MULTIVALENT PROTEIN BINDING TO METAL-COMPLEXING MATERIALS: APPLICATIONS TO SYNTHETIC RECEPTORS AND AFFINITY CHROMATOGRAPHY

Thesis by

Robert David Johnson

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, CA

1995

(Submitted September 2, 1994)

ACKNOWLEDGMENTS

I would like to thank my collaborators, Dr. Robert Todd and Dr. Sanku Mallik, for their important contributions to this work. I would also like to thank all the other members of the Arnold research group for their camaraderie, and in particular Deb Shnek and Virginie Leenknecht for their motivation and assistance. I thank Kelly Goodwin for her advice, support, and patience; I promise to reciprocate.

I would like to thank the members of my doctoral committee, Dr. Harry Gray, Dr. Jacqueline Barton, and Dr. Mark Davis, and also Dr. John Brady of my candidacy committee, for their time and their guidance.

Finally, I would like to thank my advisor, Dr. Frances Arnold, for providing me the opportunity to pursue this investigation. This work could not have been accomplished without her insight, direction, and support.

ABSTRACT

This investigation demonstrates that proteins have the capability to bind simultaneously to multiple transition metals of metal-complexing materials. This finding has important implications for the design of novel materials for protein recognition. Our approach to protein recognition, based on intermolecular metal-to-ligand interactions, matches a protein's unique pattern of histidines with a complementary arrangement of transition metal complexes.

A model system is used to demonstrate the validity of this approach in the simplest case by matching the distance between metal ions of rationally designed bis-metal (Hg^{2+}) "receptors" to that between imidazoles of bis-imidazole "targets." This system additionally demonstrates how other features of receptor design can influence binding selectivity. A 2D NMR procedure is developed to measure directly protein surface histidine binding to copper complexes, and subsequently demonstrates that the local environment of the histidine and the structure of the copper complex can modulate individual copper-histidine interactions. Thus it may indeed be possible to design metal-containing receptors which are able to form simultaneous metal-ligand bonds with a specific arrangement of protein metal-coordinating groups.

There are two important obstacles preventing a similarly detailed description of protein binding to metal-complexing surfaces: protein adsorption may involve binding to one or more metal sites, and a detailed description of the geometry of surface metal sites would be hopelessly complex. We can, however, apply the microscopic concept of simultaneous metal-ligand interactions to interpret the macroscopic phenomena of protein partitioning in immobilized metal affinity chromatography (IMAC). In this context, the ability of commercially available IMAC materials to support multiple protein-surface interactions is shown to be dependent on three factors: the number of

histidines on the protein (as manipulated by site directed mutagenesis), the number of deprotonated amino groups on the protein (pH control), and the density of binding sites on the surface (copper loading). These results demonstrate that a realistic description of protein binding in IMAC must consider a heterogeneous population of surface binding sites. In IMAC this is shown to be conveniently expressed by the Temkin isotherm, making it an instructive model to explore heterogeneity displayed by other chromatographic materials and by biological systems.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

LIST OF TABLES

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

CHAPTER 1

INTRODUCTION

PROTEIN RECOGNITION

Proteins display remarkable specificity and diversity in binding smaller molecules despite the limited supply of functional groups available to them. The strength and selectivity of binding is dictated by the quality of the match between an array of amino acids on the protein and complementary functional groups on the bound molecule. Protein structural determination together with molecular modeling make it increasingly feasible to visualize the individual interactions governing protein recognition. These interactions provide an opportunity to engineer synthetic materials which will recognize individual proteins over others which may differ by only a few amino acids. Such materials would enjoy wide use, ranging from isolation of desired proteins from a mixture of cellular debris (protein separations) to preferential inhibition of enzymes critical to an invading pathogen (structure-based drug design).

The challenge of structure-based design in the formulation of novel pharmaceuticals is to specifically bind to particular cellular receptors or enzymes to the exclusion of the large population of similar proteins present in the host. In most cases, such pharmaceuticals are analogs of natural products, such as hormones or enzyme substrates, that have been modified to effect an unusual response from the target protein. If the structure of the target protein is known, then, with an appreciation of the design rules that drive the specific protein-substrate interactions, a rational approach can be taken to match individual modifications of synthetic compounds to the distribution of functional groups on the protein target [1,2].

The challenge of protein separations in the production of protein therapeutics is to isolate a desired protein from contaminating host proteins and to purify it to a degree suitable for medicinal applications. In general, recombinant proteins account for only a

small fraction of the total protein produced in a host organism. Moreover, as all proteins are comprised of the same twenty amino acids, they have similar chemical properties. The most effective purification methods, such as affinity chromatography [3], take advantage of specific protein-ligand or protein-protein interactions to preferentially extract a particular protein or class of proteins.

The challenges of drug design and protein separations make similar demands on synthetic materials. First, the materials must have a sufficiently high *binding affinity* in order to bind the desired protein even at very low concentrations. Second, the materials must display a sufficiently high *selectivity* for the target protein in order to bind it preferentially over all others. A successful strategy for protein recognition by either small molecules suitable for therapeutic applications or derivatized surfaces suitable for chromatographic applications must satisfy these demands.

PROTEIN RECOGNITION: SMALL MOLECULES

Effective drugs function by mimicking natural products and preferentially binding to and activating (or blocking) particular target cellular receptors or enzymes. For example, some compounds inhibit enzymatic reactions critical to invading pathogens by preventing the conversion of analogous natural substrates into necessary products (e.g., sulfanilamides). Similarly, some compounds serve as effective blockers of cellular receptors; the drugs effectively bind the receptor but fail to elicit the same response as the natural agonist (e.g., antihistamines). Others serve as artificial substitutes for natural agonists, stimulating cellular receptors to produce a heightened or prolonged response (e.g., morphine).

For a drug to be effective, it must bind the target protein to the exclusion of other cellular proteins. It must have a sufficiently high binding affinity to bind the desired

receptor or enzyme even when present in very low quantities. A potential drug usually is considered to meet this requirement if, at equilibrium, nanomolar concentrations are sufficient to ensure that most of the drug is bound to the protein target (equilibrium binding constant $K \sim 10^9$ M^{-1} [4]). The drug must also display a sufficiently high selectivity for binding the target protein. Compounds which inhibit broad classes of enzymes are often fatal (e.g., mercury or arsenic). There is a delicate balance among a myriad of biochemical processes in a living cell — the ideal drug would inhibit a single reaction performed by a single enzyme without disabling the entire complex network.

For therapeutic applications, a drug optimally would be specific to a particular protein from a particular species, thus leaving similar enzymes of the host organism intact. To illustrate the principles of structure-based drug design, consider the enzyme dihydrofolate reductase (DHFR). DHFR enzymes isolated from different organisms have significant differences in amino acid sequence. For example, as illustrated in Figure 1.1, the enzyme isolated from *Lactobacillus casei* has two histidines near the substrate binding pocket. The enzyme from *Neisseria gonorrhoeae* has one, and enzymes from *Escherichia coli* (also shown in Figure 1.1) and humans both have none [1].

DHFR catalyzes the reduction of dihydrofolate, shown in Figure 1.2A, to tetrahydrofolate, a reaction essential to DNA synthesis and cell division. These enzymes are inhibited by a wide range of dihydrofolate analogs, such as methotrexate shown in Figure 1.2B, which block the substrate binding pocket. The three-dimensional structure of this compound bound to *L. casei* DHFR [5] (Figure 1.1A) displays a hydrogen bond between the imidazole nitrogen of a histidine near the binding pocket and the terminal carboxylate group of methotrexate. Such structural information has been employed previously to design more selective DHFR inhibitors containing an iodoacetamide functionality (Figure 1.2C) to react irreversibly with histidine imidazoles [6,7]. As expected, such compounds inactivate DHFR from *N. gonorrhoeae* and *L. casei*

(histidines near the binding pocket), and do not react with the human and *E. coli* enzymes (no histidines near the binding pocket) [7]. Unfortunately, there are problems with the toxicity of these particular reagents *in vivo*, most likely from nonspecific, irreversible inactivation of other histidine-containing host enzymes [1]. For therapeutic applications, a more appropriate inhibitor would reversibly block DHFR enzymes containing histidines in the binding pocket.

The previous example is an excellent illustration of drug design to recognize the structural arrangement of individual functional groups on a target protein. Previous investigators have designed synthetic systems which to a limited degree reproduce the intermolecular hydrogen-bonding, ionic, and hydrophobic interactions observed in protein-protein binding [8,9,10]. Instead, we are undertaking a novel approach to protein recognition, employing intermolecular metal-to-ligand interactions between transition metal complexes and metal-coordinating functional groups on the protein surface, as illustrated in Figure 1.3.

This approach offers several potential advantages over schemes that rely upon other mechanisms such as hydrogen bonding and hydrophobic interactions [11]. First, metal-to-ligand coordination is quite strong, even in aqueous media. For example, the association constant for Cu^{2+} coordination by the imidazole side chain of histidine ($K = 5 \times 10^3$ M^{-1} [12]) translates to a binding energy of 4.8 kcal/mol. Thus, each favorable metal-to-ligand interaction would enhance protein binding by orders of magnitude. In contrast, individual hydrogen bonding interactions generally contribute less than 1 kcal/mol to binding in water [13]. A second attractive feature is the availability of a wide variety of ligands and transition metal ions, which control the strength and kinetics of binding [11]. The strength of the metal-imidazole interaction, for example, varies over a wide range, depending on the metal ion [14].

Protein recognition based on metal-to-ligand interactions exploits the protein's unique pattern of surface histidines (or other metal-coordinating functional groups), which can be simultaneously coordinated by a receptor displaying a complementary pattern of transition metal ions. The simplest receptor design, as illustrated in Figure 1.4, would match the distance between two surface-exposed histidines with the distance between a pair of complexed metal ions. The specificity of binding is then dictated by the degree to which two-point binding (Figure 1.4A) is favored over binding to a single surface histidine (Figure 1.4B) and the probability that no other protein has a pair of histidines with the same spacing. This is aided by the fact that surface exposed histidines are relatively rare, making up approximately 1% of the amino acids in globular proteins [15].

PROTEIN RECOGNITION: SURFACES

The cost of protein isolation and purification is an important obstacle to developing recombinant protein products. Traditional separation techniques such as precipitation and ion exchange chromatography are often insufficient by themselves; a series of several different techniques is usually needed to obtain a product of desired quality. With each additional technique there is a compounded loss of product and added costs of separations material and equipment.

On the other hand, developments in biological affinity separations have demonstrated that a desired protein can be purified to homogeneity in a single step. In affinity chromatography, the separations material consists primarily of an affinity ligand, specific for the target protein, covalently attached to a solid support. The requirements for a successful affinity ligand are very similar to those outlined above for a successful drug: sufficiently high binding affinity and selectivity for the target protein. In fact,

recombinant DHFR can be purified in a single step using chromatographic materials derivatized with methotrexate. Contaminating proteins which do not have a high binding affinity for methotrexate are removed during the wash step. The pure DHFR is later eluted by washing with folate [16].

The enhanced selectivity of affinity separations comes at great cost: the required materials are often fragile, prohibitively expensive, and, because of their specificity, limited in application. Thus there is a call for novel separations materials which can recognize proteins with specificities approaching those of biological affinity chromatography but also provide the durability, low operating expense, and versatility of traditional chromatographic techniques. Immobilized metal-affinity chromatography (IMAC), with a proven record as a useful and versatile technique for protein isolation and purification, is optimally positioned to bridge this gap between biological affinity and traditional separations.

As ligands for affinity separations, transition metal complexes offer important advantages over biological affinity agents such as inhibitors and antibodies [15]. Small, inexpensive metal complexes are stable under a wide range of conditions, can be recycled many times without loss of activity, and can be formulated into very high-capacity chromatographic supports. Purified protein can be eluted under relatively mild conditions, and the columns can be cleaned and regenerated easily, without reduction in protein binding capacity. The selectivity of the separation can be tailored through the choice of metal ion, solvent conditions, or by modification of the target protein (e.g., the addition of histidine-rich affinity "handles").

As previously discussed, proteins interact with metal complexes primarily through surface coordinating groups (Figure 1.3) which can donate electrons to the metal ions [17]. It is therefore possible to modulate the interactions between a particular protein and the chromatographic surface by altering properties of the immobilized metal

complexes. As mentioned, the choice of metal ion dictates which functional groups on the protein surface are targeted. For example, Cu^{2+} complexes at neutral pH bind primarily to imidazole nitrogens (histidine side chains), and to a lesser degree amine nitrogens (lysine side chains and the amino-terminus) [18,19]. Other divalent cations such as Hg^{2+} also target the sulfur atoms of cysteine and methionine [20], while Fe^{3+} primarily targets the oxygen atoms of carboxylate and phosphate groups [21,22]. Relevant properties such as overall charge, coordination geometry, and binding affinity for additional ligands change with the choice of chelating group and can alter binding selectivity for individual proteins [23]. These advantageous features have driven the recent rapid growth in IMAC applications [24,25].

Despite recent progress in elucidating IMAC adsorption mechanisms and quantifying protein binding behavior [26,27,28], the precise molecular mechanisms by which proteins are selectively retained on these matrices are not well understood. There is an important difference between protein binding to individual metal complexes in solution, as in drug design, and to those scattered on a solid surface, as in protein separations. Target proteins often contain multiple metal-coordinating surface functional groups. Therefore, upon adsorption a single protein can bind to one *or more* of the immobilized metal sites [29,30]. Such multivalent interactions are highly cooperative; binding constants for protein adsorption by simultaneous binding to two ligands are predicted to be orders of magnitude greater than those for adsorption at a single ligand [31,32].

A consequence of multiple-site interactions in IMAC is a significant increase in binding affinity which is strongly dependent on the distribution of metal complexes. The ability to control multiple-site interactions provides a unique opportunity to tailor these and other chromatographic supports for specific protein separations [33]. As illustrated in Figure 1.5, multivalent binding in metal affinity chromatography provides a

mechanism, analogous to that proposed for drug design, to recognize individual proteins based upon their pattern of surface histidines. It is, in fact, possible to target individual molecules by matching the distribution of metal ions to a spatial distribution of metal-coordinating ligands [34], and efforts to design new chromatographic supports capable of specific multiple-site binding to target molecules are underway [35].

Alternatively, the nature of protein binding to existing IMAC supports which presumably have random distributions of metal binding sites has important implications for the design of efficient separations. A material with random arrangement of copper ions can be thought to consist of individual surface "binding sites," as illustrated in Figure 1.6, each differing in ability to support multivalent protein-ligand interactions and therefore differing in binding affinity. Such a description predicts protein adsorption with a range of protein binding energies [36], with this range dependent on the distribution of immobilized affinity ligands. Indeed, equilibrium binding studies have shown that binding heterogeneity is often observed for tightly binding proteins [33,36,37] and that higher densities of immobilized ligands support stronger protein binding [33,38].

PROJECT OVERVIEW

With the ultimate goal of designing materials to allow multiple, specific interactions with individual surface groups of a particular protein, the possibility that proteins bind simultaneously to multiple metal ions has important implications for the design of novel therapeutics and chromatographic supports. As mentioned, this new approach to protein recognition targets a protein's unique pattern of surface-exposed histidines (or other metal-coordinating functional groups) with materials containing complementary patterns of complexed metal ions. The validity of this approach will be

examined using a combination of spectroscopic techniques to directly measure metal-imidazole coordination and chromatographic techniques to measure protein binding to metal-complexing materials.

A model system will be employed to explore the extent to which two metal-imidazole interactions can provide selectivity in binding a target molecule. Previously, a series of bis-imidazole "protein analogs" differing in distance between imidazole groups and a complementary series of bis-metal ion (Hg^{2+}) receptors have been synthesized in our laboratory [11] . Equilibrium titration experiments (monitored by NMR chemical shifts in DMSO-d_6) verify that each receptor binds two equivalents of a single-imidazole control, and one equivalent of bis-imidazole species. Competitive equilibrium titration experiments are analyzed to determine the selectivity the receptors show in binding a particular bis-imidazole over others with imidazoles spaced closer or further apart. These model studies demonstrate the validity of receptor design by matching the distance between complexed metal ions to that of a complementary pair of imidazoles. These studies also serve to probe how other specific features of receptor design, such as mobility and the number of available coordination sites, can influence binding selectivity.

Molecular modeling of this bis-metal receptor system enables a detailed description of molecular recognition by arranging simultaneous metal-ligand interactions. However, the arrangements of metal ligands are not so well defined on functionalized surfaces for chromatographic applications (e.g., IDA-TSK, a polymer resin derivatized with chelating iminodiacetate); the complexity of these materials makes a detailed molecular description essentially useless. We can instead apply the concepts learned from the molecular description of the bis-metal receptor system to interpret the macroscopic phenomena of protein adsorption and chromatographic retention to commercially available IMAC materials (IDA-TSK).

In this context, the number of metal-ligand bonds between protein and surface are manipulated through three parameters: (1) the number of histidines on the protein, (2) the number of deprotonated amino groups on the protein, and (3) the density of copper sites on the surface. First, a series of yeast cytochrome *c* variants with varying surface histidine content, previously constructed using site-directed mutagenesis [18], are used to probe the role of the number and placement of protein surface histidines in adsorption to IMAC materials. Secondly, at neutral pH, protonation of protein surface amine groups (e.g., lysines) inhibits their interaction with metal ions. Chromatographic retention of single-histidine yeast cytochrome *c* variants at elevated pH illustrates the role of these additional metal-coordinating groups in protein binding to IMAC materials. Finally, the arrangement of copper sites on IMAC materials is dependent on the density of surface metal ions, which can be conveniently altered by decreasing the supply of copper added to the metal-free material. Together, these three processes demonstrate that IMAC materials support simultaneous interactions between metal-coordinating groups on the protein and multiple surface metal ions.

As demonstrated in the solution studies using bis-metal receptors, simultaneous multiple interactions depend on the quality of the match between the arrangement of protein surface histidines and the arrangement of metal ions on the chromatographic support. Thus on a material with a nonuniform arrangement of metal ions there is a potential for different classes of "binding sites," each consisting of different arrangements of affinity ligands, interacting to different degrees with the pattern of surface groups on a particular protein. A realistic description of protein adsorption due to multivalent interactions must therefore consider a heterogeneous population of surface binding sites. Binding heterogeneity in metal affinity chromatography is expressed in terms of binding energy distributions, using simple isotherm expressions previously derived for equilibrium gas adsorption [39]. In this context, the binding heterogeneity

demonstrated by IMAC materials is an instructive model for heterogeneity displayed in biological systems.

Finally, the ability of metal complexes to bind individual histidines on protein surfaces is examined by NMR spectroscopy. In the case of copper complexes, the paramagnetic electronic spin greatly increases the relaxation rate of 2D NMR signals for individual protons near the specific binding sites of the metal complex [40]. 2D NMR difference spectrum can be formed by subtracting the spectrum recorded in the presence of paramagnetic species from a suitable control spectrum; those protons nearest the binding site appear as the largest peaks in the 2D difference spectra. Using this technique, the relative binding affinities of several copper complexes for individual surface amino acids of cytochrome *c* and myoglobin are quantified by relaxation effects in 2D NMR spectroscopy. These capabilities demonstrate that individual copper-histidine interactions can be modulated either by the local environment of the protein or by the structure of the copper ligand. These results, together with those of coordinating multiple metal-ligand interactions in the bis-metal receptor systems and IMAC, suggest that indeed it may be possible to target a specific arrangement of protein surface metal-coordinating groups with a material displaying a complementary array of metal complexes.

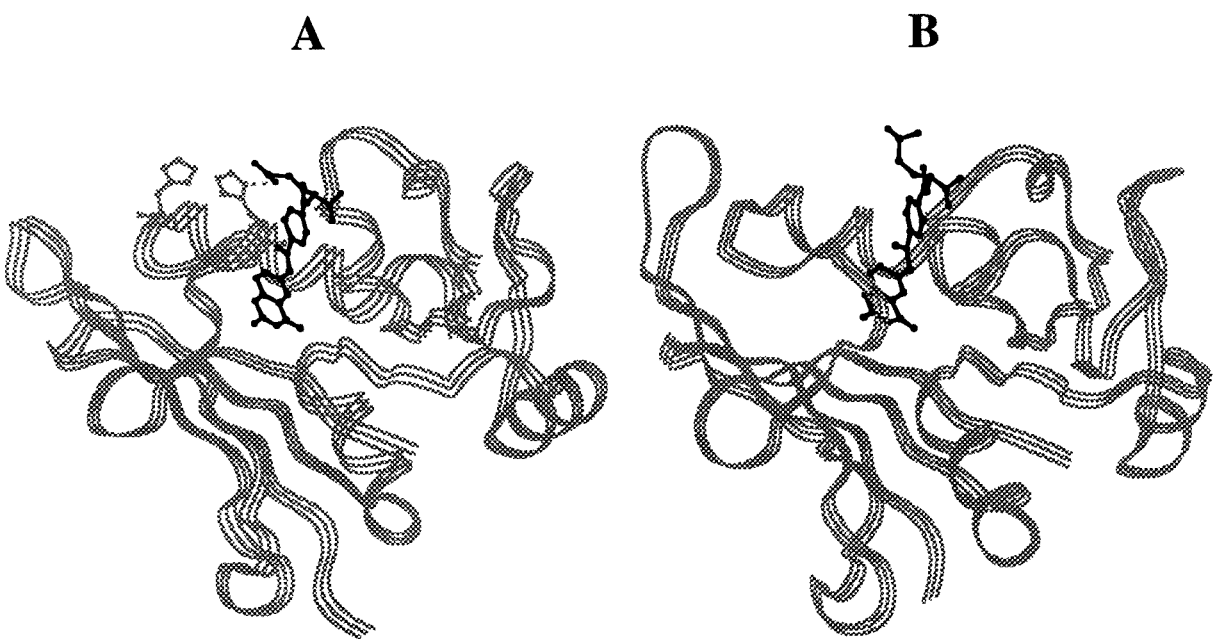
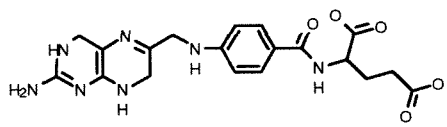
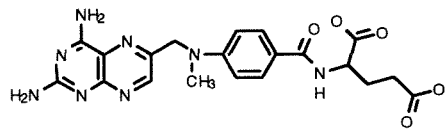


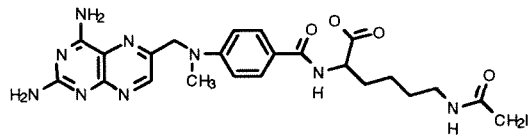
Figure 1.1. Structure-based drug design: Inhibition of DHFR by methotrexate. **(A)** Three-dimensional structure of *Lactobacillus casei* dihydrofolate reductase with methotrexate inhibitor [5]. Reversibly bound methotrexate (black) is hydrogen-bonded to one of the two histidines (gray) near the folate binding pocket. **(B)** Three-dimensional structure of *Eschericia coli* DHFR with methotrexate inhibitor [5]. There are no histidines in close proximity to the folate binding pocket.



(A) natural DHFR substrate



(B) reversible DHFR inhibitor



(C) irreversible DHFR inhibitor

Figure 1.2. Structure-based drug design: Species-specific inhibitor of *L. casei* DHFR **(A)** DHFR natural substrate (7,8-dihydrofolate). **(B)** Reversible DHFR inhibitor (methotrexate). Terminal carboxylate group participates in hydrogen bond with histidine 28 in three-dimensional structure (Figure 1.1). **(C)** Irreversible inhibitor of DHFR from *L. casei* (N^{α}-(4-amino-4-deoxy-N^{10}-methylpteroyl)-N^{ε}-(iodoacetyl)-*l*-lysine). Iodoacetamide functionality reacts covalently with imidazole nitrogen of histidines near folate binding pocket.

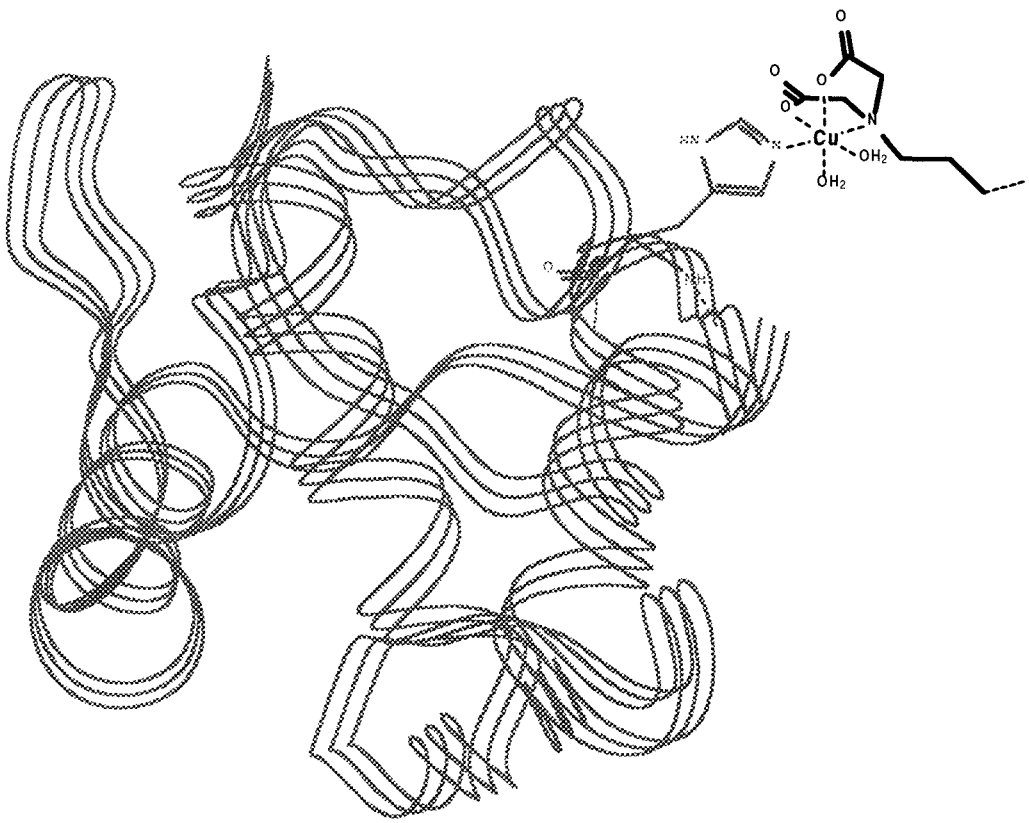


Figure 1.3. Illustration of intermolecular metal-to-ligand coordination. Metal-complexing material (e.g., derivatized with Cu(II)iminodiacetate) binds specifically to protein surface metal-coordinating groups (e.g., histidine). Vacant coordination sites of copper ion form reversible coordinate-covalent bond with imidazole nitrogen.

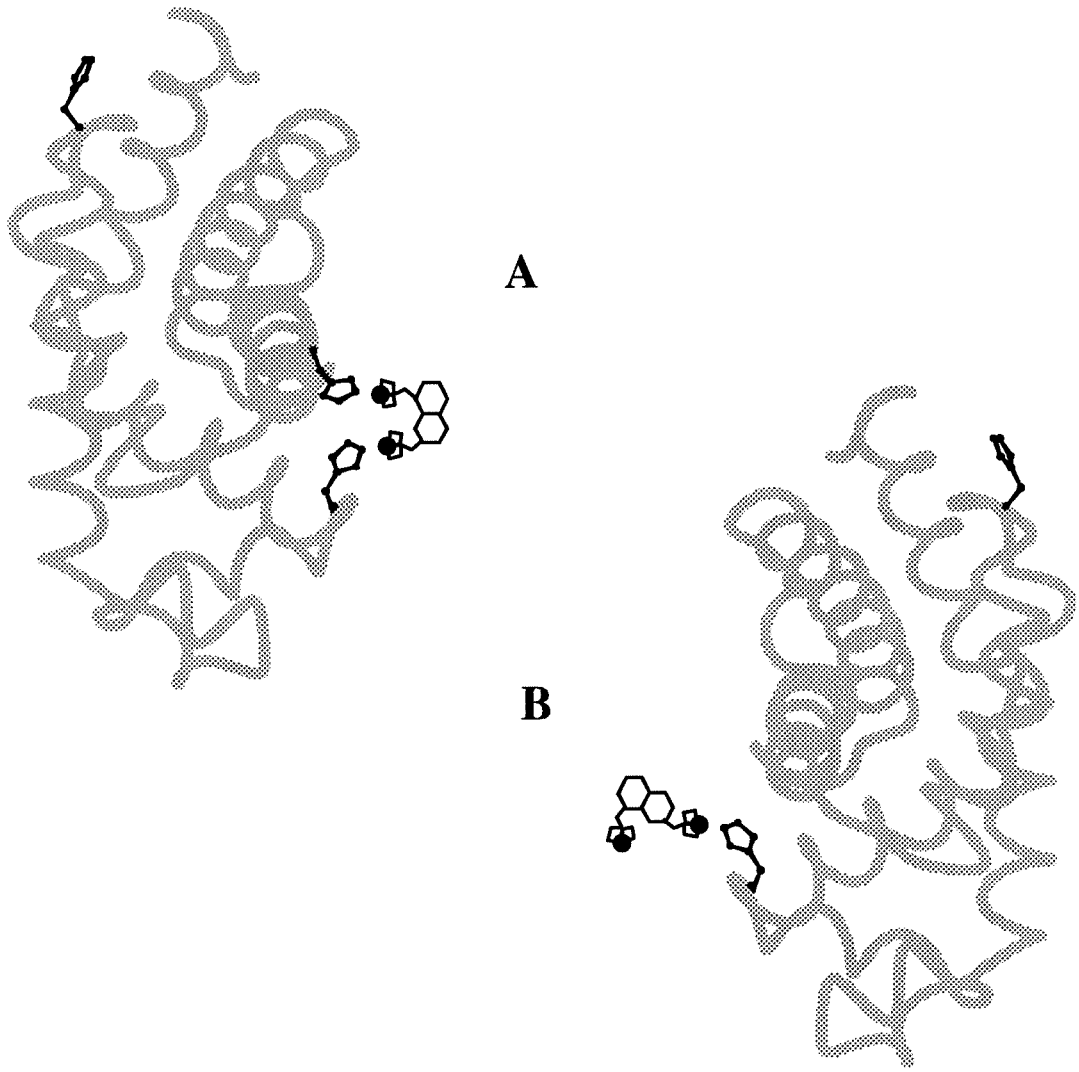


Figure 1.4. Drug design: multivalent binding to metal-complexing receptor. **(A)** Target protein has surface histidines that complement metal ions of artificial receptor, resulting in multivalent coordination of high binding affinity. **(B)** Analogous protein does not have complementary arrangement of histidines, resulting in lower binding affinity.

