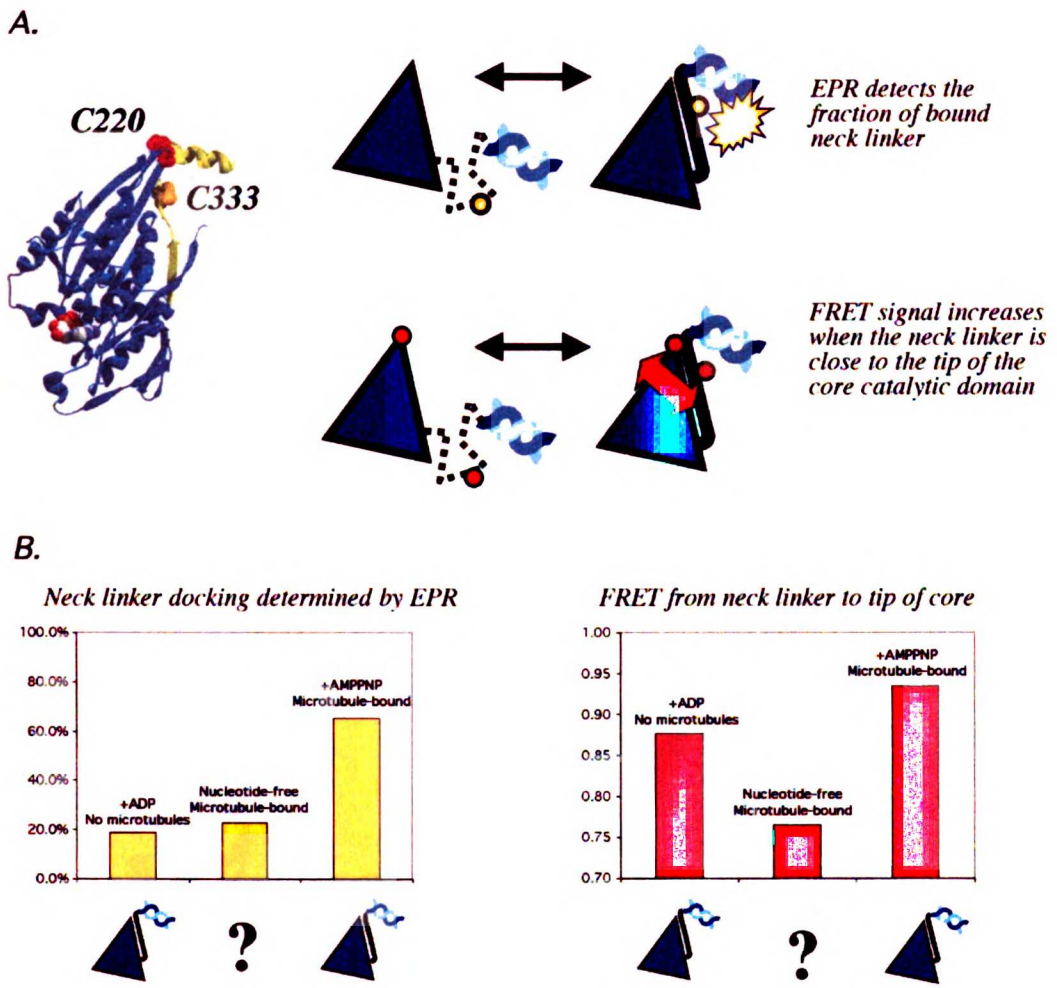



Figure 3. FRET distance monitoring of the neck linker shows a significant change(Rice, Lin et al. 1999) between two states of the motor that have similar quantities of docked neck linker (as determined by EPR).

A. The two contrasting types of information reported by the probes. EPR indicates a "bound" state of the neck linker but cannot necessarily discriminate between different conformations. FRET, on the other hand, cannot distinguish between bound or disordered conformational states, but *does* reflect positional displacements of the neck linker.

B. The nucleotide-free, microtubule-bound state of kinesin shows a partially "docked" neck linker that is apparently far from the "docked" position of the other two states presented here. Docking in these other two states (microtubule-free+ADP, and microtubule-bound+AMPPNP) has been shown with other methods to adopt a conformation like that seen in some kinesin crystal structures (see Chapter 1).

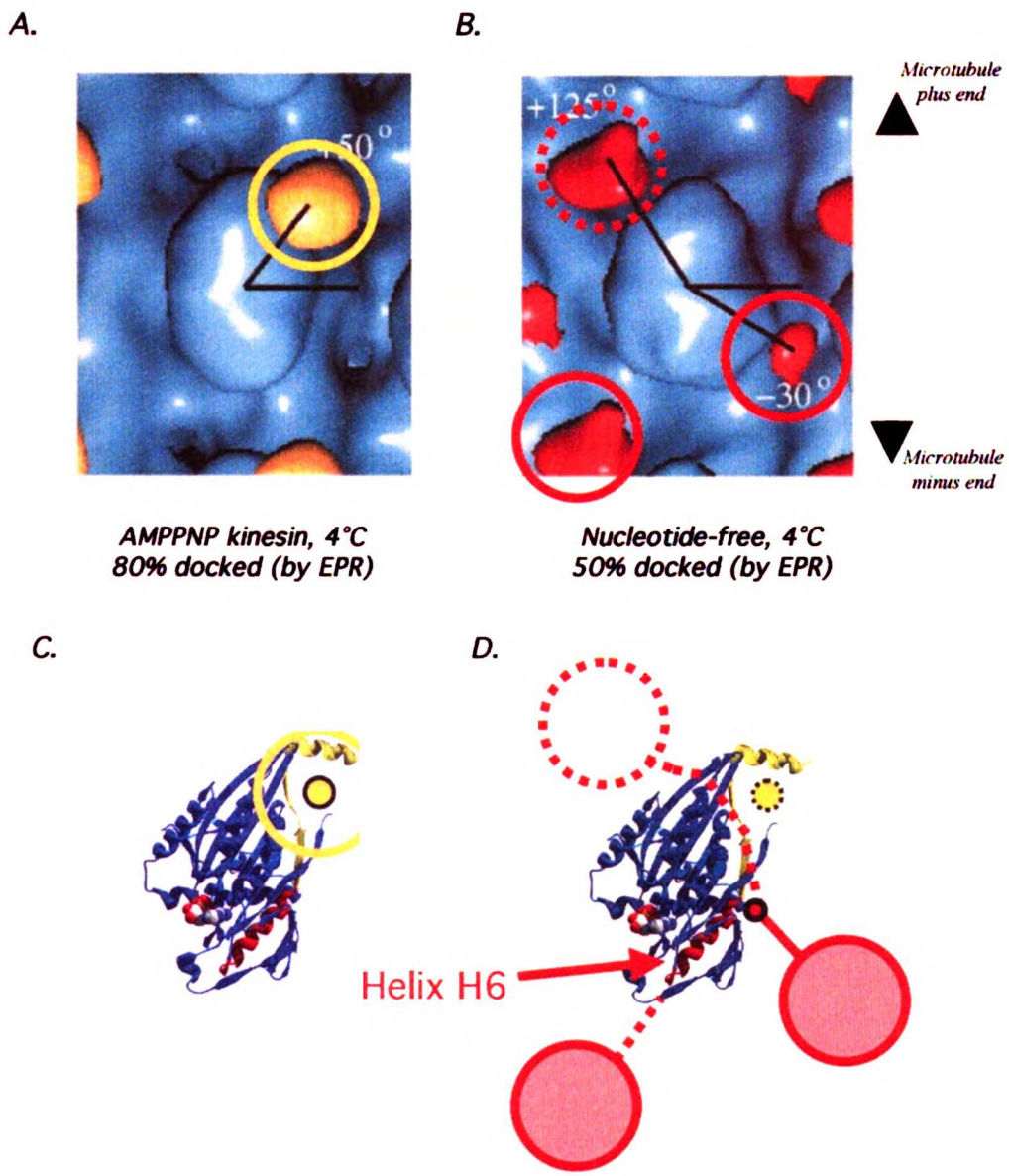
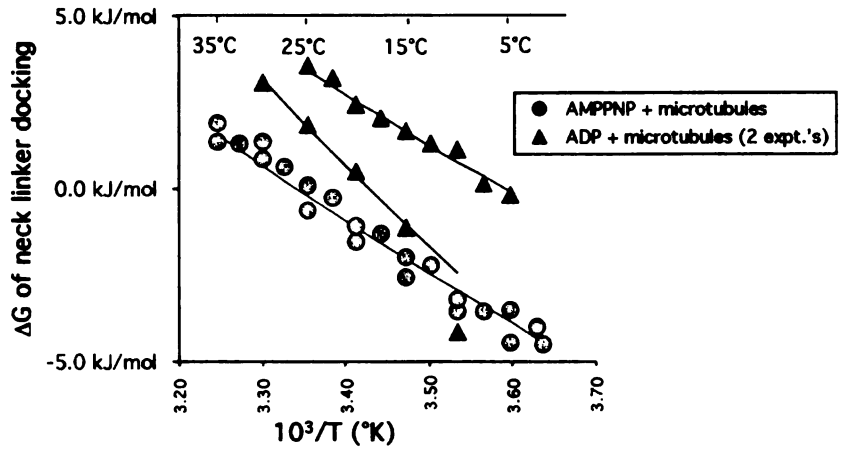


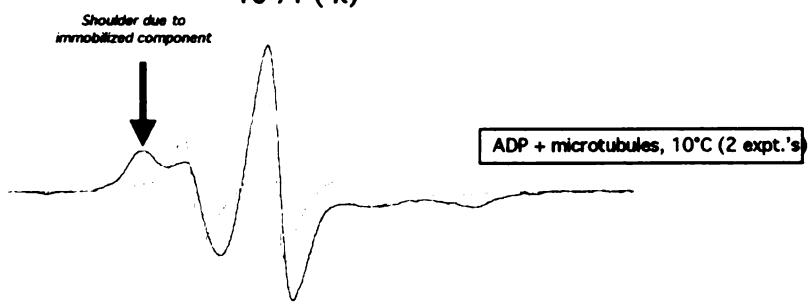

Figure 4. Cryo-EM images of kinesin•microtubule complexes indicate that ADP-bound, microtubule-complexed kinesin has *two* docked conformations of the neck linker, both distinct from the docked conformation seen in AMPPNP-bound, microtubule-complexed kinesin.

A., B. Cryo-EM images of kinesin with gold label attached to position C333 of the neck linker, taken from (Rice, Lin et al. 1999). In agreement with the data from Figure 3, these data indicate a change in the neck linker's docking mode, between the two nucleotide states shown. Note that the cryo-EM samples were equilibrated in a cold room before freezing, so that the neck linker docking (predicted by EPR) increases significantly from the conditions of Figure 3. C., D. Kinesin's crystal structure, matched approximately in orientation and scale with the images in A., B. While the gold label position for the AMPPNP state agrees well with the "docked" crystal structure (C.), the two label densities seen in D do not. Furthermore, geometry argues for that the left label density in D is *minus-end*-directed (solid red circle) relative to the kinesin monomer, as opposed to *forward*-directed (dashed circle) as was suggested by (Rice, Lin et al. 1999) (see text).

A.




B.




C.

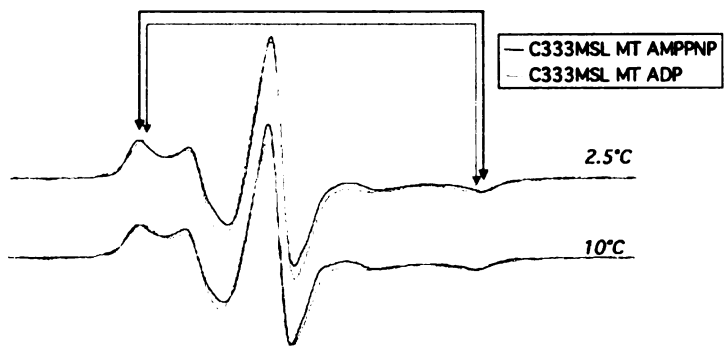



Figure 5. Two neck-linker binding modes for kinesin•ADP•microtubules are distinct from the kinesin•AMPPNP•microtubule docked state.

A. Neck linker docking data as a function of temperature (plotted as in Chapter 1, Figure 2), for experiments conducted with tubulin derived from two different sources (see text). Docking energies superimpose for all temperatures in the AMPPNP+microtubule motor state, but are greatly different (with different enthalpies, as reflected by line slope) in the ADP+microtubule state.

B. Overlay of EPR spectra for ADP+microtubule states (2 different tubulin sources) at 10°C illustrates the large docking difference between them.

C. Shifts in the positions of immobilized peaks (arrows) indicate that the probe environment has changed in the ADP+microtubule state, relative to the AMPPNP+microtubule state, suggesting a conformational change in the neck linker. This is consistent with the data from the previous two figures, which also indicate such a change between the AMPPNP+microtubule state and the ADP or nucleotide-free+microtubule states (these latter two states are probably similar in their neck linker conformations).

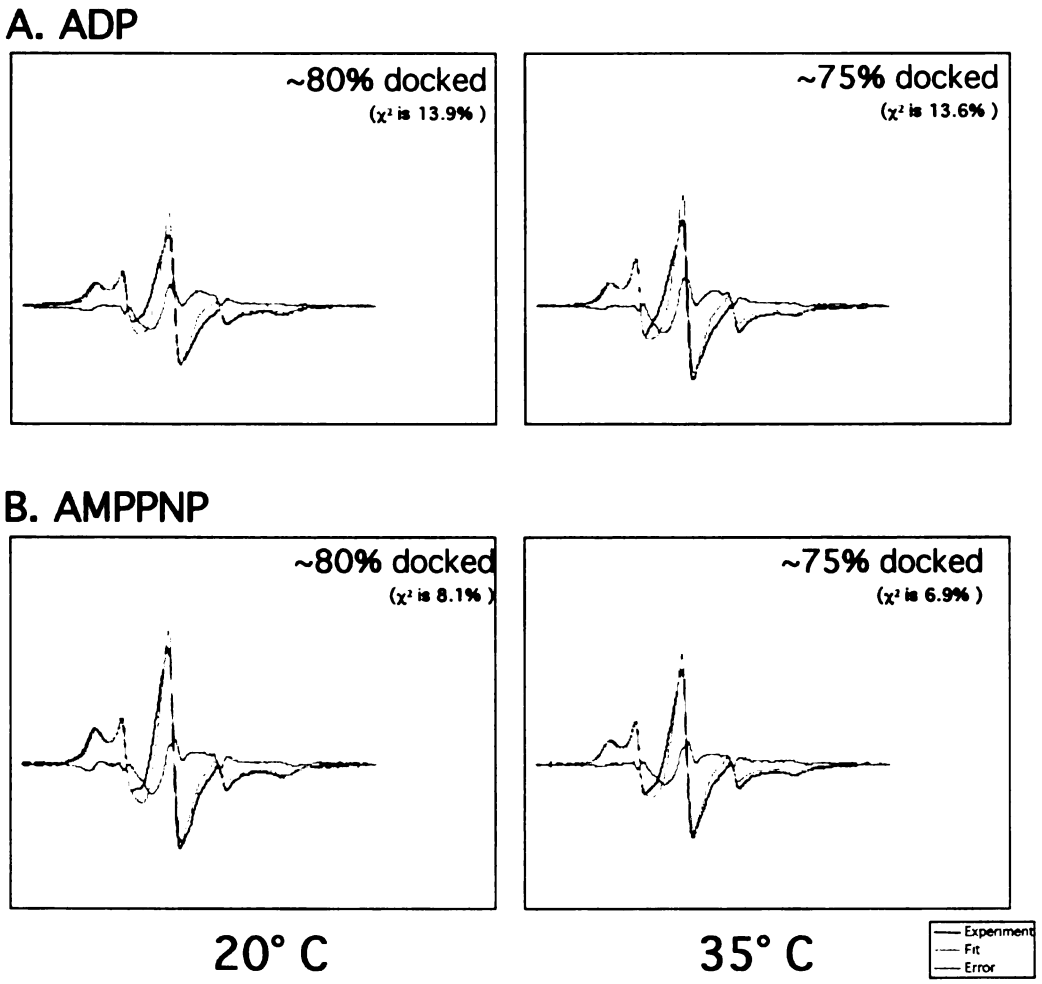



Figure 6. Neck-linker EPR spectra for the kinesin dimer (microtubule-bound)

A., B. Spectra for all temperatures can be fit with a fair degree of accuracy using the same EPR "basis set" components derived for data analysis on the monomer (see Chapter 1, Figures 1-3, 7). In the dimer, however, docking is very strong in both AMPPNP+microtubule *and* ADP+microtubule states, and is not highly affected by temperature (i.e. the docking transition is low in enthalpy).

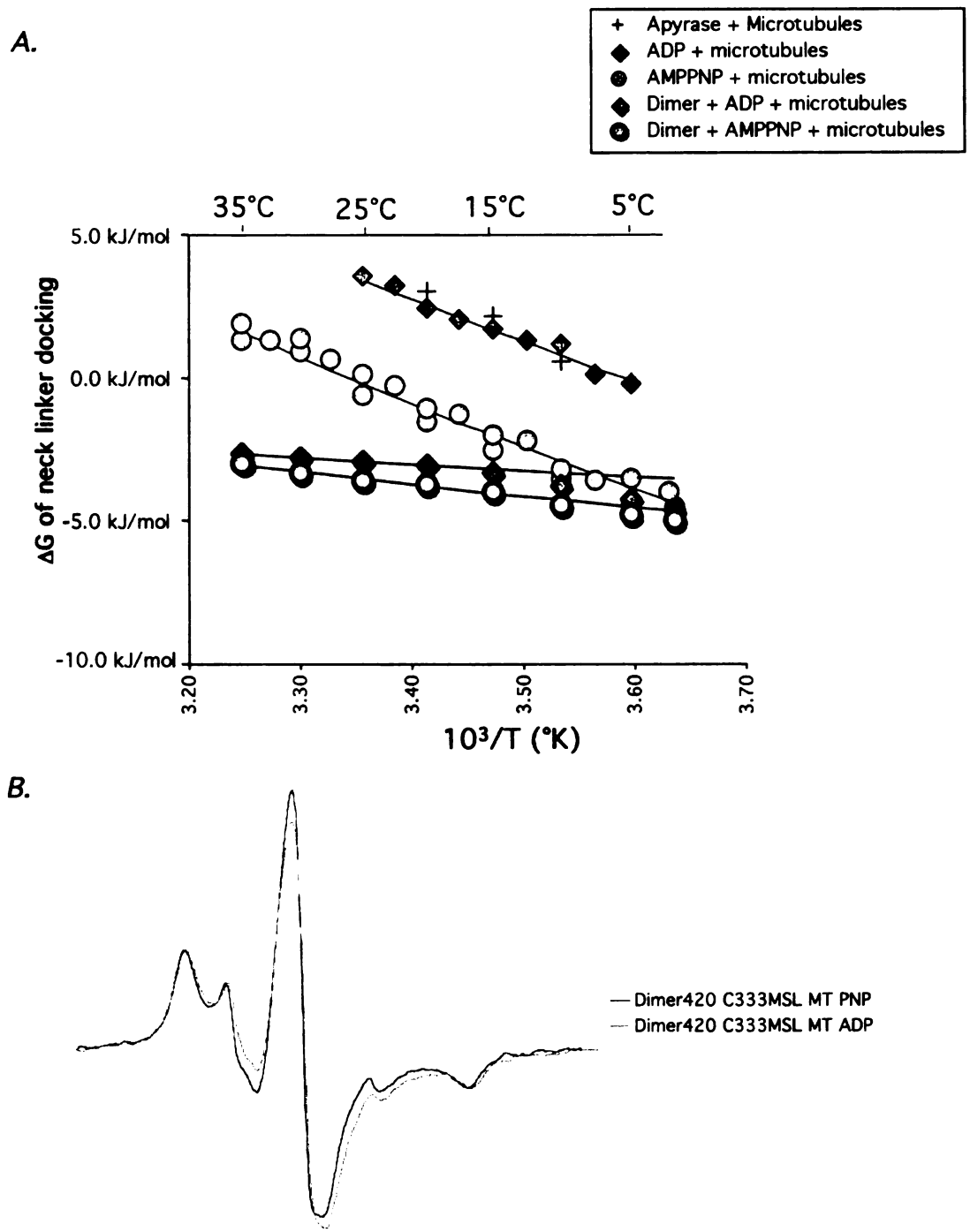



Figure 7. Free energy of neck linker docking for the kinesin dimer (microtubule-bound).

A. The 1 kcal/mol difference between ATP-bound and ADP-bound neck linker states of the microtubule-complexed monomer (upper lines) is virtually eliminated in the dimer. As can be seen (lower curves), quantifiably equivalent amounts of docked neck linker occur in either nucleotide state of microtubule-complexed kinesin.

B. Overlay of EPR spectra for microtubule-complexed, ADP or AMPPNP-bound kinesin. These spectra confirm that neck linker docking is very nearly equal in these two nucleotide states.

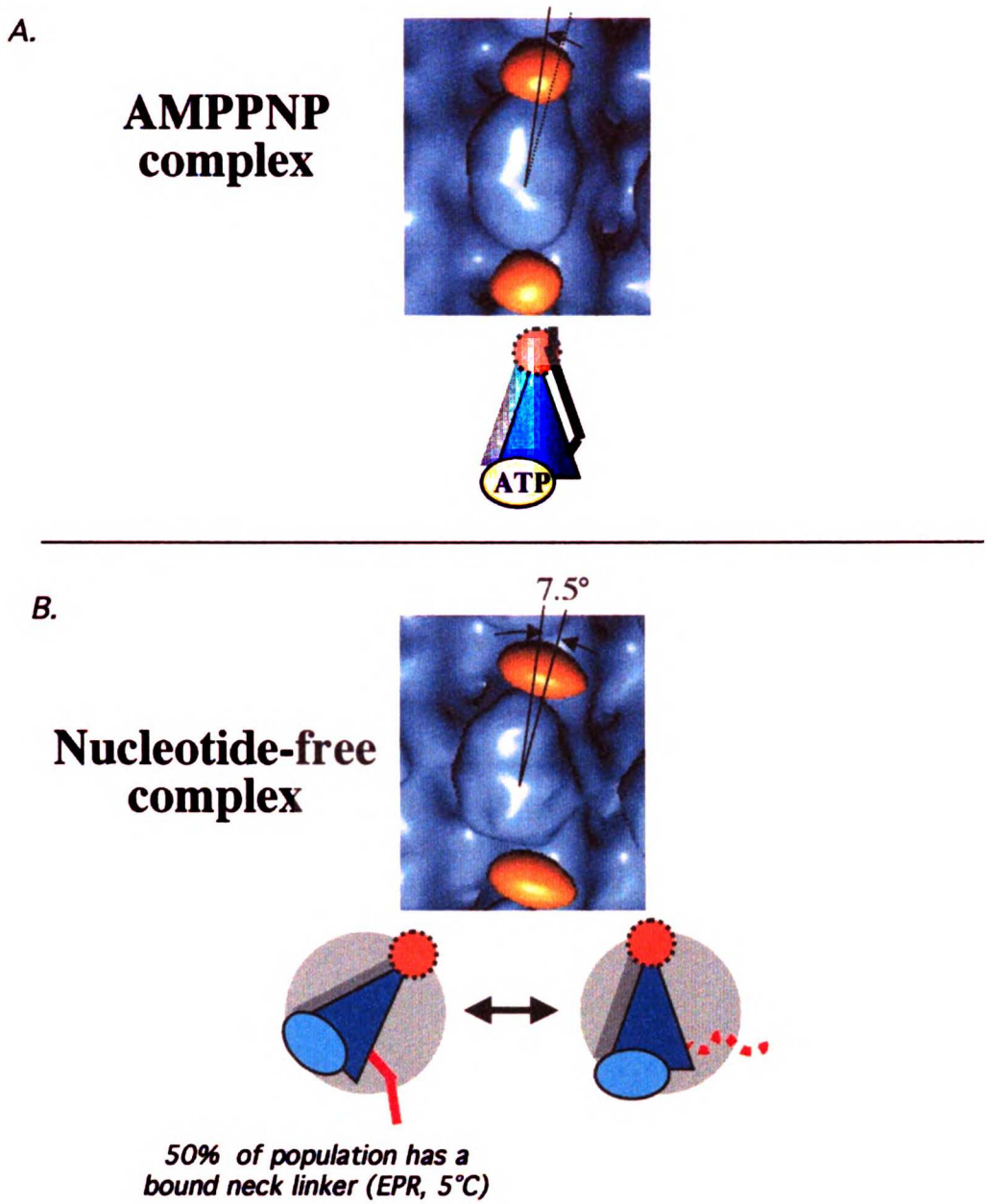



Figure 8. Evidence for rotation of kinesin's catalytic domain on the microtubule surface, driven by nucleotide exchange.

A. A gold label affixed to position 220 (see Figure 3) on the motor domain shows a mostly spherical electron density distribution in the AMPPNP complex (cryo-EM images taken from (Rice, Lin et al. 1999)).

B. The same label density becomes elongated, extending in the *clockwise* direction, in the nucleotide-free complex. This shape change is explained by rotation of kinesin's catalytic domain, with the *clockwise* position possibly compatible with a rearwards-docked neck linker (lower cartoons; see text).

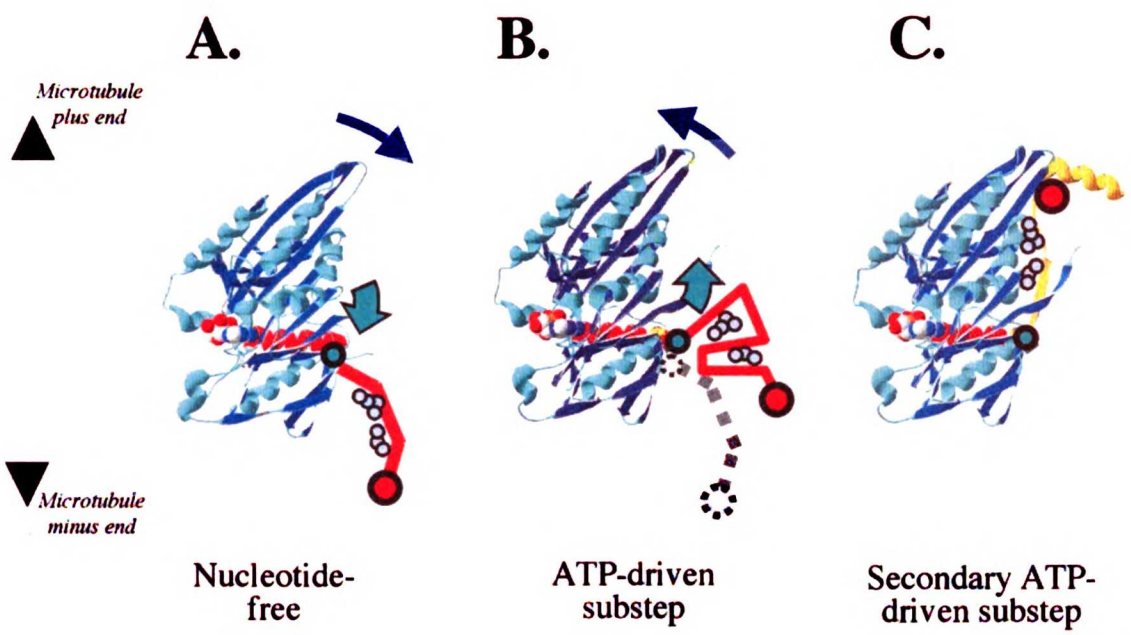



Figure 9. A model for how nucleotide-driven rotation of kinesin's catalytic domain drives a neck linker transition.

A. The nucleotide-free state permits a clockwise-oriented position of the catalytic domain, which allows rearward binding of the neck linker on a *microtubule* binding site.

B. ATP binding initiates a rotation of the catalytic core, which detaches the neck linker from its rearwards binding site. This could initially lead to a "hydrophobic collapse" of the neck linker, which has 5-6 hydrophobic residues within the span of ~15 residues—perhaps inducing a substep in the motor's walking mechanism.

C. Zippering of the neck linker into the crystallized, "docked" conformation completes the second half of the ATP-induced "powerstroke" by the neck linker.

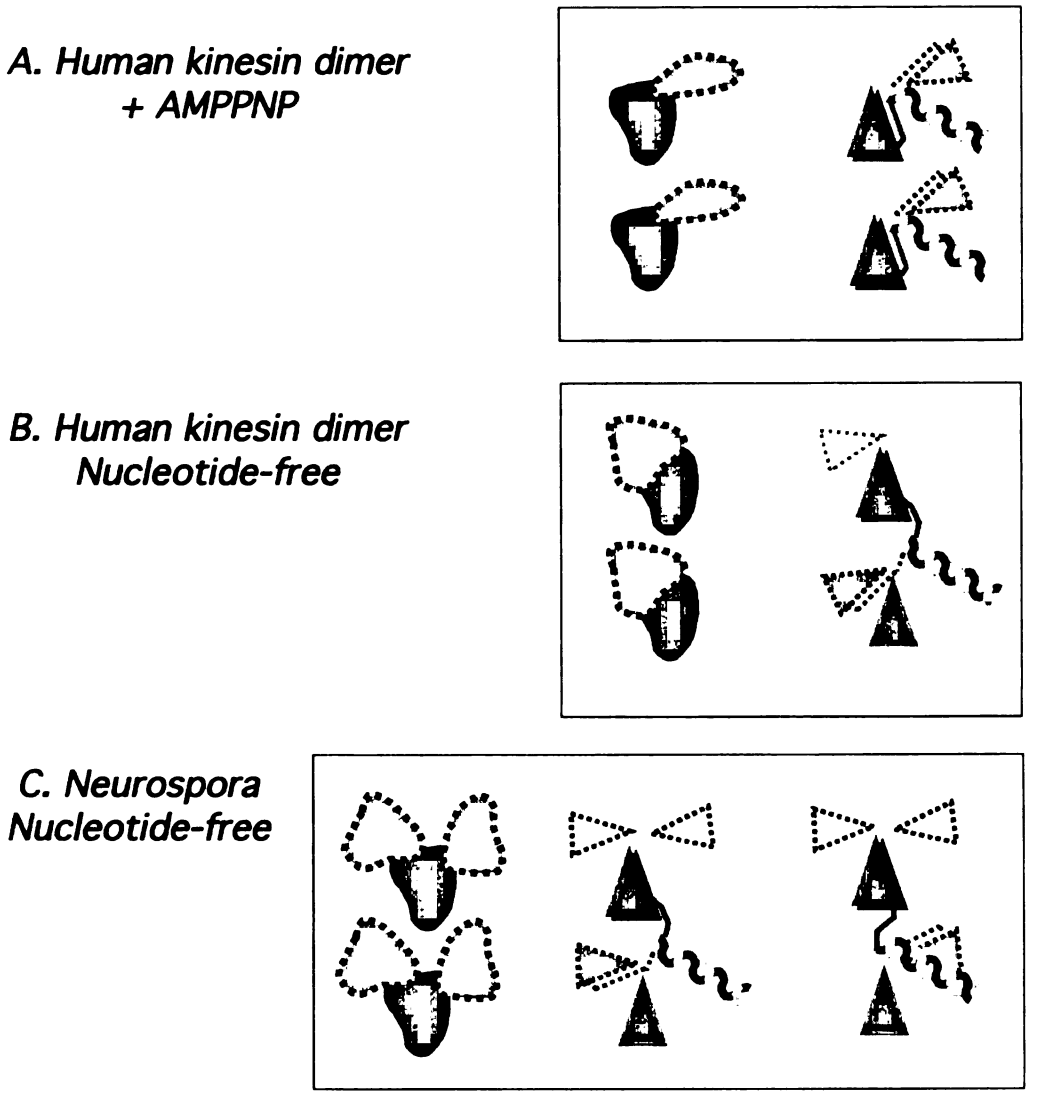



Figure 10. Reassigning the connectivity of dimeric kinesin heads in cryo-EM images of published motor-microtubule complexes(Hirose, Henningsen et al. 2000).

A. In the AMPPNP complex, the connectivity of dimeric heads is as previously proposed, with the second (microtubule-detached) head directed *forwards*, towards the microtubule plus end. This conformation may resemble the crystallized conformation of dimeric kinesin, although the neck linker of the detached head is likely to spend a significant fraction of the time "undocked" under room-temperature conditions (as inferred by our EPR data).

B. In nucleotide-free kinesin, we propose that the neck linker (red) of the tightly-associated head finds a microtubule binding site and extends towards the *minusi end*, positioning the second head so it could bind in the next "rearwards" position. Since a second dimer likely occupies the rearwards site in a significant fraction of the microtubule lattice in these experiments(Hoenger, Thormahlen et al. 2000), however, the second head could *abut* the site but not bind. This would create the misleading impression that the abutting heads are dimerically associated, but they are not.

C. In images of dimeric kinesin from Neurospora, two alternative locations are seen for the detached dimeric head in the nucleotide-free state. We propose that the two alternative positions correspond to different microtubule-associated conformations of the *attached* motor's neck linker, as shown.

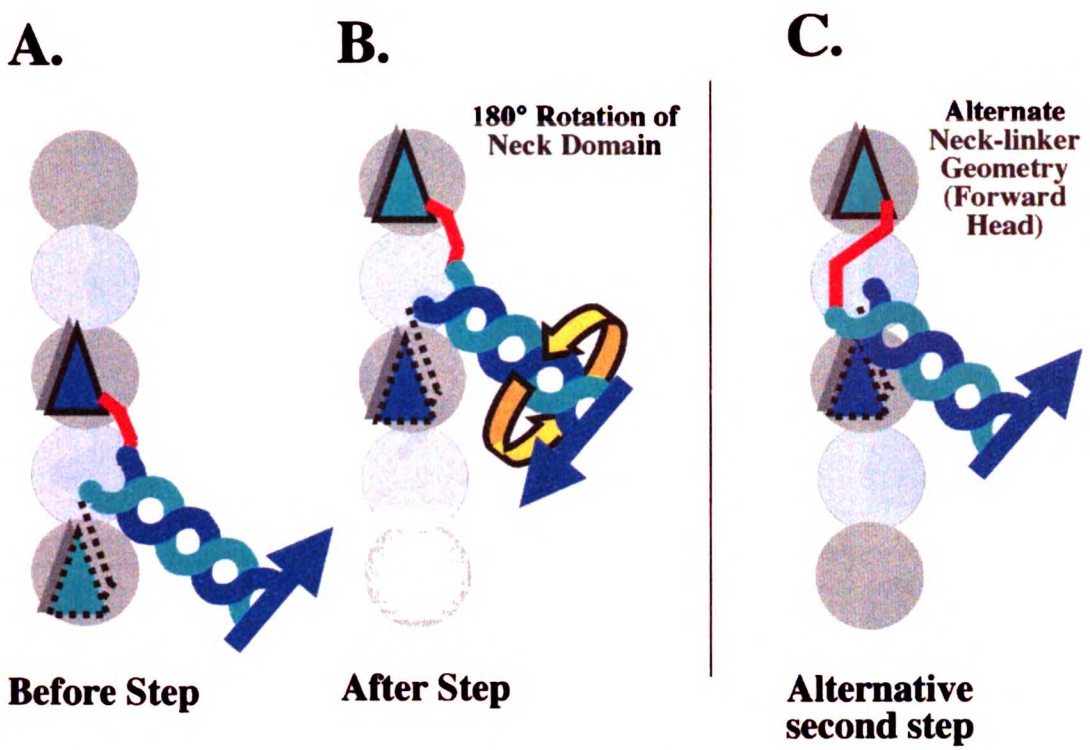



Figure 11. Symmetry restraints on a "hand-over-hand" walking mechanism, given a microtubule-bound neck linker conformation.