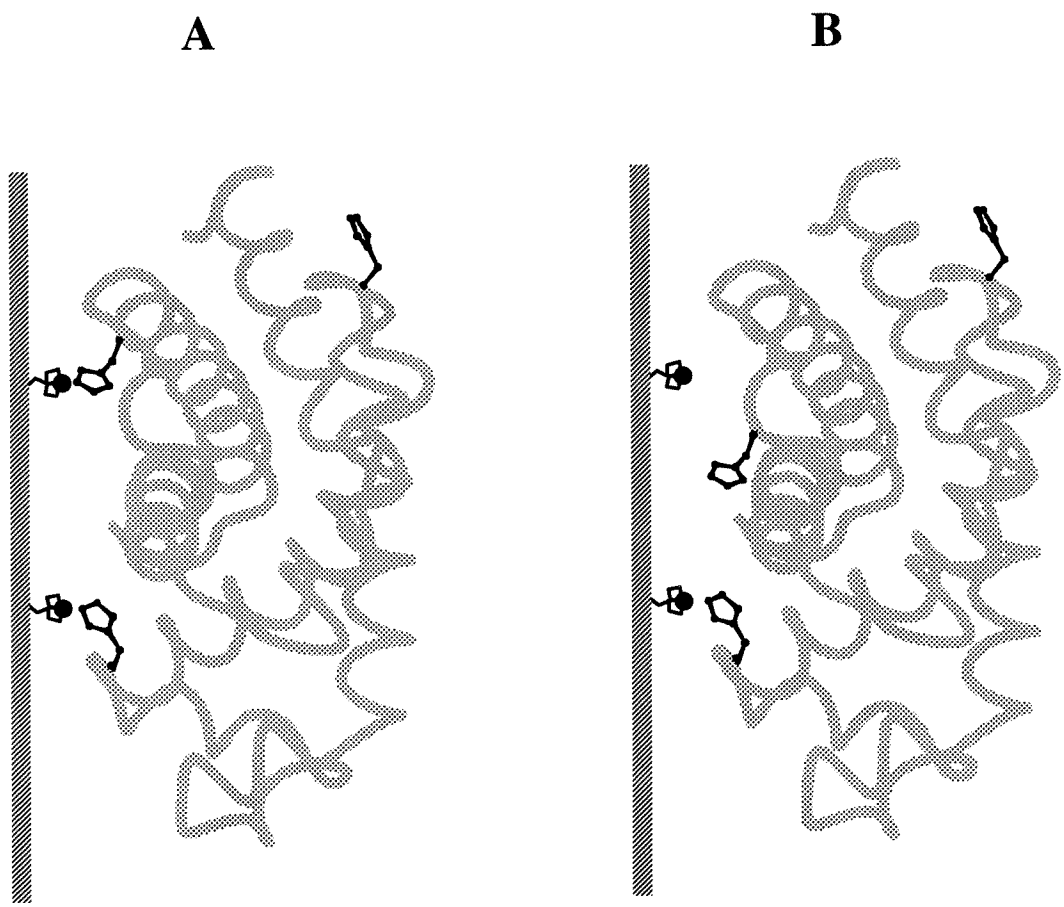


Figure 1.5. Protein separations: multivalent binding to metal-complexing surface. (**A**) Target protein has histidines that match distribution of metal ions on surface, resulting in high affinity adsorption via multivalent interactions. (**B**) Analogous protein with different arrangement of histidines does not match the pattern of affinity ligands, resulting in lower affinity adsorption.

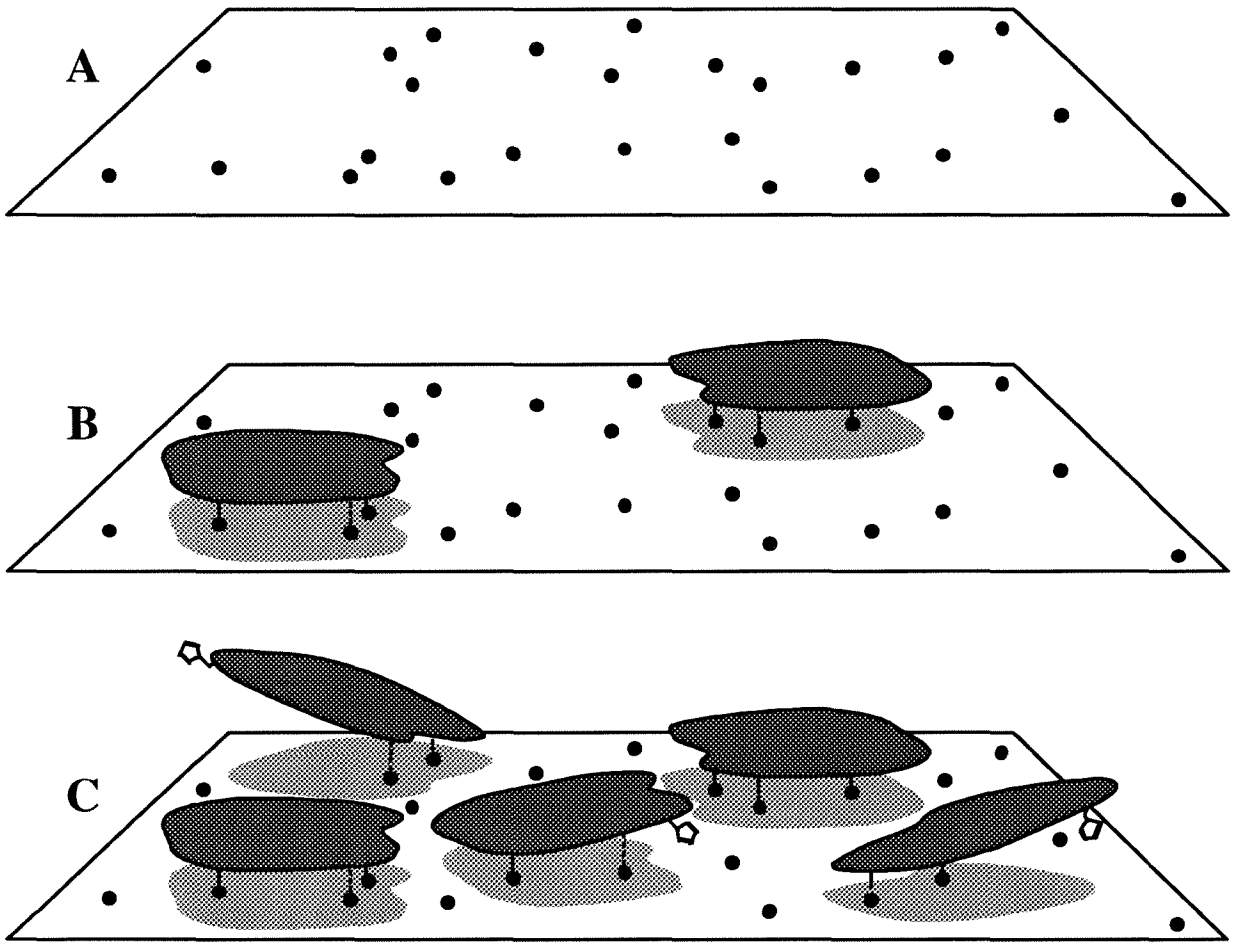


Figure 1.6. Heterogeneous protein adsorption to a random arrangement of binding sites. **(A)** Surface has a nonuniform arrangement of metal ions for protein binding. **(B)** At low coverage, protein preferentially binds to sites with optimal arrangements of metal ions, resulting in highest binding affinity. **(C)** At higher coverage, optimal sites are occupied and protein adsorbs to sites with less optimal arrangment of metal ions, resulting in lower binding affinity.

REFERENCES

1. Beddell, C. R., Editor (1992) *The Design of Drugs to Macromolecular Targets*, John Wiley and Sons, New York, NY.
2. Sutcliffe, J. A. and Georgopapakakou, N. H., Editors (1992) *Emerging Targets in Antibacterial and Antifungal Chemotherapy*, Chapman and Hall, New York, NY.
3. Chaiken, I. M., Wilchek, M. and Parikh, I., Editors (1983) *Affinity Chromatography and Biological Recognition*, Academic Press, Orlando, FL.
4. Ganellin, C. R. and Roberts, S. M., Editors (1993) *Medicinal Chemistry: The Role of Organic Chemistry in Drug Research*, Academic Press, San Diego, CA.
5. Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C. and Kraut, J. (1982) *J. Biol. Chem. 257*, 13650-13662.
6. Rosowsky, A., Wright, J. E., Ginty, C. and Uren, J. (1982) *J. Med. Chem. 25*, 960-964.
7. Tansik, R. L., Averett, D. R., Roth, B., Paterson, S. J., Stone, D. and Baccanari, D. P. (1984) *J. Biol. Chem. 259*, 12299-12305.
8. Seto, C. T., Mathias, J. P. and Whitesides, G. M. (1993) *J. Am. Chem. Soc. 115*, 1321-1329.
9. Carcanague, D. R., Knobler, C. B. and Diederich, F. (1992) *J. Am. Chem. Soc. 114*, 1515-1517.
10. Morgan Conn, M., Deslongchamps, G., de Mendoza, J. and Rebek, J. Jr. (1993) *J. Am. Chem. Soc. 115*, 3548-3557.
11. Mallik, S., Johnson, R. D. and Arnold, F. H. (1994) *J. Am. Chem. Soc. 116*, 8902-8911.

12. Sinha, P.C., Saxena, P. K., Nigam, N. B. and Srivastava, M. N. (1989) *Indian J. Chem. 28A*, 335-336.

13. Searle, M. S., Williams, D. H. and Gerhard, U. (1992) *J. Am. Chem. Soc. 114*, 10697-10704.

14. Martell, A. E. and Smith, R. M. (1975) *Critical Stability Constants*, Vol. 1-6, Plenum Press, New York.

15. F. H. Arnold (1991) *Bio/Technology 9*, 151-156.

16. Birdsall, B., Roberts, G. C. K., Feeney, J., Dann, J. G. and Burgen, A. S. V. (1983) *Biochemistry 22*, 5597-5604.

17. Sulkowski, E. (1985) *Trends Biotech. 3*, 1-7.

18. Todd, R. J. (1993) *Ph. D. Thesis*, California Institute of Technology, Pasadena, CA.

19. Yip, T. T., Nakagawa, Y. and Porath, J. (1989) *Anal. Biochem. 183*, 159-171.

20. Porath, J., Carlsson, J., Olsson, I. and Belfrage, G. (1975) *Nature 258*, 598-599.

21. Porath, J. and Olin, B. (1983) *Biochemistry 22*, 1621-1630.

22. Chung, B. H. and Arnold, F. H. (1991) *Biotech. Letters 13*, 615-620.

23. Wuenschell, G. E., Wen, E., Todd, R. J., Shnek, D. R. and Arnold, F. H. (1991) *J. Chromatogr. 543*, 345-354.

24. J. W. Wong, R. L. Albright and N.-H. L.Wang (1991) *Sep. Pur. Meth. 20*, 49-106.

25. Abelson, J. N., Simon, M. I. and Arnold, F. H., Editors (1992) *Methods: A Companion to Methods in Enzymology 4*, 1-108.

26. Belew, M., Yip, T. T., Andersson, L. and Ehrnstone, R. (1987) *Anal. Biochem. 164*, 457-465.

27. Zhao, Y., Sulkowski, E. and Porath, J. (1991) *Eur. J. Biochem. 202*, 1115-1119.

28. Mrabet, N. T. (1992) *Biochemistry 31*, 2690-2702.

29. Hogg, P. J. and Winzor, D. J. (1984) *Arch. Biochem. Biophys. 234*, 55-60.

30. Yon, R. J. (1988) *J. Chromatogr. 457*, 13-23.

31. Kyprianou, P. and Yon, R. J. (1982) *Biochem. J. 207*, 549-556.

32. Dowd, V. and Yon, R. J. (1992) *J. Chromatogr. 627*, 145-151.

33. Todd, R. J., Johnson, R. D. and Arnold, F. H. (1994) *J. Chromatogr. A 662*, 13-26.

34. Mallik, S., Johnson, R. D. and Arnold, F. H. (1993) *J. Am. Chem. Soc. 115*, 2518-2520.

35. Mallik, S., Plunkett, S., Dhal, P. K., Johnson, R. D., Pack, D., Shnek, D. and Arnold, F. H. (1994) *New J. Chem. 18*, 299-304.

36. Hutchens, T. W., Yip, T. T. and Porath, J. (1988) *Anal. Biochem. 170*, 168-182.

37. Hutchens, T. W. and Yip, T. T. (1990) *Anal. Biochem. 191*, 160-168.

38. Boyer, P. M. and Hsu, J. T. (1992) *Chem. Eng Sci. 47*, 241-251.

39. Sips, R. (1950) *J. Chem. Phys. 18*, 1024-1026.

40. Bertini, I. and Luchinat, C. (1986) *NMR of Paramagnetic Molecules in Biological Systems*, Benjamin/Cummings, Menlo Park, CA.

CHAPTER 2

ANALYSIS OF SYNTHETIC BIS-METAL ION RECEPTORS FOR BIS-IMIDAZOLE "PROTEIN ANALOGS"

PREFACE

The following chapter is adapted from a previous publication (reprinted with permission from S. Mallik, R. D. Johnson and F. H. Arnold, "Selective recognition of bis-imidazoles by complementary bis-metal ion complexes," *Journal of the American Chemical Society 115*, 2518-2520, copyright 1993 American Chemical Society), and from a manuscript in press (reprinted with permission from S. Mallik, R. D. Johnson and F. H. Arnold, "Synthetic bis-metal ion receptors for bis-imidazole 'protein analogs'," *Journal of the American Chemical Society 116*, 8902-8911, copyright 1994 American Chemical Society).

This work was done in collaboration with Dr. Sanku Mallik. Dr. Mallik performed all syntheses of bis-metal ion receptors and bis-imidazole targets. Dr. Mallik and I together designed the format of equilibrium titration and competition experiments. Dr. Mallik performed the experiments, and I performed the analysis. The following chapter focuses on the analysis and subsequent interpretation of these experiments.

INTRODUCTION

Nature successfully engineers receptors to interact with specific proteins, as demonstrated by the recognition processes of the immune system [1]. Antibodies precisely coordinate arrays of weak interactions over large contact areas (> 1000 $Å^2$) to exhibit strong binding and high selectivity for a complementary site on antigens. The rational design of "artificial antibodies" to recognize proteins or other molecules in aqueous media presents significant challenges. Design and synthesis of a receptor incorporating a large number of weak complementary interactions complementary to a particular target is a difficult task [2,3]. To simplify receptor design and preparation, but still achieve useful selectivities and binding affinities, we have chosen strong metal-to-ligand coordination as the basis for recognition. This approach for preparing small, stable artificial receptors for proteins and other molecules matches an array of metal ions in a receptor to the distributions of metal-coordinating functional groups on the molecule of interest [4].

Molecular recognition based on metal-to-ligand interactions offers several potential advantages over schemes that rely upon hydrogen bonding [5], hydrophobic [6], or other interactions [7]. For a receptor to form a complex with its target, the entropic cost of bimolecular association must be paid by the thermodynamic forces that drive binding. One obvious attractive feature of metal-to-ligand interactions is that they are quite strong, even in aqueous media. The association constant for Cu^{2+} complexation by the imidazole side chain of histidine ($K = 5 \times 10^3$ M^{-1} [8]) translates to a binding energy of 4.8 kcal/mol. Thus, fewer individual interactions between receptor and target are required to achieve tight binding. In contrast, individual hydrogen bonding interactions generally contributes less than 1 kcal/mol to binding in water [9].

A second attractive feature is the availability of a wide variety of ligands and transition metal ions, which allows tailoring of both the strength and kinetics of binding. The strength of the metal-imidazole interaction, for example, varies over a wide range, depending on the metal ion ($K \sim 10^4\ M^{-1}$ for Cu^{2+}, $10^9\ M^{-1}$ for Hg^{2+} [10]). Complexes of imidazole with exchange-inert metal ions such as Co^{3+}, Ru^{2+}, or Pt^{2+} exhibit such slow ligand exchange kinetics that they are essentially covalent for many purposes. Previous investigations have been exploiting this feature, for example, to prepare high-fidelity polymer 'imprints' of template bis-imidazoles [11,12] and to stabilize proteins by crosslinking di-histidine metal-binding sites [13]. Exchange-labile transition metal ions offer interactions that can be formed and broken under relatively mild conditions. Thus the target molecule can be released from the metal-ion containing receptor by adjusting the pH or by addition of a competing ligand. Finally, the metal center, depending on the metal ion and its oxidation state, can have magnetic, spectroscopic and chemical properties which, for example, can be used for catalysis or to extract structural information.

The affinity of the metal ion for hydroxide may restrict the choice of metal ion for aqueous systems and necessitates careful pH control. Some transition metals form insoluble hydroxides in water at pH 7; Zn^{2+}-diethylenetriamine, for example, precipitates ~ pH 6, and Fe^{3+}-iminodiacetate precipitates at pH ~ 7. Even when the metal-hydroxide is water soluble, hydroxide can compete with the target molecule for metal binding. These shortcomings can be overcome by judicious choice of the metal ion and its chelating ligand. For example, surface histidine coordination to complexes of Cu^{2+}, Ni^{2+}, and Zn^{2+} have previously been shown to provide the basis for protein separation by metal-affinity chromatography [14].

Our approach to protein recognition exploits the protein's unique pattern of surface-exposed histidines (or other metal-coordinating functional groups), which can be

recognized by a receptor displaying a complementary pattern of transition metal ions. The simplest receptor would match the distance between two surface-exposed histidines with the distance between bis-metal ions in a complementary synthetic complex. The specificity of binding would be dictated by the degree to which two-point binding is favored over binding to a single surface histidine and the probability that no other protein in the sample has a pair of histidines with the same spacing. This is aided by the fact that histidines are relatively rare: ~2% of the amino acids in globular proteins are histidines, of which ~50% are surface exposed. Metal coordination could, of course, be used in conjunction with other interactions to further increase binding specificity.

To test the validity of this simple approach, several bis-metal receptors were synthesized, and their interactions with a series of bis-imidazole targets evaluated by nuclear magnetic resonance (NMR) [15,16]. The NMR signals of imidazole ring protons are a sensitive indicator of metal coordination at the imidazole nitrogen. The equilibrium binding affinities between bis-metal receptors and bis-imidazole targets are measured by NMR spectroscopy in equilibrium titration and competition experiments. This work defines the extent to which two metal-imidazole coordination interactions can provide selectivity in binding a target molecule, and also probes specific features of the receptor design: the fidelity of distance matching and the consequences of changing receptor mobility and multiplicity of metal coordination sites.

Target and receptor design

Because force-field parameters appropriate for metal-to-ligand interactions were not readily available, Estimated conformations of bis-imidazole targets and bis-metal receptors were determined independently by molecular modeling [16].

Bis-imidazole targets ("protein analogs") were constructed to position two imidazoles at pre-determined distances using organic spacers of varying lengths, as

shown in Figure 2.1. The bis-imidazole targets are designated **T1**, **T2** and **T3**, depending on the number of phenyl rings separating the imidazole moieties. The distances between the imidazole nitrogens of the targets were estimated by molecular modeling as 7-8 Å for **T1**, 11-12 Å for **T2** and 16-18 Å for **T3** [16]. 1-Benzylimidazole (**I**), with a single imidazole group, was used as a control in the binding experiments.

A series of complementary bis-metal receptors, shown in Figure 2.2, were designed such that the distance between two coordinated metal ions was approximately the same as the distance of two imidazoles on a particular target. A diamagnetic metal ion (Hg^{2+}) was used in these model studies to monitor binding by ^1H NMR. The mercury ion was held in place by either cyclam (1,4,8,11-tetraazacyclotetradecane [17]) or dien (diethylenetriamine); the formation constants in water are 10^{23} M^{-1} for Hg^{2+}-cyclam [18] and 10^{11} M^{-1} for Hg^{2+}-dien [10]. In the case of cyclam, the large ionic radius of Hg^{2+} (1.10 Å) prevents it from remaining in plane with the four nitrogen atoms of the macrocycle. The macrocycle therefore effectively blocks one face of the metal ion [19], and each metal center in the cyclam-based receptors ($\mathbf{R_{cc}}$, $\mathbf{R_{cxc}}$, $\mathbf{R_{crc}}$, and $\mathbf{R_{cnc}}$) presents only one possible coordination site for imidazole. In contrast, each metal ion in the Hg^{2+}-dien complex has three vacant coordination sites [20]. Therefore, the bis-Hg^{2+}-dien receptor $\mathbf{R_{dxd}}$ was designed to investigate the effect on selectivity of having more than one coordination site available for imidazole binding.

The principle of using distance between the metal ions as a design parameter was evaluated using three of the cyclam-based receptors. For the series of receptors $\mathbf{R_{cc}}$, $\mathbf{R_{cxc}}$, and $\mathbf{R_{cnc}}$, the Hg-cyclam moieties are separated by increasingly long organic spacers (Figure 2.2). The distances between the mercury ions of each receptor were estimated as 13 Å for $\mathbf{R_{cnc}}$, 11 Å for $\mathbf{R_{cxc}}$, and 8 Å for $\mathbf{R_{cc}}$ [16]. The principle of distance matching predicts, for example, that the receptor $\mathbf{R_{cxc}}$ would form a 1:1 cyclic

complex with the target **T2** as illustrated in Figure 2.3B, and therefore prefer it over other imidazole species.

The series of receptors $\mathbf{R}_{crc}$, $\mathbf{R}_{cxc}$, and $\mathbf{R}_{dxd}$ was designed to evaluate the effect of conformational mobility on receptor selectivity. For example, the bis-metal ion receptors and bis-imidazole targets each have four benzylic bonds capable of rotating freely. Free rotations of these bonds of the free target and receptor are frozen in the proposed 1:1 complex, illustrated in Figure 2.3, and there is an estimated entropic cost of ~ 0.8 kcal mol^{-1} in restricting the rotation of a hydrocarbon bond [21]. Therefore, decreasing the flexibility of the free receptor could improve the final selectivity of binding by decreasing the penalty of constraining it in the 1:1 complex [2,21]. This rationale is behind the design of $\mathbf{R}_{crc}$ (Figure 2.2D), in which the four methyl groups on the phenyl ring constrain the rotation of the benzylic bonds. Alternatively, formation of the proposed 1:1 complex could be inhibited by non-optimal coordination geometry. Therefore, increasing the number of available metal coordination sites could present alternative geometries and improve the final selectivity [3]. For the receptor $\mathbf{R}_{dxd}$ the mercury ion is held in place by a dien moiety. Because of its multiplicity of coordination sites, metal-to-metal distances ranging from 8-11 Å appear to support bis-imidazole binding [16].

CALCULATIONS

The coordination of imidazoles by metal ions of the receptors is in fast exchange on the NMR time scale [15]; the observed imidazole NMR chemical shift is thus a weighted average between the signals of the metal-free and metal-coordinated species. Several simultaneous equilibria can be expected to exist in a mixture of bis-metal receptor and bis-imidazole targets. Only the equilibria illustrated in Figure 2.4 will be

included in this analysis. That is, bis-imidazole targets are considered to form 1:1 complexes with bis-imidazole receptors, and 1-benzylimidazole is considered to form complexes of 1:1 and 2:1 stoichiometry with the receptors. The observed imidazole chemical shifts in equilibrium titration and competitive titration experiments will be interpreted by equations describing the multiple equilibria of Figure 2.4. Nonlinear regression of these equations will then determine values of the three equilibrium binding constants K_1, K_2, and K_T.

Titration of bis-imidazole target with bis-mercury receptor

The analysis of equilibrium titration experiments between a bis-imidazole target and a bis-mercury receptor presumes the formation of only 1:1 complexes (see Figure 2.4). The equilibrium association constant (K_T) is defined as

$$K_T = \frac{[\mathrm{R}][\mathrm{T}]}{[\mathrm{RT}]} \quad , \tag{1}$$

where [R] and [T] are the concentrations of free receptor and target and [RT] is the concentration of the 1:1 target-receptor complex. Eliminating the concentration of free [R], this concentration can be expressed as

$$[\mathrm{RT}] = K_T[\mathrm{R}]_t[\mathrm{T}]/(1+K_T[\mathrm{T}]) \quad , \tag{2}$$

where $[\mathrm{R}]_t$ is the total amount of receptor.

In the fast exchange limit, the observed imidazole chemical shift of the target δ(T) is given by

$$\delta(T) = \delta_T \frac{[RT]}{[T]_t} \quad , \tag{3}$$

where δ_T is the change in chemical shift of the target imidazole upon coordination by the receptor metal ion. The total concentration of target $[T]_t$ must satisfy the mass balance,

$$[T]_t = [T] + [RT] \quad . \tag{4}$$

Given the change in chemical shift and the total concentration of target, the free target concentration can therefore be measured directly from its observed chemical shift,

$$[T] = [T]_t\left(1 - \delta(T)/\delta_T\right) \quad . \tag{5}$$

This results in an implicit expression (eqns. 2, 3, and 5) for the observed chemical shift $\delta(T)$ in terms of itself, the total concentrations ($[R]_t$ and $[T]_t$), the association constant (K_T), and the change in chemical shift (δ_T). A nonlinear regression procedure was developed (see Appendix) to determine the two unknown parameters δ_T and K_T from equilibrium titrations (data of the form $[R]_t$, $[T]_t$, $\delta(T)$) by minimizing the error between the set of observed chemical shifts and the set of chemical shifts calculated by the implicit expression. In this procedure, estimated errors were calculated relative to the "saturation fractions" as defined by Lenkinski *et al.* [22,23].

Titration of 1-benzylimidazole with bis-mercury receptors

Equilibrium titration experiments of 1-benzylimidazole with the bis-mercury receptors were analyzed by a variation of the method of Lenkinski for molecular complexes with 2:1 stoichiometry. In this analysis, a receptor binds two equivalents of

imidazole with equilibrium association constants (K_1 and K_2 illustrated in Figure 2.4) defined as

$$K_1 = \frac{[RI]}{[R][I]} \quad , \tag{6}$$

$$K_2 = \frac{[IRI]}{[IR][I]} \quad , \tag{7}$$

where [I] and [R] are the concentrations of free imidazole and receptor, and [RI] and [IRI] are the concentrations of receptor with one and two equivalents of imidazole bound, respectively. The resultant concentrations of the two receptor-imidazole species are

$$[RI] = K_1[R]_t[I] / \left(1 + K_1[I] + K_1K_2[I]^2\right) \quad , \tag{8}$$

$$[IRI] = K_1K_2[R]_t[I]^2 / \left(1 + \alpha_T K_1[T] + K_1[I] + K_1K_2[I]^2\right) \quad . \tag{9}$$

Assuming a fast-exchange limit, the observed chemical shift δ(I) of the imidazole control is the sum the chemical shifts from the two imidazole-containing species

$$\delta = \delta_1 \frac{[RI]}{[I]_t} + 2\delta_2 \frac{[IRI]}{[I]_t} \quad , \tag{10}$$

where δ_1 and δ_2 are the imidazole chemical shifts with one and two equivalents of 1-benzylimidazole bound to the receptor, respectively. The total concentration of 1-benzylimidazole $[I]_t$ must the satisfy the mass balance,

$$[I]_t = [I] + [RI] + 2[IRI] \quad . \tag{11}$$

If the chemical shifts δ_1 and δ_2 are identical, then the concentration of free imidazole could be calculated directly from the observed chemical shift (by analogy to eqn. 5). However, if the chemical shifts are not equal for bound substrate in 1:1 and 2:1 complexes, then this analogy does not hold and it is *not possible* to calculate the free concentration ([I]) directly from the observed chemical shift. Calculations using the larger of the two chemical shifts would overestimate the free concentration; those using the smaller of the two would underestimate it. There is still, however, a unique free concentration that satisfies the mass balance for a given set of association constants (K_1, K_2) and total species concentrations ($[I]_t$, $[R]_t$). The following analysis was developed to estimate this free concentration with sufficient accuracy for nonlinear regression.

For each set of concentrations there is a chemical shift δ_{avg} (where $\delta_2 < \delta_{avg} < \delta_1$) for which an equation analogous to eqn. 5 *would* predict the correct free concentration,

$$[I] = [I]_t\left(1 - \delta\,(I)/\delta_{avg}\right) \quad . \tag{12}$$

If the chemical shifts of bound imidazole are *approximately* the same for the species with one and two equivalents of 1-benzylimidazole bound to the receptor ($\delta_1 \sim \delta_2$), then it is possible to estimate accurately the free concentration in a single step. This procedure defines δ_{avg} as a weighted average of δ_1 and δ_2,

$$\delta_{avg} = (1 - \beta)\delta_1 + \beta\delta_2 \quad , \tag{13}$$

where $0 < \beta < 1$. In this case the two chemical shifts are given as

$$\delta_1 = \delta_{avg} - \beta\Delta \quad , \tag{14}$$

$$\delta_2 = \delta_{avg} + (1-\beta)\Delta \quad , \tag{15}$$

where Δ is the difference in chemical shifts between the 1:1 and 2:1 complexes ($\delta_2 < \delta_1$),

$$\Delta = \delta_2 - \delta_1 \quad . \tag{16}$$

Under these conditions, the free concentration of [I] can be expressed in terms of the unknown concentrations [RI] and [IRI] as

$$[\mathrm{I}] = [\mathrm{I}]_t \left(1 - \frac{\delta}{\delta_{avg}}\right) - \frac{\Delta}{2\delta_{avg}} \left(\beta\left([\mathrm{RI}] + 2[\mathrm{IRI}]\right) - 2[\mathrm{IRI}]\right) \quad . \tag{17}$$

The error in calculating [I] by eqn. 12 is therefore minimized when β is equal to the fraction of receptor-bound imidazole in the form of 2:1 complexes,

$$\beta = \frac{2[\mathrm{IRI}]}{2[\mathrm{IRI}] + [\mathrm{RI}]} = \frac{2K_2[\mathrm{I}]}{2K_2[\mathrm{I}] + 1} \quad . \tag{18}$$

These equations were used to develop a procedure for nonlinear regression (see Appendix) of the four unknown parameters δ_1, δ_2, K_1, and K_2 from equilibrium titration experiments (data of the form $[\mathrm{R}]_t$, $[\mathrm{I}]_t$, $\delta(\mathrm{I})$). For initial estimates of δ_1 and δ_2, a first estimate of [I] is calculated by (eqn. 12) using $\beta = 1/2$, providing a fractional error of order Δ. The first estimate is used to calculate a revised estimate of β (eqn. 18) and thus

more accurate estimates of δ_{avg} and of [I], providing a fractional error of order Δ^2. With the revised estimate, the predicted chemical shift is calculated (eqn. 10). The optimum values of the four parameters minimize the error in the set of predicted chemical shifts from the observed chemical shifts.

It has previously been demonstrated by Lenkinski *et al.* [22] that, for species such as 1-benzylimidazole which form 2:1 complexes, the observed chemical shift can be linearized as

$$\frac{1}{\delta(\mathrm{I})}\cdot\frac{2[\mathrm{IRI}]}{[\mathrm{I}]_t}=\frac{1}{\delta_2}-\frac{\delta_1}{\delta_2}\cdot\frac{1}{\delta(\mathrm{I})}\cdot\frac{[\mathrm{RI}]}{[\mathrm{I}]_t}\quad , \tag{19}$$

where $2[IRI]/[I]_t$ and $[RI]/[I]_t$ are termed the "saturation fractions" for each set of total concentrations. If the association constants are estimated accurately, such a plot will give a straight line (slope $-\delta_1/\delta_2$ and intercept $1/\delta_2$). This condition provides a diagnostic for the accuracy of the association constants (K_1 and K_2) and chemical shifts (δ_1 and δ_2) as determined by nonlinear regression.

Competition among bis-imidazole targets for bis-mercury receptor

Competitive titration experiments were analyzed using parameters determined in the equilibrium titration experiments. In competition among bis-imidazole targets (e.g., **T1** and **T2**), the chemical shift parameters δ_{T1} and δ_{T2} were determined from the individual equilibrium titration experiments of each target. The concentration of each receptor-target complex is given by

$$[\mathrm{RT1}]=K_{T1}[\mathrm{R}]_t[\mathrm{T1}]/\left(1+K_{T1}[\mathrm{T1}]+K_{T2}[\mathrm{T2}]\right)\quad , \tag{20}$$

$$[\mathrm{RT2}]=K_{T2}[\mathrm{R}]_t[\mathrm{T2}]/\left(1+K_{T1}[\mathrm{T1}]+K_{T2}[\mathrm{T2}]\right)\quad . \tag{21}$$

In conjunction with the relationships of the individual bis-imidazole titrations (eqn. 5), these expressions result in implicit equations for the two observed chemical shifts δ(T1) and δ(T2). These equations were used to develop a procedure (see Appendix) for nonlinear regression of the two association constants K_{T1} and K_{T2} from equilibrium competition experiments (data of the form $[R]_t$, $[T1]_t$, $[T2]_t$, δ(T1), δ(T2)). The optimum values of K_{T1} and K_{T2} minimize the error in the predicted chemical shifts relative to the observed chemical shifts.

The selectivity of the receptor for target **T2** over target **T1** is a measure of the relative amount of each target bound to the receptor. In the case of competition among bis-imidazoles, the selectivity will be the ratio of the association constants,

$$\alpha_{T2/T1} = \frac{\left([T2]_{bound}/[T2]_{free}\right)}{\left([T1]_{bound}/[T1]_{free}\right)} = \frac{K_{T2}}{K_{T1}} \quad . \tag{22}$$

The concentration of free **T1** and **T2** can be calculated from their observed chemical shifts (eqn. 5). The concentration of bound **T1** and **T2** can be determined from the mass balance (eqn. 4). Thus the selectivity can be calculated directly from the observed chemical shifts as follows:

$$\alpha_{T2/T1} = \frac{\left(\delta_{T1}/\delta\,(T1) - 1\right)}{\left(\delta_{T2}/\delta\,(T2) - 1\right)} = \frac{K_{T2}}{K_{T1}} \quad . \tag{23}$$

Competition between 1-benzylimidazole and target bis-imidazoles

In competitive titration experiments, the chemical shifts of 1-benzylimidazole and a bis-imidazole target in the presence of receptor were measured at increasing

concentrations of 1-benzylimidazole. The parameters δ_1, δ_2, δ_T, K_1, and K_2 can be taken from the equilibrium titration experiments. If the concentration of free imidazole and free target are known, the concentrations of the three receptor-bound species are given by

$$[\mathrm{RI}] = K_1[\mathrm{R}]_t[\mathrm{I}]/\left(1+K_T[\mathrm{T}]+K_1[\mathrm{I}]+K_1K_2[\mathrm{I}]^2\right) , \tag{24}$$

$$[\mathrm{IRI}] = K_1K_2[\mathrm{R}]_t[\mathrm{I}]^2/\left(1+K_T[\mathrm{T}]+K_1[\mathrm{I}]+K_1K_2[\mathrm{I}]^2\right) , \tag{25}$$

$$[\mathrm{RT}] = K_T[\mathrm{R}]_t[\mathrm{T}]/\left(1+K_T[\mathrm{T}]+K_1[\mathrm{I}]+K_1K_2[\mathrm{I}]^2\right) . \tag{26}$$

As outlined above, the concentrations of free imidazole and target can be estimated from the observed chemical shifts. Then the predicted chemical shifts can be calculated. A procedure was developed (Appendix A) for nonlinear regression of the competition experiments (data of the form $\{[\mathrm{R}]_t, [\mathrm{I}]_t, [\mathrm{T}]_t, \delta(\mathrm{I}), \delta(\mathrm{T})\}$ by minimizing the error in chemical shifts of both imidazole species.

The selectivity of the receptor for target over the control is a measure of the preference for two-site binding over single-site binding. In this case, the selectivity is dependent on the concentration of free 1-benzylimidazole, reaching a maximum at infinite dilution,

$$\alpha_{T/I} = \frac{\left([\mathrm{T}]_{bound}/[\mathrm{T}]_{free}\right)}{\left([\mathrm{I}]_{bound}/[\mathrm{I}]_{free}\right)} = \frac{K_T}{K_1}\frac{1}{1+2K_2[\mathrm{I}]} . \tag{27}$$

The concentration of free target [T] can be calculated from its observed chemical shift. As previously noted, the concentration of free imidazole [I] cannot be determined solely from chemical shift information, precluding a direct calculation of the selectivity from

chemical shift information. However, an *apparent* selectivity (α^{app}) can be calculated directly from chemical shift information presuming $\beta=0$ ($\delta_{avg}=\delta_1$) as follows:

$$\alpha_{T/I}^{app} = \frac{(\delta_1/\delta\,(\mathrm{I})-1)}{(\delta_T/\delta\,(\mathrm{T})-1)} = \frac{K_T}{K_1}\frac{1+\Delta K_1 K_2[\mathrm{R}][\mathrm{I}]/\delta_1}{1+K_2[\mathrm{I}](2-\Delta/\delta_1)} \quad . \tag{28}$$

As shown, this apparent selectivity differs from the true selectivity (eqn. 27) with a fractional error of order Δ. This error will be seen to be small relative to the observed concentration dependence of selectivity.

RESULTS AND DISCUSSION

Titration of 1-benzylimidazole

In equilibrium titration experiments between 1-benzylimidazole and the bis-metal receptors (Figure 2.5, see also Appendix), the 1-benzylimidazole control appears to form complexes of greater than 1:1 stoichiometry with the receptors. This is consistent with the equilibria presented in Figure 2.4, in which 1-benzylimidazole can bind to both metal ions (Figure 2.3). Following the approach of Lenkinski *et al.* [22] the two association constants (K_1 and K_2) were calculated from simultaneous analysis of two types of NMR titration experiments: a solution of **I** (~ 5 mM) was titrated with increasing amounts of receptor (1 - 10 mM, referred to as Type I titration), and a solution of receptor (~ 10 mM) was titrated with increasing amounts of **I** (2 - 20 mM, Type II titration). Good estimates of the first association constant and chemical shift of the 1:1 complex can be obtained from Type I titration data, while Type II titrations are more sensitive to the second association constant and chemical shift of the 2:1 complex. Thus the titration curves resulting from these two types of experiments are slightly different (Figure 2.5).

However, when appropriate values of the association constants K_1 and K_2 are chosen, these two titration curves collapsed to a single line on the linearized plot of Lenkinski *et al.*, as shown in Figure 2.6.

The 1-benzylimidazole C_2H chemical shift is shifted downfield approximately 0.7 ppm upon binding the bis-metal receptors, as presented in Table 2.1. The change is due predominately to mercury coordination at the imidazole nitrogen, although there appears to be a secondary effect due to the aromatic spacer. For example, the shift increases slightly, but consistently, with the number of aromatic rings on the receptor ($\mathbf{R}_{cc} < \mathbf{R}_{dxd}$, $\mathbf{R}_{crc}$, $\mathbf{R}_{cxc} < \mathbf{R}_{cnc}$), following the same order as aromatic ring shift currents for the spacers (no ring < benzene < naphthalene) [24]. These features and the 2:1 stoichiometry are consistent with the schematic structure for 1-benzylimidazole binding to the receptor (Figure 2.3A), in which each metal ion coordinates a single 1-benzylimidazole molecule.

As outlined above, the approximation that $\delta_1 \sim \delta_2$ was used in developing the nonlinear regression procedure for imidazole titrations. For all five receptors, the differences between the imidazole chemical shifts of the 1:1 and 2:1 complexes (δ_1 and δ_2) are small, at most 15%. Thus the free imidazole concentration calculated by the regression procedure can be estimated to have a relative error of approximately 2% (see Appendix).

The first association constants for 1-benzylimidazole are found to be approximately the same for all receptors ($\sim 10^4\ M^{-1}$), within a factor of two (Table 2.2). The second association constants, directing coordination of a second equivalent of 1-benzylimidazole (see Figure 2.4), show a somewhat wider distribution (10^2-$10^3\ M^{-1}$) among the receptors. In principle, if the first and second binding events are completely *independent*, then the second association constant would be one-fourth of the first, due to statistical effects [22]. For the bis-metal receptors studied, the second association

constant for 1-benzylimidazole is always at least an order of magnitude less than the first. Thus the presence of one imidazole *inhibits* binding of the second, by as little as a factor of 2 or as much as a factor of 30 (~ 2 kcal/mol).

Titration of bis-imidazoles

In equilibrium titration experiments between the bis-imidazole targets and the bis-metal receptors (Figure 2.7, see also Appendix), the bis-imidazole species appear to form complexes of 1:1 stoichiometry, and therefore only Type II titration experiments were performed. The imidazole C_2H NMR signals of the targets are shifted downfield by 0.7-0.8 ppm upon binding to the receptors (Table 2.2). The similarity of this chemical shift with that observed for 1-benzylimidazole is evidence that both imidazoles are simultaneously coordinated by mercury ions. Furthermore, in the case of receptor $\mathbf{R_{cxc}}$, the target bis-imidazole **T2** (with the appropriate distance between imidazoles) experiences a larger chemical shift upon binding than the other imidazoles **T1**, **T3,** and **I**. The effect is even more pronounced for receptor $\mathbf{R_{cnc}}$ with the naphthalene spacer. This apparent deshielding of **T2** by the ring current of the receptor's aromatic spacer and the measured 1:1 stoichiometry for bis-imidazole binding are consistent with the schematic structure of the target-receptor complex, in which the two imidazole moieties are simultaneously coordinated by the two mercury ions as shown in Figure 2.3B.

The changes in chemical shift of the bis-imidazoles are determined accurately by nonlinear regression of the equilibrium titration experiments (Table 2.2). However, at these concentrations (1-10 mM), such data do not determine the association constants accurately. Because the receptors can bind the bis-imidazoles with high affinity (association constants greater than 10^4 M^{-1}), titration experiments would have to be performed at concentrations less than 0.1 mM in order to obtain reliable values of these binding constants directly [25]. Such a concentration regime is difficult to probe by

NMR. Binding constants for bis-imidazoles are therefore determined indirectly, in competitive titration experiments with 1-benzylimidazole at concentrations suitable for NMR (1 - 10 mM).

Competitive titration

In competitive titration experiments, the chemical shifts of both the bis-imidazole target and 1-benzylimidazole (e.g., **T2** and **I** as shown in Figure 2.8) were followed as an equimolar mixture of target and receptor (~ 10 mM **T2** and $\mathbf{R}_{cxc}$) was titrated with increasing amounts of 1-benzylimidazole (2-20 mM **I**). These experiments illustrate the selectivity that a bis-metal receptor shows for a particular bis-imidazole over the 1-benzylimidazole control. For example, in Figure 2.8 the majority of bis-imidazole **T2** remains bound to the receptor $\mathbf{R}_{cxc}$, even in the presence of a stoichiometric amounts of **I**, indicating that the bis-metal receptor is indeed selective for its target over the single-imidazole control.

The nonlinear regression of the competitive titration experiments is sensitive to the ratio of association constants (e.g., K_{T2}/K_1, Table 2.4) and not to individual values of these parameters (e.g., K_{T2} and K_1). As discussed above, it is not possible to determine the bis-imidazole association constants (K_T) directly from bis-imidazole titration experiments (e.g., Figure 2.5). These association constants could instead be calculated indirectly from the 1-benzylimidazole association constant (K_1) and the ratio of association constants determined by the competitive titration. As shown in Table 2.4, the resultant bis-imidazole association constants reach as high as 10^6 M^{-1}. The increase in binding constant over that of 1-benzylimidazole (~10^4 M^{-1}) results from the second properly positioned metal-to-ligand interaction (Figure 2.3).

In a second set of competitive titration experiments, the chemical shifts of both bis-imidazoles were followed (e.g., **T1** and **T2** as shown in Figure 2.9) as an equimolar

mixture of one target and the receptor (~ 10 mM **T2** and $\mathbf{R}_{cxc}$) was titrated with increasing amounts of a competing target (2-20 mM **T1**). In these experiments, the bis-metal receptor selects between the two bis-imidazoles. For example, in Figure 2.9 most of the bis-imidazole **T2** remained bound to the receptor $\mathbf{R}_{cxc}$, even in the presence of excess amounts of competing targets **T3** or **T1**. Thus this particular receptor ($\mathbf{R}_{cxc}$) is indeed selective for the bis-imidazole target **T2** over the other bis-imidazoles. The ratio of the bis-imidazole association constants (e.g., K_{T2}/K_{T1}) could be calculated accurately by nonlinear regression of the observed chemical shifts of both bis-imidazole species in the competitive titration experiments (eqn. 22), using the chemical shift parameters previously determined by the bis-imidazole titration experiments (Table 2.2). The resultant selectivities for different pairs of bis-imidazoles are listed in Table 2.4.

Evaluation of equilibrium model

The simple model depicted in Figure 2.4, considering only complexes of 1:1 stoichiometry between bis-imidazoles and the bis-metal receptor, accurately describes the observed chemical shifts in the competitive titrations experiments between bis-imidazoles (e.g., Figure 2.9). This important simplification reduces the problem to only two simultaneous equilibria (eqns. 20 and 21) and assumes no significant formation of mixed complexes (e.g., **T1-R-T2**) or polymeric complexes (e.g., **T1-R-T1-R-T1 ...**). Such a simplification should be particularly relevant to protein-receptor interactions (di-histidine proteins taking the place of bis-imidazole targets), because steric interactions between proteins should prevent the formation of mixed or polymeric complexes. However, this assumption is not absolute, as demonstrated by the aggregation and precipitation of multiple-histidine proteins by some small, bis-copper complexes [26].

The observed imidazole chemical shift dependence of competitive titrations between bis-imidazoles and 1-benzylimidazole (e.g., Figure 2.8) are also accurately

described by the model depicted in Figure 2.4, which considers both 1:1 and 2:1 complexes between 1-benzylimidazole and the receptor. However, in general these three equilibria are more successful in describing the chemical shifts of the bis-imidazole and less successful in describing the chemical shifts of 1-benzylimidazole (see also Appendix). Such a discrepancy may indicate significant formation of mixed complexes (e.g., **T2-R-I**), which would tend to increase the amount of 1-benzylimidazole bound to the receptor at lower concentrations of **I**. As mentioned above, in the case of protein-ligand interactions (single-histidine proteins taking the place of 1-benzylimidazole), steric interactions between proteins would be anticipated to inhibit the formation of such mixed complexes. Nevertheless, the inclusion of 2:1 complexes of **I** is a dramatic improvement over the previous analysis [15] and reproduces more accurately the concentration dependence in the observed selectivity, as shown in Figure 2.10.

If the equilibrium model of Figure 2.4 provides a consistent description of these two types of competitive titration experiments, then the calculated ratio of bis-imidazole association constants (K_{T2}/K_{T1}) should be the same whether (a) measured directly in bis-imidazole competition experiments (Table 2.4) or (b) measured indirectly in separate competition experiments with 1-benzylimidazole (Table 2.3). Comparing the selectivities determined for target **T2** over **T1** by both methods, the two types of competition experiments are in close agreement for all five receptors. This self-consistency is an important validation of the simplified model.

Two-point binding in receptor design

The preference of the bis-metal receptors for the bis-imidazole targets over 1-benzylimidazole demonstrates that two-point binding indeed is favored over single-point binding. The fraction of bis-imidazole target (**T1**, **T2**, or **T3**) bound to the receptor compared to the fraction of 1-benzylimidazole **I** bound (eqn. 28) can be taken as a

practical measure of the selectivity that the receptor shows for two-point binding over single-point binding. As shown in Figure 2.10 for the receptor $\mathbf{R}_{CXC}$, the apparent selectivity for a bis-imidazole target over 1-benzylimidazole is strongly concentration dependent. This is to be expected because the 2:1 complex (e.g., $\mathbf{I}{:}\mathbf{R}_{CXC}{:}\mathbf{I}$, see Figure 2.4) is a more effective competitor for bis-imidazoles than the 1:1 complex (e.g., $\mathbf{I}{:}\mathbf{R}_{CXC}$). At very low concentration of **I**, the selectivity is predicted to reach a maximum value equal to the ratio of the association constants (α = 140 for **T2** and α = 10 for **T1**).

It is useful to compare these observed binding constants and selectivities can be compared to those obtainable for *ideal* two-point binding based on simple energetic arguments. For simultaneous coordination of the two imidazole moieties to the two metal ions in a receptor, the minimum overall binding energy should approximate the overall binding energy for binding two molecules of 1-benzylimidazole (Figure 2.3), and the association constant K_T would be the same order of magnitude as the product of the 1-benzylimidazole association constants K_1 and K_2 (Table 2.1). The selectivity that should be obtained for ideal two-point binding over one-point binding is therefore of the same order of magnitude as K_2 (~ 10^3). In most cases, the observed association constants for the target bis-imidazoles are considerably less than K_1K_2 (and the selectivities considerably less than K_2), indicating that two-point binding is not particularly favored. For the receptor $\mathbf{R}_{CXC}$, however, the selectivity for its target **T2** over **I** (α = 140) is approximately 1/3 the value of K_2 (420 M^{-1}). In this case two-point binding does not appear to present a severe energy penalty relative to binding two independent imidazoles. Such an energy penalty would be a minimized if, as illustrated in Figure 2.3, the conformation of the bound bis-imidazole can closely match the conformation of two bound 1-benzylimidazoles.

Distance matching in receptor design

The preference of bis-metal receptors for an appropriate bis-imidazole target over shorter or a longer competitors is evidence that distance-matching is a valid design criterion. The fraction of receptor-bound target **T2** relative to **T1** is a practical measure of the selectivity that a particular receptor shows for **T2** over **T1** (see eqn. 23). Because the receptor is assumed to form only 1:1 complexes with the bis-imidazole targets, the selectivity among targets is predicted to be independent of the concentration. As shown in Figure 2.10, the selectivity of the receptor $\mathbf{R}_{cxc}$ for the bis-imidazoles **T2** over **T1** is approximately constant ($\alpha \sim 12$) over the studied concentration regime, in contrast with the apparent selectivity for bis-imidazoles over 1-benzylimidazole. Thus the observed concentration dependence validates the assumption that the receptor forms primarily 1:1 complexes with both bis-imidazole species.

The distances between the two imidazole nitrogens for **T1**, **T2** and **T3** increase by approximately 4 Å along the series (Figure 2.1). In competitive titration experiments, $\mathbf{R}_{cxc}$ shows the highest discriminating capability, as presented in Table 2.4. It shows a selectivity of 11.5 for target **T2** over the shorter bis-imidazole **T1**, and 10.2 for **T2** over the longer bis-imidazole **T3**. Neither the longer ($\mathbf{R}_{cnc}$) nor shorter ($\mathbf{R}_{cc}$) receptors in the series demonstrate the same selectivity for any of these targets. In fact, it appears that the receptor $\mathbf{R}_{cc}$ may be unable to bind target bis-imidazoles by two-point binding; bis-imidazole binding is only marginally favored over 1-benzylimidazole (Table 2.3). Consequently, the receptor's selectivity among the pair of the bis-imidazoles is negligible ($\alpha \sim 1$).

Conformational flexibility in receptor design

Receptor flexibility is another important component of receptor design. The series of receptors $\mathbf{R}_{crc}$, $\mathbf{R}_{cxc}$, and $\mathbf{R}_{dxd}$, all of which have a single phenyl ring as a

spacer, demonstrates the intricate balance between selectivity and conformational degrees of freedom. For example, the four methyl groups in receptor $\mathbf{R}_{crc}$ result in restricted rotation around the benzylic bonds, decreasing the number of conformational degrees of freedom. On the other hand, the receptor $\mathbf{R}_{dxd}$ has multiple coordination sites available to the mercury ion, increasing the number of viable target-receptor conformations.

The formation of a cyclic complex illustrated in Figure 2.3 freezes the rotational degrees of freedom benzylic bonds and therefore may place an entropic penalty on complex formation. Computer modeling predicts that steric repulsion from the nearby methyl groups restrains the rotation of the benzylic bonds in the free receptor $\mathbf{R}_{crc}$ relative to the same bonds in receptor $\mathbf{R}_{cxc}$ [16]. Although $\mathbf{R}_{crc}$ was designed with the idea that restricting the free rotations about the benzylic bonds would increase the selectivity for **T2** over **T1**, interestingly, the net effect of the methyl groups is to decrease the selectivity from 11.5 to 2.7 (Table 2.4). The source of this decrease is a two-fold increase in affinity of $\mathbf{R}_{crc}$ for the bis-imidazole **T1** relative to binding two 1-benzylimidazoles (K_{T1}/K_1K_2 = 24/470 versus 10/420 for $\mathbf{R}_{cxc}$) as well as a reduction in affinity for the target **T2** (K_{T1}/K_1K_2 = 89/470 versus 140/420 for $\mathbf{R}_{cxc}$). Because the distances measured between metal ions of $\mathbf{R}_{crc}$ and $\mathbf{R}_{cxc}$ are identical (Figure 2.2), the shift in selectivity of the receptor $\mathbf{R}_{crc}$ towards binding the smaller bis-imidazoles could not be predicted from the simple computer modeling of the individual receptors and targets [16].

It was also anticipated that some flexibility in synthetic receptors would be desirable to allow for non-optimal matching between the receptors and their targets. Molecular modeling predicts multiple conformations of the receptor $\mathbf{R}_{dxd}$ which present two mercury coordination sites suitable for coordination to a single bis-imidazole [16]. This is a consequence of the three vacant Hg^{2+} coordination sites on each metal center left unoccupied by the tridentate diethylenetriamine. One such conformation more

closely matches the bis-imidazole **T1** while the another conformation more closely matches the bis-imidazole **T2**. The relative population of such conformers will therefore determine the binding selectivity. This additional conformational flexibility decreases the selectivity for **T2** over **T1** by a factor of 20 (relative to $\mathbf{R}_{cxc}$, see Table 2.4), so that this receptor in fact shows a small preference for **T1** over **T2** ($\alpha_{T2/T1} \sim 0.5$). Bis-imidazole binding by this receptor is not particularly favored: the association constants for both **T1** and **T2** are much smaller than those observed for binding to the analogous cyclam-based receptors $\mathbf{R}_{cxc}$ (Table 2.3). This unfavorable two-point binding is consistent with the relatively small binding affinity of receptor $\mathbf{R}_{dxd}$ for a second equivalent of 1-benzylimidazole (K_2 = 140 M^{-1}, Table 2.1).

CONCLUSIONS

Although several simultaneous equilibria can be expected to exist in a mixture of bis-metal receptor and bis-imidazole targets, equilibrium binding experiments are accurately described by a simple model considering that (a) the receptor forms only 1:1 complexes with bis-imidazole targets, and (b) the receptor forms complexes of both 1:1 and 2:1 stoichiometry with 1-benzylimidazole. The selectivity of a bis-metal receptor for two-point binding to bis-imidazole targets over one-point binding to 1-benzylimidazole control is calculated from competitive titration experiments to be as high as 140. Assuming recognition behaves comparably in an aqueous medium and that proteins are precluded from forming 2:1 complexes with small receptor complexes, then receptor binding to a specific pair of histidines could be strongly favored over binding to a single surface histidine. Competitive titration experiments also show that the bis-metal receptors select for a complementary bis-imidazole over a longer or a shorter bis-imidazole. Thus it should be possible to design a receptor to distinguish with relatively

high selectivity (>10) among pairs of protein surface histidines differing in distance by as little as 4 Å. It is also apparent that conformational mobility of the receptor (free rotations or available coordination sites) is of critical importance to the ability of the receptor to distinguish between bis-imidazole targets. These studies validate the principle of molecular recognition by as few as two properly-positioned metal-to-ligand interactions.

Table 2.1. Binding constants and chemical shifts for receptor-bound 1-benzylimidazole. First and second association constants (K_1 and K_2) and change in chemical shift of 1-benzylimidazole (**I**) for 1:1 (δ_1) and 2:1 (δ_2) complexes with the bis-metal ion receptors determined by nonlinear regression of equilibrium titration experiments (see Appendix). Error estimates are given in parentheses.

Receptor	δ_1 (ppm)	δ_2 (ppm)	K_1 (M^{-1})	K_2 (M^{-1})
$\mathbf{R}_{cnc}$	0.756 (0.02)	0.730 (0.02)	$1.69\ (0.07) \times 10^4$	$7.9\ (0.3) \times 10^2$
$\mathbf{R}_{cxc}$	0.728 (0.02)	0.632 (0.02)	$2.21\ (0.16) \times 10^4$	$4.2\ (0.3) \times 10^2$
$\mathbf{R}_{crc}$	0.694 (0.02)	0.605 (0.03)	$1.11\ (0.07) \times 10^4$	$4.7\ (0.3) \times 10^2$
$\mathbf{R}_{dxd}$	0.684 (0.02)	0.582 (0.03)	$1.70\ (0.08) \times 10^4$	$1.4\ (0.1) \times 10^2$
$\mathbf{R}_{cc}$	0.677 (0.03)	0.591 (0.03)	$1.68\ (0.18) \times 10^4$	$4.9\ (0.5) \times 10^2$

Table 2.2. Chemical shifts of receptor-bound bis-imidazole targets. Change in chemical shift of bis-imidazole targets in 1:1 complexes with the bis-metal ion receptors determined by nonlinear regression of equilibrium titration experiments (see Appendix). Error estimates are given in parentheses. N.D. = not determined.