A. Cartoon illustrating kinesin poised in between steps. The forward (dark blue) head has a "backwards," microtubule-docked neck linker (red) and the connecting geometry of the coiled coil allows the rear head (light blue; dashed lines) to bind, although this binding may be only transient following hydrolysis of the rear head's ATP (from the previous step).

B. Following a forward step, if the light blue head (including the neck linker and coiled-coil part) assumes exactly the same conformation as the dark blue head did in the previous step, the coiled-coil neck domain must rotate by 180°.

C. A forward step in which the coiled coil does *not* rotate greatly. When the dark blue head binds at the forward position, its neck linker connects backwards to the coiled-coil in a different geometry than the light blue head did in A. The coiled-coil geometry of this step requires that the neck linker of the forward head adopt a *different* microtubule-bound conformation (red) than in B, to remain compatible with binding of the rear head.

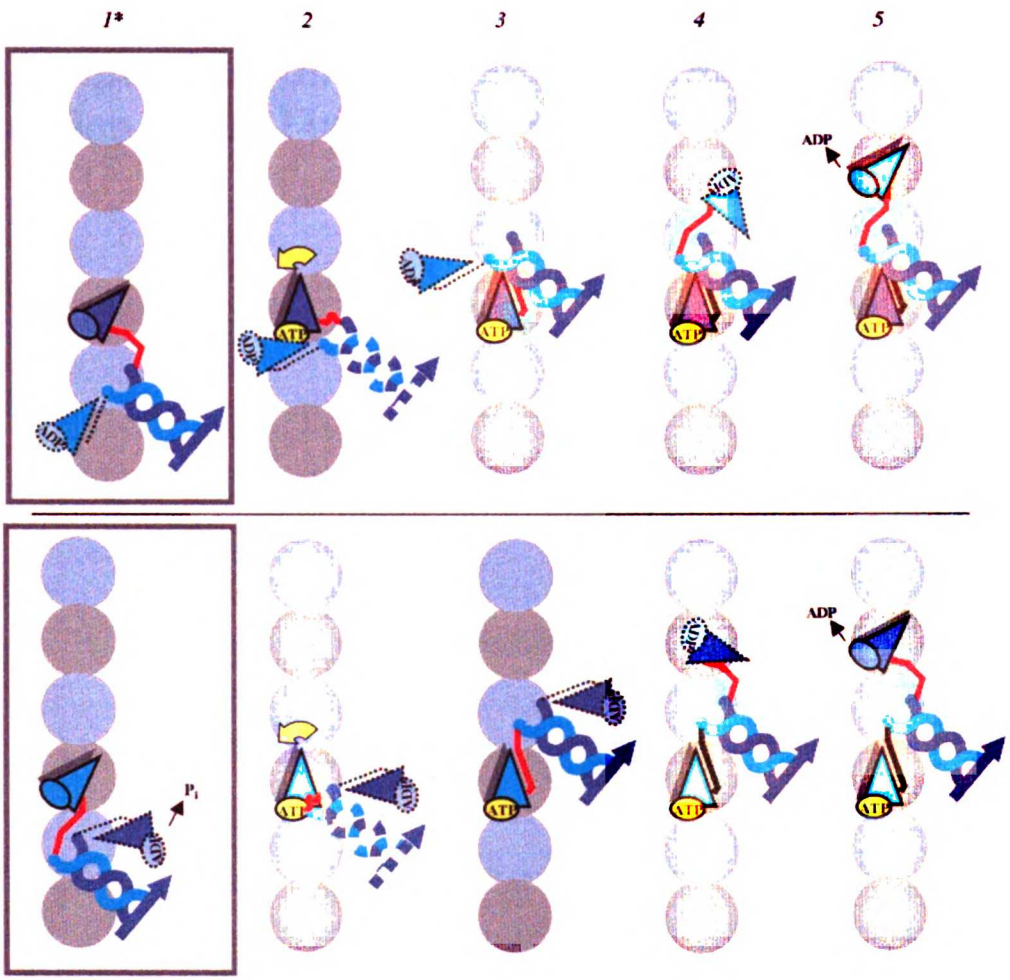



Figure 12. The "Ice-Climber" model for processive movement by the kinesin dimer.

Five discrete points in the chemomechanical cycle (numbered) are shown. In this model, the motor is powered by alternating left- and right-handed power strokes of the neck linker. The upper row in the figure depicts a powerstroke by the "right ax" (dark blue) while the lower row depicts one by the "left ax" (light blue). Step 1 (boxed) represents the static "waiting state" after the rear head has hydrolyzed its ATP, but before the forward head has bound a new ATP. The leading head's neck linker (red) can be either in a "right" (top) or a "left" (bottom) microtubule-bound conformation. We hypothesize that the coiled coil is also associated with the microtubule in this motor state. In step 2, ATP binding in the forward head unbinds this head's neck linker from its rearward-bound position, possibly leading to a substep. A rotation of the catalytic core (yellow arrow) is hypothesized to trigger this unbinding. In step 3, the neck linker (red) of the ATP-bound head docks forward on the catalytic core, triggering the final translation of the coiled coil in this enzymatic cycle. Step 4 is when the neck linker (red) of the free head finds its new microtubule binding site, directing the free head towards its next site. In step 5, the free head binds to tubulin and releases its ADP, which is followed by hydrolysis in the ATP-bound head (now the "rear" head), which leads back to a new hydrolysis cycle with the other head.

# Chapter Three:
# A Computable Analytic Expression for the Partition Function of the First Solvation Shell

## Summary

Random sampling methods, i.e. Monte Carlo or molecular dynamics simulations, are commonly used to explore configurations of aqueous solvent in computer modeling studies of biomolecules. Incompleteness of the sampling typically prevents such methods from yielding solvation free energies except in the case of very small solutes. Here we introduce an analytical formula called SWYZLE that *exhaustively* treats water configurations in the localized region surrounding the solute, thereby generating the partition function of this subvolume. We applied SWYZLE to a series of solvation test cases, employing a simple water energy function (the Mercedez-Benz, or MB model) used to model the hydrophobic effect. In these tests, water structure in the first solvation shell was replicated with high accuracy, indicating that SWYZLE provides a feasible alternative to the conventional random sampling methods. The subvolume partition function calculated by SWYZLE can be related to large-system thermodynamics, which may allow solvation free energies to be calculated in a new way.

## Introduction

Computing the solvation energy remains one of the outstanding problems in the modeling of structure and function of large biomolecules. Solvation energy is the principal force driving protein folding. All molecule-molecule interactions are also driven in large part by solvation energy, because when molecules associate they desolvate a portion of their surface, to make the interaction interface.

Conventional methods for computing the solvation energy of biomolecules are very limited. Methods for calculating the solvation energy fall into two classes: explicit solvent and implicit solvent (see, for example (Roux and Simonson 1999) and (Levy and Gallicchio 1998)). Explicit-solvent models are required for the most accurate calculations of solvation energy. Accuracy in them is ultimately determined by the quality of the molecular energy function. Available computer power, however, fundamentally restricts explicit-solvent methods to calculating solvation free energies of molecules with the size order of drugs, with perhaps tens of atoms (Levy and Gallicchio 1998). Since convergence problems increase rapidly with system size, the explicit solvent methods are

not likely to predict solvation energies of macromolecules like proteins (thousands or millions of atoms) in the foreseeable future.

Implicit solvent models, on the other hand, can typically be applied to very large systems like proteins, but suffer from accuracy problems. Not only do the implicit solvent methods fail to reach quantitative agreement with the explicit-solvent methods in many (if not most) cases, but also in general they offer no obvious methodological pathway towards greater accuracy (Southall, Dill et al. 2002). Implicit solvent models rely on approximations whose physical basis is not well understood, for example that solvation free energy is proportional to hydrophobic surface area, or that water behaves as a dielectric continuum at protein size scales (Sitkoff, Sharp et al. 1994).

The meaning of the "standard" free energy of solvation (Ben-Naim and Marcus 1984), $\mu_0$, is illustrated in Figure 1. $\mu_0$ is the log ratio of two partition functions. The first partition function is for a system with large temperature T, pressure P, containing N waters (for an isobaric system); the second is for the same N waters, with the same pressure P, but with an added single solute molecule fixed at a position $R_0$. The partition functions in Figure 1 cannot be calculated directly, because the solvent systems have far too many degrees of freedom. Therefore, calculating the partition function of even a small subvolume is all but intractable. Instead, explicit-solvent free energy calculations typically reduce the complexity of the calculation by examining difference properties between highly related ensembles, as in the commonly used free energy perturbation technique. In this way, free-energy perturbation can calculate the absolute free energy of a given solute by "growing" it from zero size. This approach only works, however, for exceptionally small solutes. Also, free energy perturbation offers little physical insight into the roles of the solvent entropy and enthalpy in governing the solute transfer process(Mark and van Gunsteren 1994; Boresch and Karplus 1995; Brady and Sharp 1995).

The purpose of this work is to develop a partition function-based approach to calculating the solvation energy. We propose to *directly* calculate the partition function of a solvated system in the critical volume near the solute, where the structural and energetic changes

in the solvent occur. It makes sense that the partition function ratio illustrated in Figure 1 should derive the solvation energy not only for macroscopic values of $N_{water}$, but for smaller systems as well. This is because a solute, at least in the (N, P, T) ensemble, is thought to perturb the structure and energetics of the solvent in only within a localized region surrounding the solute. Supporting data include the fact that the enthalpy (or "binding energy") of solvation can be derived from the first or second solvation shells only (Gallicchio, Kubo et al. 1998; Matubayasi, Gallicchio et al. 1998). Here we describe work towards retrieving the solvation energy from the partition function of the first solvation shell. While outer shells also contribute to solvation energetics, this preliminary work is designed to test the validity of the approach, and see if it can lead to a more basic understanding of solvation effects.

There is currently no available method for calculating the partition function for a system as complicated as a hydration shell of mobile solvent molecules. The problem is significant. For example, if six waters occupy a modestly sized hydration shell with as few as ~1000 available positions and orientations available to each, enumerating all water arrangements requires visiting a total of $(10^3)^6 = 10^{18}$ configurations. This level of computation, by straight enumeration at least, is not currently feasible. Nevertheless, here we present a new analytic method that explicitly treats $10^{18}$ (or more) configurations of a solvation shell, generating the partition function. The only approximation made in the method is that waters in the shell interact with their nearest neighbors only, which is suitable for the simple water model explored here as a test case. Longer-range interactions could eventually be incorporated into our method as well. Our new analytical method accurately reproduces fine structural details of solvation previously only available from monte carlo simulations or molecular dynamics calculations. More importantly, having a partition function of the solvent shell available for the first time may open up a new approach to calculating the solvation free energy, circumventing MC or MD methods entirely.

## Derivation

### *The Transfer Matrix Formalism*

For a linearly contacting chain of N molecules that only interact with their nearest neighbors, each molecule having a discrete number of conformations or states, the partition function can be calculated exactly by the transfer matrix method due to Kramers and Wannier. If the *nth* molecule is in the *i*th of $S_n$ possible states, and the (n+1)*st* molecule is in the *j*th of $S_{n+1}$ possible states , let their interaction energy be $E_n^{ij}$. The ***transfer matrix product*** is then $M_{tot} = \prod_{n=1}^{N-1} M_n$ where elements $m_n^{ij}$ (row i, column j) of each ($S_n$ by $S_{n+1}$) transfer matrix $M_n$ are defined as the partial Boltzmann weighting factor for molecules *n* and *n+1*: $m_n^{ij} = e^{-E_n^{ij}/k_B T}$.

Note that the interaction energy $E_n^{ij}$ can include not only the interaction between a shell water and its nearest neighbor, but also the shell water's environmental interactions--contacts with exterior molecules, for example. Any of the terms in $E_n^{ij}$ can depend on both *i* and *j*, although interactions of water *n* with any element of its environment besides the neighboring water *n+1* would generally be expected to depend only on its own state *i*.

The sum over all elements in the product matrix $M_{tot}$, $Z_N = \sum_{i,j} m_{tot}^{ij}$, is equal to the partition function of the system, as can readily be confirmed by expanding the matrix product. Furthermore, the terms in $M_{tot}$ are meaningful subdivisions of the partition function, as each element $m_{tot}^{ij}$ divided by $Z_N$ is the conditional probability that the system will be found with molecule 1 in state *i* and with molecule *N* in state *j*.

Figure 2 provides a simplified example of how the transfer matrix method could be applied to a model of the first solvation shell where all the waters are fixed in space but are allowed to rotate. The circular nature of the interaction in this case, however, means

that the transfer matrix multiplication must be extended by one, so that $M_{tot}^{circ} = \prod_{n=1}^{N} M_n$.

Here the elements $m_N^{ij}$ of the final transfer matrix $M_N$ are defined by the energy $E_N^{ij}$ of the *N*th molecule in its *i*th state interacting with the *first* molecule in its *j*th state. Only the diagonal elements $m_{tot}^{ii}$ of the resulting product matrix $M_{tot}^{circ}$ are meaningful in the product for the circular-interacting case, and are proportional to the probability that water 1 will be found in its state *i*. The partition function for N molecules is obtained by taking the trace of $M_{tot}^{circ}$:

$$Z_N = \sum_i m_{tot}^{ii}$$

This feature can be readily confirmed by expanding the matrix product, and the probability of water 1 being in state i is:

$$p_i^{n=1} = \frac{m_{tot}^{ii}}{Z_N}$$

An important feature of the mathematics involved is that permuting the matrix multiplication, as in:

$$M_{tot}^{circ} = \prod_{n=3}^{N} M_n \prod_{n=1}^{2} M_n$$

will give the same value for the partition function. The meaning of the diagonal elements, however, changes: in the above case, for example, $m_{tot}^{ii}$ represent the probability of the *third* water being in its state *i*. Therefore we will distinguish permutations of the transfer matrix product by the following notation:

$$M_{tot:m}^{circ} = \prod_{n=m}^{N} M_n \prod_{n=1}^{m} M_n$$

is the matrix product giving the state probabilities of water *m*, and $m_{tot:m}^{ij}$ are the elements of this matrix.

The preceding derivation is well established, and has been applied to innumerable cases where the geometry of the interacting particles (not always molecules) is fixed. Next,

however, we describe how to extend the transfer matrix machinery to calculate the partition function for a system of *mobile*, interacting particles.

***Incorporating positional mobility into the transfer matrix method***

From here on, we will frame our discussion within the context of a solvation shell model, although the method we present is general and could apply to any system with mobile particles where non-nearest-neighbor interactions could be neglected. While the discussion here will be restricted to the case of a two-dimensional system, we note (see Discussion) that this method can be generalized to the three dimensional case.

The states for each shell water molecule can be expanded to include an extent of grid points (with multiple orientations available for each grid point), as illustrated in Figure 3A. This means every water has the *same* number of states, $S_1, S_2, \ldots$ for a total of $N_{grid}N_{orient} \equiv S$, where $N_{grid}$ is the number of grid points, and $N_{orient}$ is the number of orientations per grid point. The transfer matrix calculation, as described above, applied to this expanded set of states will exhaustively treat *every* possible combination of positions for the N shell waters. However, completely enumerating water configurations like this leads to numerous pathological geometries like the one illustrated in Figure 3B, where water three is next to water one. In cases like this, the transfer matrix expression identifies nearest-neighbor interactions incorrectly (as indicated by the red arrows), and therefore assigns the wrong energy to the shell-water configuration.

In order to exclude incorrect geometries like in Figure 3B from the calculation, we therefore add a simple geometric restriction to the relative positioning of shell water molecules. Let $(r_m, \theta_m)$ be the polar coordinates of shell water *m* relative to the center of mass of the solute. Then, we restrict the coordinate $\theta_m$ of water *m* to be less than $\theta_{m+1}$ of the next consecutive water. This restriction excludes geometries like the one in Figure 3B, and can be incorporated into the transfer matrix formalism by imbuing the analytical machinery with two special features.

First, the available states $S$, in other words the gridpoints (with associated orientations), are given a special ordering, as illustrated in Figure 3A. Let $\vec{x}^{i}_{grid}$ be the polar coordinates $\left(r^{i}_{grid}, \vartheta^{i}_{grid}\right)$ of the $i$th grid point, relative the solute's center-of-mass. Each grid point is sorted by increasing value of its polar angle, $\vartheta^{i}_{grid}$. This sorting, in other words, causes the grid points to sweep around the solute shell in a counterclockwise direction when traversed in order. Second, the transfer matrix elements $m^{ij}_{m}$ are set to zero for all states $i >= j$, for all matrices $M_m$ where $1 <= m < N$ (Figure 3C), causing these matrices to become upper-triangular. Because of the sorted order of grid points $\vec{x}^{i}_{grid}$, this matrix feature forces waters 1-$N$ to traverse counterclockwise around the shell in order. In fact, because each water explores exactly the same states S, all matrices $M_m$ are identical for $1 <= m < N$.

The last transfer matrix, $M_N$, has a different restriction because it describes a wraparound interaction between water $N$ and water 1. Because water $N$ (index $i$ of the matrix) occurs *after* (that is, counterclockwise to) water 1, not before, $m^{ij}_{m}$ are set to zero for all states $i <= j$ rather than $i >= j$.

This completes our specification for incorporating mobile particles into the transfer matrix machinery. Below, this modified transfer-matrix method will be referred to as SWYZLE (Shell Waters Yield Z by Lattice Enumeration).

***SWYZLE and nearest-neighbor restrictions***

Although shell-waters are guaranteed to be ordered counterclockwise around the shell, some allowed geometries still have non-nearest-neighbor interactions (Figure 4). In the systems considered here, however, these geometries are heavily disfavored by van der Waals clashes. This is because the shell radius for a single solvation layer tends to be confining enough that these pathological geometries have unfavorable energies and do not contribute significantly to the partition function. We expect this approximation, therefore, will be valid unless either (1) the water becomes extremely dense or (2) the

shell radius considered by the grid extends significantly into the second solvation shell. Ultimately the accuracy of the approximation will depend on the form of the solvent energy function, so must be validated empirically. As we show below, the approximation proves to be an excellent one for the system, that is the MB water model, currently under study.

***Water Probability Distributions***

To obtain the overall shell-water density distribution for a given shell occupancy N, we start with the *a priori* probability $\frac{m^{ii}_{tot:m}}{Z_N}$, that a shell water *m* will be in state *i*, independent of all other shell waters. As described above, this probability is available by calculating the permuted transfer matrix product for *m*, $M^{circ}_{tot:m}$.

The set of probabilities for a given water *m* can equivalently be considered a density distribution over grid and orientation states *i*. Because the distributions for different *m* involve *different* water molecules, they can be summed to give the total density distribution for all waters. Thus, the water density $\rho(i)$ is:

$$\rho(i) = \frac{\sum_{m=1}^{N} m^{ii}_{tot\,m}}{Z_N}$$

Note that $\rho(i)$, as presented here, is normalized per *state* and not per volume. To normalize $\rho(i)$ per volume (="area" in the two-dimensional water model), a sum must be taken over all orientational states *i* that occupy the same grid point, $\vec{x}^{\,i}_{grid}$, and this sum then multiplied by the grid point density per volume.

***Handling fluctuations in the occupancy of the solvation shell***

Since in a real solvation shell of fixed radius, waters exchange in and out of bulk solvent, we extend the transfer matrix formalism to the grand canonical ensemble as follows. The shell waters transfer in and out of bulk according to a fixed chemical potential, $\mu_{bulk}$, that corresponds to the free energy of removing a water molecule from the bulk region

surrounding the explicitly represented solvent shell. The grand canonical partition function expresses the exchange equilibrium for these waters:

$$\Xi^{shell} = \sum_{N=1}^{\infty} Z_N e^{\mu_{bulk} N / k_B T}$$

where $Z_N$ is the partition function, derived above, for exactly N waters.

The probability of finding a shell occupancy of exactly *N* waters in the shell is:

$$p(N) = \frac{e^{\mu_{bulk} N / k_B T} Z_N}{\Xi^{shell}}$$

Applying the expression $\frac{m^{ii}_{tot:m}}{Z_N}$, derived above, for the shell-water probability density due to a *single* occupancy value, *N*, we obtain the following expression. The shell water density distribution (over states *i* describing the grid of positions and orientations), averaged over all possible occupancies, is:

$$\rho_{grand}(i) = \frac{\sum_{m=1}^{N} m^{ii}_{tot\,m} e^{\mu_{bulk} N / k_B T} Z_N}{\Xi^{shell}}$$

## Methods

### *A brief description of the water energy model tested with SWYZLE*

A simple, two-dimensional water model, the "MB" model designed to study hydrophobic solvation (Ben-Naim 1971; Silverstein, Haymet et al. 1998), was used as a test bed for the SWYZLE machinery. In the MB model, hydrogen bonds have orientation dependence but no polarity, so are just "sticky." The hydrogen-bonding energy is gaussian with respect to both water separation and relative angle of the bonding arms. A standard van der Waals term is also included, of the form $U_{LJ} = 4\varepsilon_{ij}\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right]$ where the well depth, $\varepsilon_{ij}$, is one tenth of $\varepsilon_{HB}$ for all species. An additive mixing rule was

applied for interactions between species of differing diameters, $\sigma = \frac{(\sigma_i + \sigma_j)}{2}$. The $\sigma_i$ for water was 0.7 of $r_{HB}$, and for solutes varied from 0.7 to 4.0 $r_{HB}$.

For simulations or calculations of a planar interface, the plane was defined at $y$=0. The water-surface interaction was defined for every water by placing a ghost solute particle ($\sigma_i$ =0.7 or 1.0) at $y$=0, with the $x$-coordinate assigned to that of the water (Southall and Dill 2000).

Temperature is defined in reduced units $T^*$, so that at $T^*$= 1, $k_BT^*=-E_{HB}$, the optimal H-bond energy. Similarly, the normalized pressure $P^*$ is defined by setting the PV work of expanding a volume by the amount $r_{HB}$ x $r_{HB}$(=1, in normalized units) against pressure $P^*$ to be $|\varepsilon_{HB}|$ (=1). In the simulations and calculations reported, $P^*$ was set to 0.19, and $T^*$ was set to values between 0.18 and 0.40. These conditions represent a range over which hydrophobic solvation in the model switches from entropically driven to enthalpically driven, in a "liquid-like" phase of the water.

***A simple approximation for the bulk interaction energy: $\mu_{effective}$, the "effective" chemical potential for bulk water***

The SWYZLE methodology allows for position- and orientation- independent interactions between shell waters and the bulk water interface. Here, however, the simplest form possible was taken for the bulk interaction term: it was assumed to be *independent* of both shell water position and orientation. The bulk interaction term was incorporated into the SWYZLE energy term as follows:

The shell-water interaction energy $E_n^{ij}$, between two shell waters $i$ and $j$, contained three terms:

$E_n^{ij} = E_{neighbor}^{ij} + E_{solute}^{i} + E_{bulk}$.

The first two terms in $E_n^{ij}$ are conformation-dependent interactions between a shell water and its explicitly represented contacts in the shell, i.e. nearest neighbor waters and the

solute. These were determined solely by geometry and the MB energy function, described above. The final term, a constant, conformation-independent energy, $E_{bulk}$, approximated a water's mean-field interaction with the "bulk" waters from outside the first solvation shell.

Because of its simple form, it is possible to factor out the bulk interaction term from the matrix expression for the partition function, as follows. The bulk term factors out of individual matrix components: $m_n^{ij} = e^{-\frac{E_n^i}{k_B T}} = e^{-\frac{E_{bulk}}{k_B T}} e^{-\left(E_{neighbor}^i + E_{solute}^i\right)/k_B T} = e^{-\frac{E_{bulk}}{k_B T}} \left(m_n^{ij}\right)'$. Here the elements $\left(m_n^{ij}\right)'$ describe a "*reduced*" transfer matrix $M_n'$, describing a system *without* a bulk interaction $E_{bulk}$. Subsequently each transfer matrix can be factored as: $M_n = e^{-\frac{E_{bulk}}{kT}} M_n'$, which allows the transfer matrix product to be expressed as

$$M_{tot}^{circ} = M_n = \prod_{n=1}^{N} e^{-\frac{E_{bulk}}{k_B T}} M_n' = e^{-\frac{NE_{bulk}}{k_B T}} \prod_{n=1}^{N} M_n' = e^{-\frac{NE_{bulk}}{k_B T}} M_{tot}^{circ'}$$

Now, since the elements $\left(m_{tot}^{ij}\right)'$ of $M_{tot}^{circ'}$ are just $m_{tot}^{ij} = e^{-\frac{NE_{bulk}}{k_B T}} \left(m_{tot}^{ij}\right)'$, we can rewrite

$$Z_N = \sum_i m_{tot}^{ii} = \sum_i e^{-\frac{NE_{bulk}}{k_B T}} \left(m_{tot}^{ii}\right)' = e^{-\frac{NE_{bulk}}{k_B T}} Z_N'$$

in other words, the $E_{bulk}$-dependent term in $Z_N$ can be factored out, leaving the "reduced" partition function $Z_N'$. This $Z_N'$ is nothing more than the partition function of the $Z_N$ shell system but with no bulk interaction.

When this expression for $Z_N$ is substituted into the grand canonical expression (allowing for fluctuations in N), the bulk interaction term can be conveniently grouped together with the bulk chemical potential:

$$\Xi = \sum_{N=1}^{\infty} Z_N e^{\frac{\mu_{bulk} N}{k_B T}} = \sum_{N=1}^{\infty} e^{-\frac{NE_{bulk}}{k_B T}} Z_N' e^{\frac{\mu_{bulk} N}{k_B T}}$$

$$= \sum_{N=1}^{\infty} Z_N' e^{\frac{N\left(\mu_{bulk} - E_{bulk}\right)}{k_B T}}$$

What this says is that, in this mean-field energy model for bulk water, the energy $E_{bulk}$ acts exactly like the bulk water chemical potential, and opposes it. Therefore, $E_{bulk}$ has no specific structural effect on the shell waters and only acts to modulate the shell occupancy. Subsequently, therefore, we will not refer directly to $\mu_{bulk}$, but instead to the sum $\mu_{bulk} - E_{bulk}$ which we will refer to as the "effective" bulk chemical potential, $\mu_{effective}$. This term will be assumed to have an implicit dependence on the bulk interaction energy of the model.

We note that in the case $E_{bulk} = 0$, the water shell energy model describes something like a mini-droplet suspended in a vacuum. The analogy is not exact, however, because the confining shell geometry and the bulk chemical potential combine to prevent waters from evaporating.

*Selecting a bulk-water chemical potential*

No expression for the bulk water chemical potential has yet been derived within the shell-water partition function machinery we have just described, so it must be determined empirically. In the calculations that we present here, the chemical potential $\mu_{effective}$ is selected so that the average shell occupancy, $<N>$, matches the integrated pair correlation function (integrated to the same shell radius) of MC simulations run with an identical solute. This is currently the best way of specifying the macroscopic system pressure (or, equivalently, density) in a SWYZLE calculation. Thus, MC simulations and SWYZLE calculations could directly compared, for systems with *identical* thermodynamic parameters (N, P, T in the isobaric system studied here). Note that the contribution of $E_{bulk}$ to $\mu_{effective}$ in the current scheme cannot be determined. Further elaboration of the solvation shell theory is required to properly represent the bulk interaction energy (see Discussion).

*Implementating the SWYZLE method*

A program was written in C to apply the SWYZLE machinery. For each number, $N$, of shell waters, the program computes the partition function and a list of probabilities for every considered shell-water position and orientation. A hexagonal grid was used to define the possible positions of water molecules. Grid points were omitted from the calculation if they closer than 90% of the van der Waals radius to the central solute. Grid spacings of between 0.1 and 0.15 of $r_{HB}$ were used, with 3 to 8 orientations considered per grid point.

The list of shell-water orientations considered at each grid position in the shell is constrained to include a reference orientation that points directly at the central solute. Other orientations at a given grid point were defined relative to this reference orientation. The three-fold symmetry of the water molecules increased the angular resolution, so $N$ orientations had a separation of $120°/N$, rather than $360°/N$ as would be required for an asymmetric molecule occupying the grid point.

A typical calculation, with an outer shell radius of 1.5 $r_{HB}$, grid spacing of 0.15 $r_{HB}$ and 7 orientations per grid point, considered 312 grid positions (for a total of 2184 states per water molecule). Such a calculation required 150MB of memory for matrix storage and took 24 minutes on a 466MHz DEC Alpha workstation. The largest system examined, with a shell radius of 2.9 $r_{HB}$, using a grid spacing of 0.13 $r_{HB}$ and 5 orientations per grid point, considered 1044 grid points, requiring 870MB for matrices and taking 9 hours on the same hardware.

In the SWYZLE calculations of the planar hydrophobic boundary, a rectangular grid extent was used, and a periodic boundary potential between shell waters was introduced in the $x$-direction. Box lengths were of between 2.0 and 5.0 $r_{HB}$ were explored, with shell radii of between 1.1 and 1.4

### *MC Simulations*

All Monte-Carlo simulations were carried out using the MC2D program (Silverstein, Haymet et al. 1998). The simulations were carried out with 60-120 MB water molecules at a "standard" pressure of 0.19 (normalized units, as above). Systems were equilibrated for 10-20 million Monte Carlo steps, and production runs for statistics were 80-300 million steps. Shell-water densities were obtained by fixing a solute (either hydrophobic or water) in the center of the periodic box, and collecting occupancy statistics over a hexagonal grid. Grid spacing (0.1 $r_{HB}$) in this case was identical to the grid used for the SWYZLE calculation.

## Results

The SWYZLE method was tested by its application to a simple, two-dimensional energy potential, the Mercedez-Benz (MB) model for hydrophobic solvation. A uniform mean-field term was used to describe the average interaction energy of the shell waters with the "bulk" waters beyond the first solvation shell (see Methods). Achieving the ultimate goal of predicting solvation free energies will almost certainly require a more elaborate treatment of the bulk solvent (beyond the first shell) than employed here. These results, however, provide a test for the basic framework of the SWYZLE formalism, and lay the groundwork for more realistic treatments of the solvation shell.

The basic mechanics of the SWYZLE calculation are illustrated in Figure 5, for a fixed value of $N$=6 shell waters, around a hydrophobic solute. Note that there is a decided lack of radial or other symmetry in individual water densities. In fact, the apportionment of density amongst waters 1, 2, 3, etc. is determined by where numbering of the gridpoints begins (see Fig. 2), but there is no physical relevance to the individual water densities. When the densities are added together, however, the total density of the water shell appears, and regains the radial symmetry of the solute.

For the calculations that follow, the grand canonical ensemble visits not just one value of $N$ (as in Figure 5), but a range that extends from zero to significantly higher (maximum value of 10-13) than the average occupancy, $\langle N \rangle$. For solutes of the size order of a water molecule, $\langle N \rangle$ was no larger than 7 under the temperature and pressure conditions explored. The average shell occupancy in these calculations was set (using the bulk water chemical potential term, $\mu_{effective}$—see Methods) to match the occupancy of the same-sized shell in corresponding large-scale MC simulations of the identical solute. The shell radius (1.5 $r_{HB}$ in both cases) was selected to match the minimum in the radial distribution function of the hydrophobic solute. In shell energy model presently examined, the shell radius is the only free adjustable parameter, as the bulk interaction energy of the shell waters is incorporated into the chemical potential $\mu_{effective}$, which merely specifies the water density of the overall system. The two quantities $\mu_{effective}$ and the shell radius were

the only parameters input into the SWYZLE machinery. Other details of the calculation were dictated solely by the MB energy model (see methods) and by matrix mechanics—although two parameters, the number of orientations per grid point and grid spacing, were adjusted to optimize the precision of the calculation.

First-shell water density distributions calculated by SWYZLE for two types of solute are shown in Figure 6, compared with the distributions produced by Monte-Carlo simulations. As shown, SWYZLE correctly captures the essential features of solvent structure seen around either a solvated water (Figure 6A) or a hydrophobic solute (Figure 6B).

Histograms of shell-water occupancies, predicted by the grand canonical ensemble as applied in SWYZLE, is shown in Figure 7, for both a hydrophobic solute and a solvated water. These graphs span the range of temperatures previously explored for the MB model in studies of hydrophobic solvation. Good quantitative agreement is seen in the histograms (particularly for the hydrophobic solute) for the lowest temperature, T=0.18, and for the two higher temperatures the agreement is excellent for both solute types investigated. At the highest temperature, T=0.40, the shell occupancy histograms approach the Poisson distribution, but the agreement is not exact (Figure 7C). The occupancy histograms of the hydrophobic solute and of solvated water have different half-width at $T$=0.18, but converge at the highest temperatures. The convergence is not entirely evident in Figure 7C because the distributions have different average values of $N$ (3.06 or 2.76 for hydrophobic solute or water, respectively). However, if the water density of the high-temperature solvated-water SWYZLE calculation is adjusted to match average occupancies with the hydrophobic solute calculation (by altering the bulk water chemical potential, $\mu_{effective}$), the resulting histograms match nearly exactly. This agreement suggests that at higher temperatures the shell occupancy distribution is independent of solute type.

The solvent structure predicted by SWYZLE was examined further by computing shell-water orientation distributions as shown in Figure 8. The predicted orientation

distribution by SWYZLE shows excellent agreement with the MC simulation. The orientation distribution of shell waters converged when at least three to five angles per grid point were used, as shown in Figure 8.