	T1	**T2**	**T3**
Receptor	δ_{T1} (ppm)	δ_{T2} (ppm)	δ_{T3} (ppm)
$\mathbf{R}_{cnc}$	0.577 (0.01)	0.953 (0.01)	0.74 (0.02)
$\mathbf{R}_{cxc}$	0.660 (0.01)	0.884 (0.01)	0.70 (0.02)
$\mathbf{R}_{crc}$	0.645 (0.04)	0.900 (0.04)	N.D.
$\mathbf{R}_{dxd}$	0.684 (0.01)	0.803 (0.01)	N.D.
$\mathbf{R}_{cc}$	0.600 (0.04)	0.640 (0.04)	N.D.

Table 2.3. Bis-imidazole binding constants determined by competition with 1-benzylimidazole. Ratio of association constants for bis-imidazoles **T1** and **T2** (K_{T1}/K_1, K_{T2}/K_1) determined by nonlinear regression of equilibrium competitive titration experiments with 1-benzylimidazole (**I**) (see Appendix). Estimated association constants (K_{T1}, K_{T2}) calculated from these ratios and the 1-benzylimidazole association constants of Table 2.1. For comparison with Table 2.4, predicted selectivity for target **T2** over **T1** (K_{T2}/K_{T1}) is calculated. Error estimates are given in parentheses.

	selectivity T1/I		**selectivity T2/I**		
Receptor	K_{T1}/K_1	K_{T1} (M^{-1})	K_{T2}/K_1	K_{T2} (M^{-1})	K_{T2}/K_{T1}
$\mathbf{R}_{cnc}$	23 (4)	3.9×10^5	140 (30)	2.4×10^6	6.3 (1.7)
$\mathbf{R}_{cxc}$	10 (1)	2.2×10^5	140 (30)	3.1×10^6	14.3 (3.4)
$\mathbf{R}_{crc}$	24 (4)	2.7×10^5	89 (20)	9.8×10^5	3.7 (1.0)
$\mathbf{R}_{dxd}$	13 (1)	2.2×10^5	4.5 (0.5)	7.7×10^4	0.35 (0.1)
$\mathbf{R}_{cc}$	1.4 (0.2)	2.4×10^4	1.3 (0.2)	2.2×10^4	1.1 (0.2)

Table 2.4. Receptor selectivity determined by competition between bis-imidazoles. Selectivity ($\alpha_{T2/T1}$, $\alpha_{T2/T3}$) determined by nonlinear regression of equilibrium competitive titration experiments (see Appendix) between **T2** and **T1** or **T3**. Error estimates are given in parentheses.

Receptor	**selectivity T2/T1** $\alpha_{T2/T1}$	**selectivity T2/T3** $\alpha_{T2/T3}$
$\mathbf{R}_{cnc}$	7.6 (0.2)	4.9 (0.3)
$\mathbf{R}_{cxc}$	11.5 (0.4)	10.9 (0.5)
$\mathbf{R}_{crc}$	2.7 (0.2)	N.D.
$\mathbf{R}_{dxd}$	0.49 (0.10)	N.D.
$\mathbf{R}_{cc}$	0.91 (0.10)	N.D.

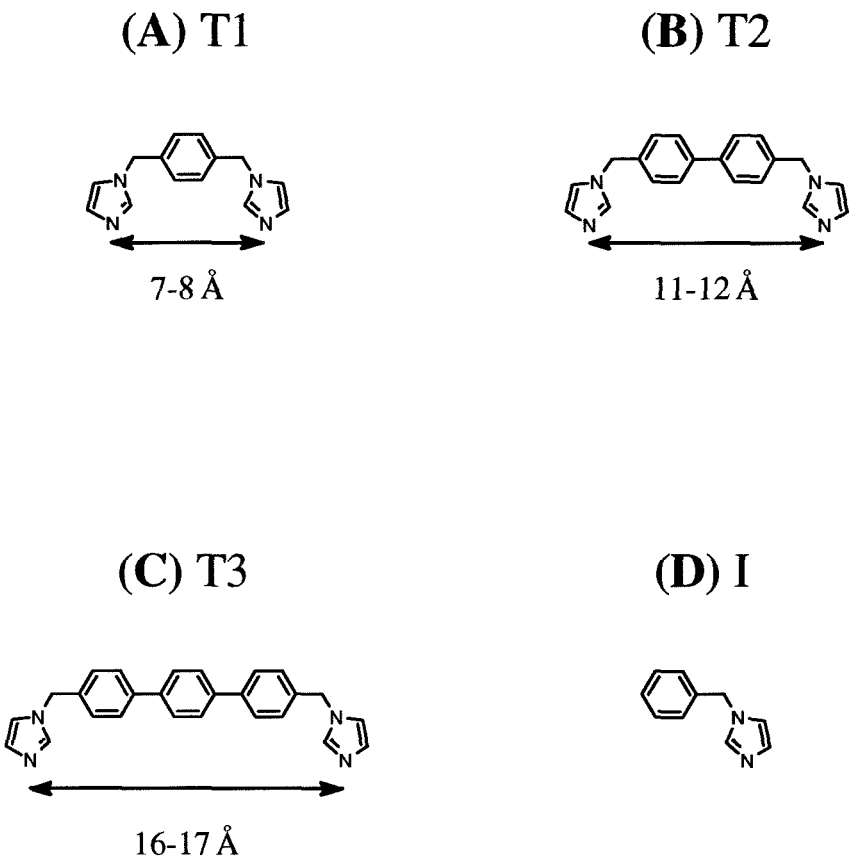


Figure 2.1. Structures of bis-imidazole targets and 1-benzylimidazole. **(A)** T1 (1,4,-bis(imidazol-1-ylmethyl)-benzene), **(B)** T2 (4,4"-bis(imidazol-1-ylmethyl)-*p*-biphenyl), (C) T3 (4,4"-bis(imidazol-1-ylmethyl)-*p*-terphenyl), (D) I (1-benzylimidazole). Distances between imidazole groups of bis-imidazole targets are estimated by molecular modeling.

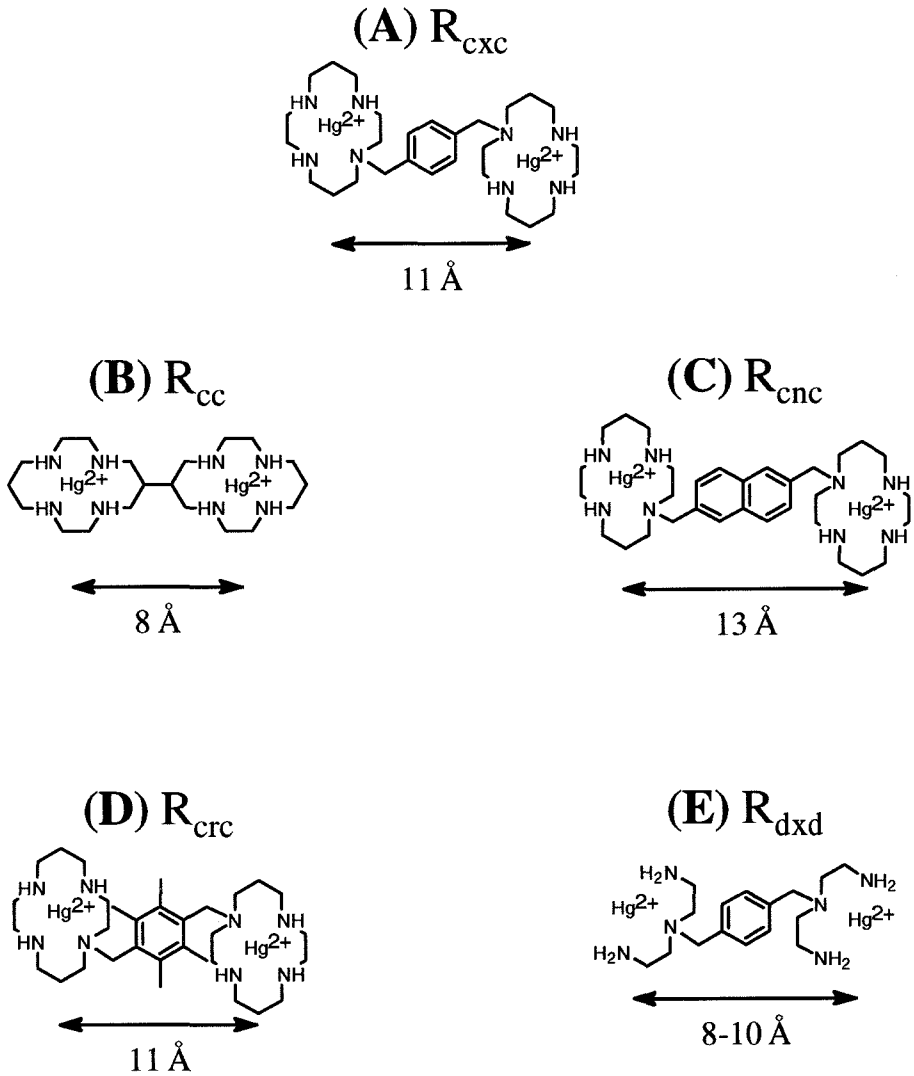
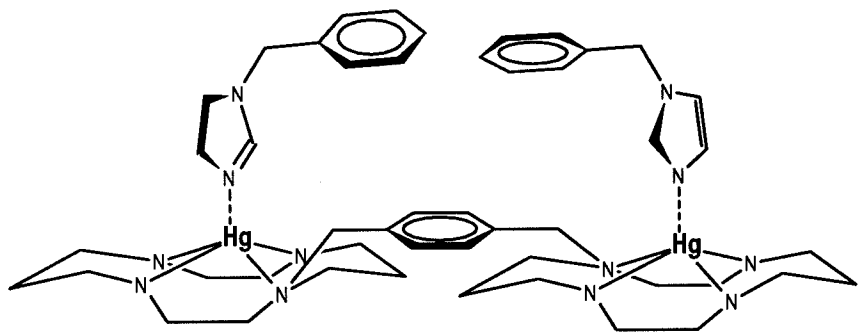
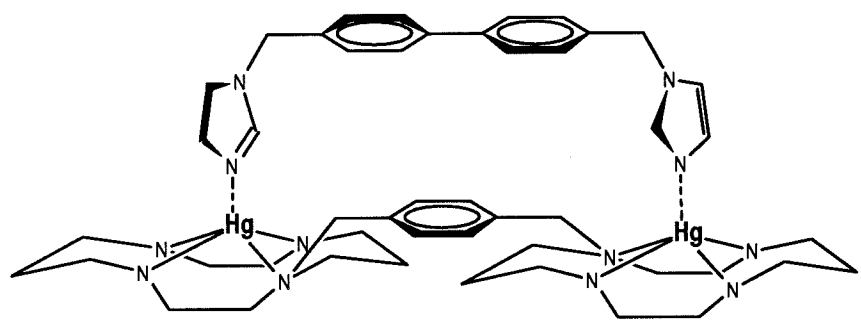


Figure 2.2. Structure of bis-metal receptors to recognize bis-imidazole targets. **(A)** R_{cxc} (1,4,-bis(1',4',8',11'-tetraazacyclotetradecan-1'-ylmethl)benzene bis mercury(II)), **(B)** R_{cc} (6,6'-bis(1,4,8,11-tetraazacyclotetradecane-mercury (II)), **(C)** R_{cnc} (2,6-bis(1'4'8'11'-tetraazacyclotetradecan-1'-ylmethyl)naphthalene bis mercury(II)), **(D)** R_{cxc} (1,4,-bis(1',4',8',11'-tetraazacyclotetradecan-1'-ylmethyl)-2,3,5,6-tetramethyl-benzene bis mercury (II)), **(E)** R_{dxd} (1,4-bis(bis(2-aminoethyl)aminoethyl)benzene bis mercury(II)). Distances (Å) between metal ions of receptors are estimated by molecular modeling.



(A) I:R_{cxc}:I



(B) T2:R_{cxc}

Figure 2.3. Proposed structure of complex between receptor and target. **(A)** 2:1 complex between two 1-benzylimidazoles (**I**) and bis-metal receptor ($\mathbf{R_{cxc}}$). **(B)** Distance-matched 1:1 complex between a bis-imidazole target (**T2**) and a bis-metal receptor ($\mathbf{R_{cxc}}$).

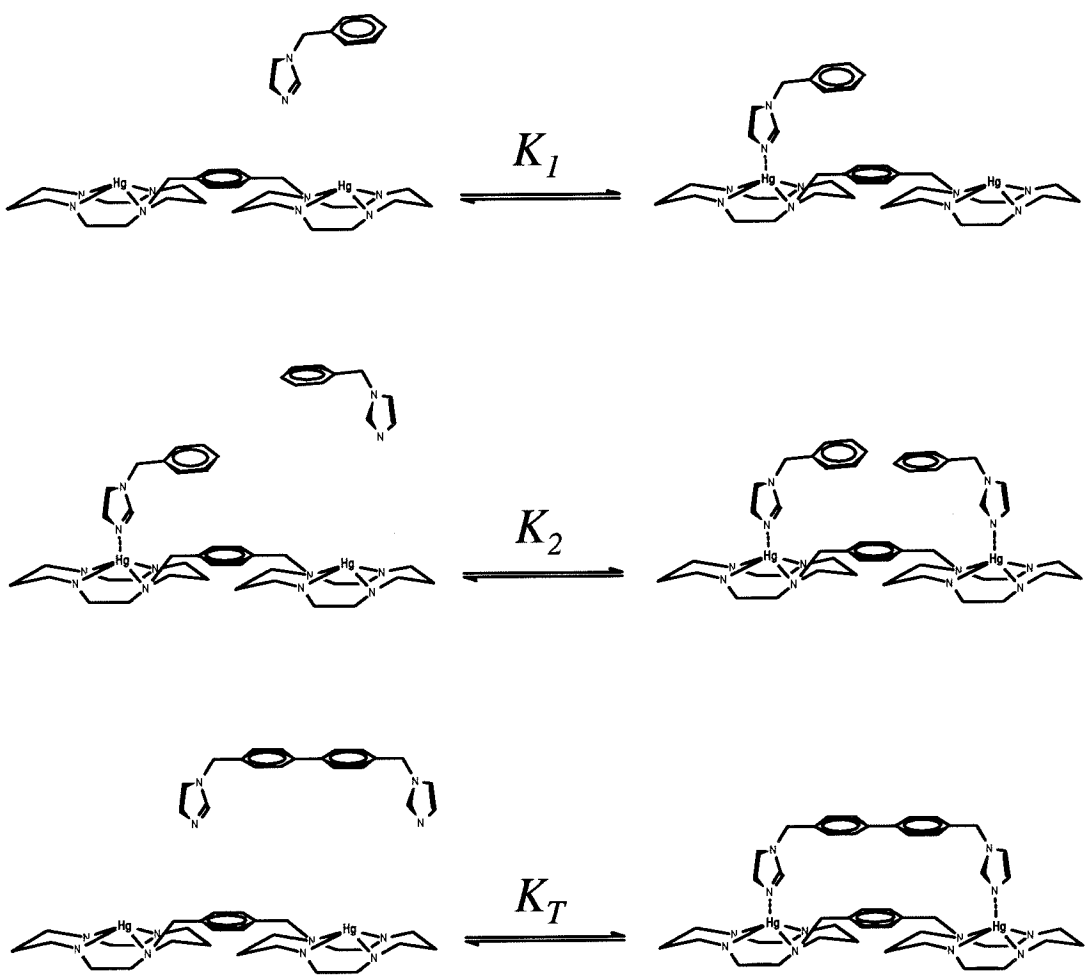


Figure 2.4. Simplified model of equilibria between receptor and imidazole species. 1-Benzylimidazole (**I**) forms complexes of both 1:1 and 2:1 stoichiometry with bis-metal receptor ($\mathbf{R}_{cxc}$). Bis-imidazole target forms only complex of 1:1 stoichiometry with bis-metal receptor ($\mathbf{R}_{cxc}$).

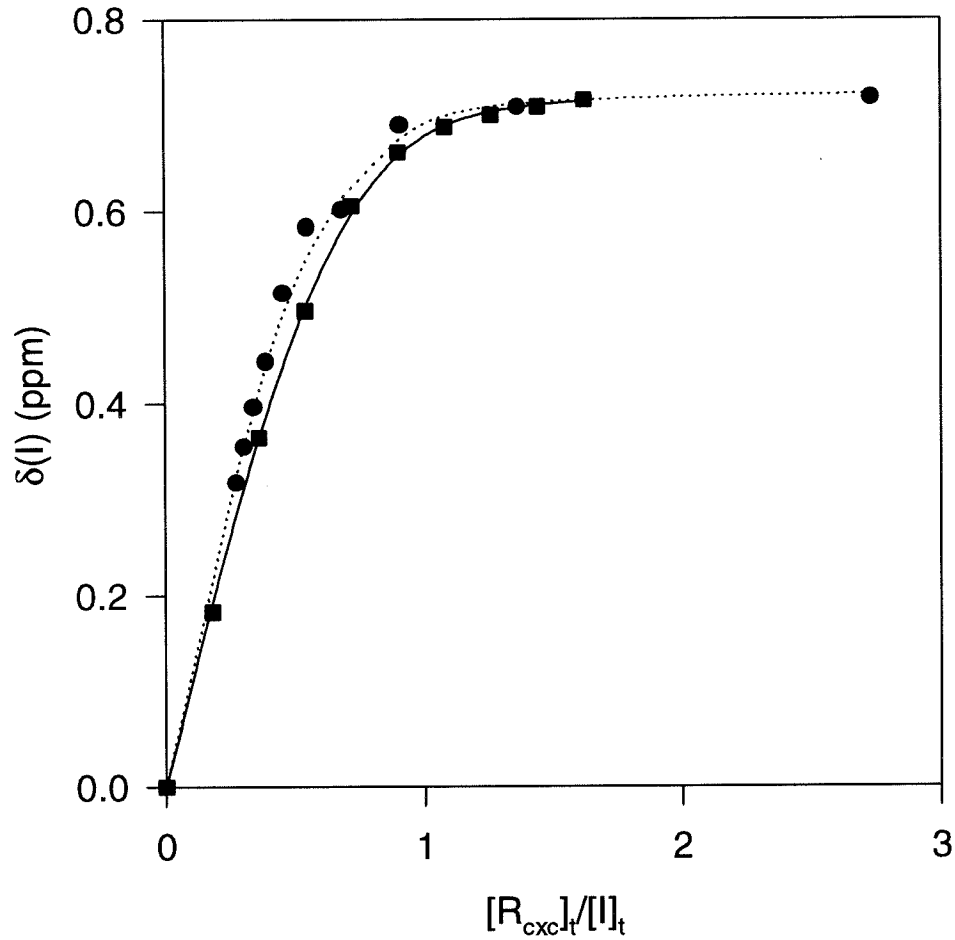


Figure 2.5. Change in chemical shift of 1-benzylmidazole in titration with receptor $\mathbf{R}_{cxc}$. Change in imidazole C_2H chemical shift (δ) of 1-benzylimidazole (**I**) is followed at increasing ratio of receptor to imidazole ($[R_{cxc}]_t/[I]_t$) of receptor $\mathbf{R}_{cxc}$ in Type I (■, —) and Type II (●, · · ·) titrations. Type I: solution of **I** ($[I]_t$ ~ 10 mM) is titrated with increasing amounts of $\mathbf{R}_{cxc}$. Type II: solution of $\mathbf{R}_{cxc}$ ($[R_{cxc}]_t$ ~ 10 mM) is titrated with increasing amounts of **I**. Curves calculated from parameters of Table 2.1.

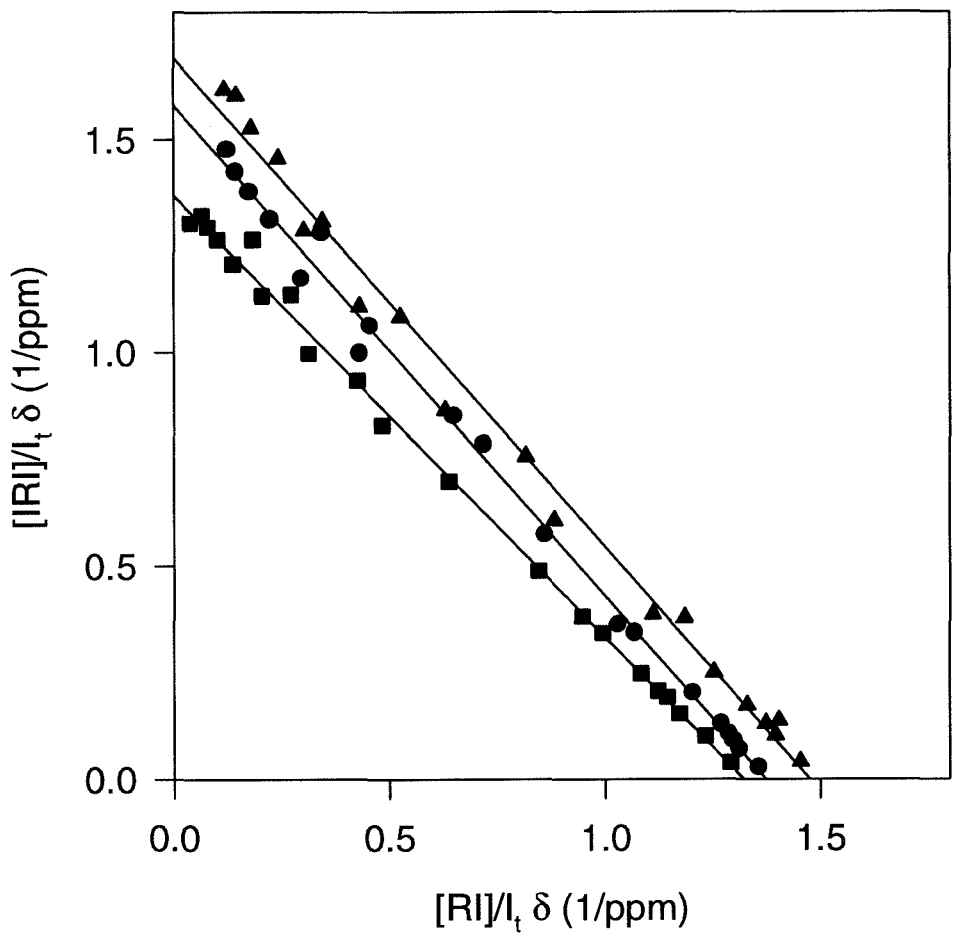


Figure 2.6. 1-Benzylimidazole titrations with receptors varying in spacer length, linearized by the method of Lenkinksi *et al.* [22]. The vertical intercept is the chemical shift of the 2:1 complex (**IRI**) and the horizontal intercept is the chemical shift of the 1:1 complex (**RI**) between 1-benzylimidazole (**I**) and receptors R_{cc} (▲), R_{cxc} (●) and R_{cnc} (■).

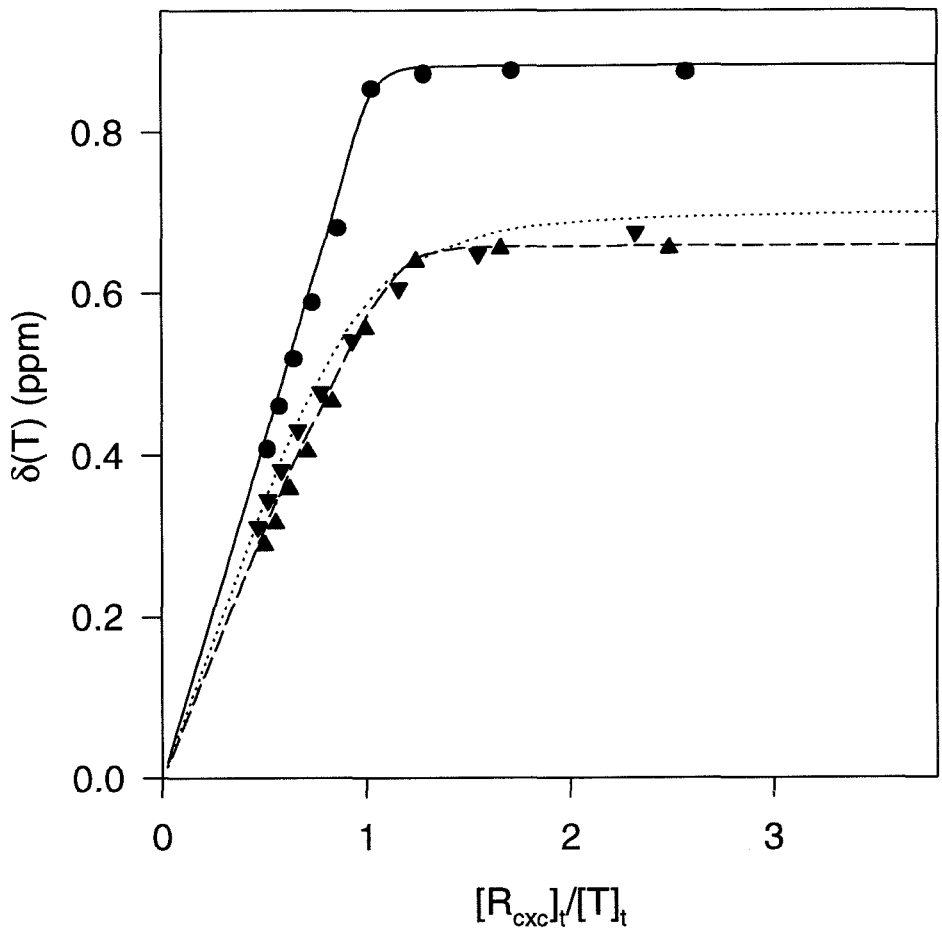


Figure 2.7. Chemical shift of bis-imidazole targets in titration with receptor $\mathbf{R}_{cxc}$. Change in imidazole C_2H chemical shifts (δ) of bis-imidazole targets **T1** (▲, - - -), **T2** (●, —), and **T3** (▼, · · ·) are followed at increasing ratio of receptor to target ($[R_{cxc}]_t/[T]_t$) in titration experiments with receptor R_{cxc} ($[R_{cxc}]_t$ ~10 mM). Curves calculated from parameters of Table 2.2.

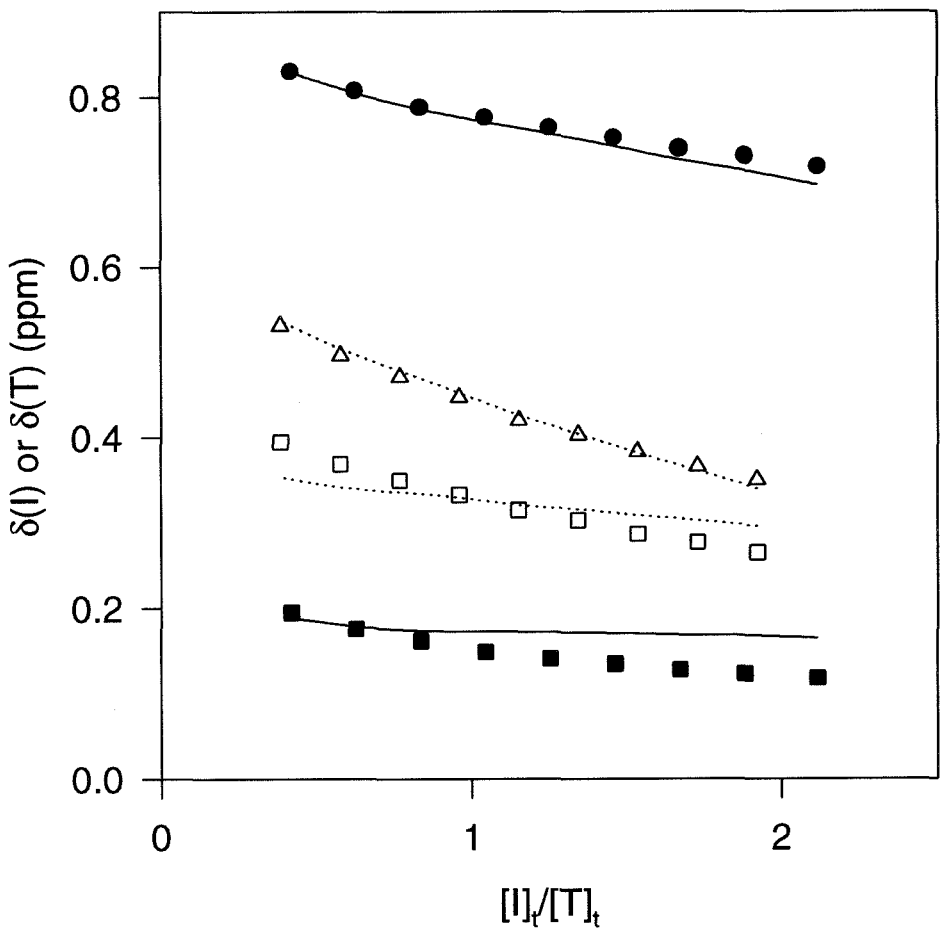


Figure 2.8. Competition of bis-imidazoles and 1-benzylmidazole for receptor $\mathbf{R}_{cxc}$. Change in imidazole C_2H chemical shifts (δ) of **T2** (●) and **I** (■) are followed as equimolar mixture of $\mathbf{R}_{cxc}$ and **T2** ($[T2]_t$, $[R_{cxc}]_t$ ~ 10 mM) is titrated with increasing amounts of **I** (—). Change in imidazole C_2H chemical shifts of **T1** (Δ) and **I** (□) are followed in analogous experiment with **T1**(· · ·). Curves calculated from parameters of Tables 2.1-2.3.

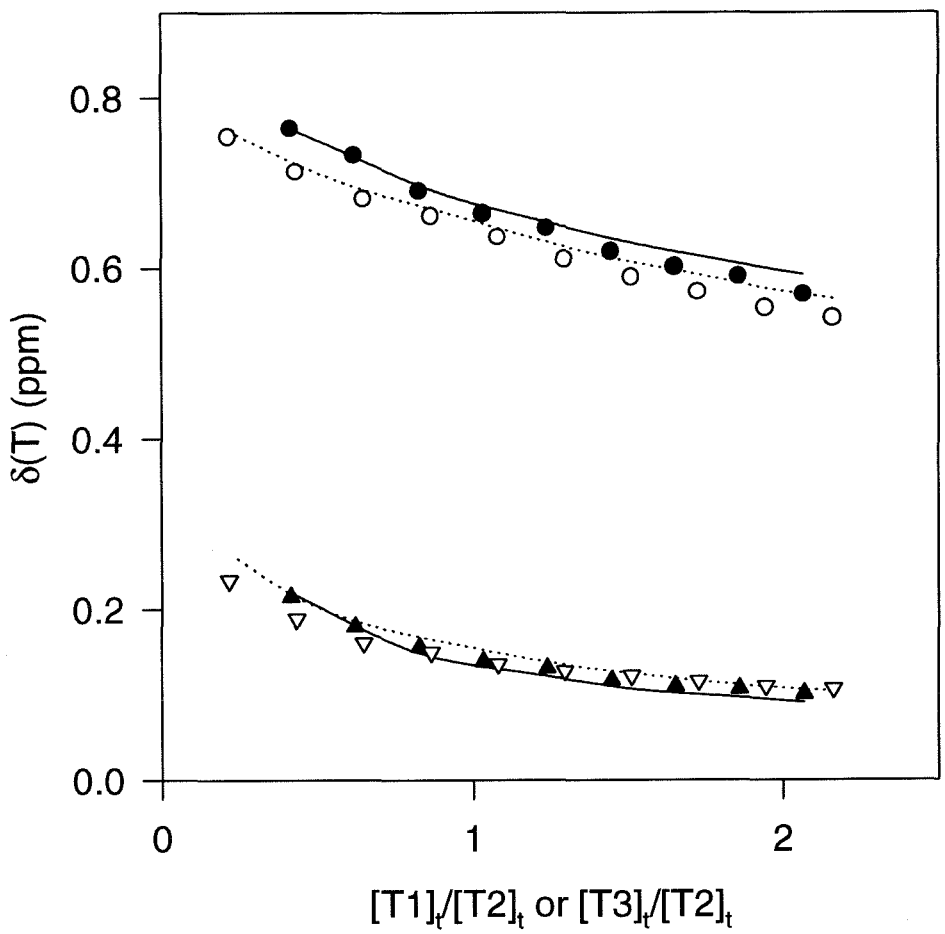


Figure 2.9. Competition of bis-imidazole targets for receptor $\mathbf{R}_{cxc}$. Change in imidazole C_2H chemical shifts (δ) of **T2** (●) and **T1** (▲) are followed as equimolar mixture of $\mathbf{R}_{cxc}$ and **T2** ($[T2]_t$, $[R_{cxc}]_t$ ~10 mM) is titrated with increasing amounts of **T1** (—). Change in imidazole C2H chemical shifts of **T2** (○) and **T3** (▽) are followed in analogous experiment with **T3** (· · ·). Curves calculated from parameters of Tables 2.1 and 2.4.

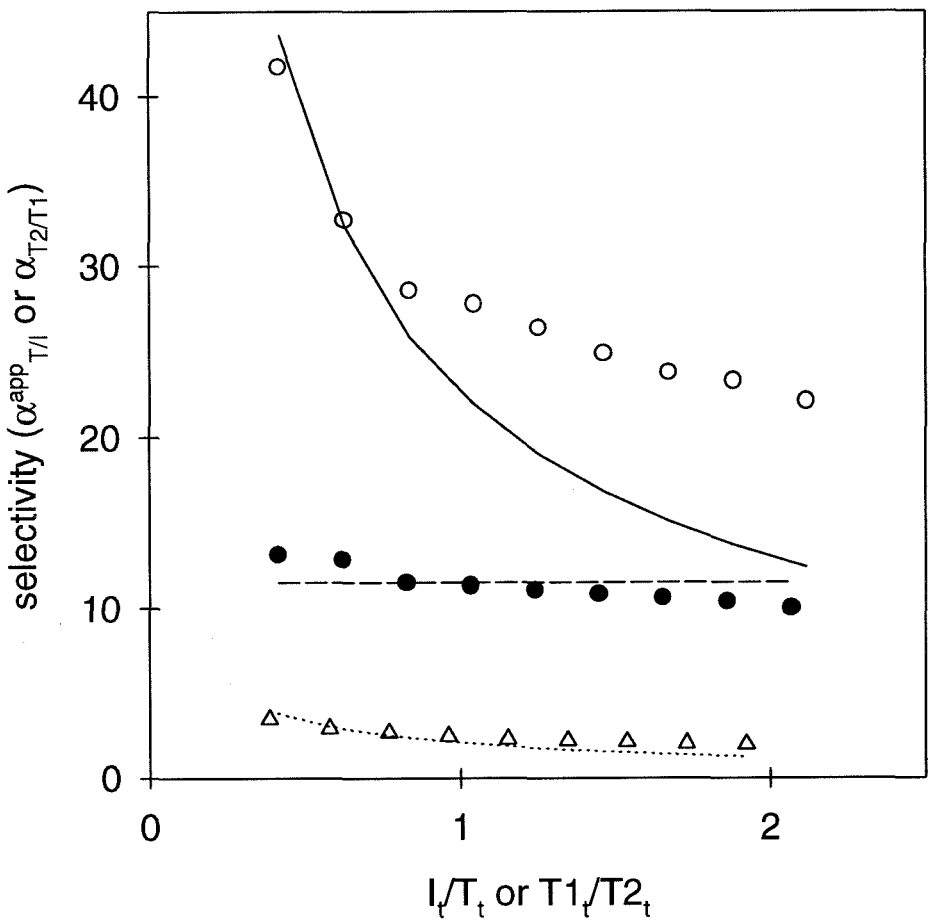


Figure 2.10. Concentration dependence of apparent selectivity of receptor $\mathbf{R}_{cxc}$. Apparent selectivity ($\alpha^{app}_{T/I}$) of $\mathbf{R}_{cxc}$ for bis-imidazole (**T2** or **T1**) over 1-benzylimidazole (**I**) calculated by eqn. 28 for competive titration experiments between **T2** and **I** (O, —) and **T1** and **I** (Δ, · · ·). Selectivity ($\alpha_{T2/T1}$) of $\mathbf{R}_{cxc}$ for T2 over T1 calculated by eqn. 23 for competitive titration experiment between **T2** and **T1** (●, - - -). Curves calculated by parameters of Tables 2.3 and 2.4 .

APPENDIX

Analysis of bis-imidazole titrations.

The following nonlinear regression routine (SigmaPlot, Jandel Scientific) was written to analyze bis-imidazole target (T) titration with receptor (R). There are two parameters to be optimized, δ_T and K_T.

```
[Parameters]
del=0.66                    ;del is chemical shift of T in RT complex
K1=10                       ;K1 is first association constant (1/mM)

[Variables]

At=col(1)                   ;At is total concentration of target T (mM)
Mt=col(2)                   ;Mt is total concentration of receptor R (mM)
ppm=col(3)                  ;ppm is observed chemical shift of T

[Equations]
Afree=At*(1-ppm/del)        ;calculate free T (Afree)
Mfree=Mt/(1+KA*Afree)       ;calculate predicted chemical shift (calc)
MA=K1*Mfree*Afree
calc=(MA*del)/At
fit calc to ppm             ;minimize difference between calc and ppm
```

Analysis of 1-benzylimidazole titrations.

The following nonlinear regression routine (SigmaPlot, Jandel Scientific) was written to analyze 1-benylimidazole (I) titrations with receptor (R). There are four parameters to be optimized, δ_1, δ_2, K_1, and K_2.

```
[Parameters]
del1=0.703                  ;del1 is chemical shift of I in RI complex
del2=0.648                  ;del2 is chemical shift of I in IRI complex
K1=20                       ;K1 is first association constant of I (1/mM)
K2=0.4                      ;K2 is second association constant of I (1/mM)

[Variables]
It=col(1)                   ;It is total concentration of I (mM)
Mt=col(2)                   ;Mt is total concentration of receptor R (mM)
ppm=col(3)                  ;ppm is observed chemical shift of I
```

```
[Equations]
dbar1 = (del1+del2)/2                    ;calculate first estimate of free I (Ifr1)
Ifr1  = It*(1-ppm/dbar1)
beta = 2*K2*Ifr1/(1+2*K2*Ifr1)           ;calculate second estimate of free I (Ifree)
dbar2 = (1-beta)*del1+beta*del2
Ifree = It*(1-ppm/dbar2)
Mfree = Mt/(1+K1*Ifree+K1*K2*Ifree^2)    ;calculate predicted chemical shift (calc)
MI1=K1*Mfree*Ifree
MI2=K1*K2*Ifree^2*Mfree
calc=(MI1*del1+2*MI2*del2)/It
fit calc to ppm                          ;minimize difference between calc and ppm
```

Error in free concentration of [I] calculated by eqn. 12:

$$[\mathrm{I}]_t\left(1-\frac{\delta}{\delta_{avg}}\right)=[\mathrm{I}]-\frac{\Delta}{2\delta_{avg}}\left(\beta\left([\mathrm{RI}]+2[\mathrm{IRI}]\right)-2[\mathrm{IRI}]\right)$$

If the second term is zero, then this equation is exact. For a first estimate (β=1/2):

$$\Delta_1=\frac{\delta_1-\delta_2}{(\delta_1+\delta_2)/2}$$

$$[\mathrm{I}]_1^{est}=[\mathrm{I}]\left[1-\frac{\Delta_1 K_1[\mathrm{R}]}{2}\left(1/2-K_2[\mathrm{I}]\right)\right]$$

For a more accurate estimate of [I], make a better estimate of β:

$$\beta^{est}=\frac{K_2[\mathrm{I}]_1^{est}}{1+K_2[\mathrm{I}]_1^{est}}$$

$$\Delta_2=\frac{\delta_1-\delta_2}{(1-\beta^{est})\delta_1+\beta^{est}\delta_2}$$

$$[\mathrm{I}]_2^{est}=[\mathrm{I}]\left[1-\frac{\Delta_1\Delta_2K_1^2K_2[\mathrm{I}][\mathrm{R}]^2}{2}\left(\frac{2K_2[\mathrm{I}]-1}{2(2K_2[\mathrm{I}]+1)-\Delta_1K_1K_2[\mathrm{R}][\mathrm{I}](2K_2[\mathrm{I}]-1)}\right)\right]$$

Analysis of bis-imidazole competition experiments.

The following nonlinear regression routine (SigmaPlot, Jandel Scientific) was written to analyze bis-imidazole target (T) titration with receptor (R). There is one parameter to be optimized, K_{T1}/K_{T2}; three parameters are previously determined from equilibrium titration experiments.

```
[Parameters]
alpha=1                    ;alpha is ratio Ka/Kb

[Variables]
At=col(1)                  ;At is total concentration of target T2 (mM)
Mt=col(2)                  ;Mt is total concentration of receptor R (mM)
Bt=col(3)                  ;Bt is total concentration of target T1 (mM)
appm=col(4)                ;appm is observed chemical shift of T2
bppm=col(5)                ;bppm is observed chemical shift of T1
AB=col(6)                  ;AB=1 for ppm=ppm(T2), AB=0 for ppm=ppm(T1)
ppm=col(7)                 ;ppm is either ppm(T2) or ppm(T1)

[Equations]
adel=0.90                  ;adel is chemical shift of T2 in RT2 complex
bdel=0.645                 ;bdel is chemical shift of T1 in RT1 complex
Kb=24*11.1
Ka=alpha*Kb
Afree=At*(1-appm/adel)     ;calculate free T2 (Afree)
Bfree=Bt*(1-bppm/bdel)     ;calculate free T1 (Bfree)
Mfree=Mt/(1+Ka*Afree+Kb*Bfree)
MA=Ka*Mfree*Afree
MB=Kb*Bfree*Mfree
apc=MA*adel/(MA+Afree)     ;calculate predicted T2 chemical shift (apc)
bpc=MB*bdel/(MB+Bfree)     ;calculate predicted T1 chemical shift (bpc)
calc=if(AB=1,apc,bpc)      ;calc is either T2 shift or T1 shift
fit calc to ppm            ;minimize difference between calc and ppm
```

Analysis of bis-imidazole and 1-benzylimidazole competition experiments.

The following nonlinear regression routine (SigmaPlot, Jandel Scientific) was written to analyze bis-imidazole target (T) and imidazole control (I) competition experiment with receptor (R). There is one parameter to be optimized, K_T/K_I; five parameters are previously determined from equilibrium titration experiments.

```
[Parameters]
alpha=1                          ;alpha is ratio Ka/K1

[Variables]
It=col(1)                        ;It is total concentration of I (mM)
Mt=col(2)                        ;Mt is total concentration of receptor R (mM)
At=col(3)                        ;At is total concentration of target T2 (mM)
ippm=col(4)                      ;ippm is observed chemical shift of I
appm=col(5)                      ;appm is observed chemical shift of T2
IA=col(6)                        ;IA=1 for ppm(I), IA=0 for ppm(T2)
ppm=col(7)                       ;ppm is either ppm(I) or ppm(T2)

[Equations]
idel1=0.756                      ;idel1 is chemical shift of I in RI complex
idel2=0.730                      ;idel2 is chemical shift of I in IRI complex
adel=0.577                       ;adel is chemical shift of T2 in RT2 complex
K1=16.9                          ;K1 is first association constant of I (1/mM)
K2=0.79                          ;K2 is second association constant of I (1/mM)
dbar1 = (idel1+idel2)/2          ;calculate first estimate of free I (Ifr1)
Ifr1  = It*(1-ippm/dbar1)
beta1 = 2*K2*Ifr1/(1+2*K2*Ifr1)
dbar2 = (1-beta1)*idel1+beta1*idel2
Ifree = It*(1-ippm/dbar2)        ;calculate second estimate of free I (Ifree)
Afree=At*(1-appm/adel)           ;calculate free T2 (Afree)
Mfree=Mt/(1+K1*Ifree+K1*K2*Ifree^2+alpha*K1*Afree)
MI1=K1*Mfree*Ifree
MI2=K1*K2*Ifree^2*Mfree
MA=alpha*K1*Afree*Mfree
ipc=(MI1*idel1+2*MI2*idel2)/(Ifree+MI1+2*MI2) ;calculate predicted I chemical shift (ipc)
apc=MA*adel/(MA+Afree)           ;calculate predicted T2 chemical shift (apc)
calc=if(IA=1,ipc,apc)            ;calc is either I shift or T2 shift
fit calc to ppm                  ;minimize difference between calc and ppm
```

Table 2A.1. Imidazole C_2H chemical shifts of 1-benzylimidazole and bis-imidazole targets in DMSO-d_6.

species	**chemical shift** (ppm)
I	7.737
T1	7.719
T2	7.772
T3	7.787

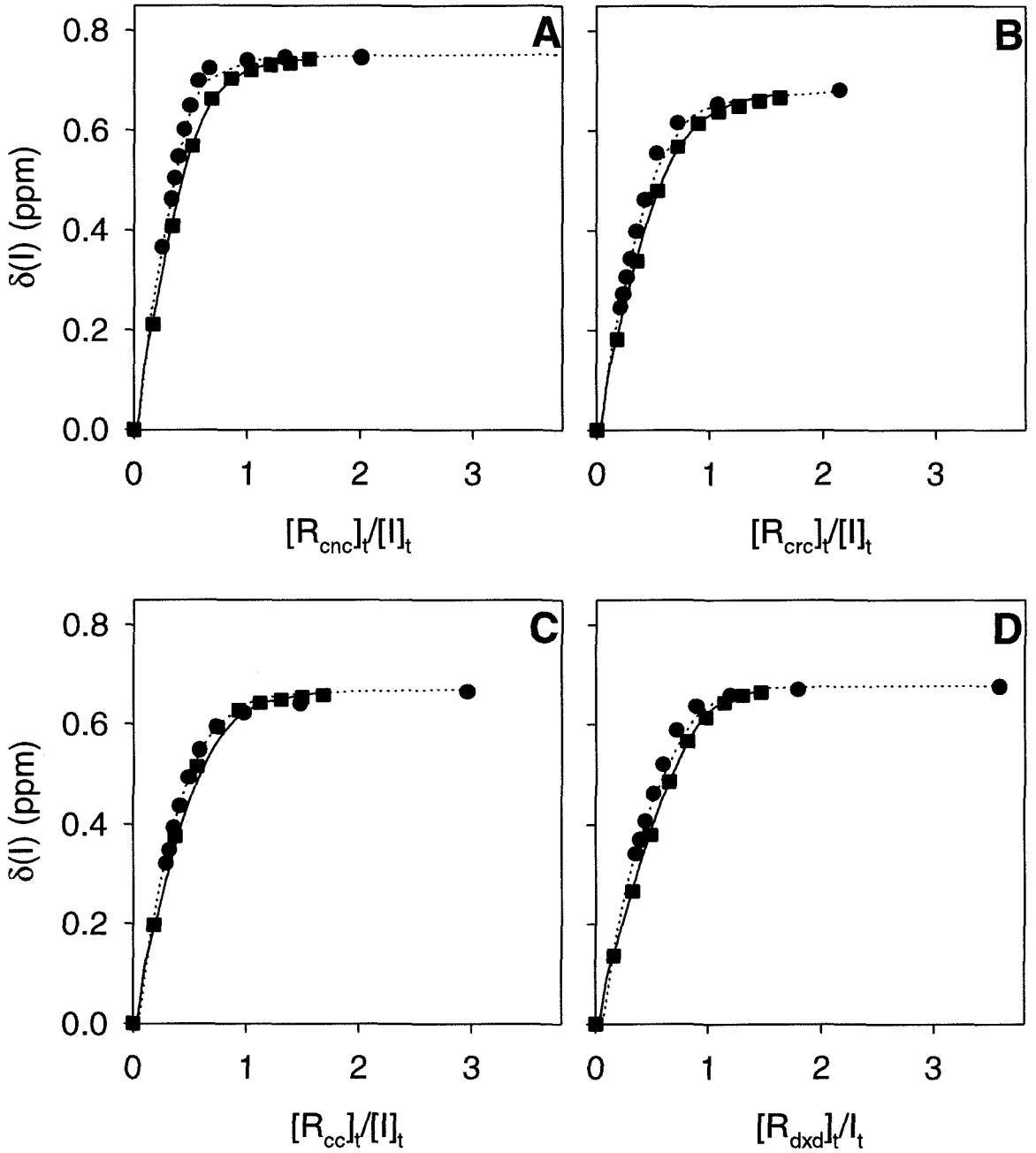


Figure 2A.1. Change in chemical shift of 1-benzylimidazole in titration with receptors. Change in imidazole C_2H chemical shift of 1-benzylimidazole (**I**) is followed at increasing ratio of receptor to imidazole of receptor in Type I and Type II (●, · · ·) titrations with (**A**) receptor $\mathbf{R_{cnc}}$, (**B**) receptor $\mathbf{R_{crc}}$, (**C**) receptor $\mathbf{R_{cc}}$, (**D**) receptor $\mathbf{R_{dxd}}$. Curves calculated from parameters of Table 2.1.

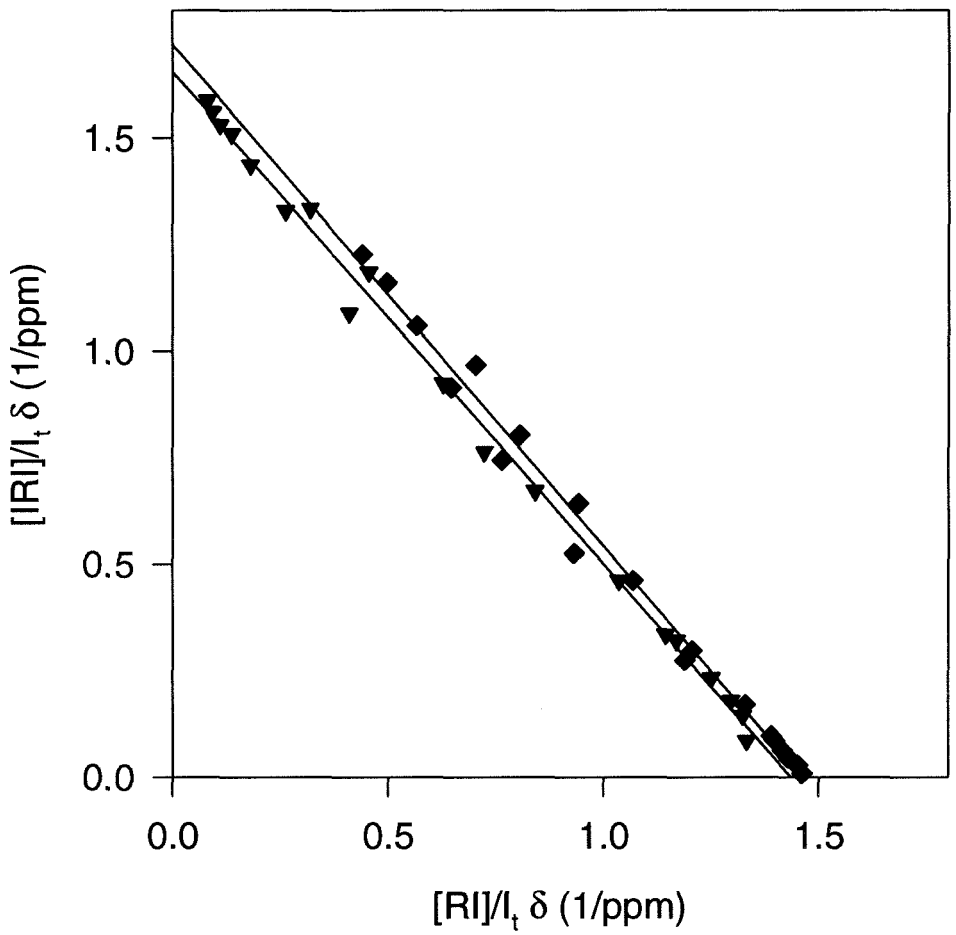


Figure 2A.2. 1-Benzylimidazole titration with receptors, linearized by method of Lenkinski *et al.* [22]. The vertical intercept is the chemical shift of the 2:1 complex (**IRI**) and the horizontal intercept is the chemical shift of the 1:1 complex (**RI**) between 1-benzylimidazole (**I**) and receptors $\mathbf{R}_{crc}$ (◆) and $\mathbf{R}_{dxd}$ (▼). Lines calculated from paramaters of Table 2.1.

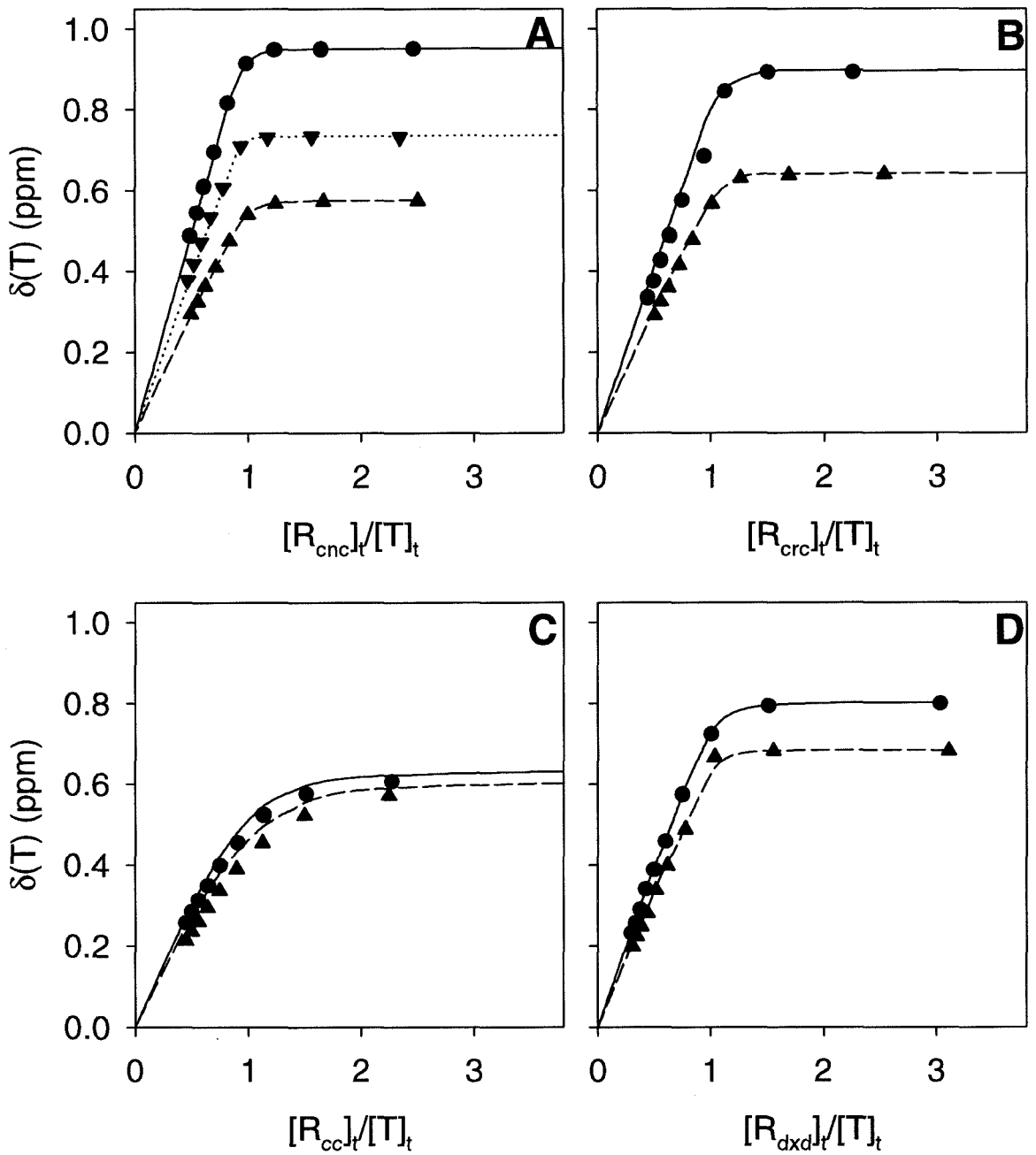


Figure 2A.3. Change in chemical shift of bis-imidazole targets in titration with receptors. Change in imidazole C_2H chemical shifts for bis-imidazoles **T1** (▲, - - -), **T2** (●, —), and **T3** (▼, · · ·) are followed at increasing ratio of receptor to target in titration experiments with **(A)** receptor $\mathbf{R_{cnc}}$, **(B)** receptor $\mathbf{R_{crc}}$, **(C)** receptor $\mathbf{R_{cc}}$, and **(D)** receptor $\mathbf{R_{dxd}}$. Curves calculated from parameters of Table 2.2.