Varying the SWYZLE calculation's sole adjustable parameter, the shell radius, significantly influences the predicted water shell structure (Figure 9). The shell-water orientation distribution surrounding the hydrophobic solute flattens somewhat with increasing shell radius. This effect depends on temperature, however, and disappears nearly completely at the highest temperature examined, where the orientation distribution is independent of the shell radius in the SWYZLE calculation. Furthermore, the orientation distribution converges to the one seen in MC simulations at these high temperatures. For the solvated water, the effect of shell radius is more significant. Indeed, for shell radii sufficiently below the optimum value, the shell waters adopt a new distribution around a central water that is not even qualitatively correct. At high temperatures, however, the shell-water orientation distribution becomes independent of the shell radius choice and converges on that found in MC simulations, just as for the hydrophobic solute.

For all solutes tested, including the "infinitely"-sized hydrophobic solute represented by the planar interface, an appropriate shell radius could be found that *quantitatively* reproduces the orientation distribution, as shown in Figure 9. Orientation distributions for the optimum choice of shell radius (1.5 $r_{HB}$ for a hydrophobic solute, 1.6 $r_{HB}$ for a solvated water, and 1.3 $r_{HB}$ for a planar hydrophobic interface) are presented in a reduced form in Figure 10. In this figure, shell-water populations are grouped according to whether a hydrogen bonding arm points at the solute surface or not.

The effects of temperature and solute size on the orientation distribution for a hydrophobic solute are shown in Figure 11. Increasing the solute size increases the relative fraction of shell waters that "waste" a hydrogen bond by directing an hydrogen bond arm directly at the solute. In the planar case, more waters adopt this "radial" orientation than they do the alternative, "straddling" orientation. Increasing temperature

tends to neutralize orientational preference, indicated by the convergence of radial and straddling orientation populations in Figure 11 at high $T$. The SWYZLE calculations capture these trends semi-quantitatively, as can be seen in the Figure.

Overall features of the pair correlation functions, such as position and relative heights of the first maximum, are in good agreement with the Monte Carlo simulation (Figure 12). Some discrepancies can be seen, however. For both solutes the distribution is flattened in the SWYZLE calculation relative to the MC simulation, and the water density is somewhat exaggerated at the outer extent of the water shell. The flattening effect is more pronounced in the case of the hydrophobic solute, while the outer density is inflated (relative to the MC simulation) by a similar ratio in either case. Pair correlation functions predicted by SWYZLE are well converged at all grid spacings tested.

Like the water orientation distributions, water-solute pair correlation functions were found to depend on the choice of shell radius in the SWYZLE calculation (Figure 13). Increasing the shell radius, as shown in the right-hand graphs, tends to sharpen the pair correlation peaks from the SWYZLE calculations, but at the same time further exaggerates the density seen at the outer shell extent. While the peak positions in the pair correlation functions were consistently accurate (relative to MC simulation), no choice of SWYZLE radius allows the MC pair correlation function to be captured with full quantitative accuracy. Furthermore, while agreement between MC simulation and SWYZLE improved at higher temperatures, a small but persistent difference remains in the pair correlation functions (as can be seen in the Figure). SWYZLE calculations consistently flattened the pair correlation function for all solutes tested, including the "infinite-radius" planar boundary, and they also overestimated solvent densities near the shell outer radius.

The effects of temperature and solute size on the pair correlation function for a hydrophobic solute are shown in Figures 14-15. Increasing either solute size or temperature flattens the pair correlation functions in MC simulations, where it is seen that increasing $\sigma$ from 0.7 to 2.0 (at the lower temperature, $T = 0.18$) decreased the first peak

height of the pair correlation function from ~2.5 to ~1.5. Increasing the temperature from 0.18 to 0.28 decreases the peak height by a similar amount in the MC simulations. These temperature- and size- dependent trends are captured, although only qualitatively, by the SWYZLE calculations. Increasing $\sigma$ from 0.7 to 2.0 (at $T = 0.18$) decreased the height of the first peak height from 1.6 to 1.5 in the SWYZLE-predicted pair correlation function. Increasing temperature from 0.18 to 0.28 decreased the peak height from 1.6 to 1.4 for the $\sigma = 0.7$ solute. SWYZLE's tendency to flatten the pair correlation function, relative to the MC simulations, extends throughout the range of solute sizes tested.

## Discussion

The analytical machinery we have introduced, SWYZLE, is capable of representing many fine details of a solute-solvent system, previously available only through Monte-Carlo or molecular dynamics simulations. In contrast to these simulation techniques, the matrix multiplication used here is an *exhaustive* calculation—sampling is carried out completely over a predefined configuration space. Our ability to accurately calculate the partition function itself means that any number of structural and energetic quantities (see the last part of the Derivation section) can be obtained in addition to the water density distribution. In our initial assessment of the SWYZLE machinery, however, we have focused primarily on water density only, including the related radial and orientational distribution functions of shell waters. These most basic structural readouts were used to assay SWYZLE's ability to reproduce the solvation characteristics of a very simple water energy model, the two-dimensional MB energy function.

***The analytic method reproduces solvation features found in Monte Carlo simulations***

Figures 6-15 show that our analytical calculations captured the essential features of the first solvation shell, as modeled by simple solutes of two types, immersed in water represented by MB disks. As will be discussed below, the accuracy of the SWYZLE calculations was hindered by a fairly simplistic treatment of the first solvation shell's boundary to the outer, "bulk" solvent (notably, this is not an inherent limitation of the analytical machinery). Despite these "boundary issues," however, the analytical calculations made solvent structure predictions that in many cases quantitatively agreed with full-scale Monte-Carlo simulations that explicitly represented many solvent layers.

The occupancy distributions of Figure 7 show good-to-excellent agreement between the SWYZLE calculations and MC simulation data. At the lowest temperature, while the agreement was only semi-quantitative, the SWYZLE calculation clearly captures the hydrophobic solute's tendency to group exactly six waters in the first solvation shell. The solvated water, by contrast, has a significantly broader distribution—a feature also captured by SWYZLE, although the analytical calculations somewhat exaggerate the peak-broadening effect. The agreement between analytical and MC calculations improves strikingly with increasing temperature, so that at the highest temperature examined the SWYZLE results are seen to superimpose exactly with the simulation data. These results clearly validate the grand canonical treatment of shell water occupancy as implemented in SWYZLE.

The differing water structure around a hydrophobic solute, as compared to solvated water, is well reflected in the shell water orientation distributions of Figures 9-11. Essentially, waters surrounding solvated *water* tend to point their bonding arms inward to make strong interactions with the "solute," while those around a hydrophobic solute form "clathrate cages," inverting their orientation to the solute surface on average (relative to solvated water). The precise nature of the SWYZLE prediction is seen to depend, in Figure 9, on the outer radius of the solvation shell used in the calculations—this choice is related to the "boundary issue" mentioned above. Nevertheless, for the correct radius

choice, it is clear that the SWYZLE calculations are capable of quantitatively reproducing the water orientation distributions of MC simulations. Interestingly, at the highest temperatures examined the orientation distributions predicted by SWYZLE become *independent* of solvent shell radius, and converge on the simulation result regardless of this parameter.

The pair correlation functions between solutes and shell waters, as predicted by SWYZLE, also capture some qualitative differences of shell water structure that depend on solute type. As seen in Figures 12-15, when the central solute is a water, the first pair correlation peak is quite sharply centered at the hydrogen-bonding distance, and is accompanied by a smaller "shoulder" indicating some waters that approach to within van der Waals contact distance. With a hydrophobic solute, by contrast, the pair correlation peak is centered at the van der Waals contact distance, and is significantly less sharp. Also, increasing the temperature tends to flatten the pair correlation function with all solute sizes; and increasing the size of a hydrophobic solute tends to *further* flatten the pair correlation function. All of these trends, as found in MC simulations, are also seen in the analogous SWYZLE calculations.

***Discrepancies between simulated correlation functions and those predicted by SWYZLE***

Unlike other structural measures discussed so far, however, the pair correlation functions predicted by SWYZLE fail to entirely match the shape predicted by the MC simulations, even at the highest temperatures tested. In particular, SWYZLE predicts pair correlation values at the outer shell boundary that are consistently higher than found in the simulations, and tends to flatten the first pair correlation peak of the hydrophobic solutes (this latter effect is evident even in Figure 6). Also, as can be seen in Figure 13, the overall shape of the first pair correlation peak, as well as the position and depth of the minimum between first and second peaks, depends significantly on the placement of SWYZLE's "bulk water boundary" (i.e. the outer shell radius). Indeed, *no* choice of this parameter quantitatively reproduces the simulation-determined pair correlation character. The error reduces with increasing temperature, but even at the highest temperature tested

the pair-correlation discrepancy between SWYZLE and MC simulation (at the outer shell boundary of the analytical calculations) is ~10% for both solutes tested (Figure 14). In light of the excellent high-temperature agreement between SWYZLE calculations and MC simulations reflected by other structural monitors, this disagreement is somewhat surprising. We now turn to a discussion of the modeling assumptions that may have led to this discrepancy.

***Controlling for shortcomings in the analytical machinery***

As discussed in the derivation section, a potential inaccuracy could arise in the SWYZLE calculations if important configurations of the first solvation shell happened to include non-nearest neighbor interactions (see Figure 4). This possibility has not yet been directly ruled out, although we note that it is quite straightforward to do so. MC simulations can be run of the identical, first-shell-only system that is modeled by the analytical calculations. Such simulations (to be carried out in the near future) will provide an exact control for the structure predictions of SWYZLE.

In the absence of this exact control, it nevertheless seems unlikely that non-nearest-neighbor interactions could explain the discrepancies between MC-simulated and SWYZLE-predicted pair correlation functions. The reason is that the shell-water densities, under the conditions currently explored, are too low. Even at the lowest temperature (and highest solvent density) examined, the calculations are dominated by first-shell occupancies of 6-8 waters (see Figure 7). Such a density is sufficient to form bonded water rings around the solute, but *not* crowded enough to produce "pathological" non-nearest-neighbor interactions (as in Figure 4B) by packing forces alone. Furthermore, at the higher temperatures examined, the water shell occupancy is significantly less (dropping to 2-4 waters at the highest temperature examined, T=0.40), while pair-correlation discrepancies persist. Non-nearest neighbor interactions in the first solvation shell seem *very* unlikely at these low densities, so we expect the SWYZLE machinery to give highly accurate predictions of the first-shell water-water and water-solute interactions. Therefore, "bulk boundary" issues are much more likely to explain

the pair-correlation differences between SWYZLE calculations and MC simulations, as discussed below.

***Improving the model of a "bulk boundary" at the surface of the first solvation shell***

Our initial choice of a boundary potential between shell waters and the outer, "bulk" solvent was motivated by simplicity. In fact, the "boundary potential" in the SWYZLE calculations was essentially a hard wall, so shell waters could freely to migrate right to the outer edge of the solvation shell volume without any change in the bulk interaction energy, $E_{bulk}$ (see Methods section). This is probably not the best way to handle a surrounding bath of "bulk" waters. A more realistic treatment, for example, would be to integrate over degrees of freedom for a large number of bulk waters, in order to generate an "average" interaction potential between bulk and inner-shell waters. The boundary potential generated by this approach would include average contributions from van der Waals interactions, as well as electrostatics and hydrogen bonds, from bulk water molecules. Indeed, Roux and others have derived boundary potentials in this way (see (Im, Berneche et al. 2001)), in order to perform efficient molecular dynamics simulations of solutes covered by only limited amounts of explicit solvent—very similar to the first-shell situation modeled by the SWYZLE calculations presented here. These workers have argued that a *harmonic* restraint is an appropriate functional form to approximate the bulk boundary potential, at least in the case of constant-pressure simulations.

A harmonic boundary restraint at the outer edge of the solvation shell differs significantly from the hard boundary currently applied in the SWYZLE calculations described here. The effect of an altered bulk boundary would be especially significant on shell water structure near the boundary. It will therefore be interesting to see whether harmonic (or functionally related) boundary potentials, applied to the SWYZLE model, reduce the discrepancy (discussed above) between analytically predicted pair correlation functions and those from MC calculations. This question will be addressed in future calculations.

***Towards solvation free energy prediction, using SWYZLE***

The primary advantage of SWYZLE over simulation methods is that it derives an accurate value for the partition function of the solvation shell subsystem—a *grand canonical* partition function in this case, because waters in the shell are allowed to freely exchange with the bulk. We have proceeded with preliminary efforts to directly calculate free energies of solute insertion, based on SWYZLE calculations of single solvation shells. Briefly, our rationale for deriving the solvation free energy from single-shell partition functions is as follows. The free energy of solvation (for the constant-pressure case) is defined by the partition function ratio of two very large systems, as in Figure 1. However, it has been shown (for the constant-pressure case) that the solvation energy is dominated by local perturbations to the solvent structure and energetics near the solute. We therefore propose that "bulk" waters (the ones not near a solute) in the two systems of Figure 1 should make large, and *equivalent*, contributions to their respective partition functions. If this hypothesis is correct, then it follows that, after making the appropriate cancellations in the partition function ratio of Figure 1, the resulting expression should be reduced to a partition function term involving only *localized* interactions of waters that are near to the solute.

Following these ideas, an expression has been derived (Appendix 3) that relates the solvation shell partition function (derived by SWYZLE) to the free energy, $\overline{\mu}_0$, of solute insertion. The result is straightforward, and involves no additional parameters beyond what was input into the SWYZLE calculation:

$$\overline{\mu}_0 = pv - kT\ln(\Xi)$$

Essentially, this equation describes a two-step solvation process, diagrammed schematically in Figure 1 of Appendix 3. First, a cavity of size $v$ is opened up in bulk solvent, where $v$ is the volume of the ***entire first solvation shell*** (used by SWYZLE). This first step corresponds to the unfavorable work ($+pv$) of pushing the bulk waters out of this cavity. It is important to note that this large cavity (which includes both solute and a shell of surrounding waters) is *different* than the cavity employed in the Scaled Particle Theory of solvation, whose size is simply equal to the "volume" of the solute itself. Indeed, our ability to explicitly treat the first-shell waters, whose structure is strongly perturbed relative to "bulk" water, is a key advantage of our new approach over the older

methods like Scaled Particle Theory. In the final step of our solvation cycle, the solute is placed at a fixed location in the center of the cavity, and bulk waters are allowed back in according to the grand-canonical equilibrium expression, $\Xi$, produced by SWYZLE. The second step leads to a favorable contribution of $-kT\ln(\Xi)$ to the solvation energy.

Preliminary work (not shown) indicates that the above expression does not yet yield accurate solvation energies, at least with the current incarnation of the water shell model that has been applied to SWYZLE. One likely contributor to this inaccuracy is the unrealistic treatment of shell waters' interaction with the bulk medium (see the previous section). Investigations to improve SWYZLE's solvation free energy prediction capabilities are ongoing.

***Precision and speed of the SWYZLE calculation***

The results presented above demonstrate that our new matrix technique is capable of well-converged calculations, and captures the salient features of shell water structure as defined by the test water energy function that was used here. For smaller solute sizes, the resources required were quite reasonable. As was shown in Figures 8 and 12, it was possible to use only 5 orientations per grid point, and a relatively coarse grid spacing equal to 15% of the hydrogen-bonding length, and still obtain excellent convergence in the calculation. For a solute with the van der Waals radius of water, this coarseness level translated to a 312-grid-point representation of the first solvation shell—for a resulting calculation that used ~80MB of computer memory and took 9 minutes on our 450MHz DEC Alpha workstations.

The SWYZLE calculations grew more costly as the solute size increased. A rough estimate is that the number of grid points required goes up linearly with the solute radius, assuming the thickness of the first solvation shell stays approximately the same as the solute grows. This means that memory requirements will increase as approximately as the *square* of the solute radius, because matrix size in the method is proportional to the number of grid points squared. Calculation time is penalized even more greatly with

larger solute size, because time required for matrix multiplications grows as the matrix dimension *cubed.* With the current implementation of the SWYZLE method, therefore, a practical limit was reached for solutes with ~6 times the van der Waals radius of water—these calculations took ~12 hours of DEC Alpha time and required ~800MB of memory for matrix storage.

***Potential Optimizations***

Optimizations are possible that could reduce the computational requirements of SWYZLE and allow solutes of much larger size to be analyzed. One of the most dramatic possibilities is suggested by the distribution of individual water densities shown in Figure 5. It is obvious from the figure that none of these *individual* waters explores significantly more than a third of the solvation shell's total volume (="area" in 2D). However, in the current calculation, each successive water in the solvation shell is allowed to wander over the *entire* grid. Instead, individual waters in the shell calculation could be restricted to only the sectors they had a significant probability of occupying, which would significantly reduce the number of grid points needed per water. Thus, in Figure 5 for example, the first water could be restricted to the grid points in the sector from 0° to ~45° (origin defined at the solute center, with angle defined as in Figure 3) with very little loss of density. Similarly, the second water would explore the sector from ~30° to ~135°, and so on (note that these sectors will have significant overlap).

Restricting the shell water positions to sectors like this would break the symmetry of the matrix calculation, because each successive water in the shell would no longer be represented by the same matrix as the last ("$M_m$," see the Derivation section). Thus, at first glance this "optimization" would seem to *increase* memory requirements of the computer algorithm, because each water in the shell would now need its own matrix storage space, as opposed to the mere *two* matrices ($M_m$, and $M_N$) needed in the current implementation. However, the new matrices would be much smaller than the original $M_m$ and $M_N$. For example, if the sector optimization reduced the number of grid points per water by a factor of ~3, on average (which seems reasonable, looking at Figure 3),

matrix size per water would be reduced by a factor of $3^2$=9. Thus, in this example of six shell waters, the total space requirement for the six "sector" matrices would still be significantly smaller than the currently used matrix pair of the SWYZLE algorithm. Furthermore, only *one pair* of the sector matrices would have to be present in computer memory at any one time—leading to a many-fold reduction in the algorithm's memory requirements.

The greatest performance gain of the sector method, however, would be in computation speed. In our hypothetical example above, a three-fold reduction in grid points per shell water would lead to a $3^3$=27-fold decrease in time spent multiplying matrices. The performance and memory gains made possible by using the sector method will undoubtedly allow the SWYZLE algorithm to analyze solutes several times larger than the current size limit. It is important to note that this sector optimization, as described, does not make any assumptions about symmetry of the solute, etc. Optimizations that incorporated symmetry of the solute would be likely to speed up the calculation far more.

Another optimization may allow significantly more coarse-grained grids to be used in the SWYZLE calculations. Currently, nothing is done to account for positions and orientations of waters *between* the explicitly described geometries defined by the grid. This means that coarse-grained grids will frequently "miss" the shell waters' optimal geometries – for example, if two waters are optimally positioned to make a hydrogen bond, but the explicitly considered orientations don't permit the bonding arms to point directly at each other. The way to fix this problem, without adding more grid positions and orientations, is to average the interaction energies over intermediate locations and orientations *not* explicitly considered by the grid. For example, if a water was located at (x,y) with orientation $\theta$, with a grid spacing dx and orientational granularity $d\theta$, its interactions with other molecules would be averaged for all the water's possible positions between (x-dx/2) and (x + dx/2), (y-dy/2) and (y + dy/2), and orientations between ($\theta$ – d $\theta$ /2) and (q + d $\theta$ /2). It is quite likely that matrix size in the SWYZLE calculation could be reduced by another factor of 4 (or more) by employing a grid-averaging scheme like this, reducing calculation time by another order of magnitude.

Even without optimization, the SWYZLE method is quite efficient at predicting water structure around small solutes. To handle the water-sized solute, large-scale Monte-Carlo simulations need at least 1-2 orders of magnitude more computer time than a SWYZLE calculation, in order to reach convergence of hydration shell structure (this, of course, depends on the standard for "convergence"). The SWYZLE calculation, furthermore, provides a quantity that Monte-Carlo simulations cannot calculate—a value for the partition function of the localized solvation shell. As has been discussed, this additional information may allow direct estimation of solute insertion free energies, without recourse to free energy perturbation simulations.

## Conclusions and Future Directions

We have presented a new approach towards calculating solvation free energies. Our analytical formula, SWYZLE, efficiently and accurately replaces molecular dynamics and monte carlo simulation methods as a predictor of solvation shell structure, in the idealized MB water model used as a "test bed" for studying hydrophobic solvation issues. Even in the simplified context of the MB model, the SWYZLE formalism promises to be a useful tool. If SWYZLE can be developed to the point where it predicts solvation free energies, it will almost certainly be capable of handling large, complex solutes—for which conventional free energy techniques become intractable even in the reduced framework of the MB model. Applied to the MB model, SWYZLE could explore these untested solute regimes and correlate the effects of solute size and shape to the solvation energy. Research along these lines might begin to address the longstanding controversy of how hydrophobic solvation energy scales with a solute's surface area, volume, and shape— and may suggest alternatives to the commonly used "surface area" approximation for calculating hydrophobic contributions to solvation energies.

The SWYZLE methodology is not limited to two-dimensional model systems, or to energy models (like MB model studied here) where nearest-neighbor interactions predominate in solvation effects. By expanding the transfer matrices we have described

into tensors, it is possible to represent three-dimensional systems and to include interactions beyond nearest-neighbor ones. Long-range electrostatic interactions could even be handled separately, in a mean-field way. With suitable computational optimizations, therefore, it is quite conceivable that SWYZLE (or a related method) could eventually be applied to the chemically more realistic water models (for example, TIP3P) used in biomolecular simulations. This could prove to be an invaluable aid in the study of solvation free energies for complex systems.

***Abbreviations***

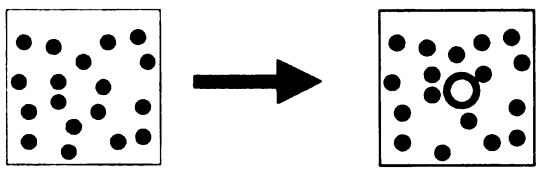
MB, Mercedez-Benz two-dimensional water energy function

MC, Monte Carlo

MD, molecular dynamics

PCF, Pair Correlation Function

SWYZLE, Shell Waters Yield Z by Lattice Enumeration



$(N_{water}, P, T)$ **Fixed solute in** $(N_{water}, P, T)$

$$\overline{\mu}^{0}_{solute} = -kT \ln\left(\frac{Q(N_{water}, P, T; solute @ R_0)}{Q(N_{water}, N_{solute} = 0, P, T)}\right)$$

Figure 1. The definition of solvation free energy using Ben-Naim's "standard state," for the fixed-pressure (isobaric) ensemble (Ben-Naim and Marcus 1984). A large volume of pure solvent (left) is perturbed by the addition of a fixed solute (right); taking (-kT) times the log ratio of these two partition functions yields the free energy of solvation.

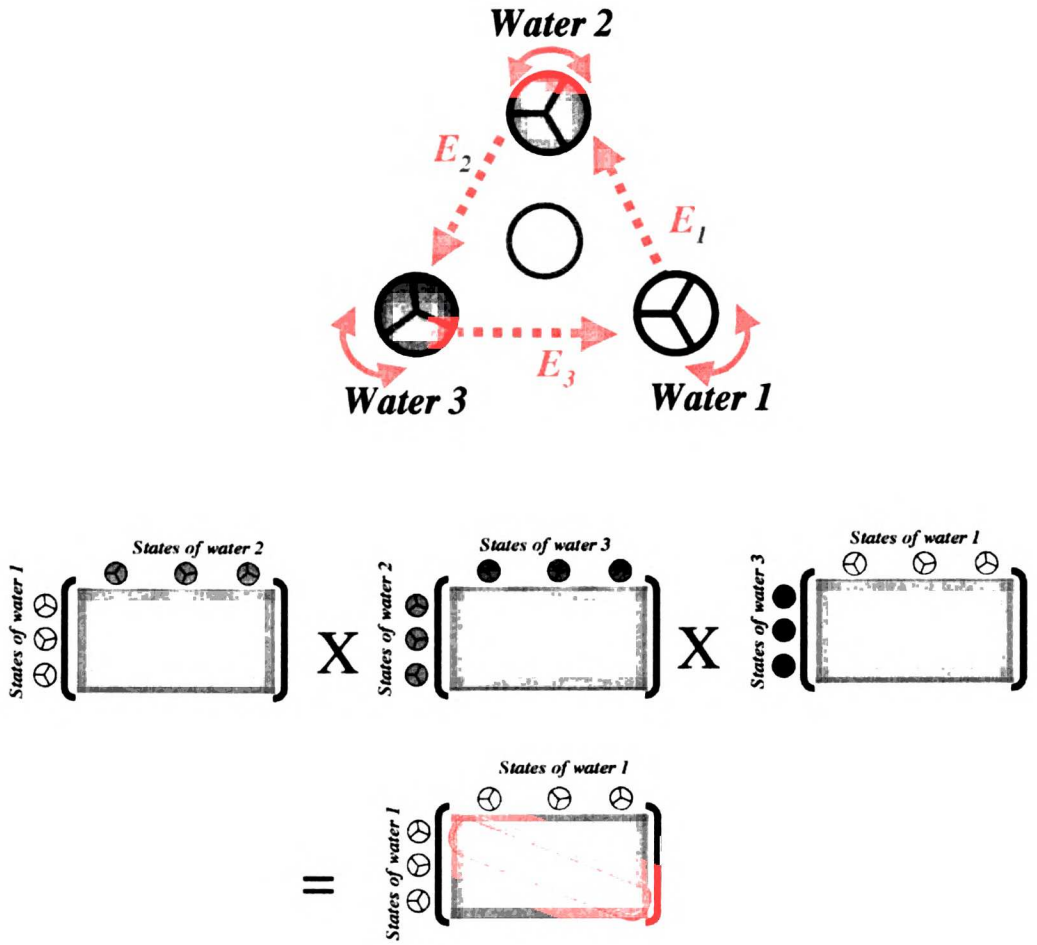



Figure 2. Illustrating the transfer matrix method applied to a very crude model of the first solvation shell. Three water molecules, at fixed positions in space, surround a central solute and each water can explore three orientations. The partition function of the system can be described by the product of three 3x3 transfer matrices (illustrated as green squares). As described in the text, diagonal elements of the final product matrix $m_{tot}^{ii}$ (indicated by the red box) are proportional to the probability that water A will be in state i.

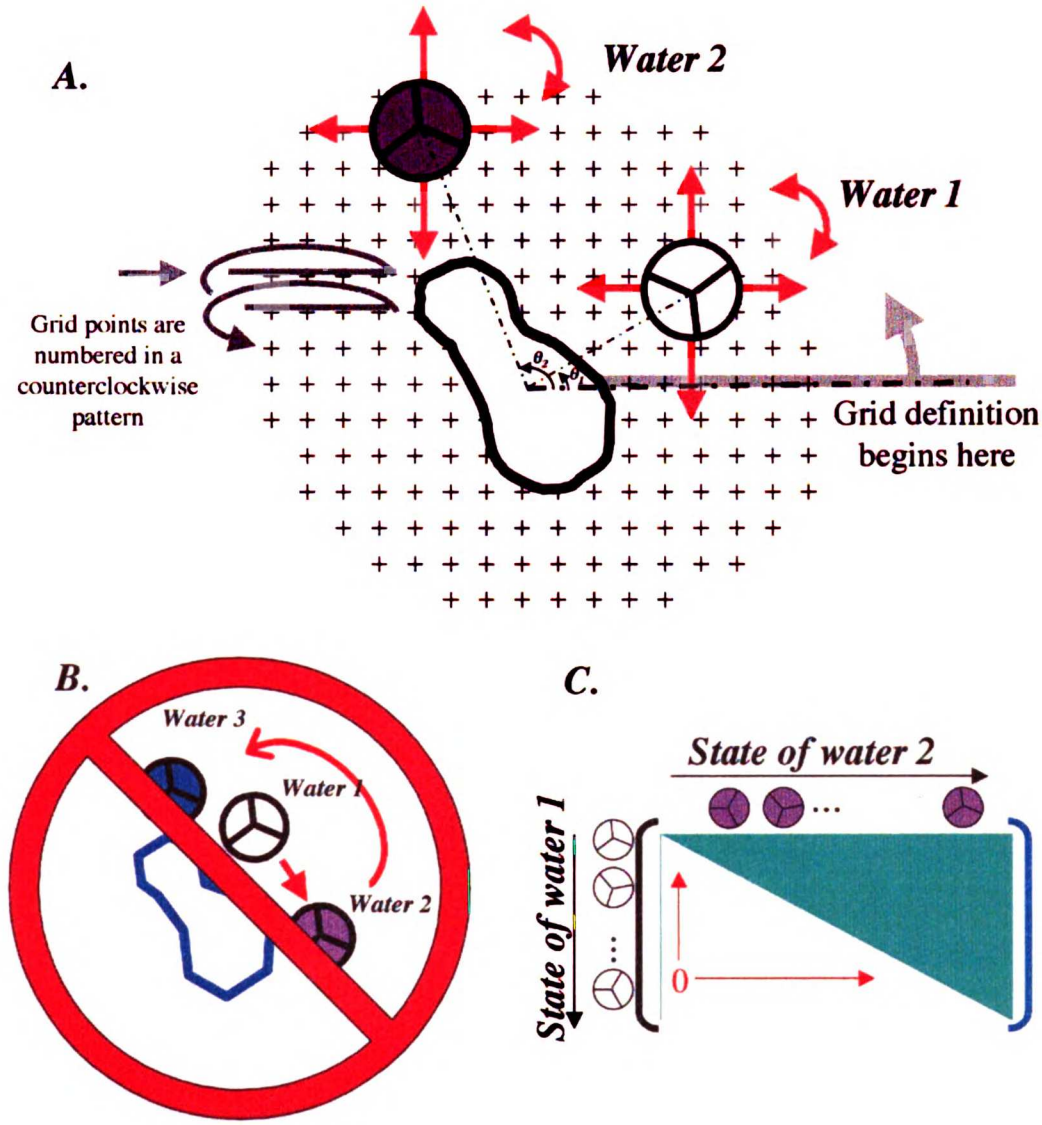



Figure 3. Modifying the transfer matrix method to allow positions of interacting waters to vary.

(A) Each water is allowed to exhaustively explore of a predefined set of grid points that fills the extent of the solvation shell. In the definition of possible water states, the numbering of grid points is chosen so that they proceed in a counterclockwise direction around the solute. In this way, for example, if water 1 is at the first grid point, and water 2 is at successive grid point, water 2 is guaranteed to be positioned counterclockwise relative to water 1. (B) By restricting each successive water going around the shell to always have a higher grid number than the last, we therefore guarantee that no out-of-order shell-water contacts occur. This geometric restriction is embodied in the individual transfer matrices by setting all lower-diagonal elements of the matrix to zero for matrices $M_1$, $M_2$, $M_3$,... $M_{N-1}$ (C). However, the final transfer matrix $M_N$ describing the interaction of the last water N back to the first water 1 has this restriction reversed, so the upper-diagonal elements are zeroed (not shown). The solute is drawn in a peanut shape to emphasize that it may have arbitrary shape and interaction character; the matrix method we describe does not place any restrictions on the solute.

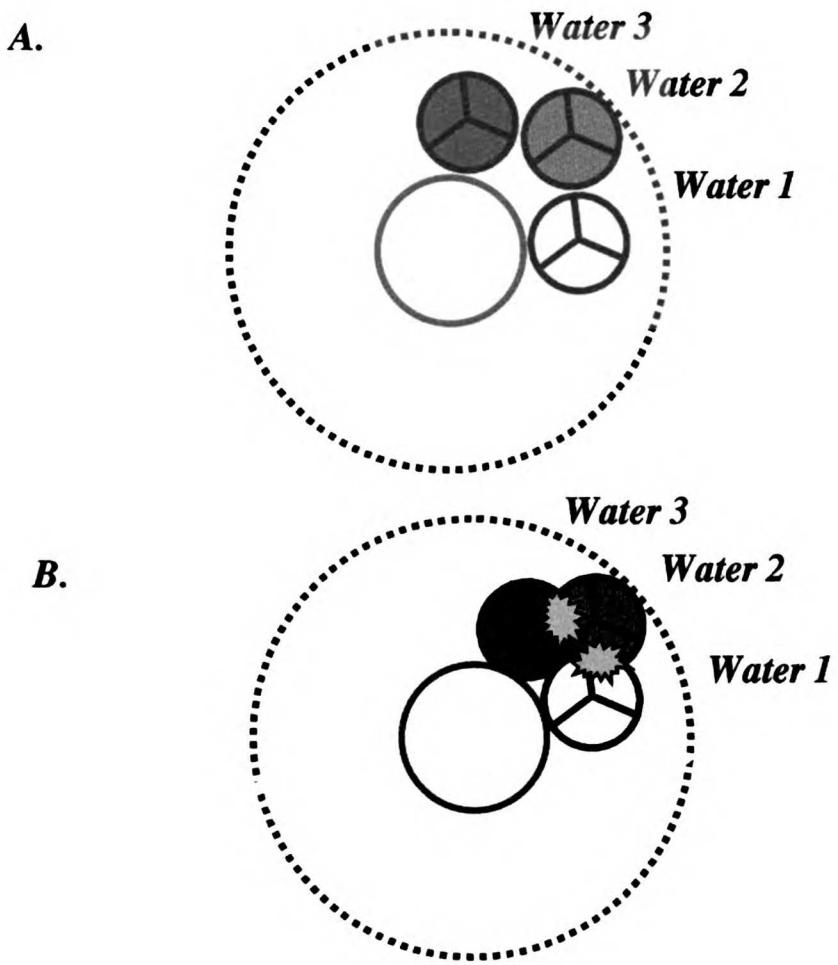



Figure 4. Non-nearest-neighbor issues in the matrix formalism

**A.** Waters ordered correctly, in a counterclockwise direction, normally do not have non-nearest-neighbor interactions in the first solvation layer.

**B.** A case where a non-nearest-neighbor interaction can occur. Waters *1* and *3* are close enough to interact. However, one or both of these waters must penetrate within the van der Waals radius of water 2 (yellow stars) in order to make the contact. For sufficiently restricted shell radii (i.e. a single solvation layer), an out-of-order shell-water contact like this will always produce such a clash, giving a highly unfavorable energy. Ignoring configurations such as this, as is done by SWYZLE, should have a negligible effect on the partition function because of this high energy..