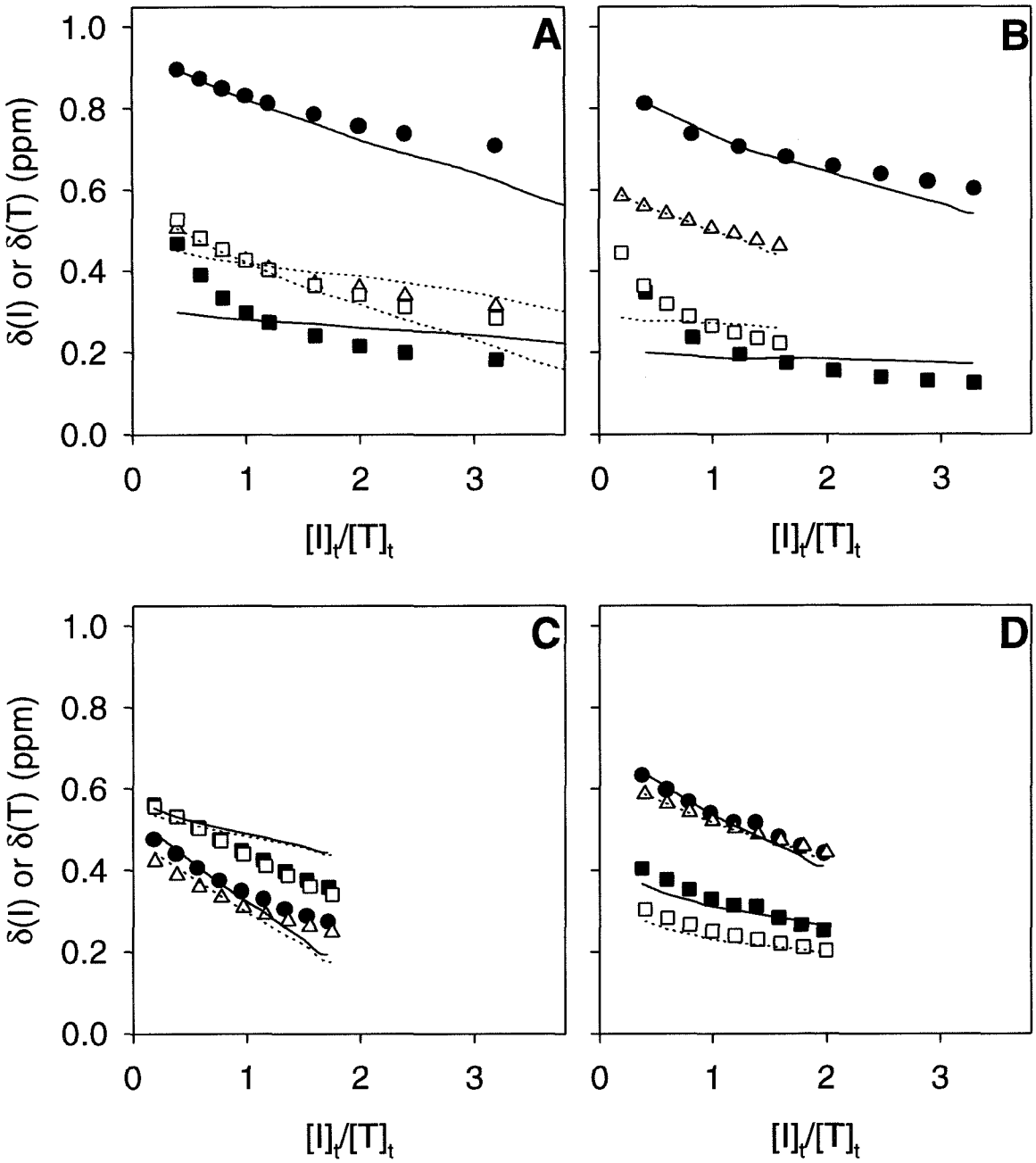


Figure 2A.4. Competitive titration of bis-imidazole targets and 1-benzylimidazole for receptors. Imidazole C_2H of **T2** (●) and **I** (■) are followed as receptor and **T2** is titrated with increasing amounts of **I** (—). Imidazole C_2H of **T1** (Δ) and **I** (□) are followed in analogous experiment with **T1** (· · ·). **(A)** receptor $\mathbf{R_{cnc}}$, **(B)** receptor $\mathbf{R_{crc}}$, **(C)** receptor $\mathbf{R_{cc}}$, **(D)** receptor $\mathbf{R_{dxd}}$. Curves calculated from parameters of Tables 2.1 - 2.3

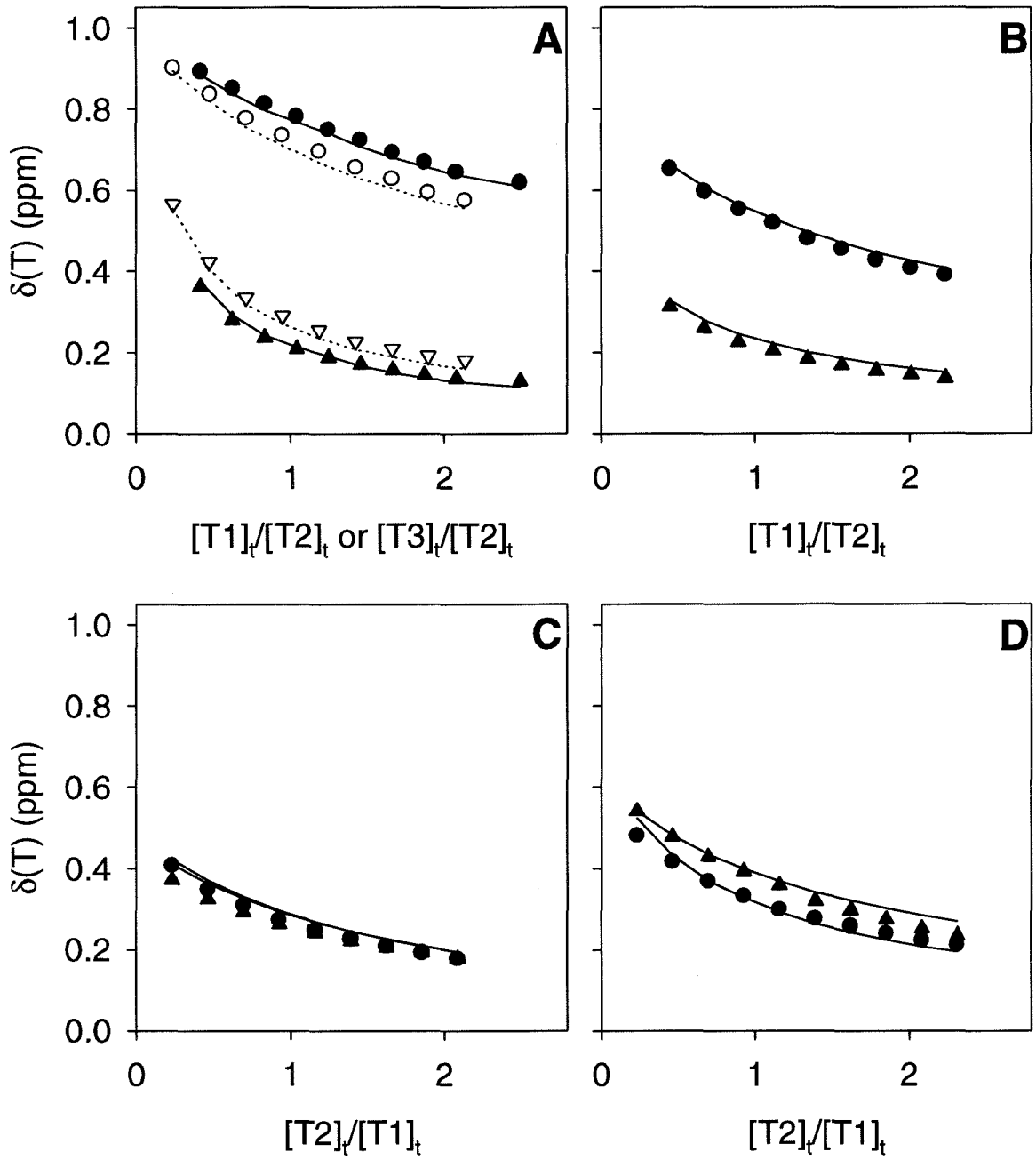


Figure 2A.5. Competition of bis-imidazole targets for receptors. **(A)** Mixture of $\mathbf{R_{cnc}}$ and **T2** (●,○) titrated with **T1** (▲, —) and **T3** (▽, · · ·). **(B)** Mixture of $\mathbf{R_{crc}}$ and **T2** (●) titrated with **T1** (▲). **(C)** Mixture of $\mathbf{R_{cc}}$ and **T1** (▲) titrated with **T2** (●). **(D)** Mixture of $\mathbf{R_{cc}}$ and **T1** (▲) titrated with **T2** (●). Curves calculated from parameters of Tables 2.2 and 2.4.

REFERENCES

1. Owen, M.J. and Lamb, J. R. (1988) *Immune Recognition*, IRL Press, Oxford, U.K.
2. Fan, E., van Arman, S. A., Kincaid, S. and Hamilton, A. D. (1993) *J. Am. Chem. Soc. 115*, 369-370.
3. Schneider, H. J. (1991) *Angew. Chem. Int. Ed. Engl. 30*, 1417-1436.
4. Mallik, S., Plunkett, S., Dhal, P. K., Johnson, R. D., Pack, D., Shnek, D. and Arnold, F. H. (1994) *New J. Chem. 18*, 299-304.
5. Seto, C. T., Mathias, J. P. and Whitesides, G. M. (1993) *J. Am. Chem. Soc. 115*, 1321-1329.
6. Carcanague, D. R., Knobler, C. B. and Diederich, F. (1992) *J. Am. Chem. Soc. 114,* 1515-1517.
7. Morgan Conn, M., Deslongchamps, G., de Mendoza, J. and Rebek, J. Jr. (1993) *J. Am. Chem. Soc. 115*, 3548-3557.
8. Sinha, P.C., Saxena, P. K., Nigam, N. B. and Srivastava, M. N. (1989) *Indian J. Chem 28A*, 335-336.
9. Searle, M. S., Williams, D. H. and Gerhard, U. (1992) *J. Am. Chem. Soc. 114*, 10697-10704.
10. Martell, A. E. and Smith, R. M. (1975) *Critical Stability Constants*, Vol. 1-6, Plenum Press, New York.
11. Plunkett, S. D., Sundaresan, V. and Arnold, F. H. (1994) *Polymer Preprints*, in press.
12. Dhal, P. K. and Arnold, F. H. (1992) *Macromolecules 25*, 7051-7059.
13. Muheim, A., Todd, R. J., Casimiro, D. R., Gray, H. B. and Arnold, F. H. (1993) *J. Am. Chem. Soc. 115*, 5312-5313.

14. Abelson, J. N., Simon, M. I. and Arnold, F. H., Editors (1992) *Methods: A Companion to Methods in Enzymology 4*, 1-108.

15. Mallik, S., Johnson, R. D. and Arnold, F. H. (1993) *J. Am. Chem. Soc. 115*, 2518-2520.

16. Mallik, S., Johnson, R. D. and Arnold, F. H. (1994) *J. Am. Chem. Soc. 116*, 8902-8911.

17. Bradshaw, J. S., ed. (1993) *Heterocyclic Compounds*, v. 51, Wiley & Sons, New York, NY.

18. Boudon, C., Gisselbrecht, J. P., Gross, M. and Hoseini, M. W. (1989) *Electroanalysis 1*, 493-499.

19. Alcock, N. W., Curson, E. H., Herron, H. and Moore, P. (1979) *J. Chem. Soc. Dalton Trans. 1979*, 1987-1993.

20. Sharma, R. R., Bhat, B. L. and Dubey, K. P. (1987) *J. Indian Chem Soc. 64*, 437-440.

21. Williams, D. H., Searle, M. S., Mackay, J. P., Gerhard, U. and Maplestone, R. A. (1993) *Proc. Natl. Acad. Sci. USA 90*, 1172-1178.

22. Lenkinski, R. E., Elgavish, G. A. and Ruben, J. (1978) *J. Mag. Reson. 32*, 367-376.

23. Deranleau, D. A. (1969) *J. Am. Chem. Soc. 91*, 4044-4049.

24. March, J. (1992) *Advanced Organic Chemistry*, John Wiley and Sons, New York.

25. Connors, K. (1987) *Binding Constants: The Measurement of Receptor Complex Stability*, John Wiley and Sons, New York.

26. van Dam, M. E., Wuenschell, G. E. and Arnold, F. H. (1989) *Biotech. Appl. Biochem. 11*, 492-502.

CHAPTER 3

MULTIVALENT BINDING I: EQUILIBRIUM ADSORPTION OF CYTOCHROME *C* IN IMMOBILIZED METAL AFFINITY CHROMATOGRAPHY

PREFACE

The following chapter is adapted from a previous publication (reprinted with permission from R. J. Todd, R. D. Johnson and F. H. Arnold, "Multiple-site binding interactions in metal-affinity chromatography: I. Equilibrium binding of engineered histidine-containing cytochromes *c*," *Journal of Chromatography A 662*, 13-26, copyright 1994 Elsevier Science B.V.).

This work was done in collaboration with Dr. Robert J. Todd. Dr. Todd performed the site-directed mutagenesis of *S. cerevisiae* iso-1-cytochrome *c* variants and the subsequent equilibrium adsorption experiments to the CuIDA-TSK Guardgel material. Dr. Todd's original analysis of these experiments appears elsewhere (R. J. Todd, *Ph.D. Thesis*, California Institute of Technology, Pasadena, CA, 1993).

I developed the procedures to prepare the IDA-TSK Guardgel material and IDA-TSK HPLC column at reduced copper loading. Subsequently, I performed the equilibrium binding experiments at reduced copper loading for horse heart cytochrome *c* and *N*-acetylhistidine and revised the analysis of the entire set of equilibrium binding experiments. Finally, the imidazole gradient chromatography of yeast cytochrome *c* variants at reduced copper loading was performed with the assistance of Virginie Leenknecht.

INTRODUCTION

Immobilized metal-affinity chromatography (IMAC) has proven to be a useful and versatile technique for the isolation and purification of proteins. As ligands for affinity separations, metal ion complexes offer important advantages over biological affinity agents such as inhibitors and antibodies [1]. Small, inexpensive metal complexes are stable under a wide range of conditions, can be recycled many times without loss of activity, and can be formulated into very high-capacity chromatographic supports. Elution can be effected under relatively mild conditions, and the columns can be cleaned and regenerated easily, without reduction in protein binding capacity. The selectivity of the separation can be tailored through the choice of metal ion, solvent conditions, or by modification of the target protein (i.e., the addition of histidine-rich affinity "handles"). These advantageous features have driven the recent rapid growth in IMAC applications [2,3].

Rapidly reversible interactions with metal ions immobilized on a hydrophilic chromatographic support (e.g., copper(II) chelated by an iminodiacetate-derivatized resin) result in the retention of proteins with metal-coordinating ligands on their surfaces, primarily histidine at neutral pH. Depending on the elution conditions, selectivity can be derived from the multiplicity or local environment of metal-coordinating residues [4,5]. Thus IMAC is effective for isolating proteins from crude mixtures; it can also be used to effect highly selective separations of closely-related proteins.

Despite recent progress in elucidating IMAC adsorption mechanisms and quantifying protein binding behavior [6,7], the precise molecular mechanisms by which proteins are selectively retained on IMAC matrices are not well understood. Equilibrium binding studies have shown that a simple Langmuir-type model may not

adequately describe protein binding to IMAC supports: binding heterogeneity as evidenced by nonlinear Scatchard plots is often observed for tightly binding proteins [6,7]. Based on these observations, Hutchens *et al.* proposed that protein binding may involve simultaneous interactions between multiple sites on the protein and the IMAC support [6]. The nature of protein binding has important implications for the design of efficient separations as well as for the design of new materials for IMAC supports. In particular, interactions at multiple sites can greatly enhance the binding affinity and dramatically alter specificity for certain classes of proteins. It is, in fact, possible to target individual molecules by matching the distribution of metal ions to a spatial distribution of metal-coordinating ligands [8, 9], and efforts to design new chromatographic supports capable of specific multiple-site binding to target molecules are underway [10,11].

Homologous proteins that differ in histidine content have proven useful for elucidating molecular bases of IMAC retention [4,12]. The ability to use protein engineering to add or delete specific amino acid residues, while leaving the remaining protein surface unchanged, allows a more complete and unambiguous study of the contributions of individual metal-coordinating sites [5]. In this study, we compare the equilibrium binding isotherms of variants of a small globular protein, iso-1-cytochrome *c* from *S. cerevisiae*, on a polymer matrix derivatized with copper(II)-iminodiacetate (CuIDA-TSK). Using these well-characterized protein variants, we are able to evaluate how the number and placement of surface histidines influences binding and IMAC separation. The engineered histidine-containing cytochrome *c* proteins are ideally suited to investigating the extent to which binding to IMAC supports involves coordination at multiple sites. This information is used to develop a simple but useful description of protein binding to IMAC supports that is consistent with our understanding of analogous metal ion complexes in solution.

If protein adsorption requires multiple interactions between the protein and immobilized affinity ligands, then adsorption will be strongly dependent on the density of these ligands. A convenient feature of metal affinity chromatography is that the affinity ligand (e.g., copper ion) is not directly attached to the polymer support; rather, it is indirectly immobilized via a chelating group covalently attached to the polymer. Thus the affinity ligand can be removed easily by the addition of a competing chelating group (e.g., EDTA), and replaced by adding additional metal ion. The ability to completely replace the immobilized affinity ligands provides important advantages over other affinity separations. In particular, the distribution of metal affinity ligands can be conveniently adjusted by reconstituting the polymer with less than one equivalent of metal ion per immobilized chelating group. In this investigation, protein adsorption will be described in terms of the availability of immobilized copper sites. Protein adsorption via multiple-site coordination has important consequences for the design of new chromatographic supports, not only for IMAC but for any separations dominated by multiple-site interactions.

CALCULATIONS

Capacities and binding constants from isotherms

Simple Langmuir binding is described by eqn. 1, where Q is the amount of protein adsorbed (mol/ml TSK) and c is the liquid-phase protein concentration (M) at equilibrium. The isotherm is fully described by two parameters, a binding constant K (M^{-1}) and the maximum capacity for the adsorbed protein, Q_{max} (mol/ml TSK),

$$Q(c) = \frac{Q_{max} K c}{1 + Kc} \quad . \tag{1}$$

This simple model cannot describe the isotherms of the cytochrome *c* variants containing multiple histidines. The simplest model to which these isotherms can be fit is a bi-Langmuir isotherm,

$$Q = \frac{Q_{max,A} K_A c}{1 + K_A c} + \frac{Q_{max,B} K_B c}{1 + K_B c} \quad , \tag{2}$$

where K_A and K_B represent the binding constants to two types of binding sites, strong and weak, respectively. The corresponding maximum capacities for binding at these two types of sites are $Q_{max,A}$ and $Q_{max,B}$.

To describe the binding of imidazole to the IMAC support, it is necessary to include the possibility that more than one imidazole will coordinate with different affinities to a single metal. Assuming two imidazoles can coordinate to a single copper, the adsorption isotherm for imidazole to a surface derivatized with CuIDA is described by

$$Q = \frac{Q_{max,A} K_A c + \left(Q_{max,A} + Q_{max,B}\right) K_A K_B c^2}{1 + K_A c + K_A K_B c^2} \quad , \tag{3}$$

where K_A refers to the first imidazole binding to the coordination sites of CuIDA, and K_B refers to a second imidazole binding to the remaining coordination sites. The maximum capacity for imidazole is $Q_{max,A} + Q_{max,B}$.

RESULTS

Cytochrome c surface histidine variants

A series of *S. cerevisiae* cytochrome *c* variants were designed and constructed for these studies (Table 3.1). The label assigned to each variant indicates the surface histidines. Amino acids 4 and 8 were chosen as sites for replacement with histidine, based on their high degree of surface accessibility. In addition, naturally-occurring histidines at positions 26, 33 and 39 were replaced with other amino acids in order to determine the contributions these residues make to the protein's interaction with the IMAC matrix. All the natural surface histidines have been replaced by other amino acids in the H(-) control.

In each of the iso-1-cytochrome *c* variants, the lone cysteine at position 102 was replaced by a serine to prevent oxidative dimerization at the surface sulfhydryl group [13]. The structural integrity of each mutant was confirmed by UV-visible spectroscopy, and the histidine content was confirmed by ^{1}H NMR spectroscopy [14]. Expression of these variants in a strain of yeast lacking cytochrome *c* guarantees that each is biologically functional, which in turn ensures that a conformation very similar to the native one has been maintained.

Formation of the ternary IDA-Cu^{2+}-protein complex that leads to retention on the IMAC support is expected to depend on the accessibilities of individual surface histidines. The accessibilities of the ε-nitrogens of each histidine residue to a probe approximately the size of CuIDA (1.93 Å radius) were determined using the coordinates of the crystal structure of *S. cerevisiae* cytochrome *c* [15] (Table 3.2). Engineered histidines were incorporated into a modified cytochrome *c* structure using the molecular graphics software InsightII (v. 2.2). Based on this structure, the histidines at positions 4

and 8 are expected to be fully accessible to CuIDA. The remaining histidines exhibit varying, lower degrees of accessibility.

Binding isotherms for single-histidine proteins and imidazole at maximum copper loading

Protein binding isotherms were measured at pH 7.0, where surface histidines are largely deprotonated and free to coordinate to the metal [14]. Isotherms for cytochrome *c* variants with differing number of surface histidines are shown in Figure 3.1. The single histidine proteins (e.g., H_8) exhibit simple Langmuir-type binding behavior (linear Scatchard plots), and the capacities and binding constants obtained by fitting these data to the Langmuir isotherm (eqn. 1) are listed in Table 3.3. The H(-) variant with no surface-accessible histidines exhibits almost no affinity for the CuIDA matrix at pH 7.0. Tuna cytochrome *c* and the H_{26} *S. cerevisiae* variant, both of which have only histidine 26, exhibit similar binding isotherms with apparent binding constants of $2 - 4 \times 10^3$ M^{-1} [14]. Their maximum capacities (extrapolated from the data) are ~1.2 μmol protein per ml of gel (~15 mg/ml). Variants with fully-accessible histidines at positions 4 and 8 bind significantly more tightly than H_{26}, whose histidine 26 has less than 10% of the accessibility of histidines 4 and 8 (Table 3.2).

Binding isotherms were also measured for imidazole [14]. While histidines at a protein surface are sterically excluded from occupying two coordination sites on one metal ion, two equivalents of imidazole can bind to matrix-bound CuIDA. Maximum capacity ($Q_{max,A}$ and $Q_{max,B}$) and binding constants for the first and second imidazoles (K_A and K_B) were obtained by fitting eqn. 3 to the experimental isotherms. The resulting values are reported in Table 3.3. The binding constant K_A for the first imidazole to matrix-bound CuIDA is similar to the binding constant for imidazole to CuIDA in solution (3.4×10^3 M^{-1} at 25 C) [16,17]. The observed maximum capacity for

imidazole ($Q_{max,A} + Q_{max,B}$) equals twice the copper loading and is much higher than the capacities observed for the single-histidine proteins. In fact, while all of the copper sites are available to imidazole, only 5 - 10% can be occupied by cytochromes *c* with single histidines.

The apparent binding constants of the single-histidine proteins obtained by fitting eqn. 1 to the isotherms are almost ten times greater than that of imidazole (e.g., 5×10^4 M^{-1} for H_8 versus 6×10^3 M^{-1} for imidazole). Similarly low capacities and high apparent binding constants have been reported for lysozyme and ovalbumins binding CuIDA-Sepharose and CuIDA-TSK gels [6,18].

Binding isotherms for multiple-histidine proteins at maximum copper loading

Proteins with multiple exposed histidines bind significantly more tightly and have higher maximum capacities than single-histidine proteins, as shown in Figure 3.1 for the $H_{26}H_8$ and $H_{26}H_{33}H_8$ variants. The sole difference between $H_{26}H_8$ and the H_8 variant is a second histidine only partially exposed on the surface of the protein (histidine 26). This partially exposed histidine acts to increase the initial slope of the binding isotherm by more than a factor of 10 (Table 3.3).

Scatchard plots of the isotherm data for the $H_{26}H_8$ two-histidine variant, shown in Figure 3.2, indicate at least two modes of binding. Binding data for all the multiple-histidine proteins were fit to the bi-Langmuir isotherm (eqn. 2), and the resulting binding constants and capacities are given in Table 3.3. The first binding constants for the two-histidine variants are approximately 10^6 M^{-1}, while the second binding constants (weak site) are similar to those of the single-histidine variants (~ 10^4 M^{-1}). The extrapolated maximum capacities for the proteins with two surface histidines are approximately 30% greater than the capacities for single-histidine cytochrome *c* variants. Isotherms of the three-histidines variants (e.g., $H_8H_{26}H_{33}$) also indicate multiple binding modes, one of

which corresponds to a very strong interaction. Fitting the bi-Langmuir model to the isotherms yields values for the first binding constant of 5-10 $\times$ 10^6 M^{-1} and capacities that are approximately twice those of the single-histidine proteins (Table 3.3).

Binding isotherms for acHIS at decreased copper loading

Equilibrium binding isotherms were measured for *N*-acetylhistidine (acHIS) on the TSK Guardgel Chelate 5PW at copper loading varying from 18.6 μmol/ml TSK (maximum loading) to 5.7 μmol/ml TSK (Figure 3.3). As demonstrated in the Scatchard plots (Figure 3.4), the adsorption isotherms of this single-imidazole species follow the Langmuir model. These data were analyzed using the Langmuir isotherm (eqn. 1), and the binding parameters are presented in Table 3.4. The Langmuir binding constant K (~ 800 M^{-1}) is not affected by decreasing the copper loading; however, the extrapolated limiting capacity Q_{max} is strongly dependent on the copper loading. At maximum copper loading, the limiting capacity is the same as the total copper loading (~18.6 μmol/ml TSK), but at lower copper loading the limiting capacity is much less than the total copper loading. This indicates that at lower copper loading much of the immobilized copper is unavailable for imidazole coordination.

Binding isotherms for a multiple-histidine protein at decreased copper loading

Equilibrium binding isotherms were measured for horse cytochrome *c* on the TSK Guardgel Chelate 5PW at copper loading varying from 18.6 μmol/ml TSK (Cu_{max}) to 6.1 μmol/ml TSK (Figure 3.5). At high copper loading (14.5 - 18.5 μmol/ml TSK) the Scatchard plots of this two-histidine protein (Figure 3.6) display multiple binding modes. These data were analyzed using the bi-Langmuir isotherm (eqn. 2), and the binding parameters are presented in Table 4.4. At lower copper loading (6.1 - 12.3 μmol/ml TSK), analysis of the data using the bi-Langmuir model resulted in negligible (< 0.02

μmol/ml TSK) values for $Q_{max,A}$, the strong binding site capacity. The data at these lower loadings were adequately described by the single Langmuir isotherm (eqn. 1).

As the copper loading is decreased, the first binding constant (strong site) of horse cytochrome *c* for the IMAC support at first decreases rapidly from 4×10^5 M^{-1} at maximum loading (18.6 μmol/ml TSK) and then levels off at 5×10^3 M^{-1} at two-thirds loading (12 μmol/ml TSK). The extrapolated maximum capacity ($Q_{max,A} + Q_{max,B}$) decreases steadily with copper loading, until there is negligible adsorption at one-third the maximum copper loading (6.1 μmol/ml TSK).

DISCUSSION

Single-site protein binding

Comparing the isotherms for the H(-) and H_8 variants (Figure 3.1) indicates that a surface histidine is a prerequisite for significant binding of cytochrome *c* to the CuIDA support at pH 7.0. In the absence of accessible histidines, the contributions from the remaining surface residues and the amino terminus in H(-) result in minimal protein adsorption. This observation mirrors those reported for other proteins and other IMAC supports at neutral pH [4,5,19].

Once there is a histidine accessible for coordination to the immobilized copper ions, protein binding is assured. However, the binding behavior of histidine-containing proteins is different from the binding of imidazole (Table 3.3). While all the copper ions appear to be accessible to imidazole, only a small fraction (< 10%) appear to be involved in protein binding. Furthermore, cytochrome *c* variants with fully-accessible histidines bind with apparent affinities ten times greater than imidazole. What is the source of this high binding affinity?

A simple model which assumes that protein adsorption blocks access to multiple metal ion sites at the surface of the chromatographic support ("site exclusion") can explain both the large apparent binding constants for single-histidine proteins and the low protein binding capacities. For a protein with a single histidine, the interaction illustrated in Figure 3.7A is described by

$$P + nCu \rightleftharpoons PCu + \{(n-1)\,Cu\} \qquad K_1 = \frac{[PCu]}{[P][Cu]}$$

where the protein (P) blocks n - 1 additional copper sites from further interactions (blocked sites designated by brackets) upon coordinating to a metal ion (Cu) on the chromatographic support. This model also results in a Langmuir-type isotherm,

$$Q = \frac{Cu_{acc} K_1 c}{1 + nK_1 c} \quad , \tag{4}$$

where Cu_{acc} is the total concentration of immobilized copper accessible for protein binding and K_1 is the binding constant for a protein bound to a single copper site via a single histidine. At saturation copper loading, the concentration of accessible copper, as measured by imidazole and *N*-acetylhistidine adsorption, is equal to the total concentration of immobilized copper (18.6 μmol/ml TSK). Parameters obtained by fitting the isotherms to eqn. 4 are given in Table 3.6. This analysis indicates that adsorption of cytochromes *c* via single histidines blocks ~15-18 copper sites.

If protein adsorption only involves coordination of the imidazole ring of a single histidine residue to an immobilized copper ion, then the intrinsic binding constants should not be greater than the binding constant of free imidazole to the IMAC support. If the solvation effects involved in a protein histidine-Cu interaction are the same as for

an imidazole-Cu interaction, then free imidazole should in fact provide an upper bound for the contribution of the first fully-accessible histidine to protein binding. In fact, binding constants for the H_4 and H_8 variants with fully-accessible histidines become comparable to the binding constant for imidazole and *N*-acetylhistidine when the protein adsorption isotherms are fit to the "site exclusion" model, as shown in Table 3.6. The relatively low binding constant for H_{26} can once again be attributed to the low accessibility of its histidine.

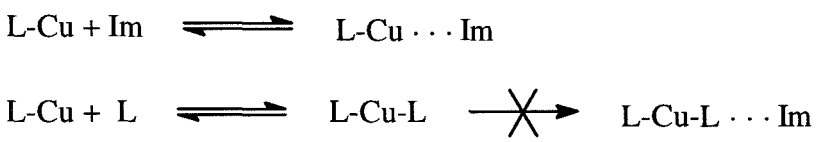
A lower bound for the contribution of a fully-accessible histidine can be estimated by comparing the binding of the H_4 and H_8 variants to the same protein without surface histidines, H(-). The binding constant of an essentially nonbinding protein, H(-), can be estimated from the initial slope of its adsorption isotherm, $Cu_{acc}\,K_1$. The single histidines in the H_4 and H_8 variants contribute more than a factor of 200 to the binding constant relative to the H(-) variant (Table 3.6), and > 3 kcal mol^{-1} to the binding free energy. For comparison, the free energy of binding of a single imidazole to aqueous Cu-IDA is 4.8 kcal mol^{-1} at 25 °C [16].

This simple "site exclusion" model can also explain the seemingly surprising result reported by Hutchens and Yip [18] that free copper does not influence protein binding to an IMAC support and therefore must differ from immobilized copper in how it interacts with the protein. Free copper at the low concentrations used in those experiments (10-100 μM) would in fact not be expected to influence binding if the intrinsic protein-to-immobilized copper binding constants are only $\sim 3 \times 10^3$ M^{-1} instead of ten times larger. Although free copper may be able to bind at sites not accessible to immobilized copper, its affinity for cytochrome *c* surface histidines is immediately apparent in ^{1}H NMR spectra, as demonstrated in Chapter 6, and is well documented for other proteins [20].

The possibility that protein binding blocks access to other copper sites was first discussed by Belew *et al.* [21], who rejected it in favor of invoking multipoint attachment of the protein to the IMAC matrix via undefined histidine, cysteine, or tryptophan residues to explain the apparent high affinities. The cytochromes *c* used in the current studies do not have surface cysteine residues, and the lone surface tryptophan (33) in tuna cytochrome *c* is not present in the proteins from *S. cerevisiae* or horse heart. While one cannot rule out other sites (e.g., lysines or free amino-terminus) for multipoint attachment of the proteins containing only one accessible histidine, the contributions of other individual surface residues is certainly small compared to that of even a poorly-accessible histidine. Therefore, we favor the "site-exclusion" explanation which yields intrinsic protein binding constants that are consistent with analogous imidazole-CuIDA complexes, both immobilized and in solution. As will be shown below, this framework is also consistent with binding data for multiple-histidine proteins and is further supported by indications of a high surface density of immobilized copper sites.

Distribution of immobilized copper sites

The CuIDA complexes must be quite densely packed at the surface of this IMAC support for adsorption of cytochrome *c* to block 15 copper ions. The adsorption behavior of *N*-acetylhistidine (acHIS) at decreased copper loading provides strong evidence for this close packing of metal sites. Consider a pair of immobilized IDA ligands (L). Under copper-limited conditions, two ligands will chelate a single copper ion, provided they can both reach the metal. The L-Cu-L complex completely cages the copper, leaving no vacant coordination sites for imidazole binding:



If the immobilized IDA ligands are close enough together to form L-Cu-L complexes, then as illustrated in Figure 3.8, the capacity to bind acHIS should decrease more rapidly than the total copper loading.

Because the immobilized metal can be blocked from coordinating free imidazole, a distinction must be made between the concentration of immobilized copper *accessible* for imidazole coordination (Cu_{acc}) and the *total* concentration of immobilized copper (Cu). Assuming equilibrium binding constants of immobilized IDA for copper that are comparable to formation constants for aqueous Cu-IDA species ($\sim 10^{11}$ M^{-1} for aqueous Cu^{2+}-IDA and $\sim 10^{17}$ M^{-2} for Cu^{2+}-$(IDA)_2$ [17]), all the immobilized copper will be coordinated by a single IDA group at saturation loading (Figure 3.8A). At lower copper loading, however, some fraction of the immobilized copper will be in the form of L-Cu-L complexes, precluded from interactions with protein or acHIS (Figure 3.8B).

In fact, the concentration of immobilized copper accessible to the protein can be measured directly by the capacity of the support to bind acHIS. The initial slope of the adsorption isotherm, shown in Figure 3.9, for acHIS decreases approximately linearly from a maximum at saturation loading to negligible adsorption at ~ 5 μmol copper per ml TSK gel. For a compound such as acHIS which interacts with only one immobilized copper, there is no effect of the copper density upon the binding affinity; the only effect is to limit the total number of copper sites accessible for coordination. The effect of decreasing copper loading is to isolate individual immobilized copper sites from one another.

The equilibrium binding isotherms of horse cytochrome *c* at decreased copper loading provide supporting evidence for this interpretation. There is in fact no measurable adsorption of horse cytochrome *c* to the IDA-TSK gel at low copper loading ($Cu < 7$ μmol/ml TSK, see Figure 3.9). Significant residual protein binding capacity would be expected if the metal ions were isolated and could remain accessible.

Multiple-site protein binding

The existence of multiple coordinating sites on a protein surface can influence binding affinity in various ways. If two histidines bind *independently* (only one at a time), the apparent binding constant of a two-histidine protein would be twice that of a protein containing only one histidine of comparable accessibility, reflecting the increased availability of binding sites (a statistical effect). If, on the other hand, two histidines (or a histidine and another functional group) can bind *simultaneously* to distinct copper sites on the matrix (two-site binding), the binding constant could be much larger.

Two-site protein binding, illustrated in Figure 3.7B, can be analyzed in a manner analogous to chelation in small metal ion complexes. The increase in the binding constant for the two-histidine protein compared to a single-histidine protein (Figure 3.7A) will depend on the enthalpy gained from forming the second histidine-CuIDA interaction relative to the entropy lost by further constraining the protein and the metal ion complexes at the support surface. If the entropic losses and the strain introduced upon binding are negligible, the increase in the apparent binding constant can be as large as the increase in affinity of a single-histidine protein over the protein with no histidines: a factor of 10^2 - 10^3.

Multiple-site binding to a support whose distribution of ligands is essentially random requires that the metal ion sites be densely packed in regions of the matrix accessible to protein. Our observations that ~15 immobilized copper ions are blocked by

adsorption of cytochrome *c* and that the immobilized IDA ligands (L) are close enough to form L-Cu-L complexes indicate a site density on this IDA-TSK gel that could support multiple-site binding.

For a protein with two potential binding sites, the reaction represented by eqn. 4 can be extended to include a second interaction,

$$P + nCu \rightleftharpoons PCu + \{(n-1)\,Cu\} \qquad K_1 = \frac{[PCu]}{[P][Cu]}\ ,$$

$$PCu + \{(n-1)\,Cu\} \rightleftharpoons PCu_2 + \{(n-2)\,Cu\} \qquad K_2 = \frac{[PCu_2]}{[PCu]}\ ,$$

where PCu_2 represents a protein bound to the matrix at two copper sites. As before, n represents the number of copper ions blocked by protein binding. The second binding constant (K_2) represents the apparent binding affinity of the second (histidine) side chain for a nearby copper site on the surface, a reaction of zero order in protein concentration. These equilibria yield the modified isotherm of eqn. 5,

$$Q = \frac{Cu_{acc} K_1 (1 + K_2) c}{1 + n K_1 (1 + K_2) c}\ . \qquad (5)$$

Because the distribution of copper ions on the support surface is not uniform, not all proteins adsorbed via a single histidine will find appropriate sites for a second interaction. Assuming that only some fraction (θ) of the surface sites can accommodate strong two-site binding ($K_2 >> 1$), while the remainder of the sites can accommodate only single-site binding ($K_2 \sim 0$), eqn. 5 can be extended to a bi-Langmuir-type equation,

$$Q = \frac{\theta Cu_{acc} K_1 K_2 c}{1 + nK_1 K_2 c} + \frac{(1-\theta) Cu_{acc} K_1 c}{1 + nK_1 c} \quad , \qquad (6)$$

where K_1K_2 is the apparent binding constant for protein binding via two-site interactions.

The binding data for the $H_{26}H_4$ and $H_{26}H_8$ variants can now be analyzed using eqn. 6, where Cu_{acc} equals the total copper loading (18.5 μmol/ml TSK) and the binding constants K_1 for fully-accessible histidines 4 and 8 are already known from the single-histidine variants H_4 and H_8. The values of K_2, θ, and n obtained from the isotherms for the multiple-histidine proteins are given in Table 3.6. Adding histidine 26 to a protein with either histidine 4 or 8 results in a 20-fold or greater increase in the apparent binding constant (relative to single-site binding) for a majority of the surface sites. This increase matches the 20-fold increase that histidine 26 provides the H_{26} variant over H(-). Similarly, the 200-fold increase histidines 4 or 8 offer the $H_{26}H_4$ and $H_{26}H_8$ variant over H_{26} is consistent with the increase of at least 200-fold these histidines offer the H_4 and H_8 variants over H(-). Thus a second, fully-accessible histidine adds > 3 kcal mol^{-1} to the binding energy, compared to a single-histidine protein. The values of n obtained by this analysis are also consistent with those for the single-histidine variants: ~15 copper sites are blocked by each protein molecule.

Addition of histidine 33 in $H_{26}H_{33}H_4$ variant increases the apparent initial slope by a factor of 8 compared to $H_{26}H_4$ (Table 3.3), indicating the possibility of three-site binding. Although the analysis represented by eqn. 6 can be extended to include simultaneous three-site binding (see Chapter 4), fitting the isotherms for the three-histidine cytochrome *c* variants leaves indeterminate the three-site binding parameters analogous to θ and K_2 for two-site binding. Because of the high affinity expected for simultaneous three-site binding (~ 10^7 M^{-1}), the initial slope of the adsorption isotherm must be measured accurately at impractically low protein concentrations (< 10^{-7} M) to

determine these parameters. Nonetheless, the apparent initial slope calculated from the first point on the adsorption isotherm represents a lower bound for the true initial slope. Thus simultaneous three-site binding is estimated to be at least an order of magnitude stronger than two-site binding.

The increased maximum capacity of the three-histidine variants correspondingly decreases the parameter n, the number of copper sites blocked by protein binding, from 15 - 18 for single-histidine proteins to only 8 - 10 for three-histidine proteins. The overly-simplified model developed here probably cannot provide a satisfactory explanation for this result. The increased capacity for tighter-binding proteins likely reflects a heterogeneous population of sites of widely different binding affinities (see Chapter 5), or, possibly, conformational rearrangement (compaction) of proteins adsorbed via high affinity three-site interactions compared to lower affinity single-site binding.

Binding mode changes with decreasing copper loading

We would expect binding via multiple-site interactions to depend strongly on the density and distribution of available copper ions at the support surface. While multiple-site binding might occur on a densely-derivatized support, binding should occur via single-site interactions on a support with a low concentration of available metal ions. This can be achieved easily by decreasing the copper loading (Figure 3.7C). As demonstrated by *N*-acetylhistidine adsorption, the immobilized IDA groups are close enough to simultaneously coordinate a single metal. A decrease in copper loading results in a significant portion of the total immobilized copper sequestered in L-Cu-L complexes. This leads to a more rapid decrease in the concentration of immobilized copper accessible to protein surface groups ($Cu_{acc} < Cu$).

As shown in Figure 3.9, the initial slope of the binding isotherm of a two-histidine protein decreases sharply with decreasing copper loading near the maximum (Cu ~ 18.6 μmol/ml TSK). At these high loadings the protein displays high-affinity binding characteristic of simultaneous, multiple-site interactions (Table 3.5). According to eqn. 6, the initial slope of the adsorption isotherm is given by the product $\theta K_1 K_2 Cu_{acc}$. According to the local equilibrium assumption, the initial slope of the isotherm is proportional to protein retention in isocratic chromatography under analytical conditions ($Q << \theta Q_{max} / n$) [22]. Decreasing the copper loading lowers the concentration of accessible copper sites (Cu_{acc}), which also affects the value of θK_2.

At lower copper loading (Cu ~ 10 μmol/ml TSK), accessible metal sites are much less densely packed, which decreases the likelihood that the surface can support high-affinity two-site binding. At these low loadings, this two-histidine protein displays simple isotherms typical of single-histidine proteins (Table 3.5). The apparent binding constant remains roughly constant with copper concentration and is comparable to that of imidazole (4×10^3 M^{-1}), suggesting that adsorption is due primarily to single-site binding to isolated copper sites. If so, the initial slope of the adsorption isotherm is $K_1 Cu_{acc}$, where K_1 does not depend on copper concentration. As expected, the initial slope of the horse cytochrome *c* adsorption isotherm is much less sensitive to the copper concentration in this range.

Implications of changing the binding mode from single to multiple-site interactions

The possibility that proteins can bind at multiple copper sites simultaneously has important implications for the use of existing chromatographic supports as well as the design of new ones. If a protein interacts with the support at more than one metal site, the availability of metals on a particular chromatographic surface will dramatically influence the retention and separation that can be achieved. For example, retention

during analytical isocratic chromatography of two proteins is governed by the ratio of the initial slopes of their respective binding isotherms. The initial slope of the isotherm is dominated by multiple-site interactions, as seen for the yeast cytochrome *c* variants with different surface histidines. The extent to which multiple-site interactions can occur is governed in turn by the density and accessibility of immobilized copper. Thus it may be possible to tailor IMAC supports for specific separations of multiple-histidine proteins by controlling the concentration of metal and how it is linked to the support.

For example, to separate a protein containing a single, high-affinity binding site (e.g., a metal-chelating site [1]) from other proteins containing multiple, low-affinity binding sites (e.g., non-chelating histidines) multiple-site interactions, as depicted in Figure 3.10A, should be avoided. Multiple-site interactions are minimized at low loadings, where the metal ions are isolated. Under these conditions protein adsorption would be governed by the sum of single-site binding constants, which favor the single high-affinity site. A support with a low density of immobilized copper ions, as depicted in Figure 3.10B, would efficiently separate a protein containing a di-histidine metal-binding site from other proteins with multiple surface histidines [23]. Because only a small fraction of the metal ions are actually involved in coordinating a protein, this could conceivably be accomplished without a drastic reduction in protein binding capacity.

This principle is demonstrated by the imidazole gradient elution of a series of yeast cytochrome *c* variants at decreased copper loading. As shown in Figure 3.11, at saturation copper loading the concentration of imidazole required to elute the protein from the metal affinity column increases with the number of surface histidines isolated on the protein surface (e.g., $H_{26}H_{33}$, $H_{26}H_{33}H_{39}$, and $H_{26}H_{33}H_{39}H_8$). Two additional cytochrome *c* variants contain di-histidine metal chelation sites, one at a beta sheet motif containing positions 39 and 58 ($H_{26}H_{33}H_{39}H_{58}$ [24]) and one at an alpha helix motif containing positions 50 and 54 ($H_{26}H_{33}H_{50}H_{54}$ [25]). At saturation copper loading the

concentration of imidazole required to elute the two metal-chelating variants, each with a total of four surface histidines, is somewhat higher than that required for an analogous four-histidine protein ($H_{26}H_{33}H_{39}H_8$). Under these conditions, protein adsorption is dominated by multiple-site binding (Figure 3.10A), which masks the relative advantage of single high-affinity site behind a background of multivalent interactions of analogous multiple-histidine proteins.

As the copper loading is decreased and the immobilized copper ions become more isolated, the proteins dependent on multiple-site interactions eventually elute with the void volume (1 mM imidazole). However, as illustrated in Figure 3.10B, the proteins with a metal chelation site still partition to the metal-affinity support phase with a relatively high affinity. Under these conditions, both the $H_{26}H_{33}H_{39}H_{58}$ and $H_{26}H_{33}H_{50}H_{54}$ variants require approximately 5 mM imidazole to elute (Figure 3.11). Under conditions of copper site isolation, protein adsorption is governed by single-site binding, which favors the single high-affinity metal chelation site.

It should also be noted that the developed model of protein adsorption limits multivalent binding to some subset of surface sites while the majority of sites participate in single-site adsorption, even for multiple-histidine proteins. Thus the low protein coverage of chromatographic experiments biases them toward those surface sites which accommodate simultaneous coordination of multiple protein functional groups. It is feasible that operation near limiting protein coverage could instead bias the separation toward those surface sites which cannot accommodate multivalent interactions. Such a bias would again favor those proteins with di-histidine metal chelation sites over those dependent on interactions with multiple copper sites.

Conversely, to optimize the separation of two-histidine proteins from those with single histidines, two-site interactions (θK_2) should be maximized relative to single-site interactions (K_1). The product θK_2 is sensitive to the density of copper sites, while the

binding constant K_1 for a single histidine protein is not. The high density of immobilized copper sites of this TSK matrix provides for relatively large values of θ and K_2, which translate into efficient separations.

A logical extension of this concept is the creation of specific high-affinity binding sites by distributing metals on the chromatographic surface to match the distribution of histidines on a target protein. For example, receptor complexes that position two metals at a fixed distance selectively bind their complementary bis-imidazole "protein analogs" in the presence of other bis-imidazoles. Selective chromatographic supports could be prepared by immobilizing appropriate bis-metal complexes. An alternative approach to preparing such "patterned" supports is template polymerization, in which the target molecule itself acts as a template to correctly position the metal ions in the polymer matrix during synthesis [26].

CONCLUSIONS

The role of surface histidines in IMAC was probed by equilibrium adsorption of a set of yeast cytochrome *c* variants differing only in surface histidine content. Proteins containing a single accessible histidine exhibit Langmuir-type isotherms with extrapolated limiting protein capacities much less than the capacity of the support to bind imidazole (~ 10%). A simple model, which assumes the copper sites were densely packed and that approximately 15 sites could be blocked by protein adsorption, yields binding constants for single-histidine proteins comparable to those for free imidazole and *N*-acetylhistidine.

Proteins containing multiple accessible histidines do not exhibit simple Langmuir-type behavior. At low protein coverage, the binding affinity of these variants increases by as much as a factor of 1000, apparently the result of simultaneous

coordination to more than one surface copper site. A consequence of multiple-site interactions is a significant increase in the binding affinity that strongly depends on the metal ion distribution. This principle is not limited to metal-affinity chromatography, but should be applicable to any separation in which multiple-site interactions can dominate binding (e.g., ion exchange [27]). The ability to control multiple-site interactions provides a unique opportunity to tailor IMAC and other chromatographic supports for specific protein separations.

EXPERIMENTAL METHOD

Variants of *S. cerevisiae* iso-1-cytochrome *c* (M_r = 12,588) containing different distributions of histidine residues were constructed by site-directed mutagenesis, expressed in yeast and purified as described previously [14,23]. Cytochromes *c* from tuna heart (M_r = 12,170) and horse heart (M_r = 12,384) were purchased from Sigma and used without further purification.

Preparation of metal-affinity gel

TSK Guardgel Chelate-5PW (17 μm macroporous beads) was purchased from Tosohaas. The immobilizing ligand, iminodiacetic acid (IDA), is attached to the TSK matrix via a spacer. The specific surface area of this polymer-based, hydrophilic support in the swollen state is not known; a pore size of ~1000 Å is reported by the manufacturer. One gram of dry TSK gel swells to 6.0 ml packed volume. From one ml of packed volume, 0.2 g of water can be removed by filtration through ASTM 10-15 fritted filter; thus approximately 20% of the packed volume is interstitial space. Approximately 70% of a TSK HPLC column is void volume (see below). From this, we estimate the porosity

of the TSK support as 50%, resulting in an estimated surface area of 95 m^2/g dry TSK, or 16 m^2/ml TSK (packed volume).

To prepare the support for use in binding studies, the TSK Chelate-5PW was packed into a column (1 cm × 10 cm) and washed extensively with 50 mM EDTA, pH 8.0. To obtain the maximum copper loading, the column was equilibrated with 50 mM $CuSO_4$. Excess copper was removed by washing with 0.1 M sodium acetate, pH 4.0. The gel was then equilibrated in 50 mM sodium phosphate, 0.5 M NaCl, pH 7.0 for measurement of binding isotherms. For copper loading less than maximum, the column was equilibrated with 0.1 M sodium acetate, pH 4.0, and 2 ml aliquots of the gel were transferred to 50 ml centrifuge tubes. Different copper loadings were obtained by adding limiting amounts of $CuSO_4$ (< 18.5 μmol Cu per ml gel) dissolved in 40 ml acetate buffer, followed by rapid vortexing and subsequent mixing by inversion for at least one hour. The gel was then washed three times with 40 ml of acetate buffer and three times with 40 ml phosphate buffer.

Measurement of copper loading

To quantify copper loading, the TSK gel was resuspended in the equilibration buffer (50 mM sodium phosphate, 0.5 M NaCl, pH 7.0; packed volume/total volume = 1/4). A 50 mM solution of EDTA, pH 8.0 (600 μl) was added to 400 μL of the gel suspension in a microfuge tube. This suspension was equilibrated for 60 min by inversion and spun at 10,000 rpm in a microcentrifuge for 5 min. The supernatant was transferred to a fresh tube, and the CuEDTA concentration was determined, in quadruplicate, by measuring the visible absorbance at 800 nm in comparison with that of CuEDTA, pH 7.0 (ε_{800} = 73.6 cm^{-1} M^{-1}). The maximum copper loading was found to be 18.5 μmol per ml of packed TSK gel, in agreement with values previously determined by repeated washing with EDTA [14].

Equilibrium binding isotherms of cytochromes c

Equilibrium binding isotherms were measured at room temperature using a modified version of the procedure developed by Hutchens and co-workers [6,7]. The TSK gel was loaded with copper ions and resuspended in the equilibration buffer (50 mM sodium phosphate, 0.5 M NaCl, pH 7.0; packed volume/total volume = 1/4) as described above. The high ionic strength, sufficient to elute the net positively charged cytochromes *c* from CM-Sephadex cation exchange resin [14], was used to minimize electrostatic interactions during IMAC. Cytochrome *c* (800 μL of 0.4 mM - 0.004 mM dilutions prepared in the same buffer) was added to 200 μL of the gel suspension in a microfuge tube. This suspension was equilibrated for 30 min. by inversion and spun at 10,000 rpm in a microcentrifuge for 5 min. The supernatant was transferred to a fresh tube, and the protein concentration was determined, in duplicate, by measuring the visible absorbance at 550 nm ($\varepsilon_{550} = 2.76 \times 10^4\ cm^{-1}\ M^{-1}$) after reduction with sodium dithionite [28].

Equilibrium binding isotherms of N-acetylhistidine

Equilibrium binding isotherms for *N*-acetylhistidine (acHIS) were measured at room temperature using a variation of the above procedure for cytochromes *c*. The IDA-TSK gel was loaded with copper and resuspended in the equilibration buffer as described above. Samples were prepared by adding 300 μl of acHIS solution (1 mM-50 mM dilutions prepared in the equilibration buffer) to 400 μl of the gel suspension plus 300 μl of equilibration buffer in a microfuge tube. This suspension was equilibrated for 30 min. by inversion and spun at 10,000 rpm in a microcentrifuge for 5 min. The supernatant was transferred to a fresh tube. At the same time, samples for a standard

curve were prepared by adding 300 µl of acHIS solution to 700 µl of equilibration buffer in a microcentrifuge tube.

The concentrations of acHIS in the sample and blank tubes were determined by reaction with diethylpyrocarbonate (DEPC) by a modification of the procedure of Roosemont [29,30]. Reaction mixtures were prepared in triplicate by adding 100 µl of acHIS sample to 800 µl of reaction buffer (25 mM $CuCl_2$, 0.125 mM EDTA, 0.125 M sodium acetate pH 6.0) plus 100 µl of 0.75% (vol/vol) DEPC diluted in 100% ethanol, with immediate vortexing. Small amounts of copper appeared to be stripped from the IDA-TSK gel at higher acHIS concentrations. The resultant CuEDTA complex absorbs strongly in the ultraviolet region. A constant background of copper (20 µM) in the reaction mixture greatly improved the reproducibility of the DEPC reaction. Reaction blanks were prepared by adding 100 µl of sample to 800 µl of reaction buffer plus 100 µl of 100% ethanol. All reactions were mixed by inversion for one hour. Concentrations of acHIS were determined by absorbance at 240 nm for reaction mixture relative to reaction blank, with standard curves determined daily ($\varepsilon_{240} = 3.5 \times 10^3\ M^{-1}\ cm^{-1}$).

Imidazole gradient chromatography

A TSK Chelate-G6000XL (10 µm macroporous beads) HPLC column (7.5 cm × 7.5 mm ID, total volume 3.31 ml, column void volume 2.3 ml) from Tosohaas (#80645) was used in all experiments. Experiments were performed at a flow rate of 1.0 ml/min as described previously [14]. For maximum copper loading, the column was regenerated by first washing with 15 ml of 50 mM EDTA pH 8.0, and then reloading with 15 ml of 50 mM $CuSO_4$ in water and finally washing with 15 ml of 0.1 M sodium acetate pH 4.0. For reduced copper loading, the column was regenerated by first washing with 15 ml of 50 mM EDTA pH 8.0. The column was rinsed with 10 ml of water and the pH adjusted

to 1.6 by washing with 15 ml of 0.2 M HCl. The column was then reloaded by washing with greater than three equivalents of $CuCl_2$ (> 250 μmol), at a concentration of 2 - 20 mM in 0.2 M HCl. The column was then washed briefly (5 ml) with water and finally washed with 15 ml of 0.1 M sodium acetate pH 4.0.

To quantify copper loading, the column was washed with 15 ml of 50 mM EDTA pH 8.0. The effluent was pooled and the concentration of CuEDTA determined by absorbance at 800 nm ($\varepsilon_{800} = 73.6\ cm^{-1}\ M^{-1}$). The saturation copper loading was found to be 81 μmol of copper (24.5 μmol per ml of column) in agreement with values previously determined [14].

The TSK Chelate column was regenerated as described above and then equilibrated between each experiment with 10 ml of 50 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole pH 7.0 followed by 10 ml of the same buffer containing 1 mM imidazole. A 25 μl sample of 3 μM cytochrome *c* in buffer containing 1 mM imidazole was injected and the column washed for six minutes. The protein was eluted with a linear imidazole gradient ranging from 1 mM to 20 mM imidazole over 54 minutes. The eluent was monitored at 410 nm. The reported imidazole concentration at elution is calculated from the elution time of the absorbance peak maximum.

Table 3.1. Engineered variants of *S. cerevisiae* iso-1-cytochrome *c*. All engineered variants contain the replacement of cysteine by serine at position 102 to prevent dimerization. Remaining mutations alter surface histidine content [14]. Number of surface histidines includes partially accessible histidine at position 26; native cytochromes *c* from tuna and horse are listed for comparison.

Label	Amino acid substitutions	Surface histidines
H(-)	C102S, H39Q, H33N, H26N	0
H_4	C102S, H39Q, H33N, H26N, K4H	1
H_8	C102S, H39Q, H33N, H26N, T8H	1
H_{26}	C102S, H39Q, H33N	1
$H_{26}H_4$	C102S, H39Q, H33Q, K4H	2
$H_{26}H_8$	C102S, H39Q, H33Q, T8H	2
$H_{26}H_{33}$	C102S, H39Q	2
$H_{26}H_{33}H_4$	C102S, H39Q, K4H	3
$H_{26}H_{33}H_8$	C102S, H39Q, T8H	3
$H_{26}H_{33}H_{39}$	C102S	3
$H_{26}H_{33}H_{39}H_8$	C102S, T8H	4
$H_{26}H_{33}H_{39}H_{58}$[a]	C102S, L58H	4
$H_{26}H_{33}H_{50}H_{54}$[b]	C102S, H39Q, D50H, K54H	4
Tuna (H26)		1
Horse (H26,H33)		2

[a] metal chelation site at H39-H58 [24]. [b] metal chelation site at H50-H54 [25].

Table 3.2. Calculated accessible surface areas of cytochrome *c* surface histidines. Calculations were performed using coordinates from the high-resolution crystal structure of *S. cerevisiae* iso-1-cytochrome *c* [15]. The surface area is reported as the total area at the ε- and δ-nitrogens accessible to a probe the size of CuIDA (1.93 Å radius [5]). The percent accessibility is the accessible surface area relative to an unhindered imidazole.

Histidine Position	**Accessible Surface Area** ($Å^2$)	**Percent Accessibility**
4	25.7	88
8	25.5	87
26	3.2	11
33	5.1	17
39	20.9	71

Table 3.3. Langmuir binding parameters for iso-1-cytochrome *c* variants, native cytochromes *c*, imidazole, and acHIS to CuIDA TSK. Adsorption isotherms of single-histidines proteins and acHIS have been fit to the Langmuir model (eqn. 1). Isotherms of multiple-histidine proteins have been fit to the bi-modal Langmuir model (eqn. 2). Imidazole isotherms have been fit to eqn. 3. Error estimates are given in parentheses.

	K_A (M^{-1})	$Q_{max,A}$ (μmol/ml TSK)	K_B (M^{-1})	$Q_{max,B}$ (μmol/ml TSK)
Imidazole	5.7 (0.7) × 10^3	18.0 (0.9)	4.0 (3.0) × 10^1	18.8 (0.6)
N-acetylhistidine	--	--	8.2 (1.0) × 10^2	18.6 (0.7)
H(-)	--	--	N.D.*	N.D.*
Tuna	--	--	2.4 (0.1) × 10^3	1.25 (0.30)
H_{26}	--	--	3.5 (0.6) × 10^3	1.12 (0.13)
H_4	--	--	5.6 (0.3) × 10^4	1.15 (0.02)
H_8	--	--	4.9 (0.5) × 10^4	1.03 (0.04)
Horse	3.8 (0.6) × 10^5	0.17 (0.02)	0.7 (0.3) × 10^4	1.33 (0.34)
$H_{26}H_4$	9.0 (1.0) × 10^5	1.10 (0.10)	1.9 (1.0) × 10^4	0.29 (0.10)
$H_{26}H_8$	1.0 (0.1) × 10^6	0.78 (0.06)	1.8 (1.0) × 10^4	0.56 (0.09)
$H_{26}H_{33}$	7.4 (0.7) × 10^5	0.52 (0.03)	2.4 (0.5) × 10^4	0.90 (0.02)
$H_{26}H_{33}H_4$	5.6 (0.8) × 10^6	1.33 (0.13)	4.8 (2.0) × 10^4	1.08 (0.13)
$H_{26}H_{33}H_8$	3.6 (0.7) × 10^6	1.32 (0.10)	1.5 (0.8) × 10^4	0.95 (0.11)
$H_{26}H_{33}H_{39}$	4.3 (0.4) × 10^6	1.27 (0.04)	2.9 (0.5) × 10^4	0.96 (0.04)

* K and Q_{max} could not be determined independently. The initial slope is 0.3 ml/ml TSK.

Table 3.4. Langmuir binding parameters for acHIS at decreased copper loading. Langmuir adsorption parameters (with standard error) determined by nonlinear regression to Langmuir isotherm (eqn. 1) of *N*-acetylhistidine equilibrium adsorption to IDA-TSK at pH 7.0 prepared at reduced copper loading (18.6 μmol/ml is maximum). Error estimates are given in parentheses.