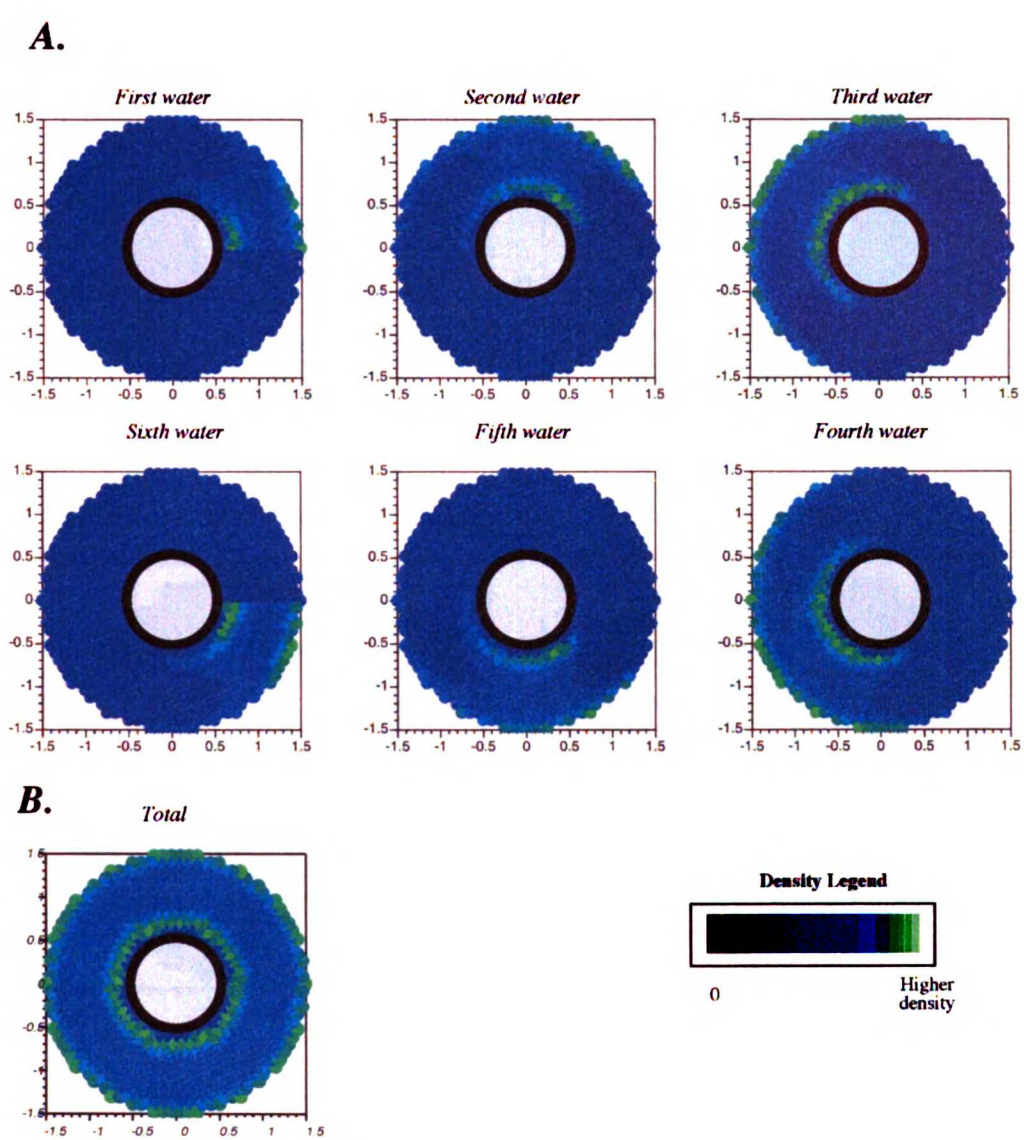



Figure 5. Illustrating the mechanics of the SWYZLE calculation.

Shown are shell-water density distributions as a function of grid position, colored so that blue is the lowest density and green is the highest.

A. Each permuted matrix product $M_{tot:m}^{circ}$ produces the density distribution of the *m*th water, as is shown for the case of six shell waters. These densities do not have a true physical meaning.

B. Summing the 6 density distributions in A yields the total water density distribution of the first solvation shell.

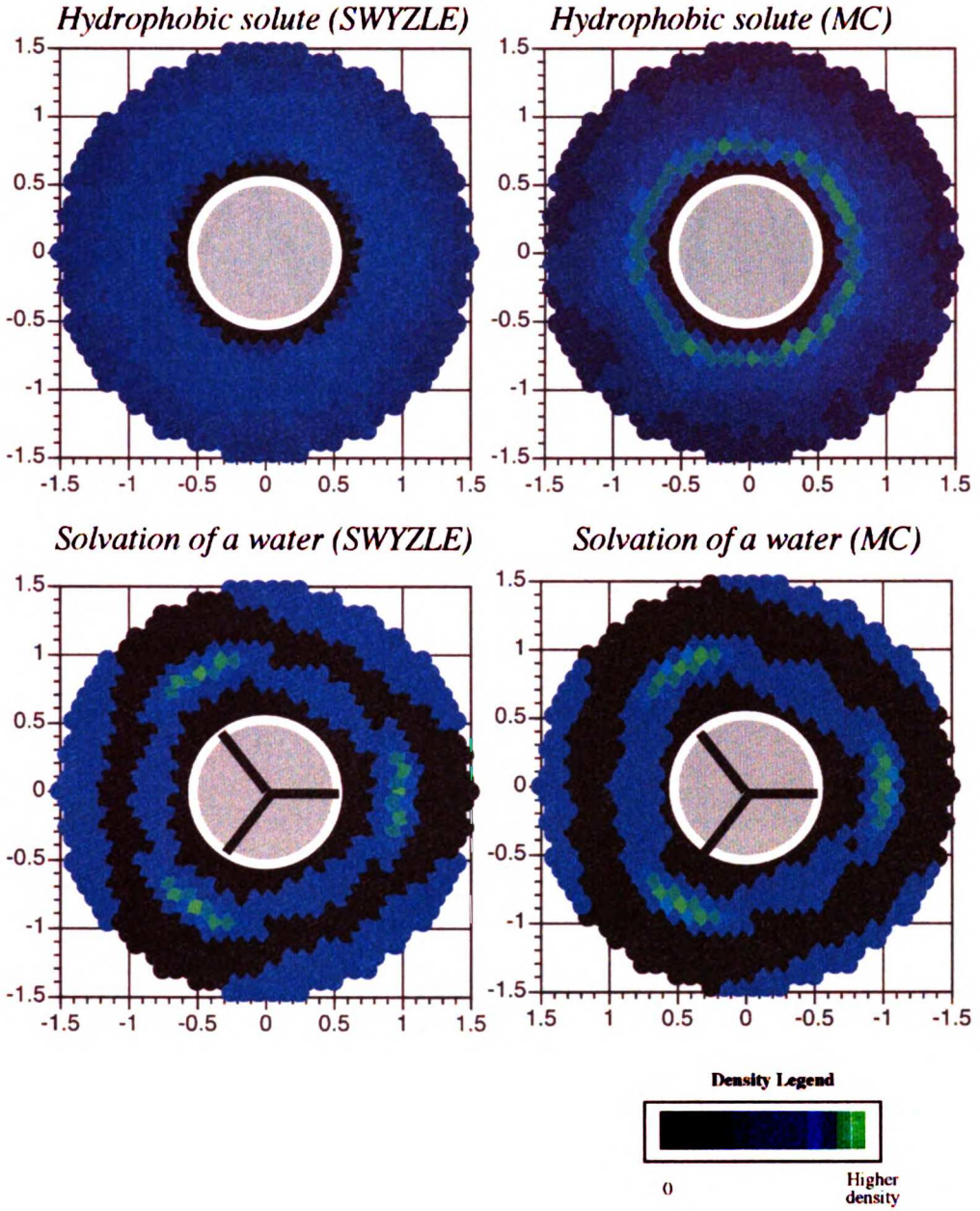



Figure 6. Shell-water density distributions for two solute types. The results of SWYZLE calculations are compared with those of MC simulations at the same temperature and water density. At top are graphs for a hydrophobic solute (gray circle). At bottom are graphs for a solvated MB water molecule (black lines represent hydrogen bonding arms). To set the water density in the SWYZLE calculations, the shell water occupancy was matched, for each solute type, to that of the corresponding MC simulation (see next Figure).

A green ring of higher density surrounding the hydrophobic solute in the MC simulation is not seen in the SWYZLE calculation, probably the result of the crude outer boundary potential employed in the current SWYZLE implementation. The text discusses probable fixes for this problem.

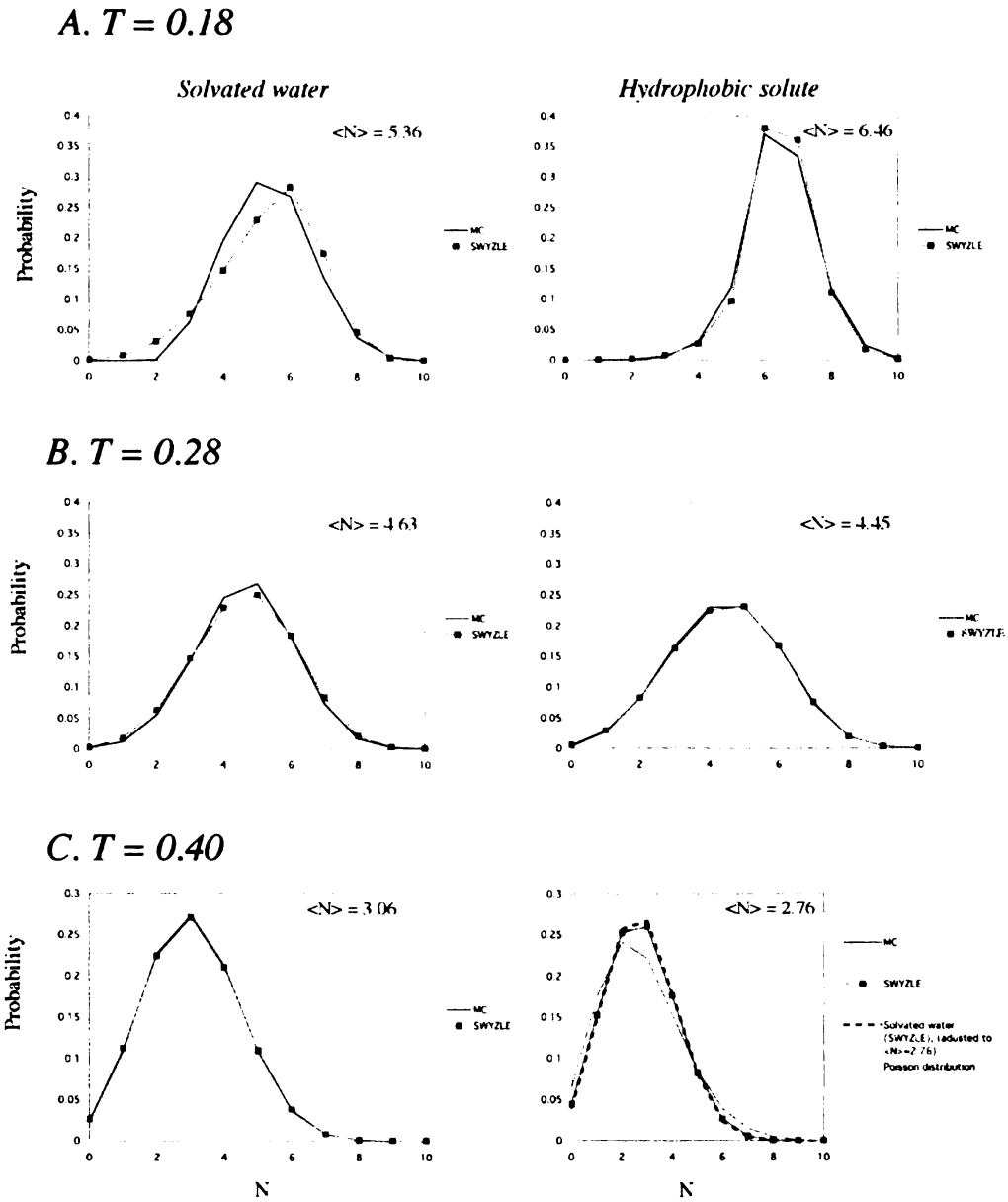



Figure 7. Shell-water occupancy distributions for a solvated water and for a hydrophobic solute, shown for three temperatures. MC simulation results are compared with SWYZLE calculations where the average occupancy, <N>, of the first solvation shell was matched to that found in the MC simulations. The high-temperature results are seen to qualitatively approach the Poisson distribution. Also, at the highest temperature, matching the occupancy of SWYZLE calculations for the two solute types (dashed line, lower right-hand panel) produces superimposable occupancy histograms.

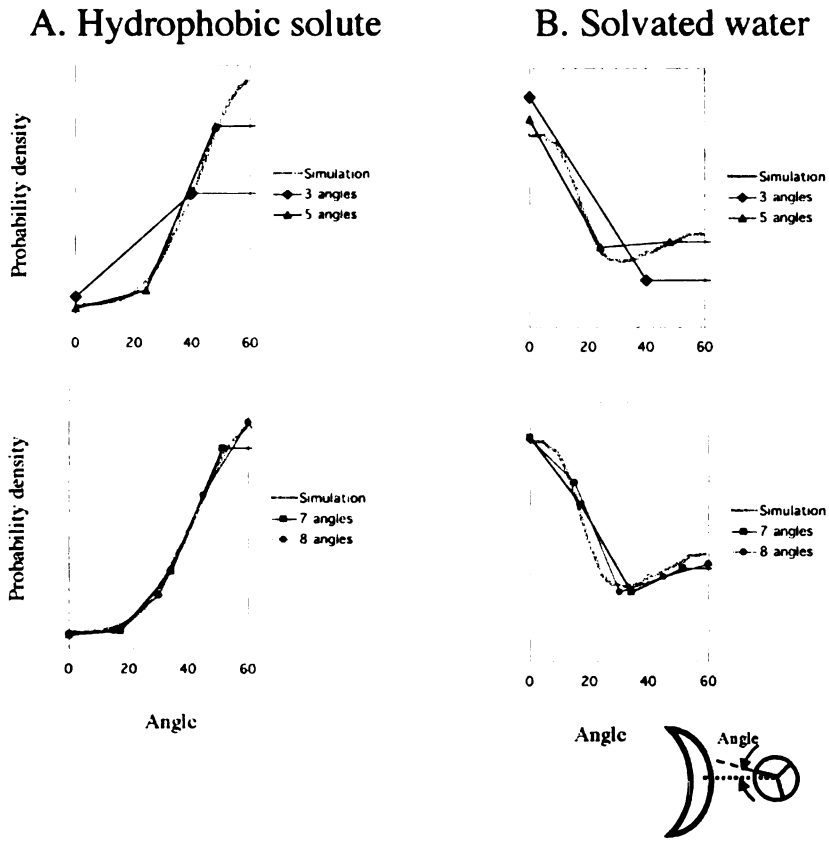



Figure 8. Shell water orientation distributions for two solute types.

The solute types are a hydrophobic solute ($\sigma = 0.7$) and a solvated water. Distributions predicted by SWYZLE are compared with those from MC simulations. Distributions for four values of the number of orientations per grid point (in the SWYZLE calculation) are shown. Orientations were averaged for all shell waters within a radius of 1.4 $r_{HB}$ of the solute, for both solute types shown. A 0° shell-water orientation is defined as the case when the angle of a hydrogen-bonding arm is coincident with the surface normal of the solute (and so, points at the solute center for a spherical solute). The shell radius of the SWYZLE calculation was 1.5 $r_{HB}$ for the hydrophobic solute and 1.6 $r_{HB}$ for the solvated water, which gave optimal agreement with the MC simulation (see Figure 9).

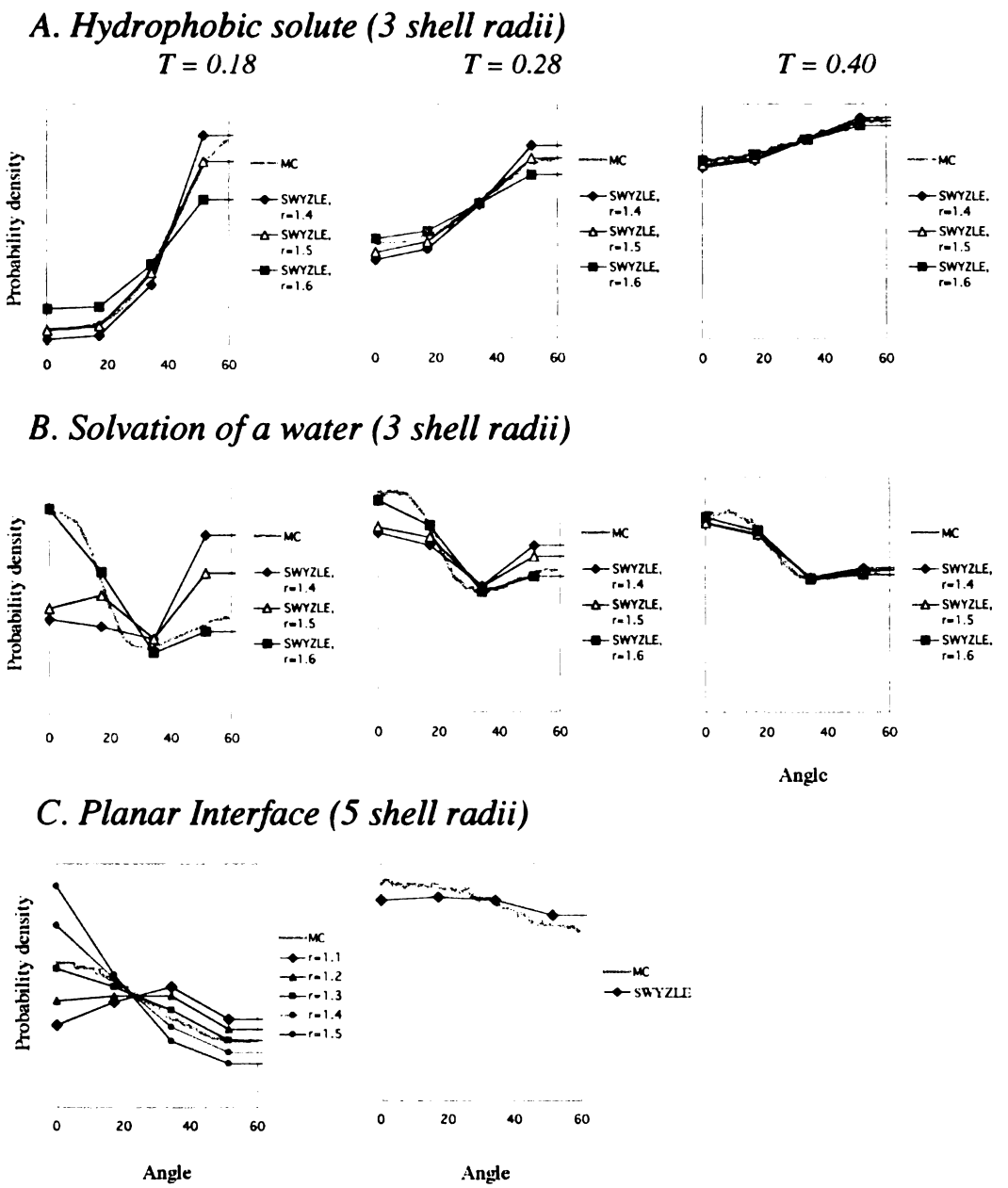
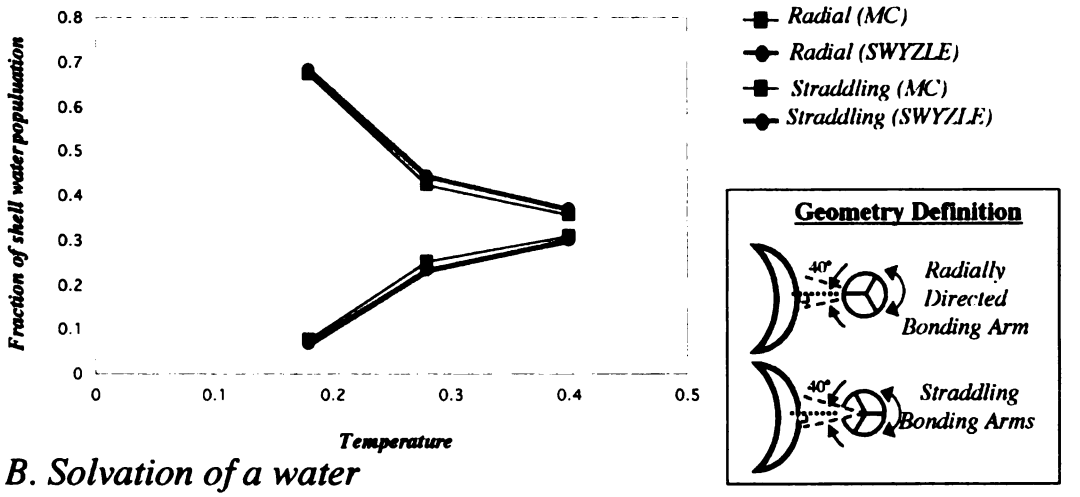



Figure 9. Shell-water orientation distributions depend on the choice of shell radius in the SWYZLE calculations. Three values of the shell radius, $r$, were considered for the two small solute types, while five values of $r$ were examined for the planar interface. The orientation angle of a shell water is defined as in Figure 8.

## A. Hydrophobic solute




## B. Solvation of a water

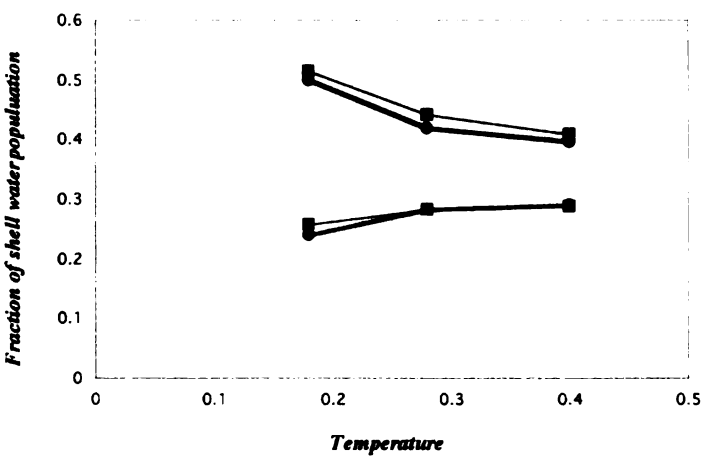



Figure 10. Overall shell-water orientation behavior as a function of temperature and solute type.

Two categories of shell water were defined by their relative orientation to the surface normal of the solute. These categories correspond to whether the water has a solute-directed bonding arm ("radial") or whether the two of the water's two bonding arms straddle the solute ("straddling"). Each category accounts for a 40° swath of the water's possible orientations, times three because of the model's 3-fold symmetry. A third category accounting for the remainder of possible orientations is left out of the graphs for clarity. The outer shell radii of the SWYZLE calculations for both solute types was the same as in Figure 7.

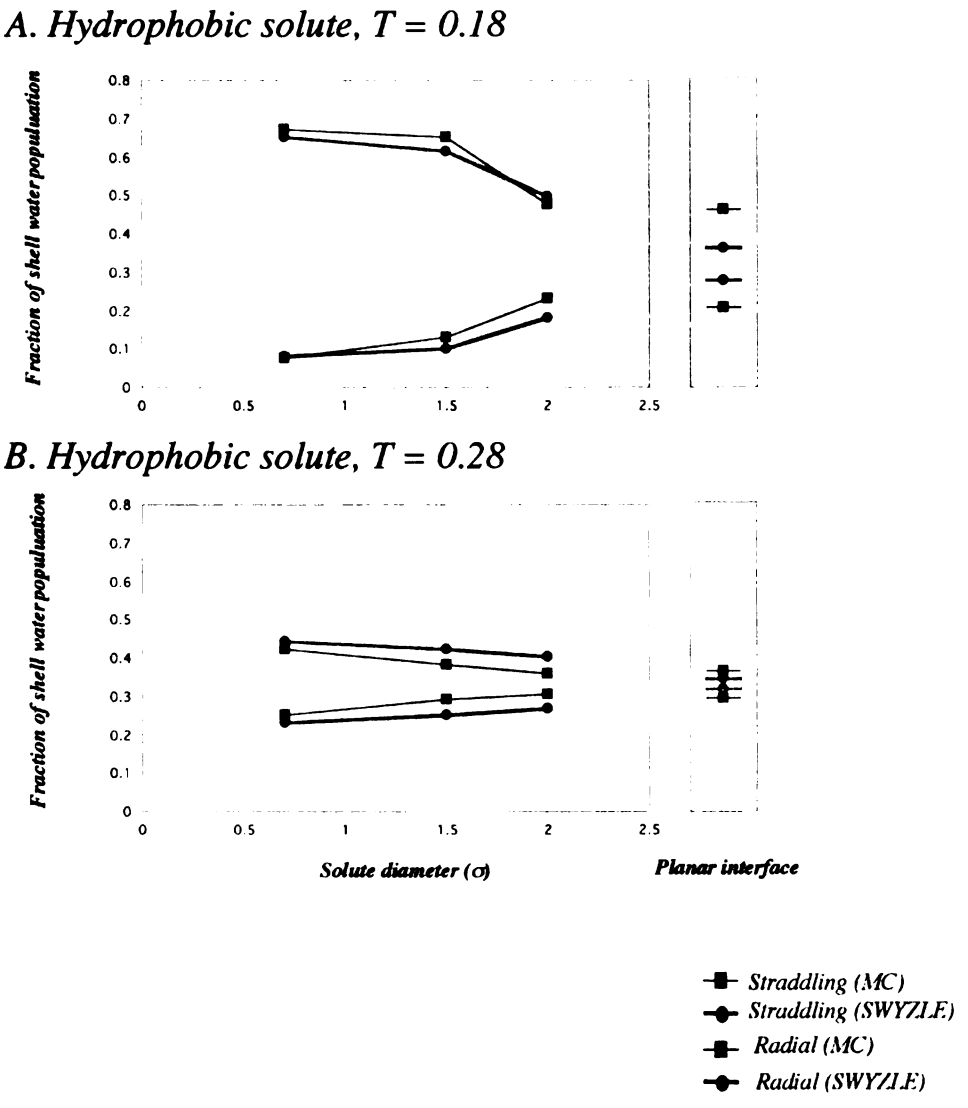
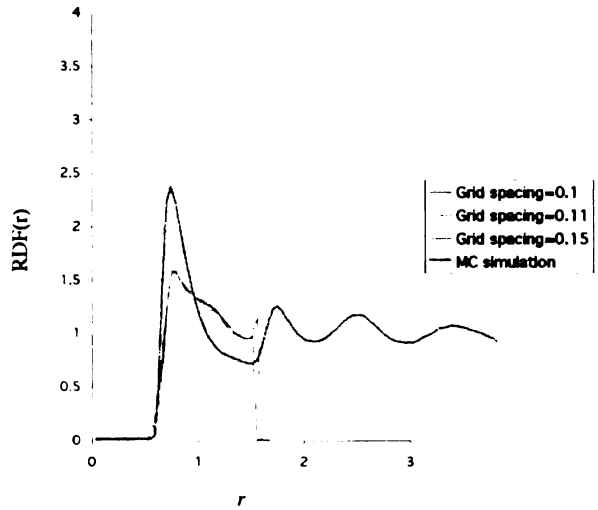



Figure 11. Shell-water orientation category, as defined in Figure 9, as a function of size for the hydrophobic solute. In all cases the outer extent of the first solvation shell (SWYZLE calculation) was set to radius of the minimum between first and second pair correlation function peaks as determined by the MC simulations.

**A.** For a temperature of $T = 0.18$. **B.** For a temperature of $T = 0.28$

## *A. Hydrophobic solute*




## *B. Solvated Water*

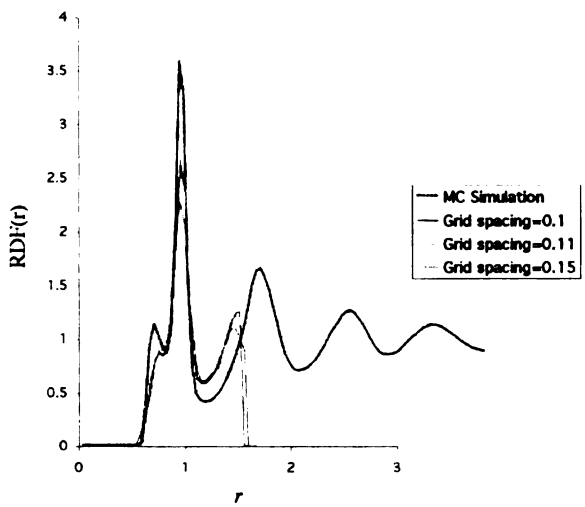
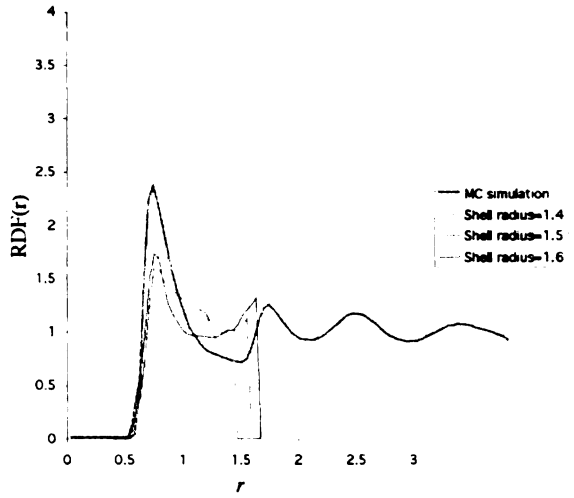



Figure 12. Solute-water pair correlation functions for a solvated MB water molecule (A), and a hydrophobic solute (B) of the same diameter. Shown are results from MC simulations and from SWYZLE calculations for three different grid spacings.

*A. Hydrophobic solute*




*B. Solvated Water*

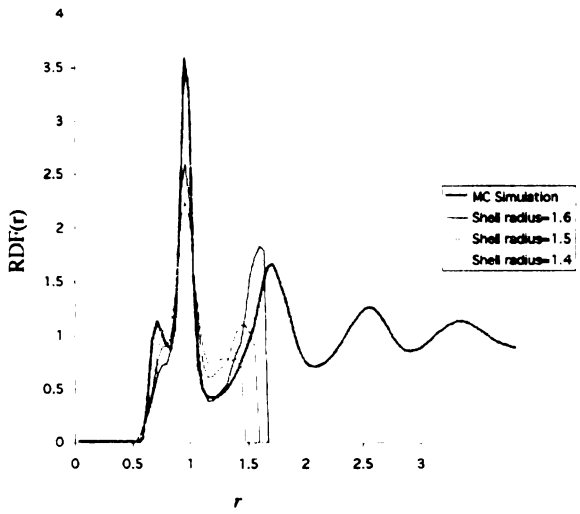



Figure 13. Solute-water pair correlation functions depend on the choice of shell radius in the SWYZLE calculations. Shown are solute-water pair correlation functions for a solvated MB water molecule (A), and a hydrophobic solute (B) of the same diameter.

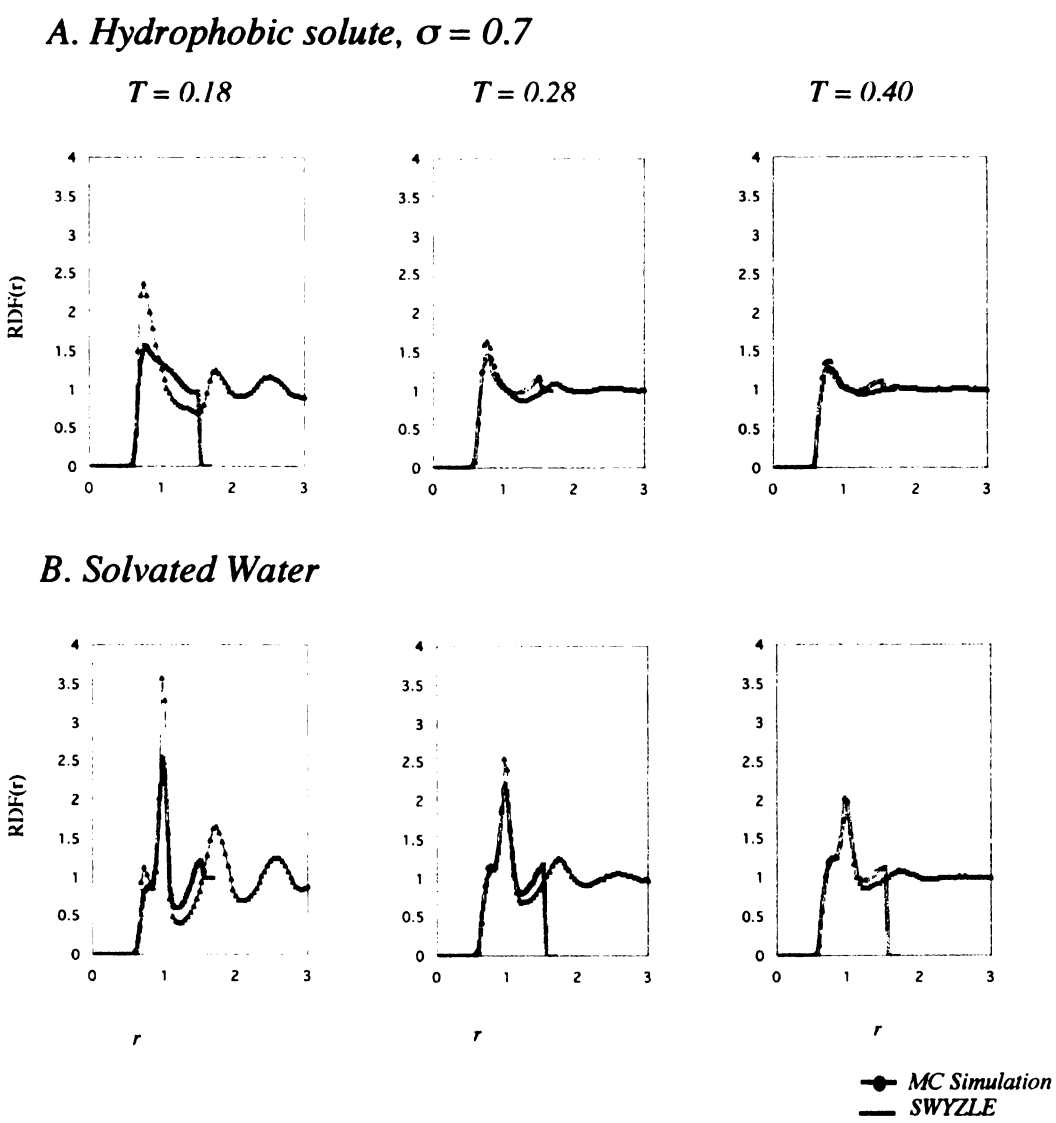



Figure 14. Pair correlation functions for a hydrophobic solute and a solvated water, for three temperature values.

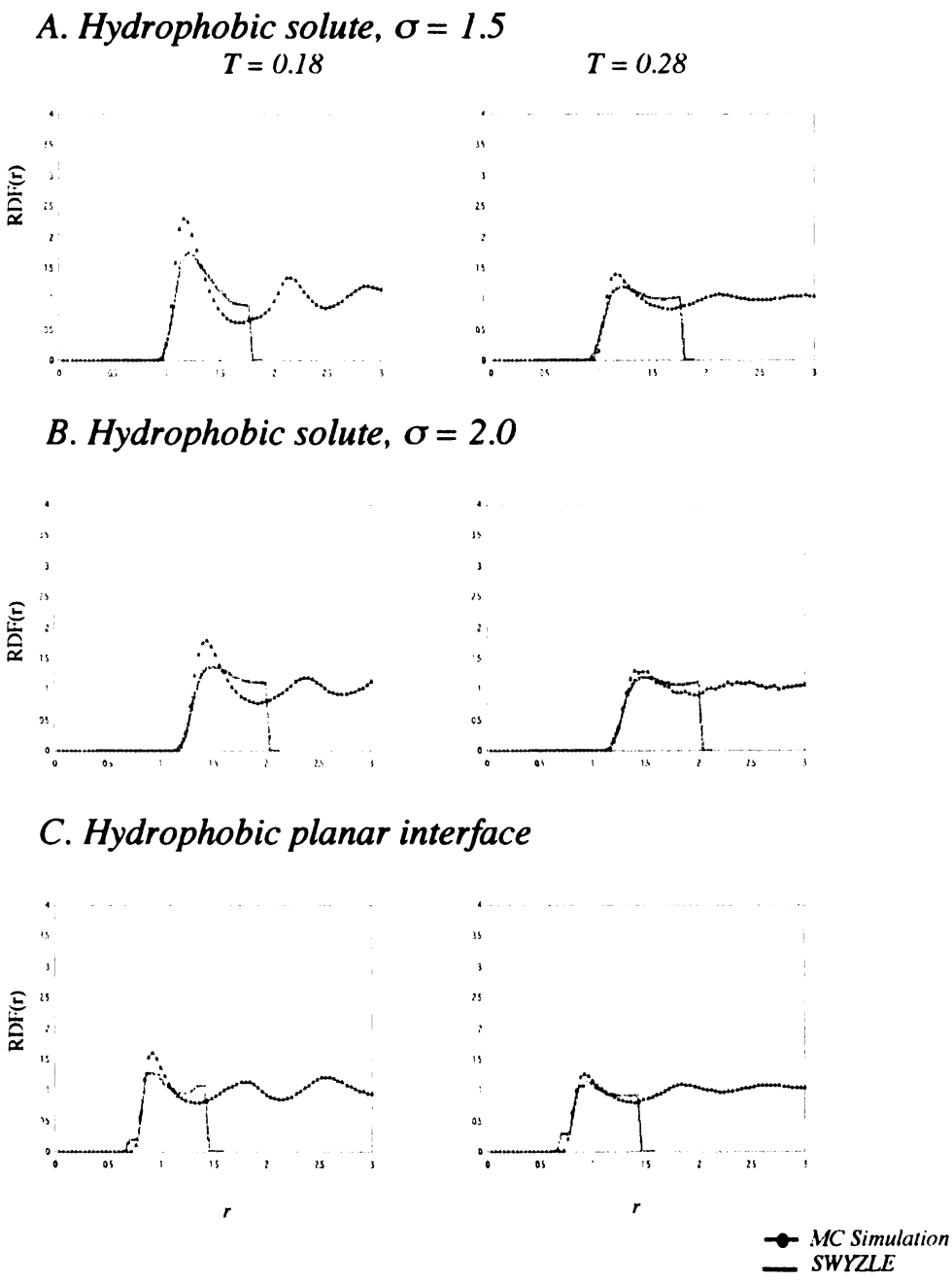



Figure 15. Pair correlation functions as a function of solute size, for two temperature values.
In all cases the outer extent of the first solvation shell (SWYZLE calculation) was set to radius of the minimum between first and second pair correlation function peaks as determined by the MC simulations.

# Appendix One:
# My early experimental meanderings with kinesin

## Initial nucleotide cocrystallization experiments by our group

Following the first crystal structures of kinesin and myosin, the course for continued structure characterization in both of these motors was quite clear. Each motor's mechanism relied on attachment and release from the respective filament tracks (actin for myosin, microtubule for kinesin), dependent on the state of active-site nucleotide in the motor. Furthermore, it seemed likely that nucleotide-dependent conformation changes in these motors, while they were attached to their filaments, would prove to drive the force-producing component of the motility mechanisms. Therefore, groups immediately began seeking X-ray crystal structures of both kinesin and myosin with alternative nucleotide forms bound in their active sites. The new structures found in this way, it was hoped, would be to reveal conformations of the motors that showed how changes in their nucleotide state caused them to either bind, or detach from, their filaments—at the same time, perhaps, delivering the "powerstrokes" that made them go.

The above experiments, of course, fell short of what might be considered the "ideal" experiment: solving *co*crystal structures of kinesin and myosin bound with components of their respective filaments—where the effects of nucleotide exchange would be potentially more correlated with a "powerstroke" in these motors. However, properties of the filament components, tubulin and actin, make such a cocrystallization approach very difficult. This is because actin and tubulin both polymerize into long, asymmetric structures when they are concentrated to strengths where crystallization might be attempted. Recently developed techniques in our laboratory and in collaboration with Jim Spudich's laboratory at Stanford may soon allow cocrystallization of both kinesins and myosins with their filament partners; in 1995, however, such options were not yet on the horizon.

By the time I joined the Fletterick laboratory and began working on kinesin, the first round of results had already come in from the nucleotide-exchange experiments—and they were disappointing, or at least, surprising. Elena Sablin, Jon Kull, and a newly recruited graduate student (Angela Newhoff) had recrystallized both of the laboratory's solved kinesin variants ("conventional" kinesin and NCD) in the presence of chemical analogs related to ATP—nonhydrolyzable versions of this nucleotide that were thought convert the motor to an "ATP-like" form. However, none of the analogs (including the commonly used AMPPNP as well as the metal complexes ADP•aluminum fluoride or ADP•beryllium fluoride) appeared in the resulting crystal structures—all contained ADP and all were indistinguishable from the originally crystallized forms. The laboratory of Eckhard Mandelkow, which had solved the structure of rat kinesin soon after Jon's and Elena's structures, would soon publish results indicating a similar inability to exchange nucleotide analogs in their system(Muller, Marx et al. 1999).

## Crystallography studies with kinesin mutants

My first project in the Fletterick lab, as a rotation student, was to subclone a mutant, nonhydrolyzing form of kinesin into a construct suitable for crystallizing. The idea was that kinesin might, in this mutant, be stuck in an "ATP-like" conformation that differed from the crystal structures yet seen. The placement of the mutated residue, Gly234-> Ala, was particularly significant in that the residue was absolutely conserved, not only in kinesin, but also in the myosins and the G-proteins—in both of whom the analogous Ala mutation also had analogous, crippling effects as were seen in kinesin. Thanks to the wonderful guidance of Jon Kull and especially Elena Sablin (who made my first experience with molecular biology a true joy), this project went amazingly quickly and was essentially complete by the end of my rotation—the fastest progress I would see, it turned out, for a long while.

The bad news was that the hoped-for conformational change did not in fact surface, and the new crystal structure was identical (in all ways we could see) to the previous forms solved. In light of my later discovery, described in Chapter 1, of a means to selectively

crystallize our *wild-type* kinesin construct in two alternative conformations, it might be interesting to reinvestigate this mutant. Another, related nonhydrolyzing mutant (Glu 236-> Ala) was also subcloned during my rotation and could be of similar interest. While various intriguing features of these mutants are now well characterized by the probe work of Sarah Rice in the Vale lab (see Appendix 1), there are no current plans to further investigate these mutants by crystallography. Certainly at the time that I solved the G234A structure, there seemed to be nowhere to go with them.

## Crystallography of nucleotide-exchanged kinesin

At the end of my rotation, I began a renewed quest to discover conditions in which ATP analogs would bind and cocrystallize with kinesin, with the goal of perhaps discovering some new analog or crystallization condition that might tip the balance in favor of exchange and conformational change. At least initially, results seemed promising. Using a previously untried nonhydrolyzable ATP analog (AMPPCP, 10mM), I found that kinesin's structure appeared to be altered sufficiently that the original crystallization conditions for our construct no longer worked—the motor precipitated instead. After some screening, I discovered modified conditions (see Appendix 5) under which kinesin would still crystallize with AMPPCP present. Protein purity issues appeared to impede the growth of very large crystals (these problems were later solved, see the Methods of Chapter 1), but nevertheless some preliminary data were collected—to 4.5Å, not sufficient for atomic characterization.

At this point, however, it became clear that the parameters describing the new crystal lattice (unit cell dimensions and space group) were very similar to those of the original, solved crystal structure. Since it seemed unlikely that significant conformational change would be seen in such a similar crystal environment, I eventually set this project aside. The true nature of these crystals would probably have remained unsettled had it not been for the strong work of Mary Jane Budny in our laboratory, who recently repeated this AMPPCP cocrystallization experiment using the enhanced kinesin purification protocols described in Chapter 1. The resulting higher-resolution data set (<3.0Å), while

incomplete (~50% completeness in the refinement statistics), was sufficient to generate difference maps that showed quite conclusively that ADP, not the analog, occupied the active site. And so, it is fair to conclude my efforts to exchange ATP analogs into kinesin's site and crystallize new conformations met with the same lack of success as the other reported efforts.

## Using peptides to mimic the kinesin-tubulin interaction

A potential breakthrough in understanding the specific interactions of kinesin to its microtubule track came in the fall of my first year in the Fletterick lab, when Eva Nogales solved the atomic-resolution structure of tubulin in collaboration with Ken Downing at the Lawrence Berkeley National Laboratories (Nogales, Whittaker et al. 1999). Docking the tubulin structure along with kinesin X-ray structures into cryo-electron microscopy images of the motor-microtubule complex provided the first look at the residues of tubulin that were close proximity to the motor. Strikingly, tubulin's C-terminal ~50 or so residues were found to project significantly from the microtubule surface, forming a "ridge" upon which kinesin binds. Furthermore, experiments had demonstrated that the 12 C-terminal residues from β-tubulin could competitively inhibit the ~1000-fold microtubule-stimulated hydrolysis of kinesin (Tucker and Goldstein 1997). These observations suggested that peptides derived from tubulin's C-terminal segment might by themselves compose a functional part of kinesin's microtubule binding site, and I began a collaboration with Mundeep Chana, a graduate student working with Professor Robert Hodges at the University of Edmunton, Alberta, to examine this possibility.

Once Mundeep had synthesized the peptides, however, limited testing soon indicated that they did not have the kind of functional activity we were looking for—namely, the ability to stimulate kinesin's ATPase. In control experiments, I was able to clearly detect kinesin's baseline ATPase activity using the highly sensitive Malachite Green assay. However, it appeared that the one peptide tested (the C-terminal 38-mer of β-tubulin) actually *reduced* kinesin's baseline ATPase to below the detectable limit. These

experiments, however, were never repeated and more careful controls would be needed to prove such a negative effect. Because I was looking for activation rather than deactivation in these experiments, though, I decided to shelve the project at the time and move on.

This topic however may be worth revisiting. In light of the results and model of Chapter 2, I now think it is likely that β-tubulin's C-terminus may directly interact with kinesin's *neck linker*. In my view, this part of tubulin is one of the prime candidates for the microtubule component that stabilizes the "microtubule-docked" neck-linker conformations we have inferred to exist in kinesin's nucleotide-free and ADP-bound states. If an interaction occurs between the neck linker and this tubulin element, then it is entirely possible that the tubulin peptide by itself retains an ability to interact with kinesin's neck linker. And it could be that, in the absence of the rest of the microtubule structure, such a neck-linker interaction might be *inhibitory* to kinesin's ATPase. For example, by preventing docking of the neck linker onto kinesin's catalytic core, the tubulin peptide might destabilize the "upstroke" position of switch II that is thought to accompany ATP binding of the microtubule-bound motor.

# Appendix two:
# Summary of EPR data

The EPR data used to build many of the conclusions in Chapters 1 and 2 contain many (to me) fascinating and poorly understood features. For this reason I am including here a fairly comprehensive survey of the majority of EPR data ever taken from kinesin's neck linker (using the monomeric kinesin construct K349). Much detail is included here that was not presented in the Chapters, for example I show the raw populations of each of the 3 signal components describing the neck linker (see Figure 7 of Chapter 1 for a description of how these populations were derived). I also show the calculated free energies and enthalpies of transition between *each* of these three populations, as well as upper and lower bounds on the final free energy of neck linker "docking" as was reported in Chapter 1.

It should be noted that data in Figures 1-4 and 18 were taken with "old-style" tubulin from *adult* cows (see "an anomalous measurement of neck linker binding" in Results of Chapter 2), so have stronger neck linker docking in the microtubule-bound, ADP states than would be expected if the experiments were repeated with new stocks of calf tubulin.

One of the more interesting details to compare in the figures that follow is how well the free energies and enthalpies agree (or do not agree) in measurements at the *three* different neck linker probe positions examined, C328, C330, C333. If you use your imagination, it is even possible to see a slight "zippering" effect in the docking free energies as you traverse down these positions—but this phenomenon (if it is real) is similar in magnitude to the error in the measurements. Trends in the enthalpies are harder to interpret— as the error in them appears be significant— but the enthalpy of docking appears to be stronger in the microtubule-bound, AMPPNP state than for the microtubule-bound ADP state at all three probe sites (barring the "anomalous" second measurement of MT•ADP at position C333, Figure 18).

Measurements of the kinesin mutants G234A, E236A, I235A, and I325A/N327A have many features that remain obscure. It is striking that E236A appears to have a nucleotide-dependent enthalpy of docking when bound to the microtubule, while G234A does not. This is likely explained by G234A's neck linker remaining "stuck" in a *microtubule*-bound conformation in both nucleotide states.while that of E236A makes the transition for "backwards-docked" to "forwards-docked" (the transition proposed to occur in wild-type kinesin, see Chapter 2). It is also striking that the enthalpy measurement in the microtubule-bound, AMPPNP-complexed state of E236A is the highest of any yet made. We currently have no of understanding effects like this.

One last surprising feature is found in the microtubule-bound, AMPPNP-complexed form of I325A—which shows a highly diminished population of the *intermediate* neck-linker EPR component, relative to all other measurements made. I have no guess for why this might be. However, it is striking that this effect goes away in the *AMPPNP* state of this mutant, suggesting that the neck linker can still make its "backwards" to "forwards" transition.

It will be interesting to see how future investigation sheds light on these most puzzling features of kinesin's neck linker.

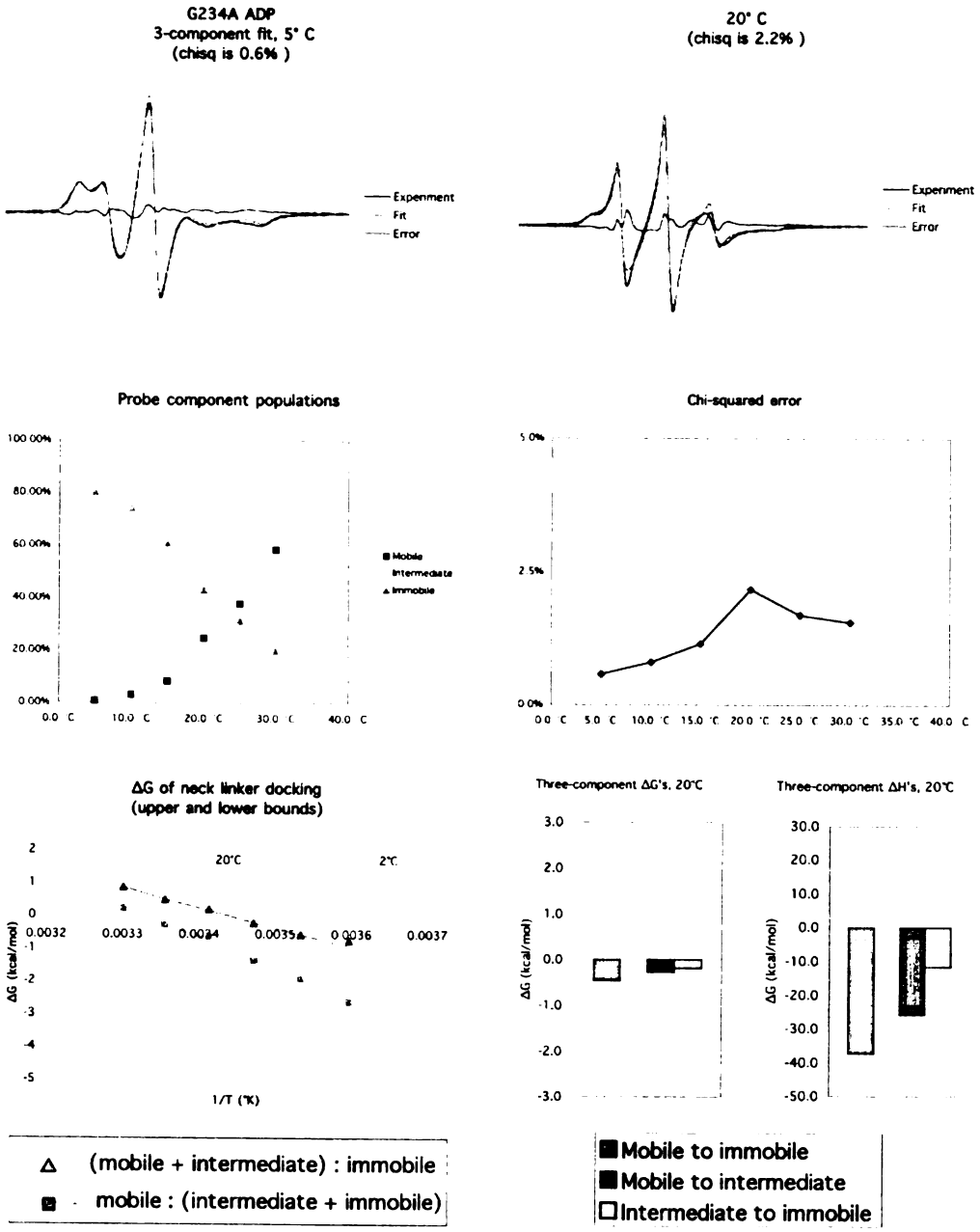

G234A ADP
3-component fit, 5° C
(chisq is 0.6% )
Experiment
Fit
Error
20° C
(chisq is 2.2% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 C
10.0 C
20.0 C
30.0 C
40.0 C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 C
5.0 C
10.0 C
15.0 C
20.0 C
25.0 C
30.0 C
35.0 C
40.0 C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
2
1
0
-1
-2
-3
-4
-5
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
1/T (°K)
Three-component ΔG's, 20°C
ΔG (kcal/mol)
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
Three-component ΔH's, 20°C
ΔG (kcal/mol)
30.0
20.0
10.0
0.0
-10.0
-20.0
-30.0
-40.0
-50.0
Δ (mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 1.

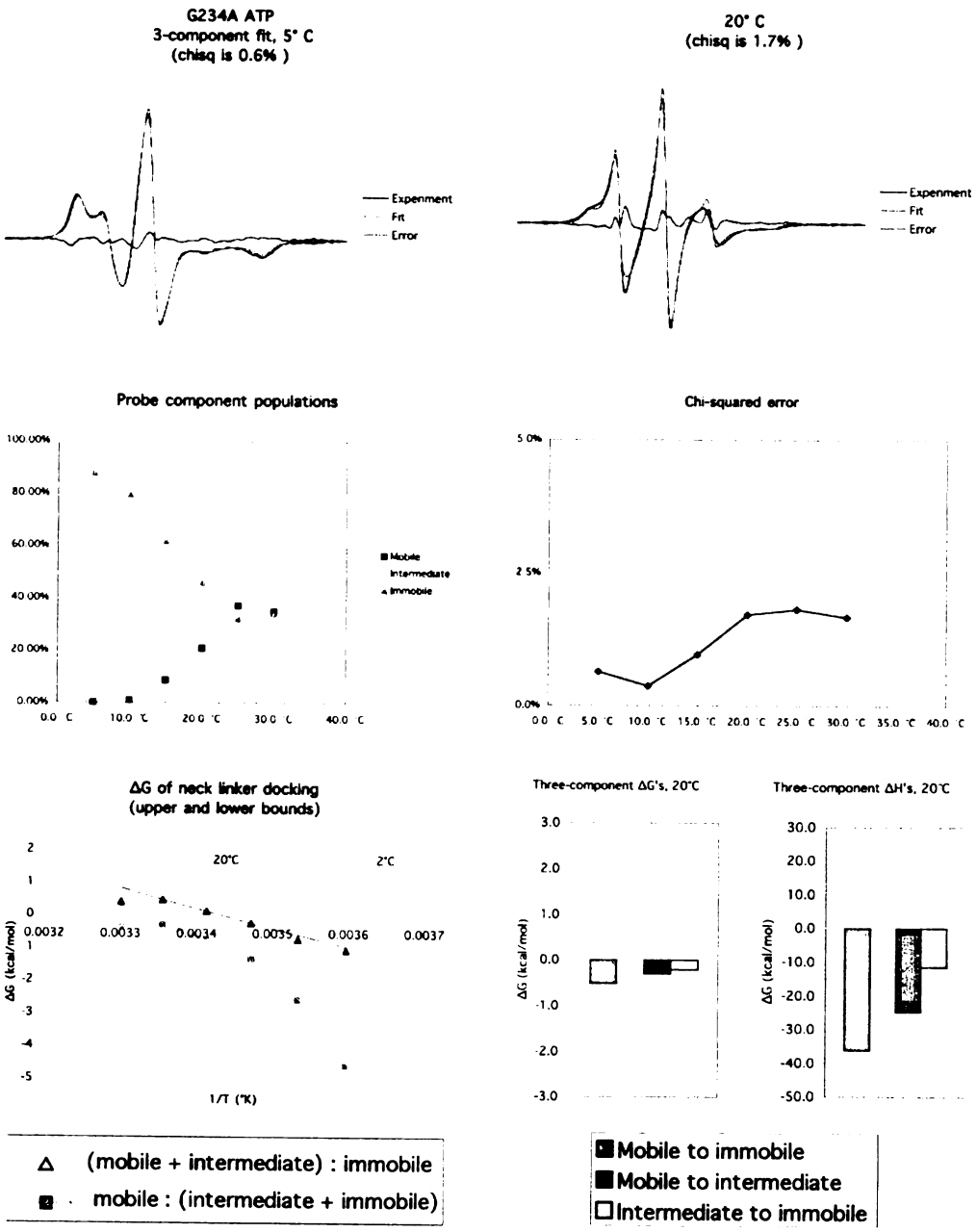

G234A ATP
3-component fit, 5° C
(chisq is 0.6% )
Experiment
Fit
Error
20° C
(chisq is 1.7% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 C
10.0 °C
20.0 °C
30.0 C
40.0 C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 C
30.0 °C
35.0 °C
40.0 C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
1/T (°K)
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 2.

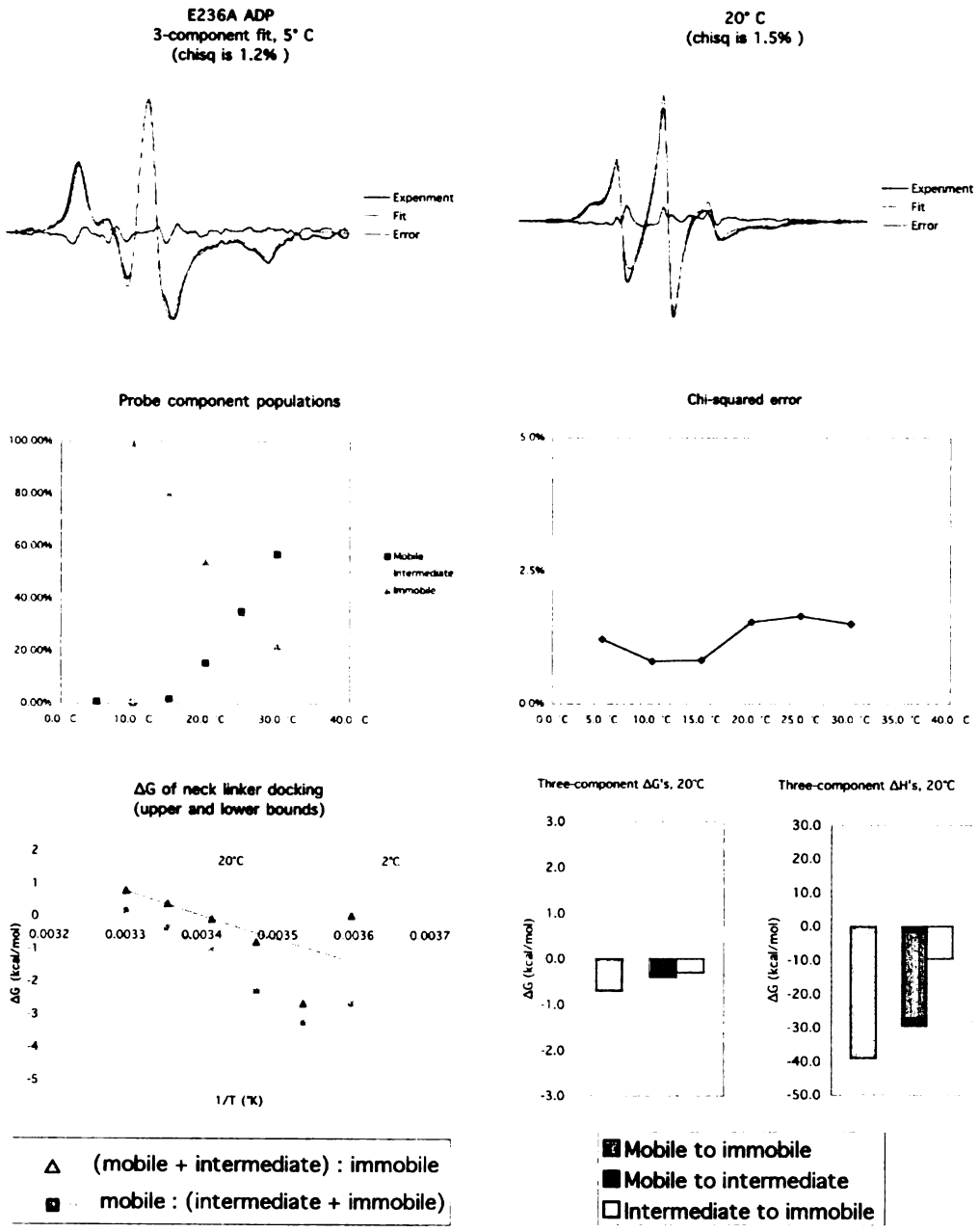

E236A ADP
3-component fit, 5° C
(chisq is 1.2% )
Experiment
Fit
Error
20° C
(chisq is 1.5% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 C
10.0 C
20.0 C
30.0 C
40.0 C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 C
5.0 C
10.0 C
15.0 C
20.0 C
25.0 C
30.0 C
35.0 C
40.0 C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
1/T (°K)
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 3.

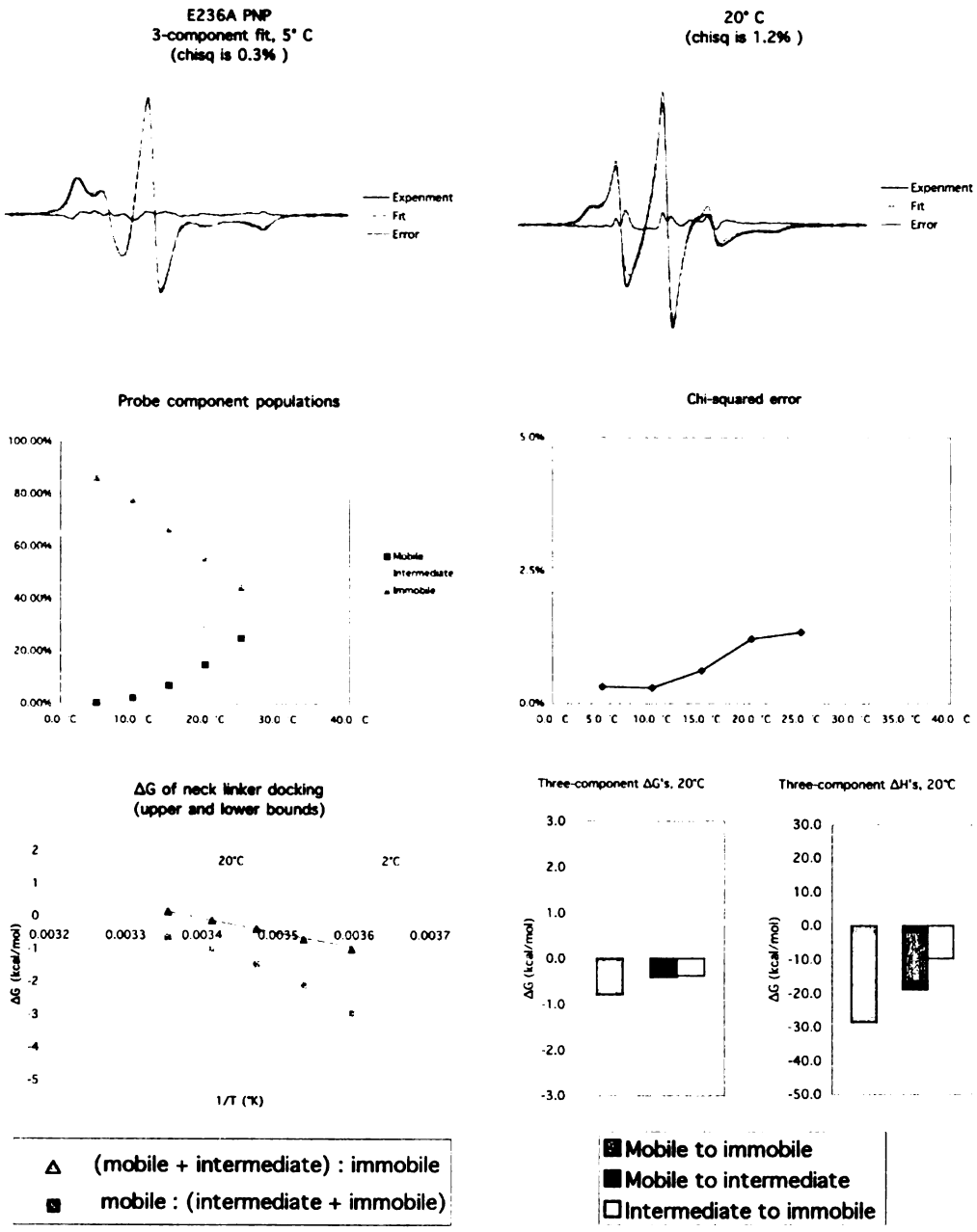

E236A PNP
3-component fit, 5° C
(chisq is 0.3% )
Expenment
Fit
Error
20° C
(chisq is 1.2% )
Expenment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 C
10.0 C
20.0 C
30.0 C
40.0 C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 C
5.0 C
10.0 C
15.0 C
20.0 C
25.0 C
30.0 C
35.0 C
40.0 C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
1/T (°K)
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
ΔG (kcal/mol)
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 4.

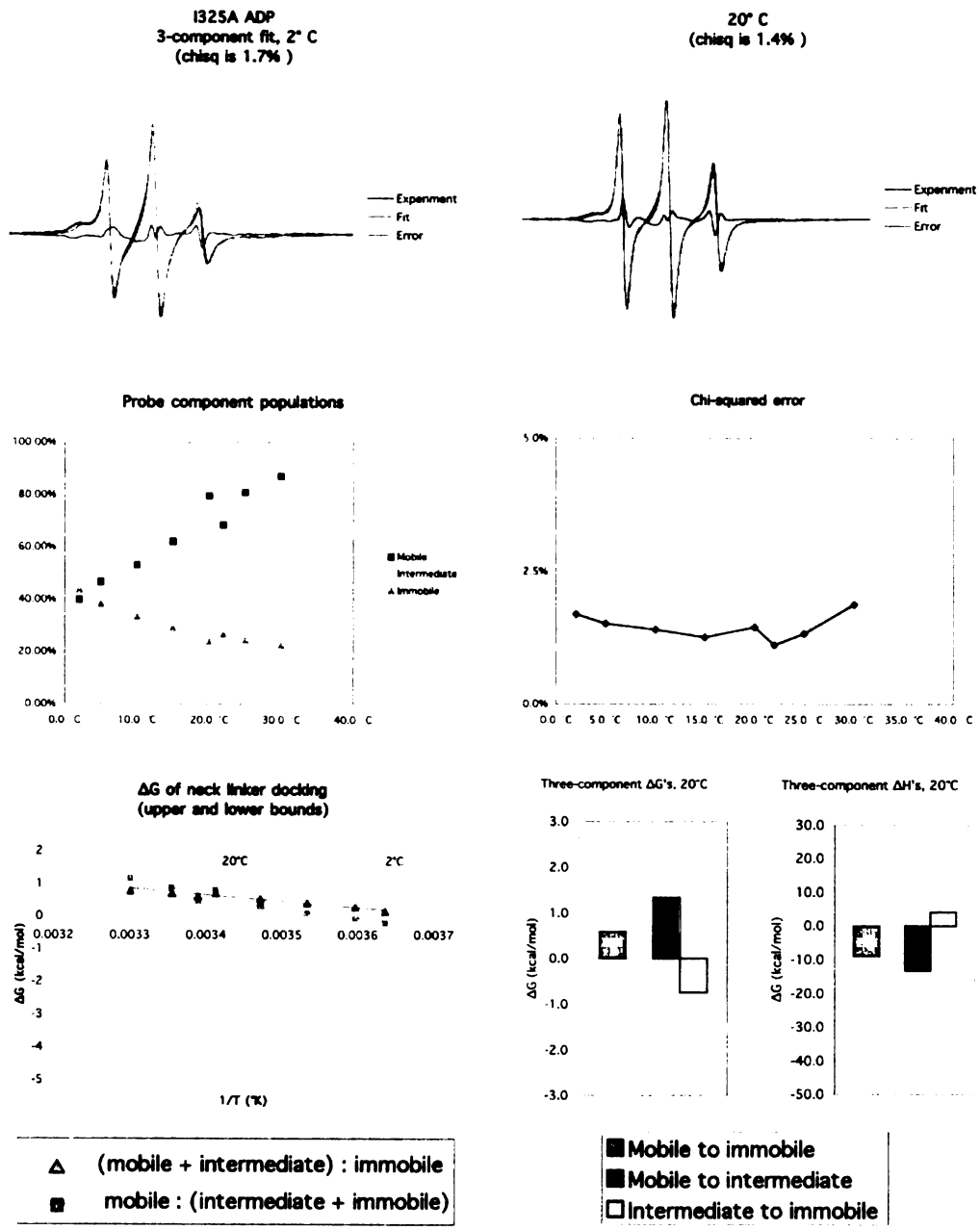

I325A ADP
3-component fit, 2° C
(chisq is 1.7% )
Experiment
Fit
Error
20° C
(chisq is 1.4% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 C
10.0 C
20.0 C
30.0 C
40.0 C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 C
5.0 C
10.0 C
15.0 C
20.0 C
25.0 C
30.0 C
35.0 C
40.0 C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
1/T (°K)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 5.

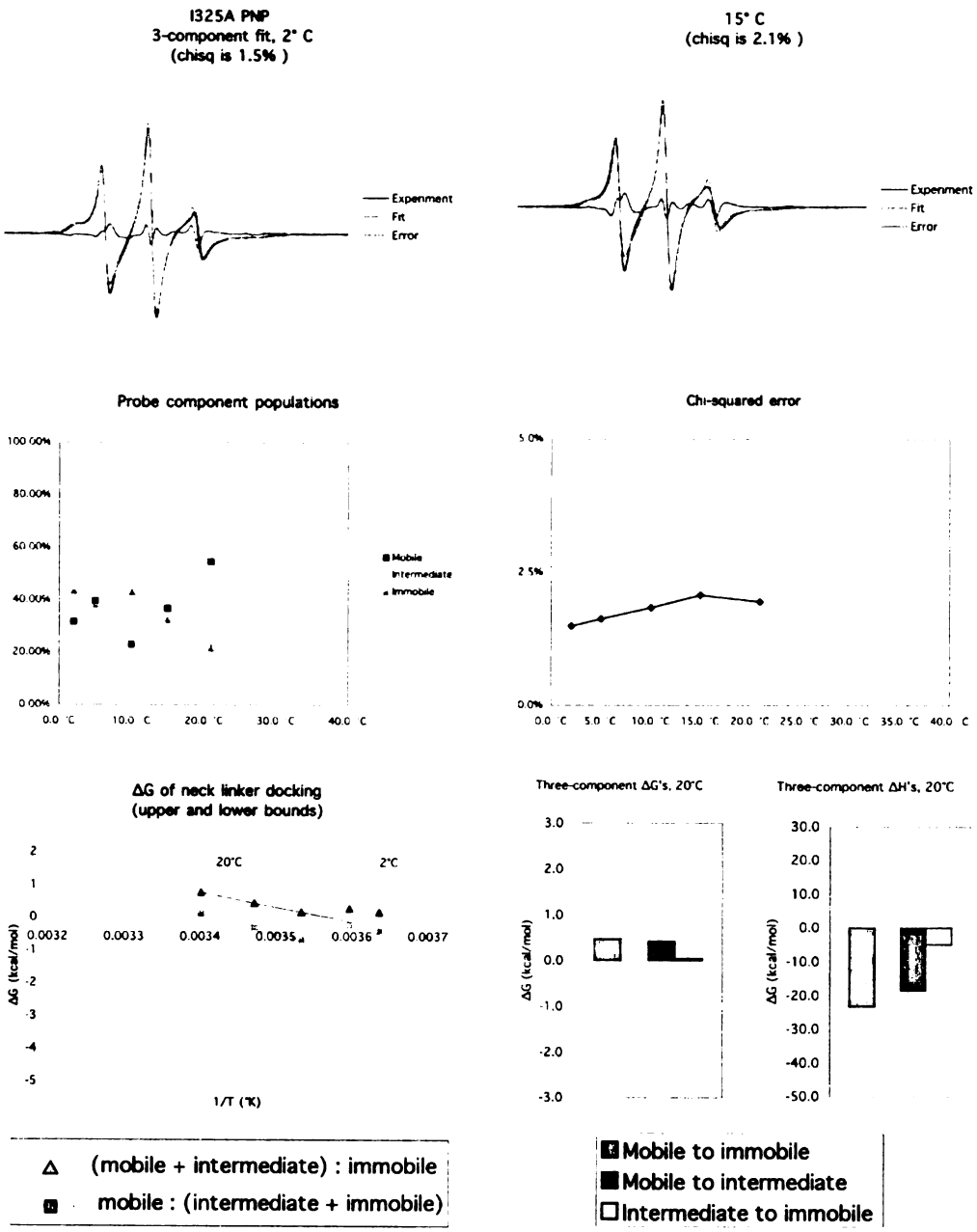

I325A PNP
3-component fit, 2° C
(chisq is 1.5% )
Experiment
Fit
Error
15° C
(chisq is 2.1% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 °C
10.0 °C
20.0 °C
30.0 °C
40.0 °C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 °C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 °C
35.0 °C
40.0 °C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
1/T (°K)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 6.

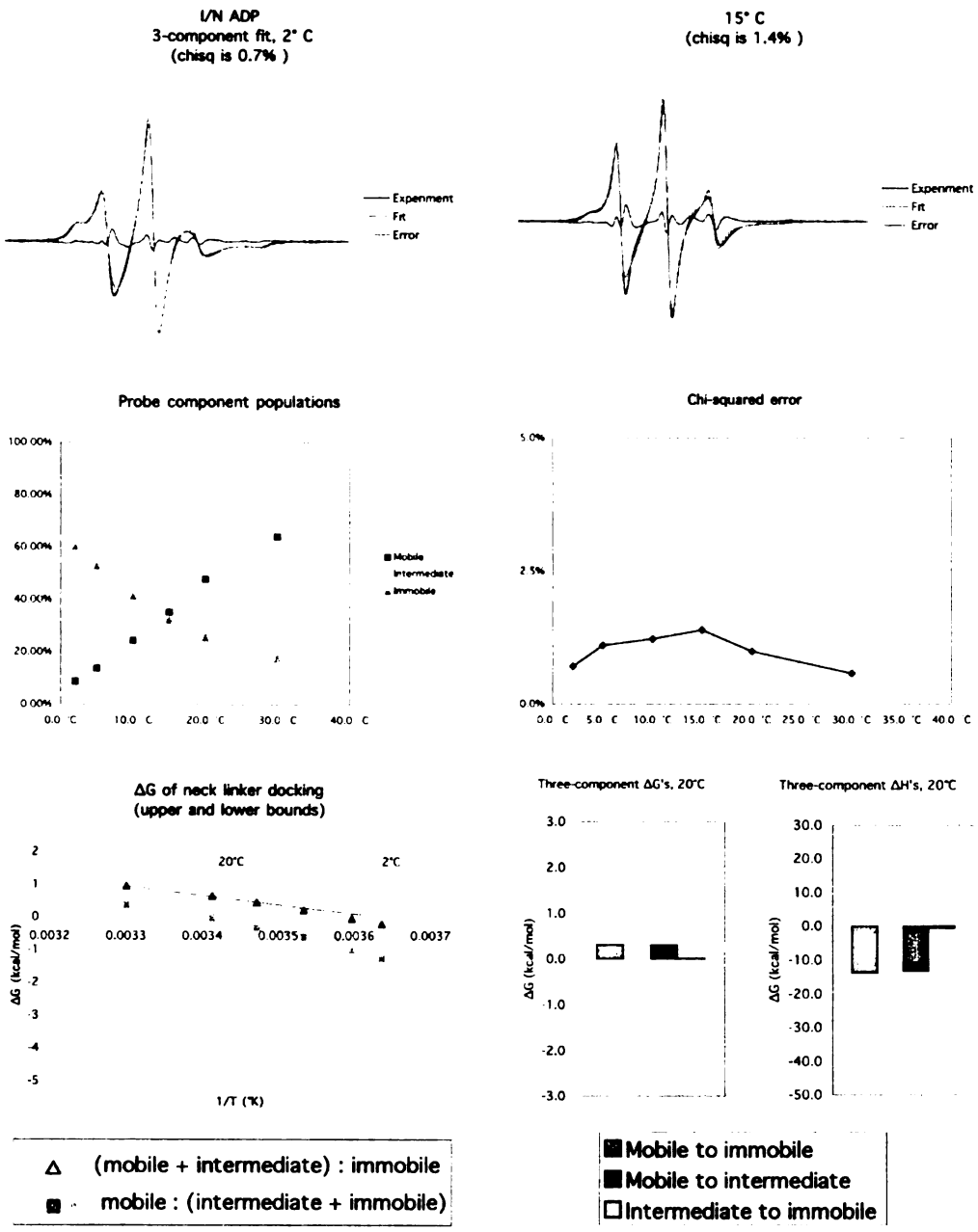

I/N ADP
3-component fit, 2° C
(chisq is 0.7% )
Experiment
Fit
Error
15° C
(chisq is 1.4% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 °C
10.0 °C
20.0 °C
30.0 °C
40.0 °C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 °C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 °C
35.0 °C
40.0 °C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
1/T (°K)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 7.

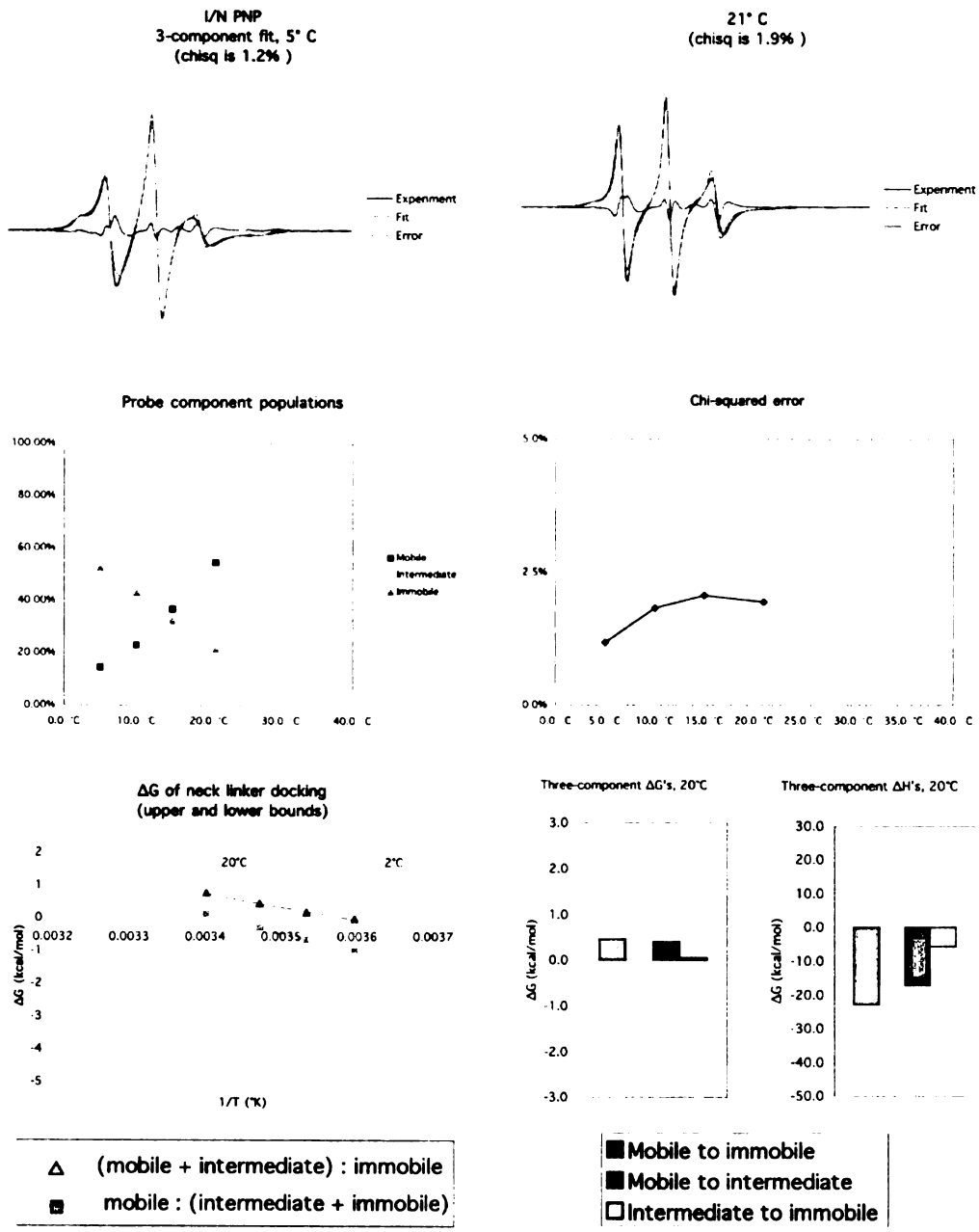

I/N PNP
3-component fit, 5° C
(chisq is 1.2% )
Experiment
Fit
Error
21° C
(chisq is 1.9% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 °C
10.0 °C
20.0 °C
30.0 °C
40.0 °C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 °C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 °C
35.0 °C
40.0 °C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
1/T (°K)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 8.

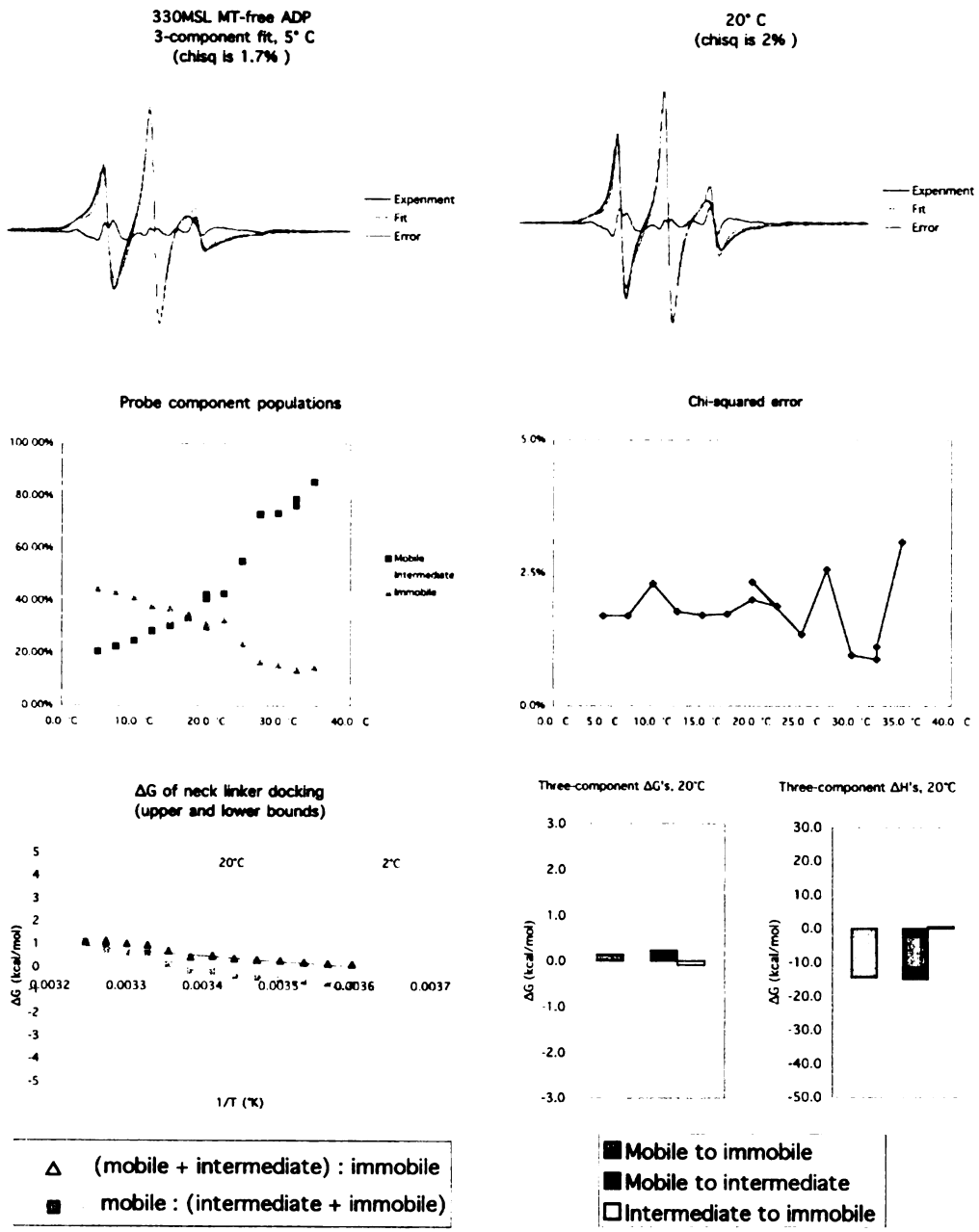

330MSL MT-free ADP
3-component fit, 5° C
(chisq is 1.7% )
Expenment
Fit
Error
20° C
(chisq is 2% )
Expenment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 °C
10.0 °C
20.0 °C
30.0 °C
40.0 °C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 °C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 °C
35.0 °C
40.0 °C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
1/T (°K)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 9.

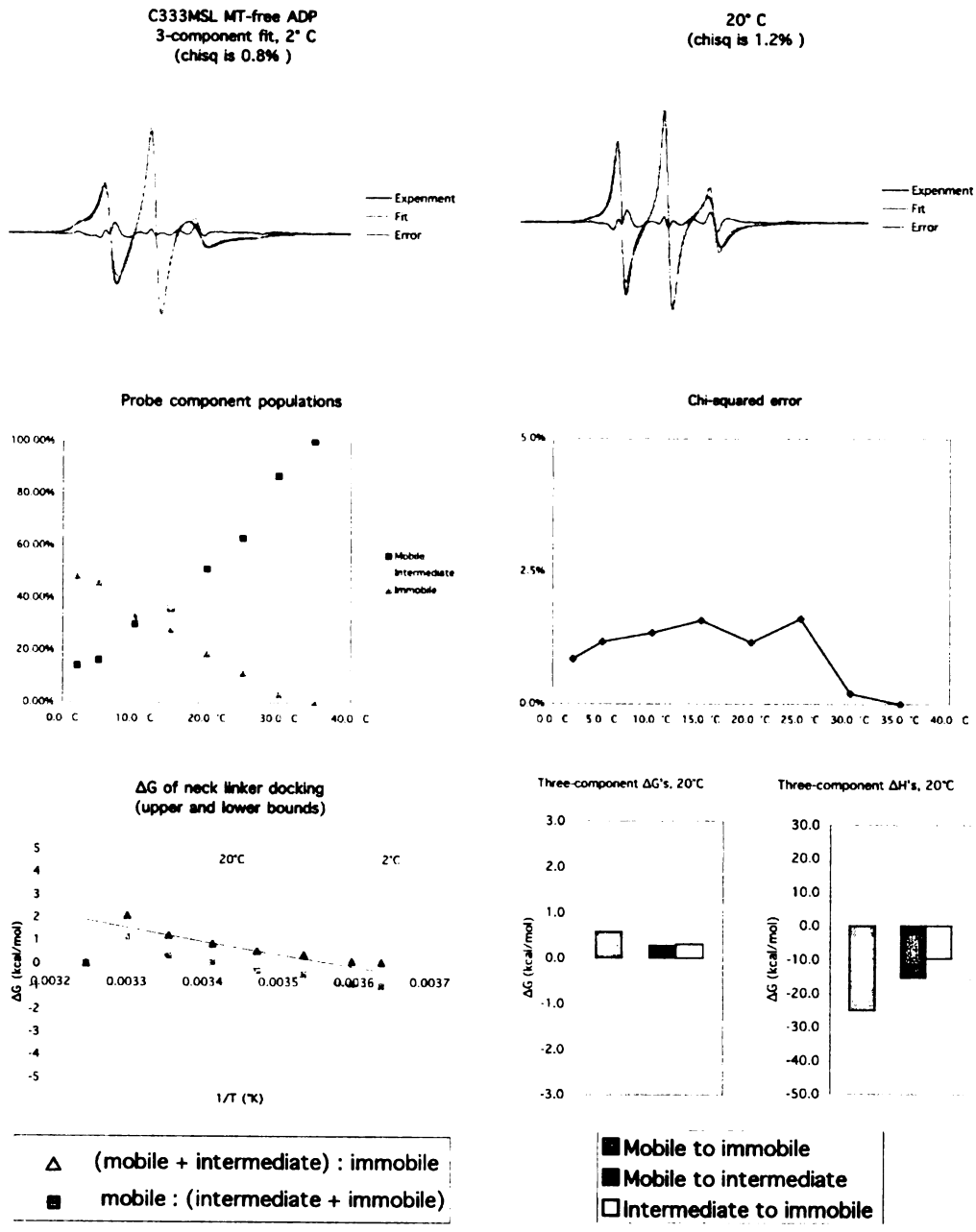

C333MSL MT-free ADP
3-component fit, 2° C
(chisq is 0.8% )
Experiment
Fit
Error
20° C
(chisq is 1.2% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 C
10.0 C
20.0 °C
30.0 C
40.0 C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 C
5.0 C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 °C
35.0 °C
40.0 C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
5
4
3
2
1
0
-1
-2
-3
-4
-5
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
1/T (°K)
Three-component ΔG's, 20°C
ΔG (kcal/mol)
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
Three-component ΔH's, 20°C
ΔG (kcal/mol)
30.0
20.0
10.0
0.0
-10.0
-20.0
-30.0
-40.0
-50.0
Δ (mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 10.

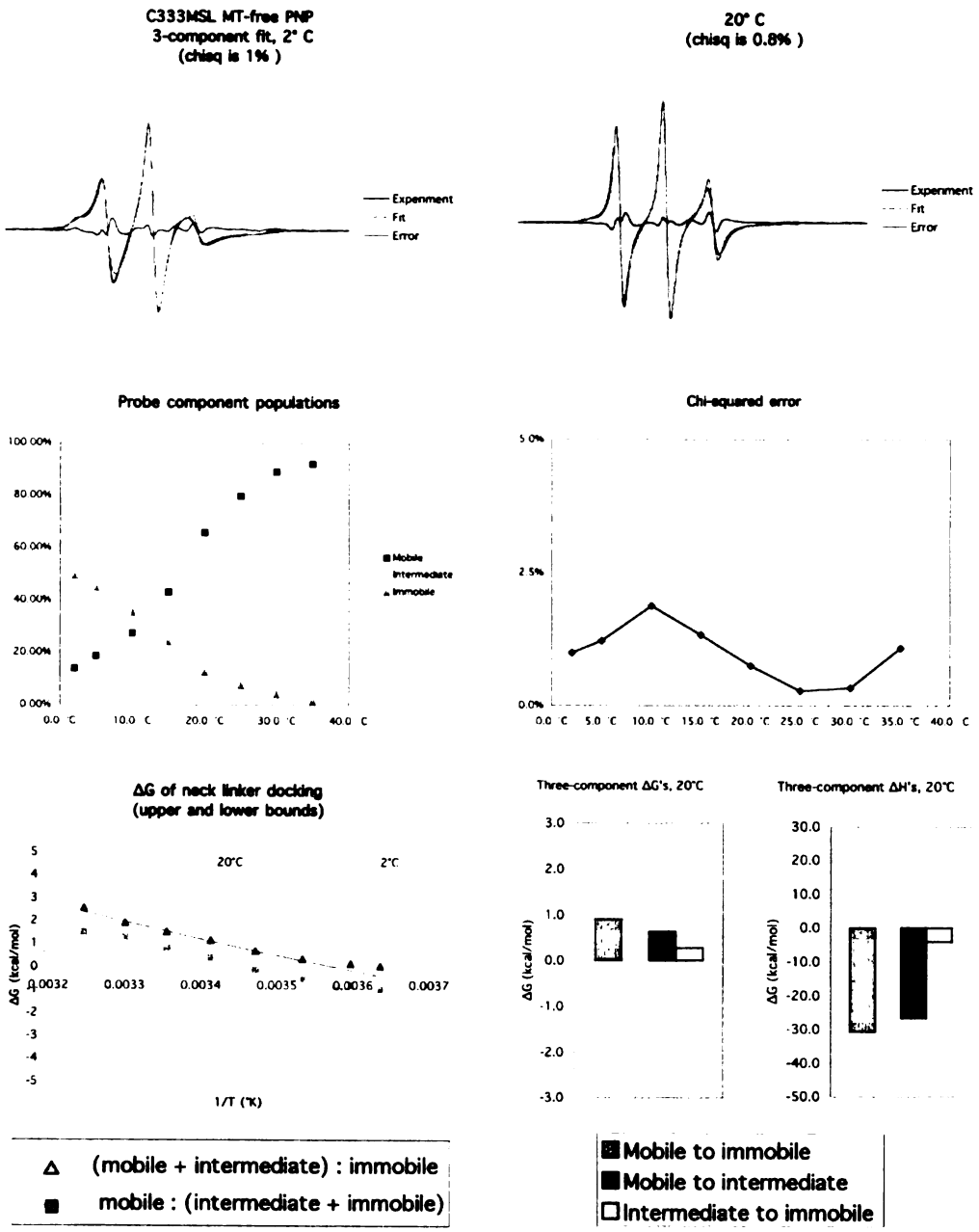

C333MSL MT-free PNP
3-component fit, 2° C
(chisq is 1% )
Experiment
Fit
Error
20° C
(chisq is 0.8% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 °C
10.0 °C
20.0 °C
30.0 °C
40.0 °C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 °C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 °C
35.0 °C
40.0 °C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
1/T (°K)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 11.

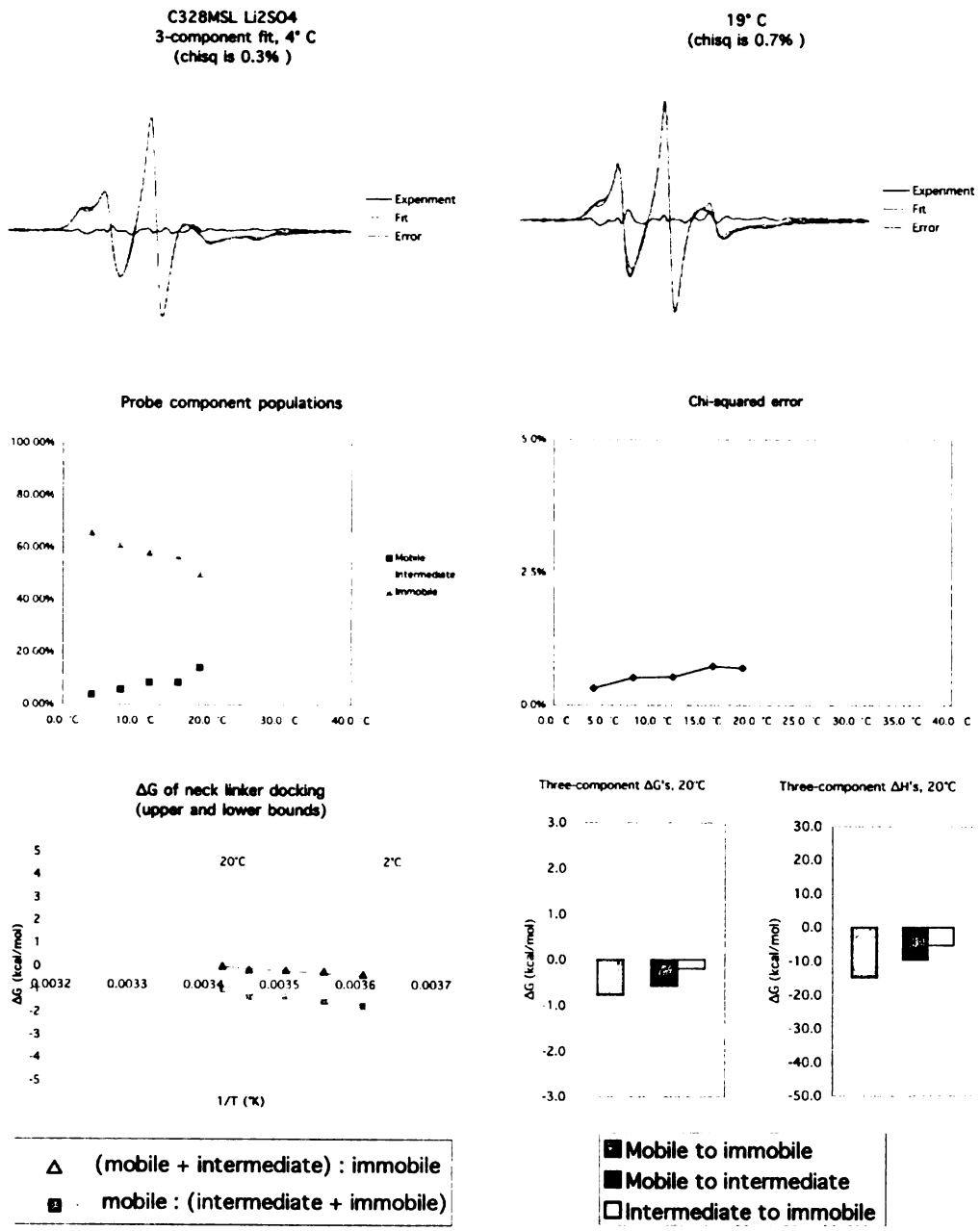

C328MSL Li2SO4
3-component fit, 4° C
(chisq is 0.3% )
Experiment
Fit
Error
19° C
(chisq is 0.7% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 °C
10.0 °C
20.0 °C
30.0 °C
40.0 °C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 °C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 °C
35.0 °C
40.0 °C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
1/T (°K)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
ΔH (kcal/mol)
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 12.

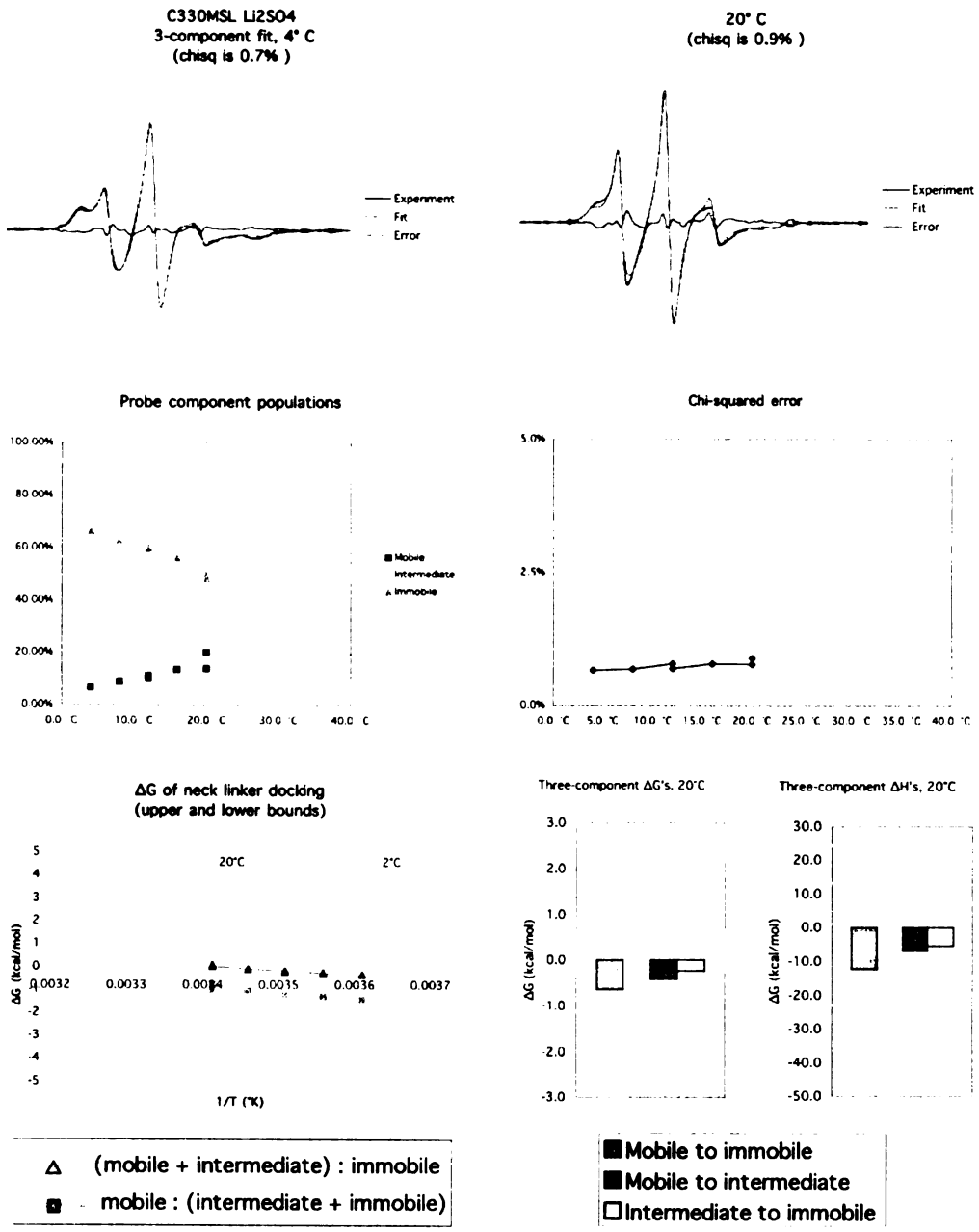

C330MSL Li2SO4
3-component fit, 4° C
(chisq is 0.7% )
Experiment
Fit
Error
20° C
(chisq is 0.9% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 C
10.0 C
20.0 C
30.0 C
40.0 C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 °C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 °C
35.0 °C
40.0 °C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
5
4
3
2
1
0
-2
-3
-4
-5
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
1/T (°K)
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
30.0
20.0
10.0
0.0
-10.0
-20.0
-30.0
-40.0
-50.0
Δ (mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 13.

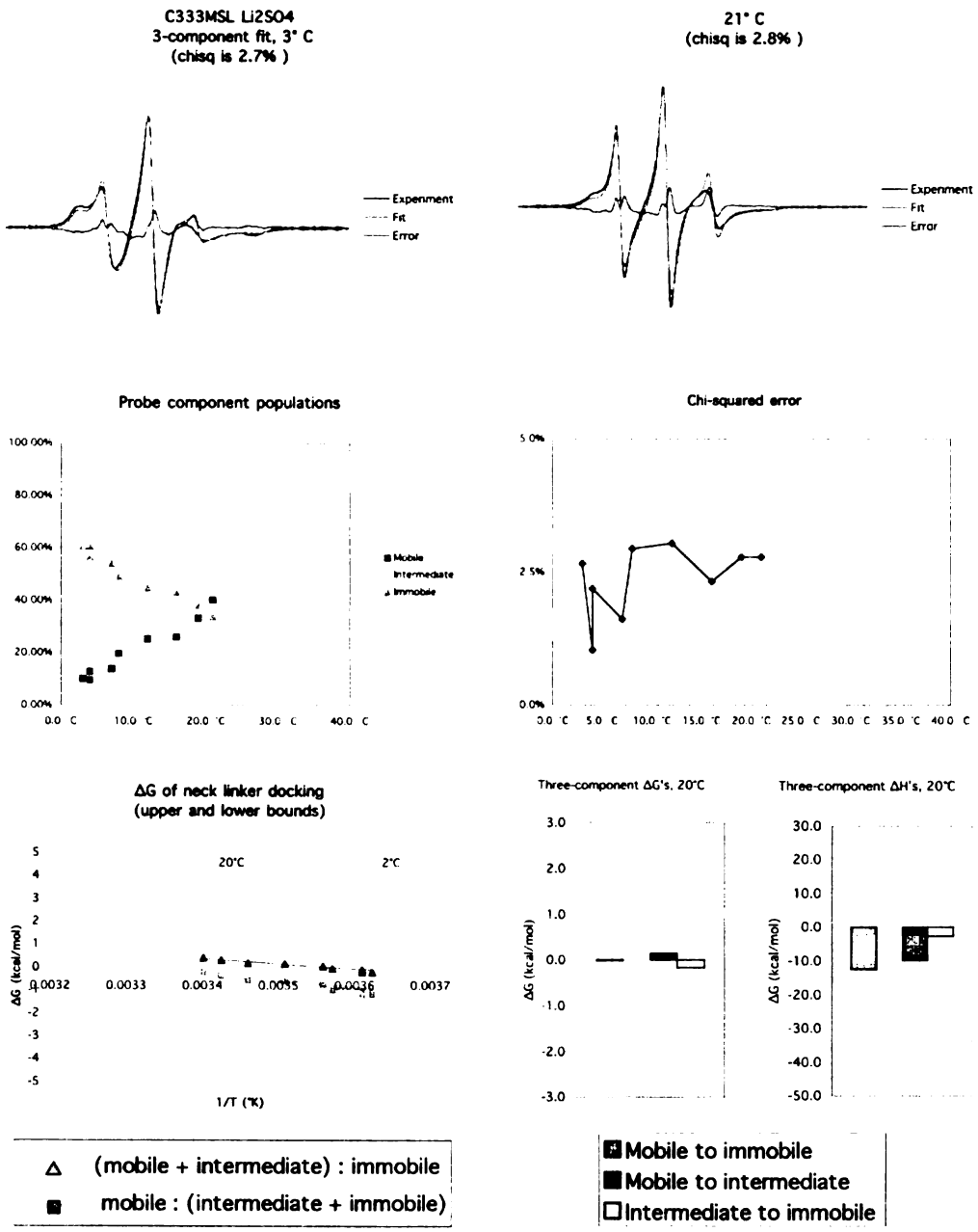

C333MSL Li2SO4
3-component fit, 3° C
(chisq is 2.7% )
Expenment
Fit
Error
21° C
(chisq is 2.8% )
Expenment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 C
10.0 C
20.0 C
30.0 C
40.0 C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 C
5.0 C
10.0 C
15.0 C
20.0 C
25.0 C
30.0 C
35.0 C
40.0 C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
1/T (°K)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 14.

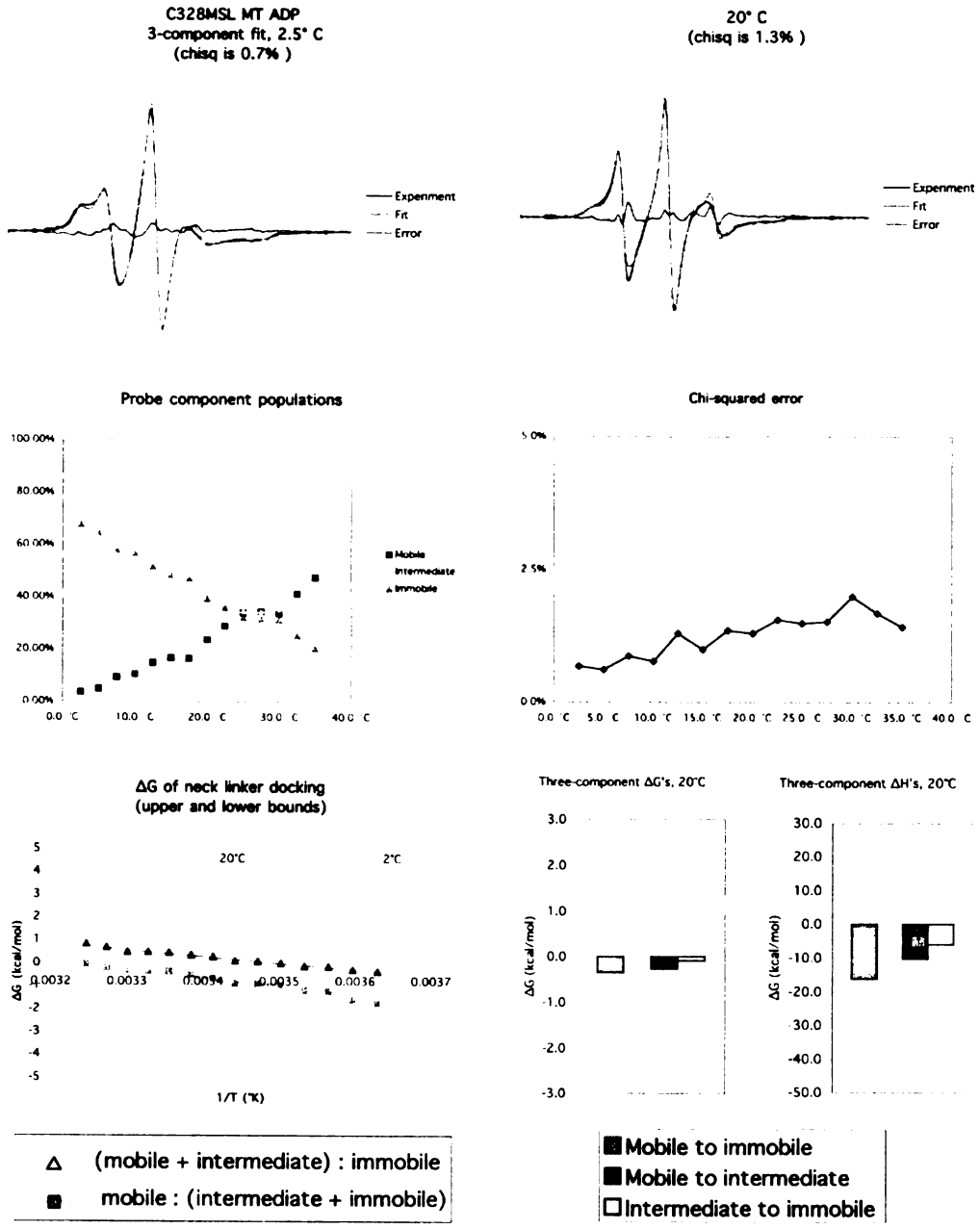

C328MSL MT ADP
3-component fit, 2.5° C
(chisq is 0.7% )
Experiment
Fit
Error
20° C
(chisq is 1.3% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 °C
10.0 C
20.0 C
30.0 C
40.0 C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 °C
5.0 C
10.0 °C
15.0 °C
20.0 °C
25.0 C
30.0 °C
35.0 °C
40.0 C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
5
4
3
2
1
0
-1
-2
-3
-4
-5
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
1/T (°K)
Three-component ΔG's, 20°C
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
ΔG (kcal/mol)
Three-component ΔH's, 20°C
30.0
20.0
10.0
0.0
-10.0
-20.0
-30.0
-40.0
-50.0
ΔG (kcal/mol)
Δ (mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 15.

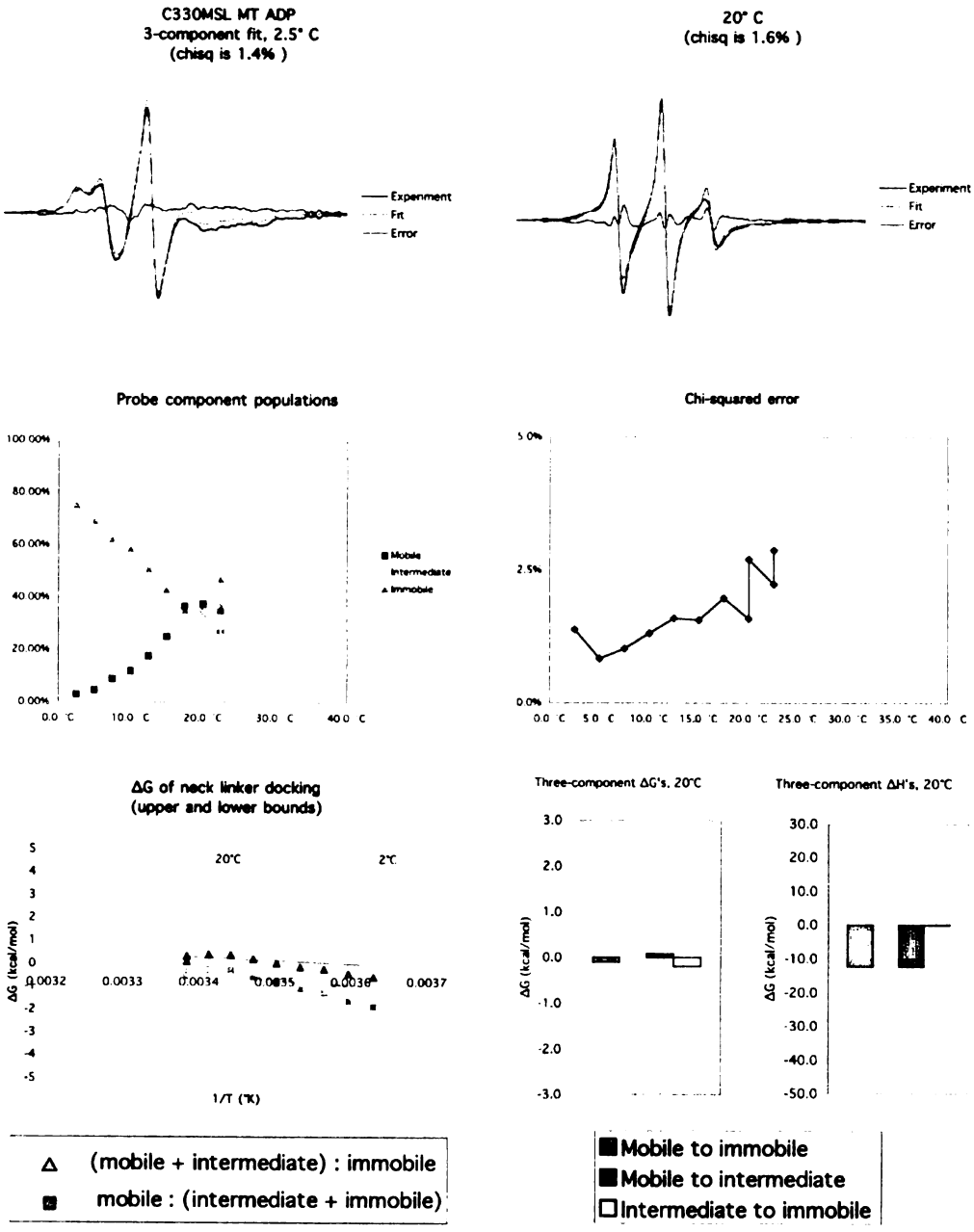

C330MSL MT ADP
3-component fit, 2.5° C
(chisq is 1.4% )
Experiment
Fit
Error
20° C
(chisq is 1.6% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 °C
10.0 °C
20.0 °C
30.0 °C
40.0 °C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 °C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 °C
35.0 °C
40.0 °C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
5
4
3
2
1
0
-1
-2
-3
-4
-5
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
1/T (°K)
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Three-component ΔG's, 20°C
ΔG (kcal/mol)
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
Three-component ΔH's, 20°C
ΔG (kcal/mol)
30.0
20.0
10.0
0.0
-10.0
-20.0
-30.0
-40.0
-50.0
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 16.

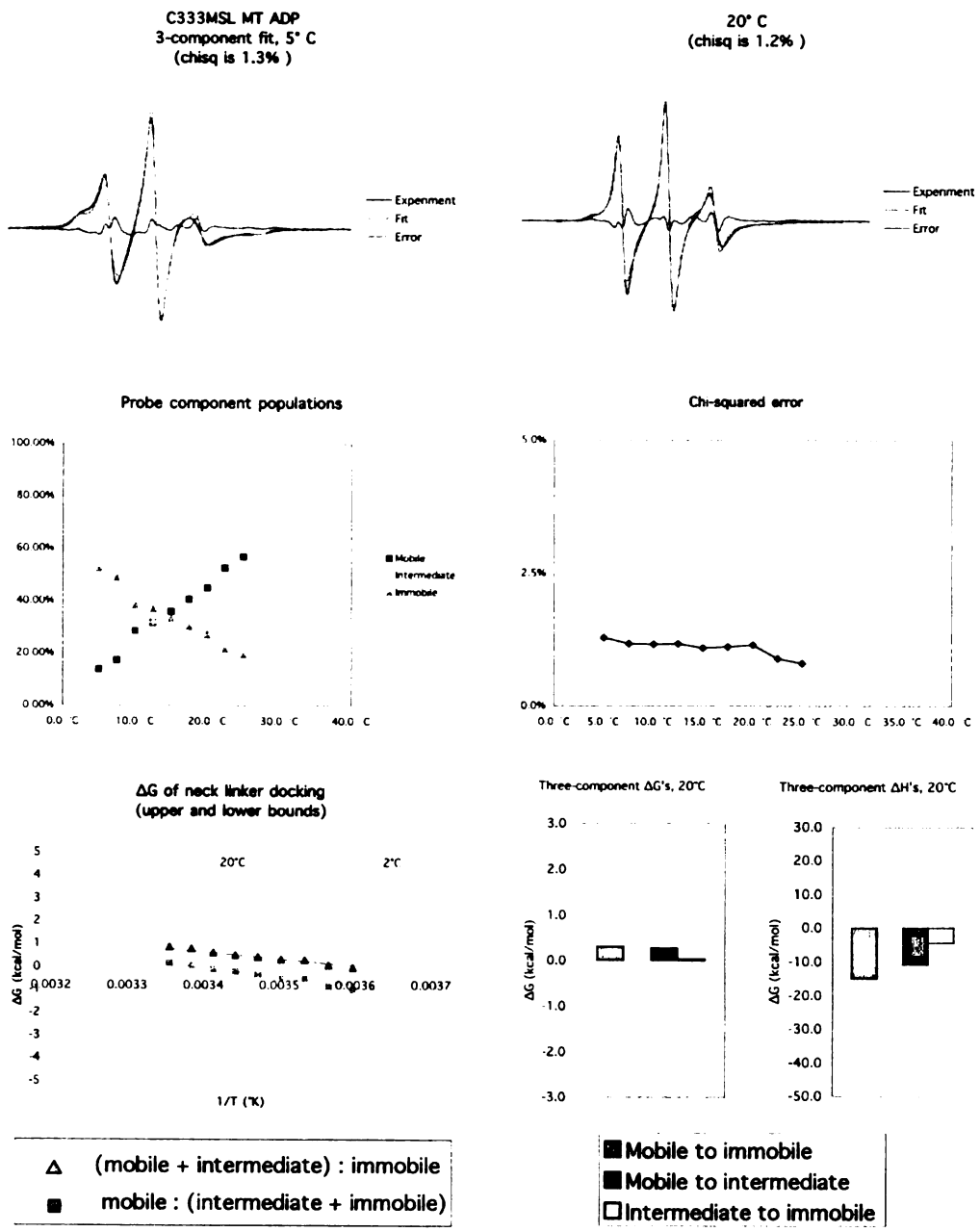

C333MSL MT ADP
3-component fit, 5° C
(chisq is 1.3% )
Expenment
Fit
Error
20° C
(chisq is 1.2% )
Expenment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 °C
10.0 C
20.0 C
30.0 C
40.0 C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 C
35.0 °C
40.0 C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
1/T (°K)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 17.

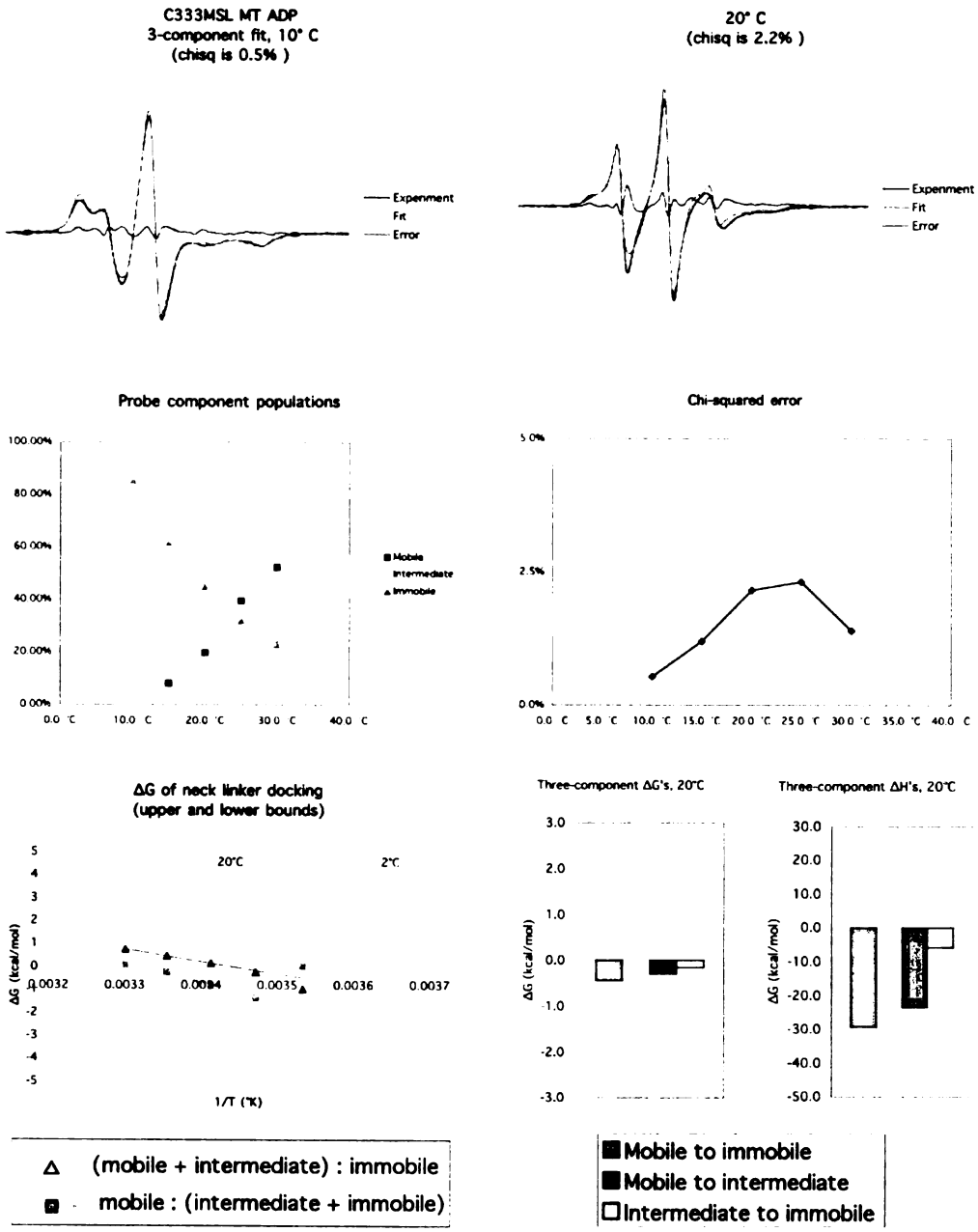

C333MSL MT ADP
3-component fit, 10° C
(chisq is 0.5% )
Expenment
Fit
Error
20° C
(chisq is 2.2% )
Expenment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 °C
10.0 °C
20.0 °C
30.0 °C
40.0 °C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 °C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 °C
35.0 °C
40.0 °C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
5
4
3
2
1
0
-1
-2
-3
-4
-5
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
1/T (°K)
Δ (mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Three-component ΔG's, 20°C
ΔG (kcal/mol)
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
Three-component ΔH's, 20°C
ΔG (kcal/mol)
30.0
20.0
10.0
0.0
-10.0
-20.0
-30.0
-40.0
-50.0
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 18.

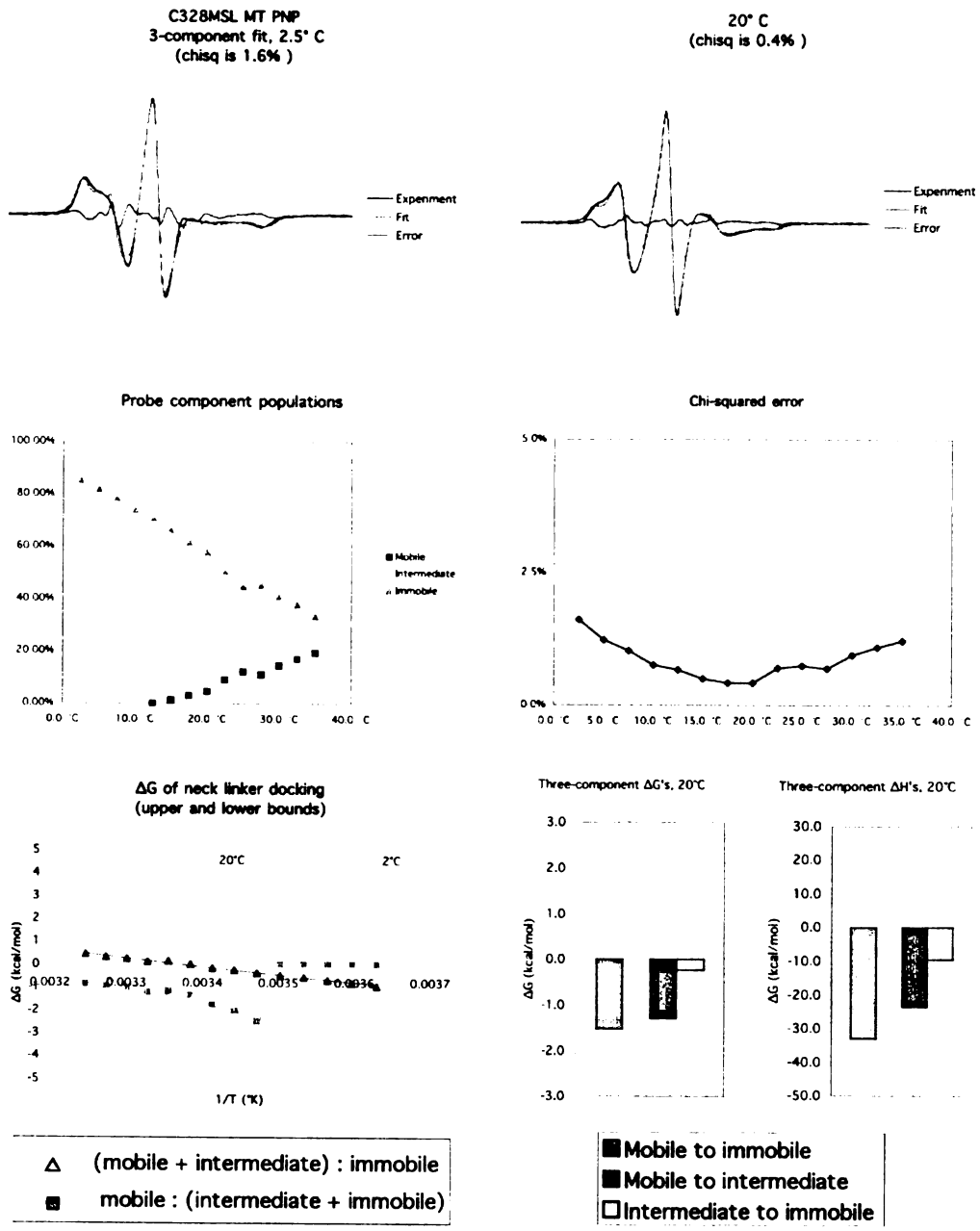

C328MSL MT PNP
3-component fit, 2.5° C
(chisq is 1.6% )
Expenment
Fit
Error
20° C
(chisq is 0.4% )
Expenment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 °C
10.0 °C
20.0 °C
30.0 °C
40.0 °C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 °C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 °C
35.0 °C
40.0 °C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
1/T (°K)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
Δ (mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 19.

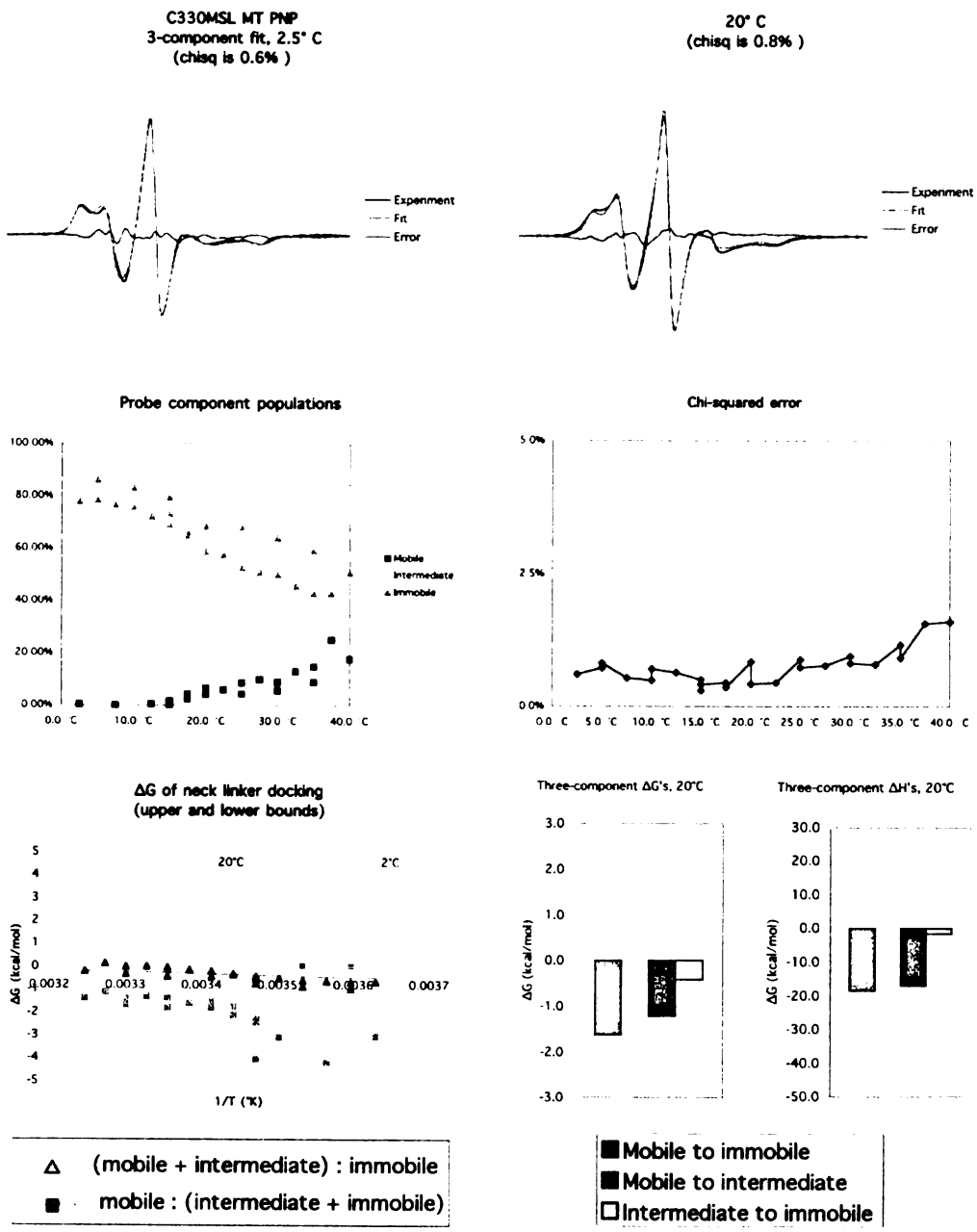

C330MSL MT PNP
3-component fit, 2.5° C
(chisq is 0.6% )
Experiment
Fit
Error
20° C
(chisq is 0.8% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 C
10.0 °C
20.0 °C
30.0 C
40.0 C
Mobile
Intermediate
Immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 °C
35.0 °C
40.0 C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
1/T (°K)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
ΔG (kcal/mol)
Δ (mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 20.

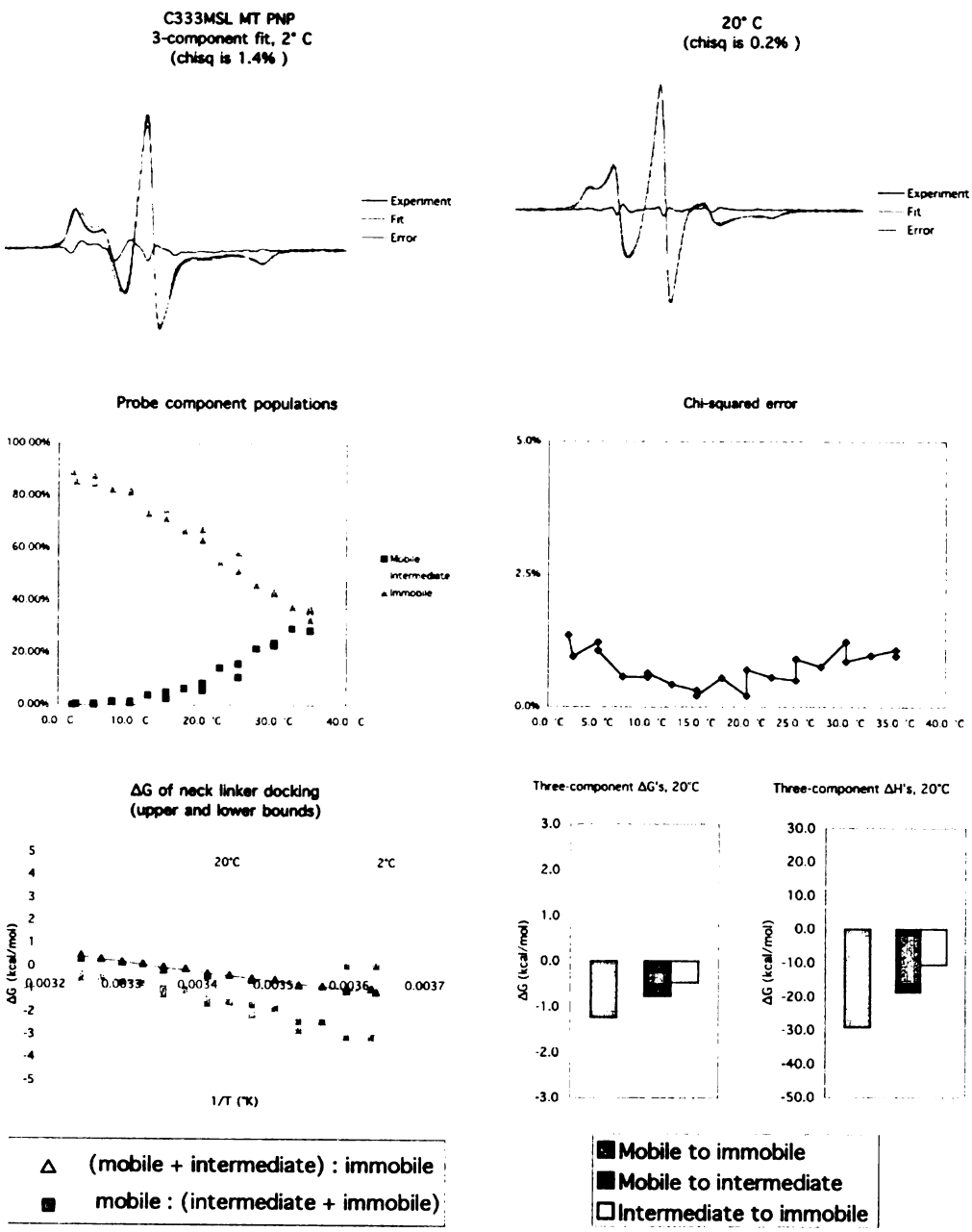

C333MSL MT PNP
3-component fit, 2° C
(chisq is 1.4% )
Experiment
Fit
Error
20° C
(chisq is 0.2% )
Experiment
Fit
Error
Probe component populations
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
0.0 C
10.0 C
20.0 C
30.0 C
40.0 C
Mobile
intermediate
immobile
Chi-squared error
5.0%
2.5%
0.0%
0.0 °C
5.0 °C
10.0 °C
15.0 °C
20.0 °C
25.0 °C
30.0 °C
35.0 °C
40.0 °C
ΔG of neck linker docking
(upper and lower bounds)
20°C
2°C
ΔG (kcal/mol)
1/T (°K)
0.0032
0.0033
0.0034
0.0035
0.0036
0.0037
(mobile + intermediate) : immobile
mobile : (intermediate + immobile)
Three-component ΔG's, 20°C
Three-component ΔH's, 20°C
ΔG (kcal/mol)
Mobile to immobile
Mobile to intermediate
Intermediate to immobile

Figure 21.

# Appendix three:
# Deriving solvation free energies from the subvolume partition function generated by SWYZLE

Below we briefly sketch an approach for calculating the solvation free energy of a particle using the first-shell partition function calculated by SWYZLE. This derivation represents a work in progress; i.e. not all of the approximations (such as separability of integrals) applied in it may prove to be workable. Also, some issues related to the difference between (N,V,T) and (N, P, T) ensembles remain unresolved (below we show our work for an (N, V, T) ensemble). Nevertheless, our hope is that by "co-evolving" these analytical formulas with enhancements in the models applied to SWYZLE, we will soon arrive at an accurate and powerful new technique for calculating and understanding solvation free energies.

We begin with Ben-Naim's expression for the standard free energy of solvation (Ben-Naim and Marcus 1984):

$$e^{-\bar{\mu}^{0}_{solute}/kT} = \frac{Q(N_{water}, N_{solute} = 1, V, T; R_0)}{Q(N_{water}, N_{solute} = 0, V, T)}$$

It is possible to expand the numerator into an expression that explicitly represents the waters inside a shell of volume $v$, and those outside it. This is done as a sum over all possibilities $m$ for the number of waters within the shell (with $N$-$m$ waters remaining outside) (see (SotoCampos, Corti et al. 1998) for a very well-written derivation describing a system related to the one described here). The integrals for these two groups of waters are then carried out over their respective volumes, as shown below:

$$Q(N_{water}, N_{solute} = 1, V, T; R_0) = \sum_{m=0}^{N_{water}} \frac{1}{(N_{water} - m)!m!\gamma^{N_{water}}} \left[ \int_{V-v} e^{-\beta E_{ext}} dR^{N_{water}-m} \right] \left[ \int_{v} e^{-\beta (E^{sol}_{int} + E^{sol}_{surf})} dR^{m} \right]$$

Here there are three energy terms: $E^{sol}_{int}$, the interaction energy of all particles within the solvation shell (waters and solute); $E_{ext}$, the interaction of all $N$-$m$ "bulk" waters with each other; and $E^{sol}_{surf}$, the interaction *between* the bulk waters and the molecules interior

to the shell. The term $\gamma$ represents the integrated momentum partition function for a single water molecule. Collecting terms together into a useful form, we rewrite the above expression as:

$$Q(N_{water}, N_{solute} = 1, V, T; R_0)$$

$$= \sum_{m=0}^{N_{water}} \left[ \frac{1}{(N_{water} - m)! \gamma^{N_{water} - m}} \int_{V-v} e^{-\beta E_{ext}} dR^{N_{water} - m} \right] \left[ \frac{1}{m! \gamma^m} \int_v e^{-\beta (E_{int}^{sol} + E_{surf}^{sol})} dR^m \right]$$

The above expression can now be rewritten *if* the two integrations within the brackets can be carried out separately. Separability of the integrals is motivated by the fact that the waters outside of $v$ should behave mostly like "bulk"waters, unaffected by the presence of the solute—which will be true for large enough $v$. However, even in the case of sufficiently large $v$ it remains unclear to me how clean the separability of the integrals really is. This is a fundamental question we hope to clarify in future theoretical and computational investigations.

If the two integrals *can* be separated, then the left-hand bracketed expression is immediately recognized as the partition function of $N$-$m$ bulk waters, in the reduced volume $V$-$v$. :

$$\left[ \frac{1}{(N_{water} - m)! \gamma^{N_{water} - m}} \int_{V-v} e^{-\beta E_{ext}} dR^{N-m} \right] = Q(N_{water} - m, N_{solute} = 0, V - v, T; R_0)$$

This expression can be converted to the equivalent partition function for $N_{water}$ waters by multiplying by a chemical potential term, as follows:

$$Q(N_{water} - m, N_{solute} = 0, V - v, T; R_0) = Q(N_{water}, N_{solute} = 0, V - v, T; R_0) e^{m\mu_{wat}/kT}$$

Meanwhile, the *second* bracketed term in the original partition function expression for $Q(N_{water}, N_{solute} = 1, V, T; R_0)$ is an integral over the first-shell volume $v$. This is precisely the quantity $Z_m$ (for $m$ shell waters) calculated by SWYZLE (see the Derivation in Chapter 3):

$$\left[ \frac{1}{m! \gamma^m} \int_v e^{-\beta (E_{int}^{sol} + E_{bulk})} dR^m \right] = Z_m$$

We note here that the interaction energy of the shell waters to the bulk ( $E_{surf}^{sol}$ ) is approximated as an *average* energy(independent of shell-water configuration), $E_{bulk}$, which constitutes an additional assumption in our derivation. As discussed in the Methods and Results of Chapter 3, SWYZLE has the capability to incorporate conformation-dependent values of $E_{surf}^{sol}$. Here, however, we will not discuss this possibility here further, except to note that adding a configuration dependence to $E_{surf}^{sol}$ could allow more accurate representations of shell-water surface energies in future work.

Substituting our new expressions into the bracketed components of our solute partition function expression for $Q(N_{water}, N_{solute}=1, V, T; R_0)$, we get:

$$Q(N_{water}, N_{solute}=1, V, T) = \sum_{m=0}^{N_{water}} Q(N_{water}, N_{solute}=0, V-v, T) e^{m\mu_{wat}/kT} Z_m$$
$$= Q(N_{water}, N_{solute}=0, V-v, T)\Xi^{shell}$$

and substituting this expression back into the original expression for the standard free energy of solvation yields:

$$e^{-\bar{\mu}^0_{solute}/kT} = \frac{Q(N_{water}, N_{solute}=1, V, T; R_0)}{Q(N_{water}, N_{solute}=0, V, T)}$$
$$= \frac{Q(N_{water}, N_{solute}=0, V-v, T)}{Q(N_{water}, N_{solute}=0, V, T)}\Xi^{shell}$$

The partition function ratio in this last expression is for two systems of pure water that have slightly different volumes, but are otherwise identical. The free energy difference between them is therefore just the work of expanding the system, so the following expression is true:

$$\frac{Q(N_{water}, N_{solute}=0, V-v, T)}{Q(N_{water}, N_{solute}=0, V, T)} = \frac{e^{-A(N,V-v,T)/kT}}{e^{-A(N,V,T)/kT}} = e^{-[A(N,V,T-v)-A(N,V,T)]/kT} = e^{-pv/kT}$$

We can therefore substitute this expression back into our result for the standard free energy:

$$e^{-\mu^0_{solute}/kT} = \frac{Q(N_{water}, N_{solute}=1, V, T; R_0)}{Q(N_{water}, N_{solute}=0, V, T)}$$
$$= e^{-pv/kT}\Xi^{shell}$$

which gives the following very simple result:

$$\bar{\mu}^{0}_{solute} = pv - kT\ln\left(\Xi^{shell}\right)$$

The process that this equation represents is illustrated schematically in Figure 1. First, in Figure 1B a void is opened up in pure water—but notably, the structure of the water immediately surrounding the void *is not perturbed at all* in this nonphysical, but thermodynamically meaningful, step. Conveniently, this feature allows us to write characterize this step as having free energy equal to the work (*pv*) of expanding the void.

This void formation, as we define it, contrasts with the free energy of cavity formation as defined for example in Scaled Particle Theory, where the water *responds* to the presence of a void next to it. There is a reason for us to define a void formation in our differing, somewhat unconventional manner. The fundamental assumption in our derivation here is that the bulk waters outside the first shell are *not* perturbed significantly in structure or energetics by the presence of a solute. We take advantage of this assumed property by keeping the ensemble of bulk waters the *same* in every step of our derivation.

In the final steps (Figures 1C and 1D), a solute is introduced into the void, and waters from the bulk region are allowed back into the void, interacting with the solute. This process is what is modeled by the SWYZLE calculation, and so is characterized by a free energy of $-kT\ln\left(\Xi^{shell}_{solute}\right)$. This step completes a thermodynamic cycle representing the insertion of a solute into a bulk water system.

While we have presented our derivation for the constant-*volume* ensemble, the same arguments should also apply to a constant-pressure ensemble—which brings up a final, significant point. Introducing a solute into a constant-volume environment, as we describe here, may result in significant non-local contributions to the solvation energy (i.e. by waters far from the first solvation shell). This is because *every* water in the system, in the constant-volume case, will have a reduced accessible volume once the solute is inserted—potentially leading to a cumulative energetic effect. Thus, the constant-*pressure* ensemble is really the appropriate one to derive solvation energies

from a localized shell method like SWYZLE—because the system volume is free to relax following solute insertion. Thus, the "bulk" waters in the constant pressure case are able to explore the same volume either with or without solute present. We are currently reworking the above derivation in the constant pressure ensemble, to obtain a more formally correct result.

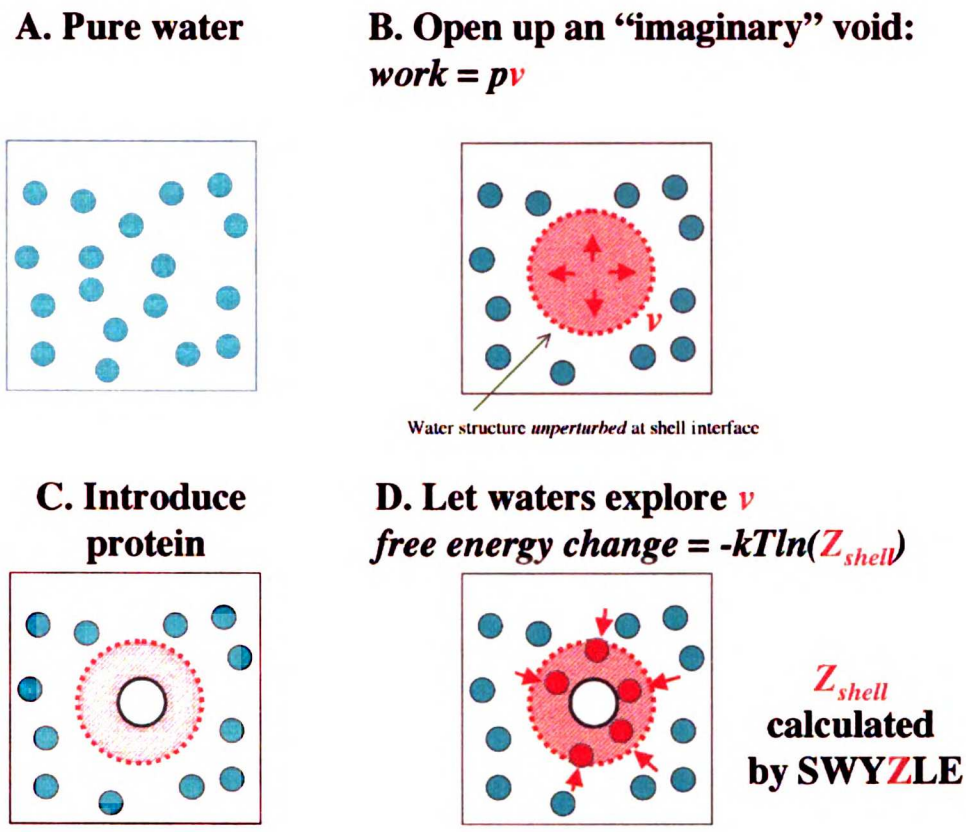



Figure 1. Schematic illustration of the cycle used to derive the free energy of solvation from the partition function calculated by SWYZLE.

# Appendix four:
# Retrieving structural and energetic quantities from the partition function

The SWYZLE formula of Chapter 3 directly calculates the partition function of a solvent shell system. This allows a great deal of structural and energetic detail to be extracted, at relatively low computational cost. Here we describe the generic formula that allows the derivation of such quantities from the SWYZLE calculation.

Let $F(\vec{x}_i)$ be any quantity of interest (energetic or structural) that depends on the coordinates $\vec{x}_i$ of the solvent-shell system components. Then the average value $\overline{F}(\vec{x}_i)$ of this quantity can be obtained in the following way. First, introduce the artificial parameter $\lambda$ into the definition of shell-water energy function:

$$E'(\vec{x}_i,\lambda) = E(\vec{x}_i) + \lambda F(\vec{x}_i)$$

so that at $\lambda = 0$, the system energy $E'(\vec{x}_i,0)$ is unperturbed. Plugging this energy definition into the partition function definition gives the following definition of a 'perturbed' Z:

$$Z(\lambda) = \sum_i e^{\left[E(\vec{x}_i)+\lambda F(\vec{x}_i)\right]/kT}$$

Taking the derivative with respect to $\frac{\lambda}{kT}$ gives:

$$\frac{\delta Z(\lambda)}{\delta(\lambda/kT)} = \sum_i F(\vec{x}_i) e^{\left[E(\vec{x}_i)+\lambda F(\vec{x}_i)\right]/kT}$$

so that taking the value of this derivative at $\lambda = 0$, and dividing by Z results in the desired quantity:

$$\frac{1}{Z}\frac{\delta Z(\lambda)}{\delta(\lambda/kT)}\bigg|_{\lambda=0} = \frac{\sum_i F(\vec{x}_i) e^{\left[E(\vec{x}_i)\right]/kT}}{\sum_i e^{\left[E(\vec{x}_i)\right]/kT}} = \overline{F}(\vec{x}_i)$$

Applying this formula within the matrix formalism of SWYZLE is straightforward, because Z can be directly calculated, and the derivative $\frac{\delta Z(\lambda)}{\delta(\lambda/kT)}$ can be found by a finite difference method. To calculate $\frac{\delta Z(\lambda)}{\delta(\lambda/kT)}$, all that needs to be done is to compute $Z(\lambda)$ for two values of $\lambda$ (0 and a very small, nonzero value) and divide by the difference of the $\lambda$ values.

Sample applications of this formula (within the MB water shell framework described in Chapter 3) have produced excellent results, and will be described in a future publication.

# Appendix Five:
# Summary of useful crystallization conditions

## Optimizing PEG-containing crystallization conditions

Notably, all PEG conditions examined with kinesin could be optimized by adding between 1%-5% isopropyl alcohol. This in many cases made the difference between smallish, "urchinlike" clusters of clusters (much too small to diffract), and the larger, diffracting single "blades."

## Crystallizing kinesin in PEG at a non-acidic pH

One of the potential criticisms of the original crystallization conditions for wild-type human kinesin K349 was that at the low pH of the crystallization buffer (pH 4.6), unusual protonations might occur on some of the sidechains. Therefore, the following condition (discovered during screening trials) is of interest, because crystals appeared with very similar morphology ("urchin-like" crystals) as are found in the original (pH 4.6) conditions. These new crystals have not been positively verified as protein crystals, but I strongly suspect they are.

~4% PEG 8000, 0.05M TrisHCl (pH 8.5), 100mM KCl

This condition crystallized wild-type kinesin under the "standard" buffer/crystallization conditions, i.e. ~10 mg/mL kinesin, 5mM ATP or ADP, 20mM MgCl, other buffer conditions as in Chapter 1.

Another similar "lead" condition (crystals were seen, but not verified as protein) is:

21% PEG 1000, 25mM HEPES (pH 7.0), 150mM KCl

### Crystallizing kinesin in the presence of different nucleotides

The following condition crystallized kinesin in the presence of AMPPCP (as described in Appendix 1).

~12% PEG 2000 monomethylether, ~50-100mM sodium acetate pH 4.6, ~100-150mM ammonium sulfate, ~100-150mM KCl

# Appendix six:
# Quantifying kinesin's switch II movement using GEM

Here I present a simple script, to demonstrate how the program GEM(Browner, Fauman et al. 1992) can be used to quantify the movement of the switch II domain in kinesin. This script is meant as an example, to assist others in making future comparisons—I found GEM to be a truly useful tool in this case, and hope to encourage its continued use.

The essential features of the script are that it reads in two pdb files (previously globally aligned, although GEM could also have done this), selects the backbone atoms in the switch II region, and executes the "axis" command which gives movement of switch II in "principle axis" terms. The important output of the script is the angular rotation of the selected domain (highlighted in gray), and also the movement of its center of mass (the output line immediately following). Note that the "newchain" command executed after reading the first PDB file is simply to get rid of the chain A and B identifiers present in this, but not the second, PDB file. A control "axis" command is also executed, to show that the central beta sheet portion of kinesin does not move.

File "k349_helix.in"

```
out k349helix.log
#   /k349_rms_breakdown/kin_kull_aln.pdb
newchain " "
apply replace
#   /k349_rms_breakdown/bindividual2_newadp.pdb
atom CA N C O
range @betasheet.dat
axes
reset range
range @swIIhelix.dat
axes
super
reset atom range
apply replace
range @swIIhelix.dat
atom CA N C O
pdbrms
```

File "betasheet.dat"

| | |
|---|---|
| 131 | 138 |
| 204 | 210 |
| 227 | 233 |
| 81 | 84 |
| 298 | 300 |
| 11 | 16 |

File "swIIhelix.dat"

| | |
|---|---|
| 258 | 270 |

Command to execute: "/public/apps/bin/gem < k349_helix.in > k349_helix.out"

Output file "k349_helix.out"

```
Good evening, charless

This is the program Gem  Multi-faceted Protein Examiner
Documentation available:  /oscar/faumanv/gem/gem.doc
Enter HELP or ? for more information.  Enter NEWS for latest functions (8/19/92)
Gem:          K349HELIX.LOG is open for output

.

Apply:        Thinking...
              1) 140 of 2733 atoms selected in:  /K349_RMS_BREAKDOWN/KIN_KULL_ALN.PDB [pr]
              2) 142 of 2709 atoms selected in:  /K349_RMS_BREAKDOWN/BINDIVIDUAL2_NEWADP.PDB [se]
              For PDB file 1 -  /K349_RMS_BREAKDOWN/KIN_KULL_ALN.PDB
              center of mass located at (  69.61,   46.26,   86.13)
              Total mass      = 1890.98 grams/mole
              Components of Rgyr =   8.98,    5.42,    2.58 Angstroms
              For PDB file 2 -  /K349_RMS_BREAKDOWN/BINDIVIDUAL2_NEWADP.PDB
              center of mass located at (  69.57,   46.18,   86.16)
              Total mass      = 1922.98 grams/mole
              Components of Rgyr =   9.21,    5.47,    2.57 Angstroms
              Angles to last axes =   1.24,    0.68,    1.14 Degrees
              C.O.M. moved    =  0.093

.

Axes:         Thinking...
              1) 52 of 2733 atoms selected in:  /K349_RMS_BREAKDOWN/KIN_KULL_ALN.PDB [pr]
              2) 52 of 2709 atoms selected in:  /K349_RMS_BREAKDOWN/BINDIVIDUAL2_NEWADP.PDB [se]
              For PDB file 1 -  /K349_RMS_BREAKDOWN/KIN_KULL_ALN.PDB
              center of mass located at (  55.00,   40.20,   90.46)
              Total mass      = 702.37 grams/mole
              Components of Rgyr =   5.68,    1.41,    1.33 Angstroms
              Angles to last axes =  69.26,   68.03,   77.66 Degrees
              C.O.M. moved    = 16.329
              For PDB file 2 -  /K349_RMS_BREAKDOWN/BINDIVIDUAL2_NEWADP.PDB
              center of mass located at (  55.58,   38.64,   92.88)
              Total mass      = 702.37 grams/mole
              Components of Rgyr =   5.66,    1.40,    1.34 Angstroms
              Angles to last axes =   7.57,   21.38,   20.70 Degrees
              C.O.M. moved    =  2.940

.
.
```

Figure 1. Input files to GEM, and output result, to obtain the angle change of the switch II helix relative to the core domain. Highlighted in gray are the lines of the output file, resulting from the "axis" commands (highlighted in the input file), that show the angle change between selected residues of the two PDB input files. The first axis command compares a control region of kinesin, the immovable beta-sheet core, to show that the angle change (<1.3°) is very small (PDB files that were input to GEM were globally aligned beforehand, although GEM is also capable of doing such an alignment). Three angles are output by the axis command of GEM, in order of decreasing length of principle axis. For the case of an alpha helix, only the first angle, 7.5° (corresponding to the principle axis that goes along the helical axis), is reliable.

# Bibliography


Accelrys, inc. (www.accelrys.com).

Abrahams, J. P., A. G. W. Leslie, et al. (1994). "Structure at 2.8-Angstrom Resolution of F-1-Atpase from Bovine Heart Mitochondria." *Nature* **370**(6491): 621-628.

Arnal, I. and R. H. Wade (1998). "Nucleotide-dependent conformations of the kinesin dimer interacting with microtubules." *Structure* **6**(1): 33-8.

Baldwin, R. L. (1996). "How Hofmeister ion interactions affect protein stability." *Biophys J* **71**(4): 2056-63.

Ben-Naim, A. (1971). "Statistical mechanics of 'waterlike' particles in two dimensions. I. Physical model and application of the Percus-Yevick equation." *J. Chem. Phys.* **54**(9): 3682-95.

Ben-Naim, A. and Y. Marcus (1984). "Solvation thermodynamics of nonionic solutes." *J. Chem. Phys.* **81**(4): 2016-2027.

Berman, H. M., J. Westbrook, et al. (2000). "The Protein Data Bank." *Nucleic Acids Res* **28**(1): 235-42.

Block, S. M. (1996). "Fifty ways to love your lever: myosin motors." *Cell* **87**(2): 151-7.

Block, S. M. (1998). "Leading the procession: new insights into kinesin motors." *J Cell Biol* **140**(6): 1281-4.

Boresch, S. and M. Karplus (1995). "The meaning of component analysis: decomposition of the free energy in terms of specific interactions." *J Mol Biol* **254**(5): 801-7.

Brady, G. P. and K. A. Sharp (1995). "Decomposition of interaction free energies in proteins and other complex systems." *J Mol Biol* **254**(1): 77-85.

Browner, M. F., E. B. Fauman, et al. (1992). "Tracking conformational states in allosteric transitions of phosphorylase." *Biochemistry* **31**(46): 11297-304.

Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." *Acta Crystallogr D Biol Crystallogr* **54**(Pt 5): 905-21.

Case, R. B., D. W. Pierce, et al. (1997). "The directional preference of kinesin motors is specified by an element outside of the motor catalytic domain." *Cell* **90**(5): 959-66.

Case, R. B., S. Rice, et al. (2000). "Role of the kinesin neck linker and catalytic core in microtubule-based motility." *Curr Biol* **10**(3): 157-60.

Coppin, C. M., J. T. Finer, et al. (1996). "Detection of sub-8-nm movements of kinesin by high-resolution optical-trap microscopy." *Proc Natl Acad Sci U S A* **93**(5): 1913-7.

Fisher, A. J., C. A. Smith, et al. (1995). "X-ray structures of the myosin motor domain of Dictyostelium discoideum complexed with MgADP.BeFx and MgADP.AlF4." *Biochemistry* **34**(28): 8960-72.

Gallicchio, E., M. M. Kubo, et al. (1998). "Entropy-enthalpy compensation in solvation and ligand binding revisited." *Journal of the American Chemical Society* **120**(18): 4526-4527.

Goldstein, L. S. (2001). "Kinesin molecular motors: transport pathways, receptors, and human disease." *Proc Natl Acad Sci U S A* **98**(13): 6999-7003.

Guex, N. and M. C. Peitsch (1997). "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling." *Electrophoresis* **18**(15): 2714-23.

Hackney, D. D. (1994). "Evidence for alternating head catalysis by kinesin during microtubule- stimulated ATP hydrolysis." *Proc Natl Acad Sci U S A* **91**(15): 6865-9.

Hancock, W. O. and J. Howard (1999). "Kinesin's processivity results from mechanical and chemical coordination between the ATP hydrolysis cycles of the two motor domains." *Proc Natl Acad Sci U S A* **96**(23): 13147-52.

Henningsen, U. and M. Schliwa (1997). "Reversal in the direction of movement of a molecular motor." *Nature* **389**(6646): 93-6.

Hirokawa, N. (1998). "Kinesin and dynein superfamily proteins and the mechanism of organelle transport." *Science* **279**(5350): 519-26.

Hirose, K., U. Henningsen, et al. (2000). "Structural comparison of dimeric Eg5, Neurospora kinesin (Nkin) and Ncd head-Nkin neck chimera with conventional kinesin." *Embo J* **19**(20): 5308-14.

Hoenger, A., M. Thormahlen, et al. (2000). "A new look at the microtubule binding patterns of dimeric kinesins." *J Mol Biol* **297**(5): 1087-103.

Houdusse, A., V. N. Kalabokis, et al. (1999). "Atomic structure of scallop myosin subfragment S1 complexed with MgADP: a novel conformation of the myosin head." *Cell* **97**(4): 459-70.

Howard, J. (2001). *Mechanics of motor proteins and the cytoskeleton.* Sunderland, MA, Sinauer Associates, Inc.

Howard, J., A. J. Hudspeth, et al. (1989). "Movement of microtubules by single kinesin molecules." *Nature* **342**(6246): 154-8.

Hua, W., J. Chung, et al. (2002). "Distinguishing inchworm and hand-over-hand processive kinesin movement by neck rotation measurements." *Science* **295**(5556): 844-8.

Im, W., S. Berneche, et al. (2001). "Generalized solvent boundary potential for computer simulations." *Journal of Chemical Physics* **114**(7): 2924-2937.

Kawaguchi, K. and S. Ishiwata (2001). "Nucleotide-dependent single- to double-headed binding of kinesin." *Science* **291**(5504): 667-9.

Kikkawa, M., E. P. Sablin, et al. (2001). "Switch-based mechanism of kinesin motors." *Nature* **411**(6836): 439-45.

Kissinger, C. R., D. K. Gehlhaar, et al. (1999). "Rapid automated molecular replacement by evolutionary search." *Acta Crystallogr D Biol Crystallogr* **55**(Pt 2): 484-91.

Kozielski, F., S. Sack, et al. (1997). "The crystal structure of dimeric kinesin and implications for microtubule-dependent motility." *Cell* **91**(7): 985-94.

Kull, F. J., E. P. Sablin, et al. (1996). "Crystal structure of the kinesin motor domain reveals a structural similarity to myosin." *Nature* **380**(6574): 550-5.

Kull, F. J., R. D. Vale, et al. (1998). "The case for a common ancestor: kinesin and myosin motor proteins and G proteins." *J Muscle Res Cell Motil* **19**(8): 877-86.

Levy, R. M. and E. Gallicchio (1998). "Computer simulations with explicit solvent: recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects." *Annu Rev Phys Chem* **49**: 531-67.

Mark, A. E. and W. F. van Gunsteren (1994). "Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies." *J Mol Biol* **240**(2): 167-76.

Matubayasi, N., E. Gallicchio, et al. (1998). "On the local and nonlocal components of solvation thermodynamics and their relation to solvation shell models." *Journal of Chemical Physics* **109**(12): 4864-4872.

Muller, J., A. Marx, et al. (1999). "The structure of the nucleotide-binding site of kinesin." *Biol Chem* **380**(7-8): 981-92.

Naber, N., R. Cooke, et al. (1997). "Binding of ncd to microtubules induces a conformational change near the junction of the motor domain with the neck." *Biochemistry* **36**(32): 9681-9.

Nogales, E., M. Whittaker, et al. (1999). "High-resolution model of the microtubule." *Cell* **96**(1): 79-88.

Otwinowski, Z. and W. Minor (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Macromolecular Crystallography, part A*, Academic Press. **276:** 307-326,.

Rayment, I., W. R. Rypniewski, et al. (1993). "Three-dimensional structure of myosin subfragment-1: a molecular motor." *Science* **261**(5117): 50-8.

Rice, S., A. W. Lin, et al. (1999). "A structural change in the kinesin motor protein that drives motility." *Nature* **402**(6763): 778-84.

Rosenfeld, S. S., J. J. Correia, et al. (1996). "Structural studies of kinesin-nucleotide intermediates." *J Biol Chem* **271**(47): 30212-21.

Rosenfeld, S. S., G. M. Jefferson, et al. (2001). "ATP reorients the neck linker of kinesin in two sequential steps." *J Biol Chem* **276**(43): 40167-74.

Rosenfeld, S. S., B. Rener, et al. (1996). "Equilibrium studies of kinesin-nucleotide intermediates." *J Biol Chem* **271**(16): 9473-82.

Roux, B. and T. Simonson (1999). "Implicit solvent models." *Biophysical Chemistry* **78**(1-2): 1-20.

Sablin, E. P., F. J. Kull, et al. (1996). "Crystal structure of the motor domain of the kinesin-related motor ncd." *Nature* **380**(6574): 555-9.

Sack, S., F. J. Kull, et al. (1999). "Motor proteins of the kinesin family. Structures, variations, and nucleotide binding sites." *Eur J Biochem* **262**(1): 1-11.

Sack, S., J. Muller, et al. (1997). "X-ray structure of motor and neck domains from rat brain kinesin." *Biochemistry* **36**(51): 16155-65.

Schief, W. R. and J. Howard (2001). "Conformational changes during kinesin motility." *Curr Opin Cell Biol* **13**(1): 19-28.

Silverstein, K. A. T., A. D. J. Haymet, et al. (1998). "A simple model of water and the hydrophobic effect." *Journal of the American Chemical Society* **120**(13): 3166-3175.

Sitkoff, D., K. A. Sharp, et al. (1994). "Correlating solvation free energies and surface tensions of hydrocarbon solutes." *Biophys Chem* **51**(2-3): 397-403; discussion 404-9.

Song, Y. H., A. Marx, et al. (2001). "Structure of a fast kinesin: implications for ATPase mechanism and interactions with microtubules." *Embo J* **20**(22): 6213-25.

Sosa, H., D. P. Dias, et al. (1997). "A model for the microtubule-Ncd motor protein complex obtained by cryo-electron microscopy and image analysis." *Cell* **90**(2): 217-24.

Sosa, H., E. J. Peterman, et al. (2001). "ADP-induced rocking of the kinesin motor domain revealed by single- molecule fluorescence polarization microscopy." *Nat Struct Biol* **8**(6): 540-4.

SotoCampos, G., D. S. Corti, et al. (1998). "A small system grand ensemble method for the study of hard-particle systems." *Journal of Chemical Physics* **108**(6): 2563-2570.

Southall, N. T. and K. A. Dill (2000). "The mechanism of hydrophobic solvation depends on solute radius." *Journal of Physical Chemistry B* **104**(6): 1326-1331.

Southall, N. T., K. A. Dill, et al. (2002). "A view of the hydrophobic effect." *Journal of Physical Chemistry B* **106**(3): 521-533.

Thorn, K. S., J. A. Ubersax, et al. (2000). "Engineering the processive run length of the kinesin motor." *J Cell Biol* **151**(5): 1093-100.

Tucker, C. and L. S. Goldstein (1997). "Probing the kinesin-microtubule interaction." *J Biol Chem* **272**(14): 9481-8.

Turner, J., R. Anderson, et al. (2001). "Crystal structure of the mitotic spindle kinesin Eg5 reveals a novel conformation of the neck-linker." *J Biol Chem* **276**(27): 25496-502.

Vale, R. D., R. Case, et al. (2000). "Searching for kinesin's mechanical amplifier." *Philos Trans R Soc Lond B Biol Sci* **355**(1396): 449-57.

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

Vriend, G. (1990). "WHAT IF: A molecular modeling and drug design program." *J. Mol. Graph.* **8**: 52-56.

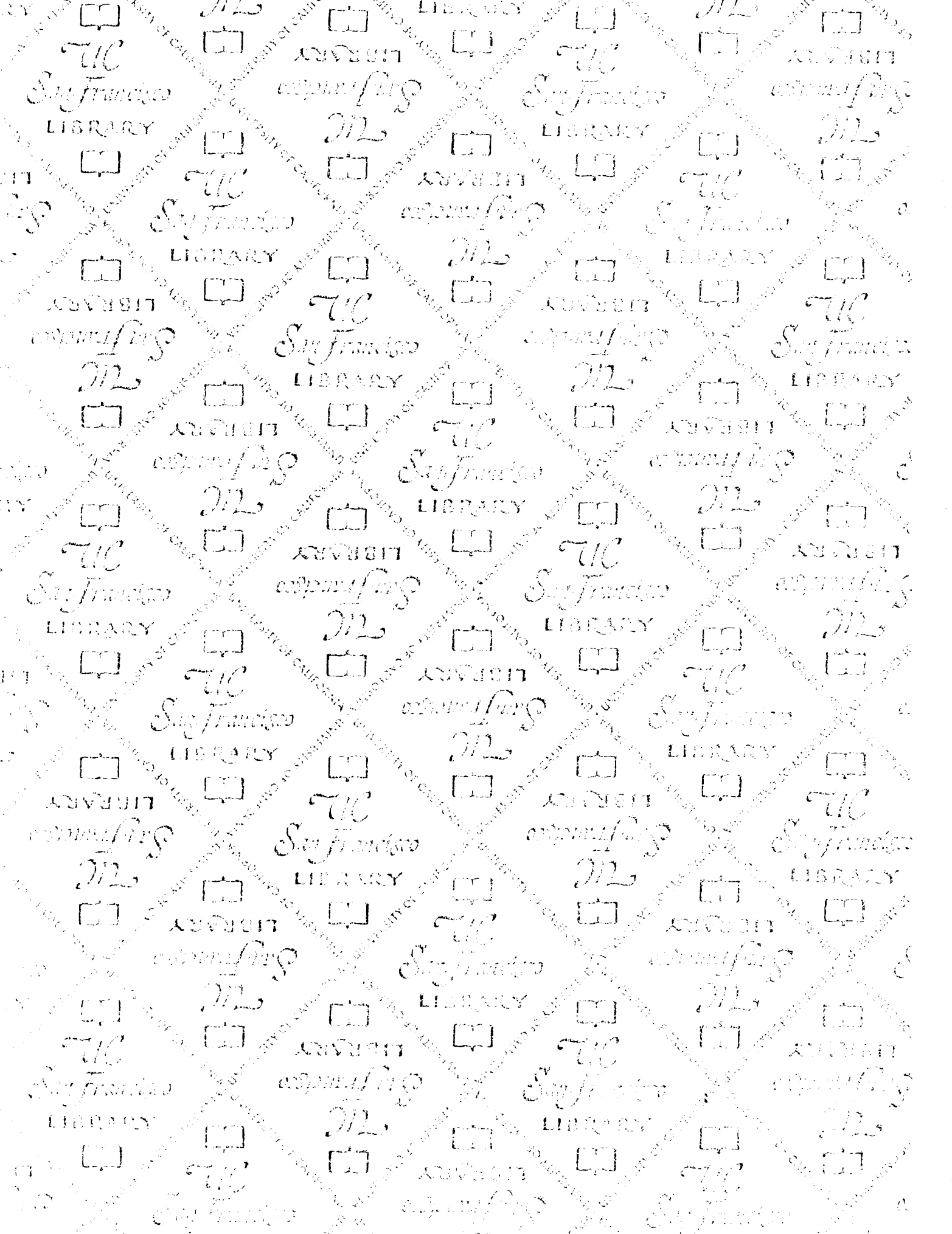
Woehlke, G., A. K. Ruby, et al. (1997). "Microtubule interaction site of the kinesin motor." *Cell* **90**(2): 207-16.

Woehlke, G. and M. Schliwa (2000). "Walking on two heads: the many talents of kinesin." *Nat Rev Mol Cell Biol* **1**(1): 50-8.

Wriggers, W. and K. Schulten (1998). "Nucleotide-dependent movements of the kinesin motor domain predicted by simulated annealing." *Biophys J* **75**(2): 646-61.

Xing, J., W. Wriggers, et al. (2000). "Kinesin has three nucleotide-dependent conformations. Implications for strain-dependent release." *J Biol Chem* **275**(45): 35413-23.

Yun, M., X. Zhang, et al. (2001). "A structural pathway for activation of the kinesin motor ATPase." *Embo J* **20**(11): 2611-8.


The San Francisco Public Library