Cu (μmol/ml TSK)	***K*** (M^{-1})	Q_{max} (μmol/ml TSK)
18.5 (0.4)	$8.2\ (1.0) \times 10^2$	18.4 (0.7)
14.5 (0.3)	$8.5\ (1.1) \times 10^2$	11.3 (0.4)
12.0 (0.3)	$4.6\ (0.8) \times 10^2$	14.2 (0.9)
8.7 (0.6)	$4.6\ (1.9) \times 10^2$	5.6 (0.8)
5.7 (0.3)	$6.9\ (1.6) \times 10^2$	2.4 (0.3)

Table 3.5. Langmuir binding parameters for horse heart cytochrome *c* at decreased copper loading. Bi-modal Langmuir adsorption parameters (with standard error) determined by nonlinear regression to bi-modal Langmuir isotherm (eqn. 2) of equilibrium adsorption data at pH 7.0 for horse heart cytochrome *c* to IDA-TSK prepared at reduced copper loading (18.6 μmol/ml is maximum). For $Cu < 14$ μmol/ml TSK, the limiting capacity of the strong site was negligible ($Q_{max,A} < 0.02$ μmol/ml TSK) and the adsorption data were fit to the Langmuir isotherm (eqn. 1). Error estimates are given in parentheses.

Cu (μmol/ml TSK)	K_A (M^{-1})	$Q_{max,A}$ (μmol/ml TSK)	K_B (M^{-1})	$Q_{max,B}$ (μmol/ml TSK)
18.5 (0.4)	$3.8\ (0.6) \times 10^5$	0.17 (0.02)	$7.3\ (3.0) \times 10^3$	1.33 (0.34)
15.6 (0.1)	$2.1\ (0.4) \times 10^5$	0.23 (0.05)	$7.4\ (4.9) \times 10^3$	0.99 (0.35)
14.8 (0.1)	$1.2\ (0.4) \times 10^5$	0.19 (0.08)	$3.8\ (1.9) \times 10^3$	1.19 (0.18)
12.3 (0.2)	--	--	$5.0\ (1.4) \times 10^3$	0.97 (0.25)
12.0 (0.1)	--	--	$5.3\ (0.9) \times 10^3$	0.80 (0.12)
10.5 (0.1)	--	--	$4.2\ (1.3) \times 10^3$	0.53 (0.15)
9.5 (0.1)	--	--	$4.0\ (3.1) \times 10^3$	0.43 (0.31)
7.6 (0.4)	--	--	$4.3\ (2.8) \times 10^3$	0.41 (0.15)
6.1 (0.3)	--	--	N. D.*	N. D.*

* K and Q_{max} could not be determined independently. The initial slope is 0.4 ml/ml TSK.

Table 3.6. Binding parameters for selected *S. cerevisiae* iso-1-cytochrome *c* variants based upon "site-exclusion" model. Adsorption isotherms for single-histidine variants have been fit using eqn. 4. Isotherms for two-histidine proteins have been fit to eqn. 6, using K_1 from analogous single-histidine variants. Total copper loading (Cu_{acc}) is 18.5 μmol/ml TSK. Error estimates are given in parentheses.

variant	K_1 (M^{-1})	n	K_2	θ
H(-)	$1.6\ (0.4) \times 10^1$	N. D.*	--	--
H_{26}	$2.2\ (0.2) \times 10^2$	16 (2.2)	--	--
H_4	$3.5\ (0.1) \times 10^3$	16 (0.2)	--	--
H_8	$2.7\ (0.2) \times 10^3$	18 (0.7)	--	--
$H_{26}H_4$	3.5×10^3	14 (0.3)	22 (3)	0.76 (0.05)
$H_{26}H_8$	2.7×10^3	15 (0.4)	27 (4)	0.58 (0.05)

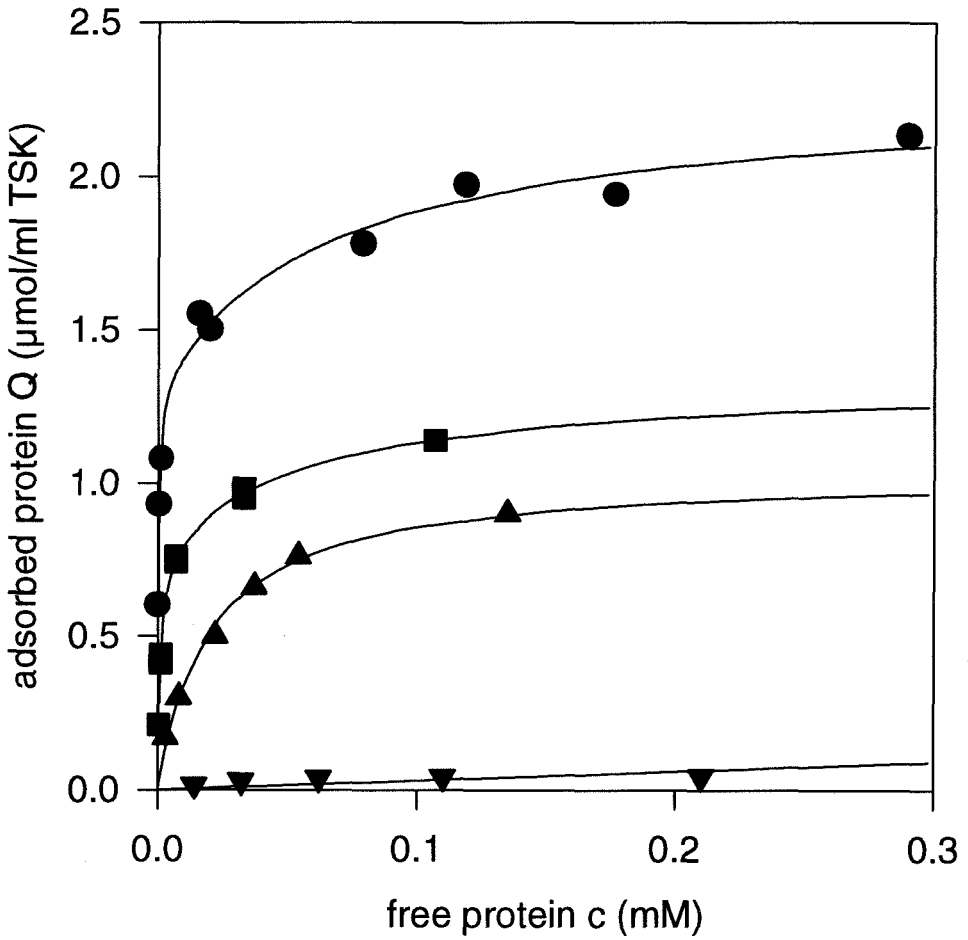


Figure 3.1. Equilibrium adsorption isotherms of yeast cytochrome *c* variants. Lines through data for (▼) H(-) and (▲) H_8 variants of *S. cerevisiae* iso-1-cytochrome *c* represent fit of Langmuir isotherm (eqn. 1) with parameters given in Table 3.3. Lines through data for (■) $H_{26}H_8$ and (●)$H_{26}H_{33}H_8$ variants represents the fit to a bi-modal Langmuir model (eqn. 2) using the parameters given in Table 3.3.

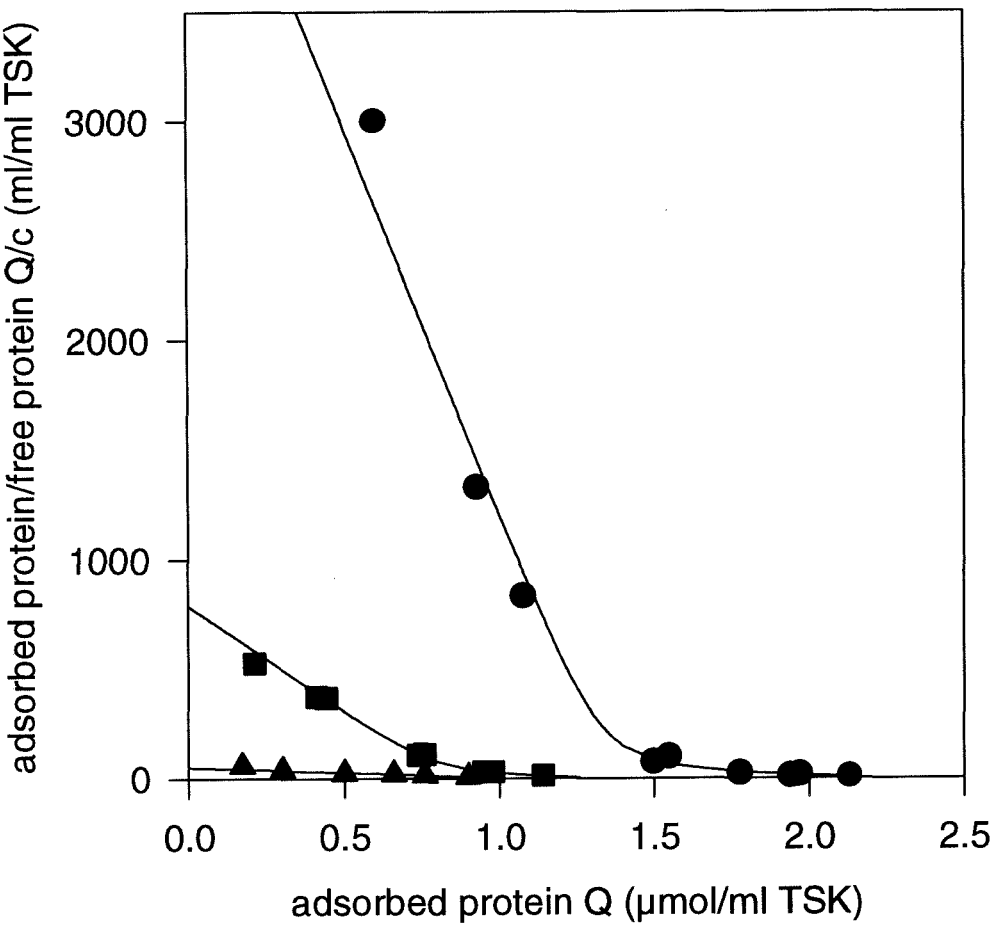


Figure 3.2. Scatchard plots for adsorption data of yeast cytochrome *c* variants. Line through data of (▲) H_8 variant of *S. cerevisiae* iso-1-cytochrome *c* represents fit of Langmuir isotherm (eqn. 1). Lines through data of (■) $H_{26}H_8$ and (●) $H_{26}H_{33}H_8$ variants represent fit to bi-modal Langmuir isotherm (eqn. 2).

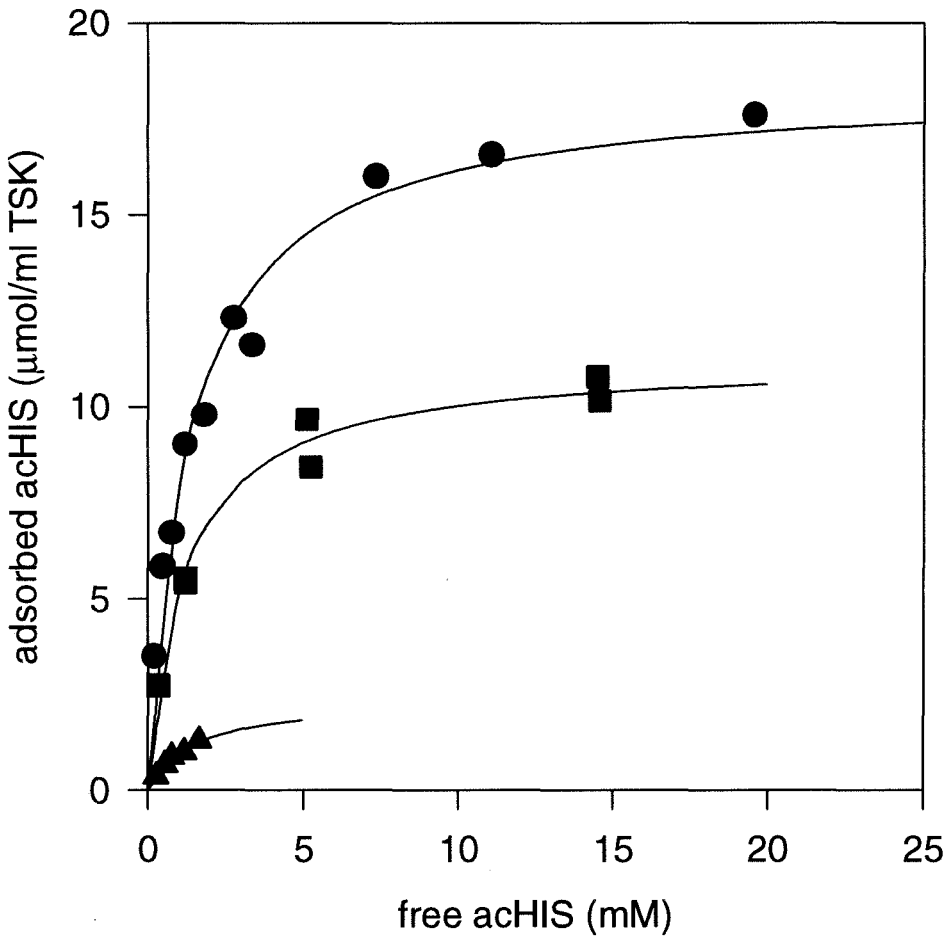


Figure 3.3. Equilibrium adsorption isotherms of acHIS at reduced copper loading. The amount of *N*-acetylhistidine adsorbed to IDA-TSK 5PW prepared with copper loading of (●) 18.6 µmol/ml TSK gel (saturation), (■) 14.5 µmol/ml, and (▲) 5.7 µmol/ml. Lines are best fit to Langmuir isotherm (eqn. 1) with parameters of Table 3.4.

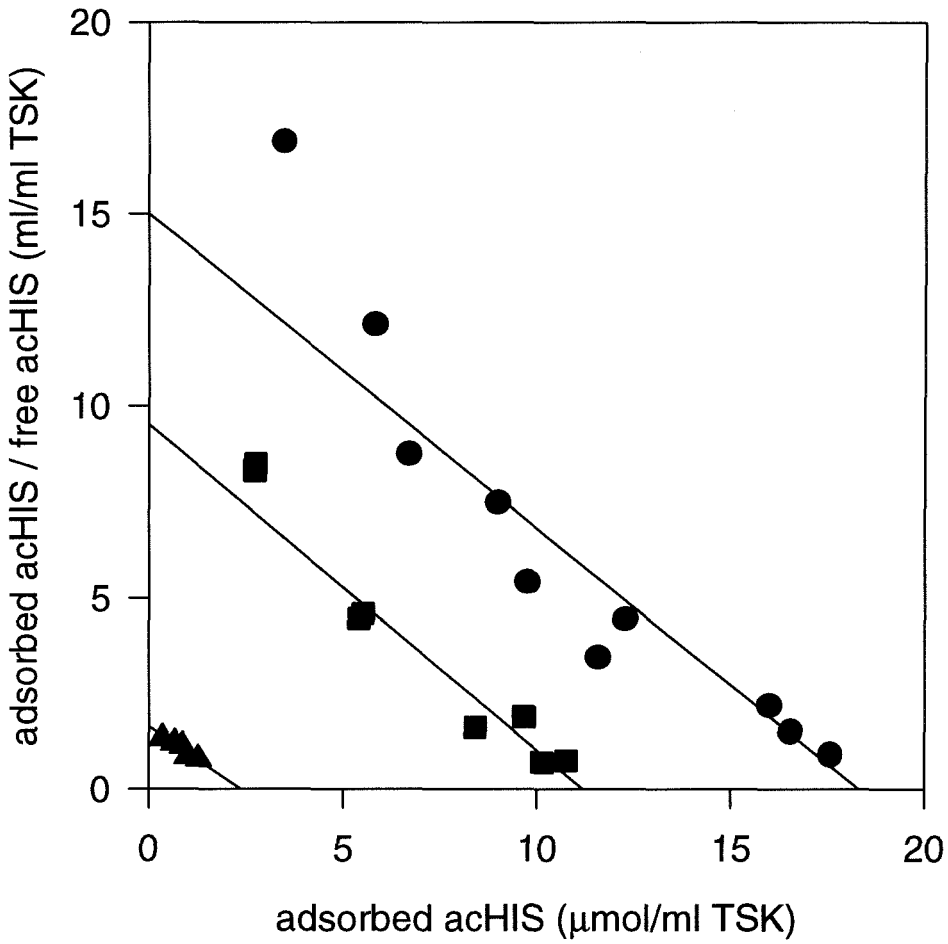


Figure 3.4. Scatchard plots of adsorption data for acHIS at reduced copper loading. The amount of *N*-acetylhistidine adsorbed to IDA-TSK support prepared with copper loading of (●) 18.6 µmol/ml TSK gel (saturation), (■) 14.5 µmol/ml, and (▲) 5.7 µmol/ml. Lines are best fit to Langmuir isotherm (eqn. 1) with parameters of Table 3.4.

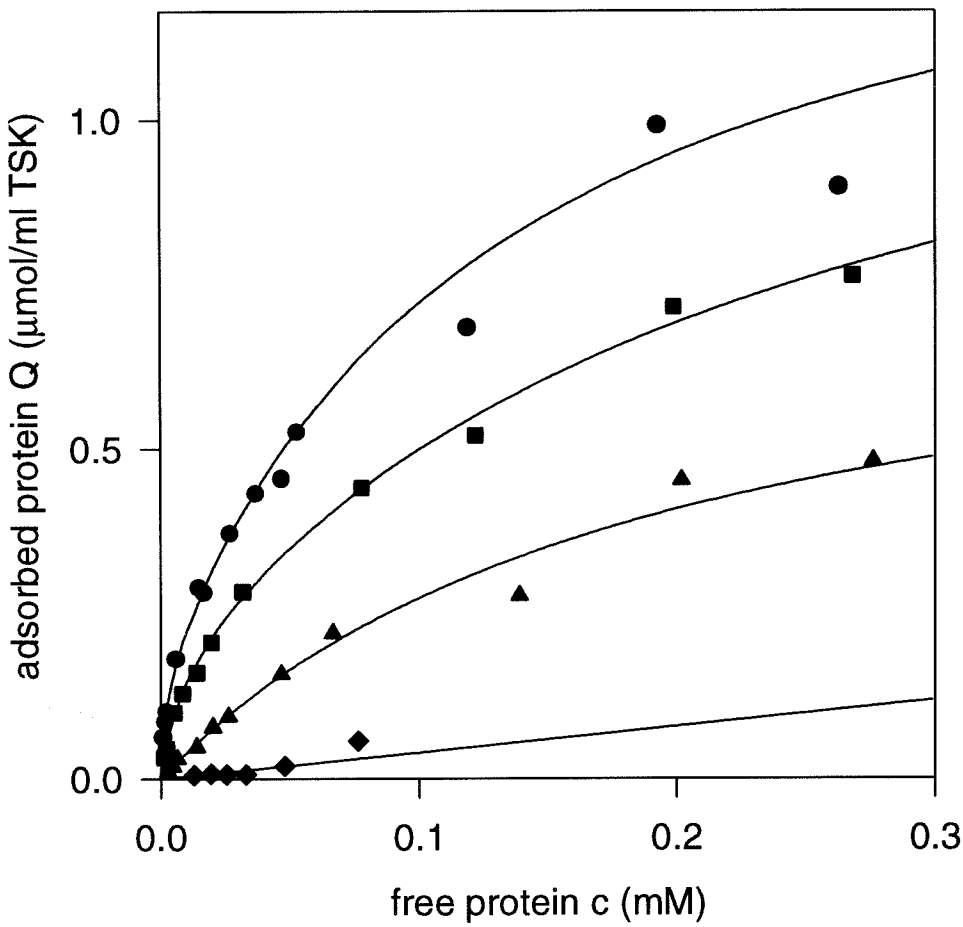


Figure 3.5. Equilibrium adsorption of horse cytochrome *c* at reduced copper loading. The amount of horse heart cytochrome *c* adsorbed to IDA-TSK 5PW prepared with copper loading of (●) 18.6 µmol/ml TSK gel (saturation), (■) 14.4 µmol/ml, (▲) 12.0 µmol/ml, and (▼) 6.1 µmol/ml. Lines are best fit to bi-modal Langmuir isotherm (eqn. 2) with parameters of Table 3.5.

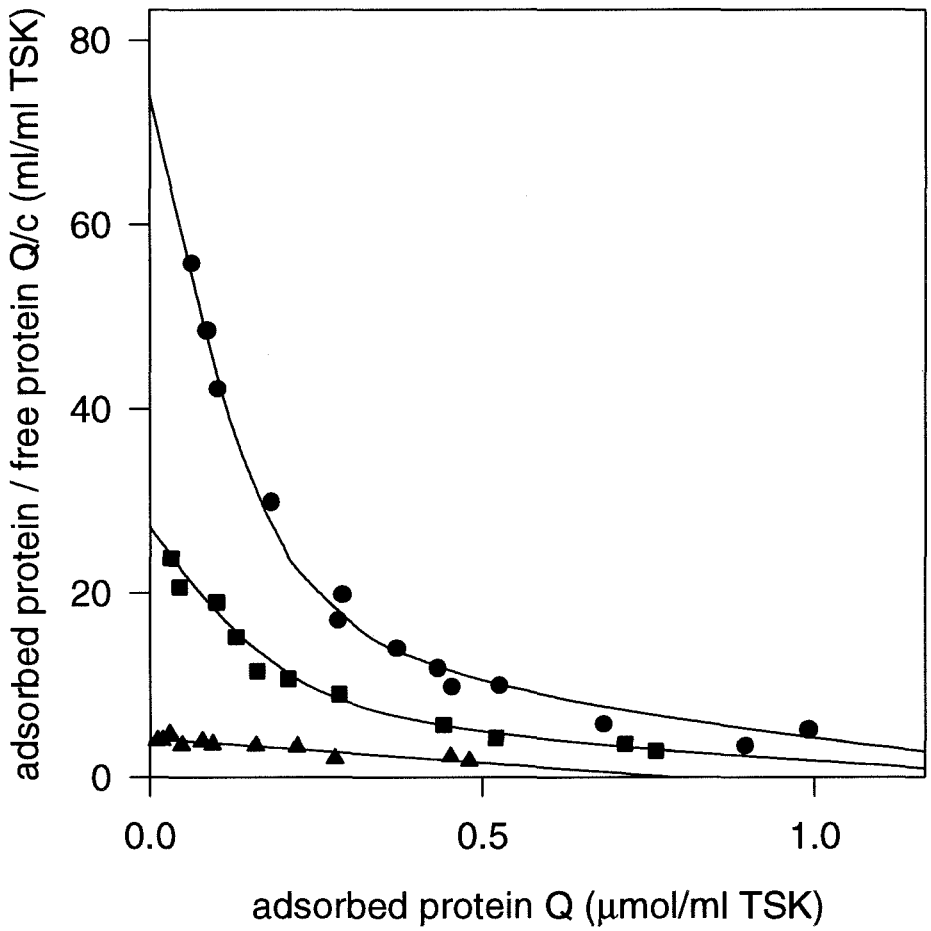


Figure 3.6. Scatchard plots of adsorption data for horse cytochrome *c* at reduced copper loading. The amount of horse heart cytochrome *c* adsorbed to IDA-TSK 5PW prepared with copper loading of (●) 18.6 µmol/ml TSK (saturation), (■) 14.4 µmol/ml, and (▲) 12.0 µmol/ml. Lines are best fit to bi-modal Langmuir isotherm (eqn. 2) with parameters of Table 3.5.

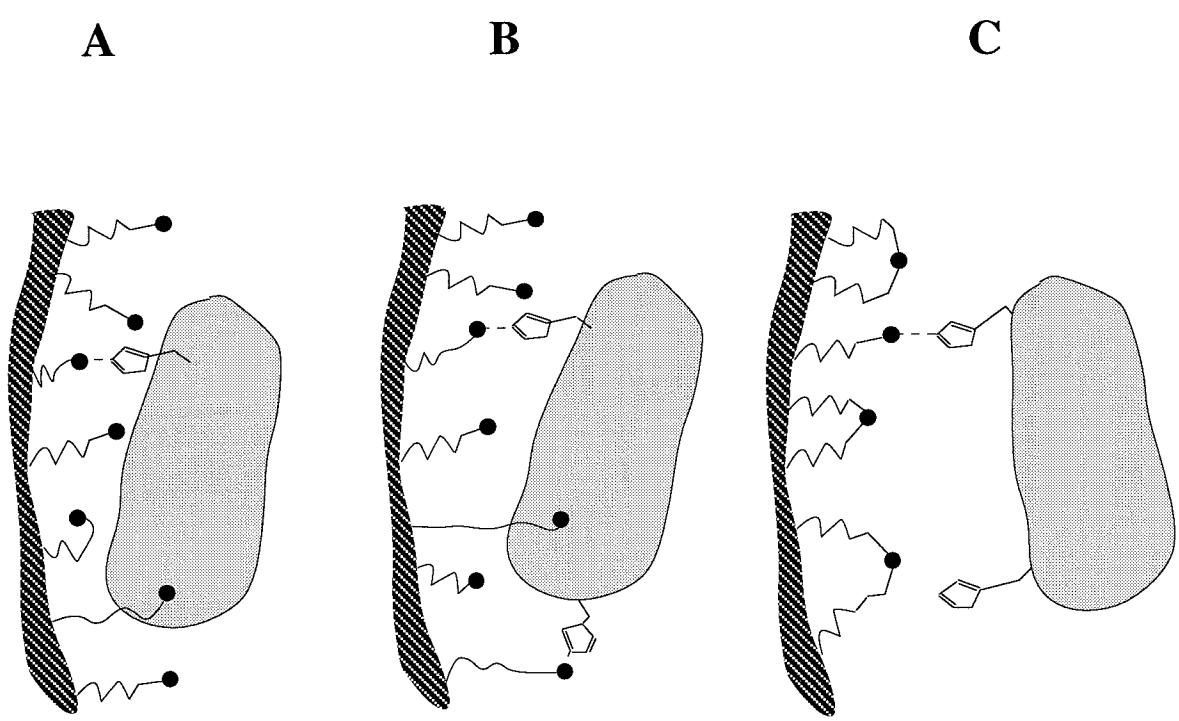


Figure 3.7. Proposed protein binding interactions in IMAC. **(A)** Protein binds IMAC surface at a single copper site and blocks access to other sites. **(B)** Multiple-site binding to IMAC surface: two histidines simulataneously coordinate to individual immobilized metal ions. **(C)** Multiple-histidine protein exhibiting single-site binding to IMAC surface at reduced copper loading. In some cases, adjacent immobilizing ligands simultaneously chelate a singel metal ion, preventing coordination by protein.

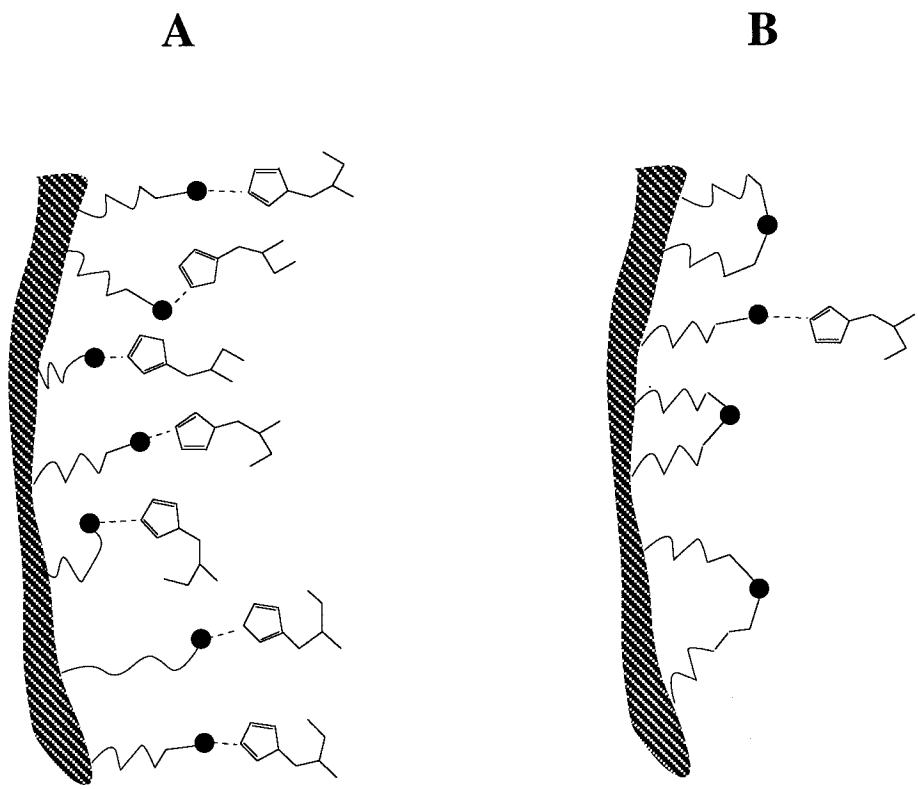


Figure 3.8. Proposed acHIS binding interactions in IMAC. **(A)** At saturation copper loading, acHIS binds surface at individual immobilized metal sites. **(B)** At reduced copper loading, acHIS exhibits reduced binding capacity to IMAC surface. In some cases, adjacent chelating ligands simultaneously coordinate a single metal ion, preventing interaction with acHIS.

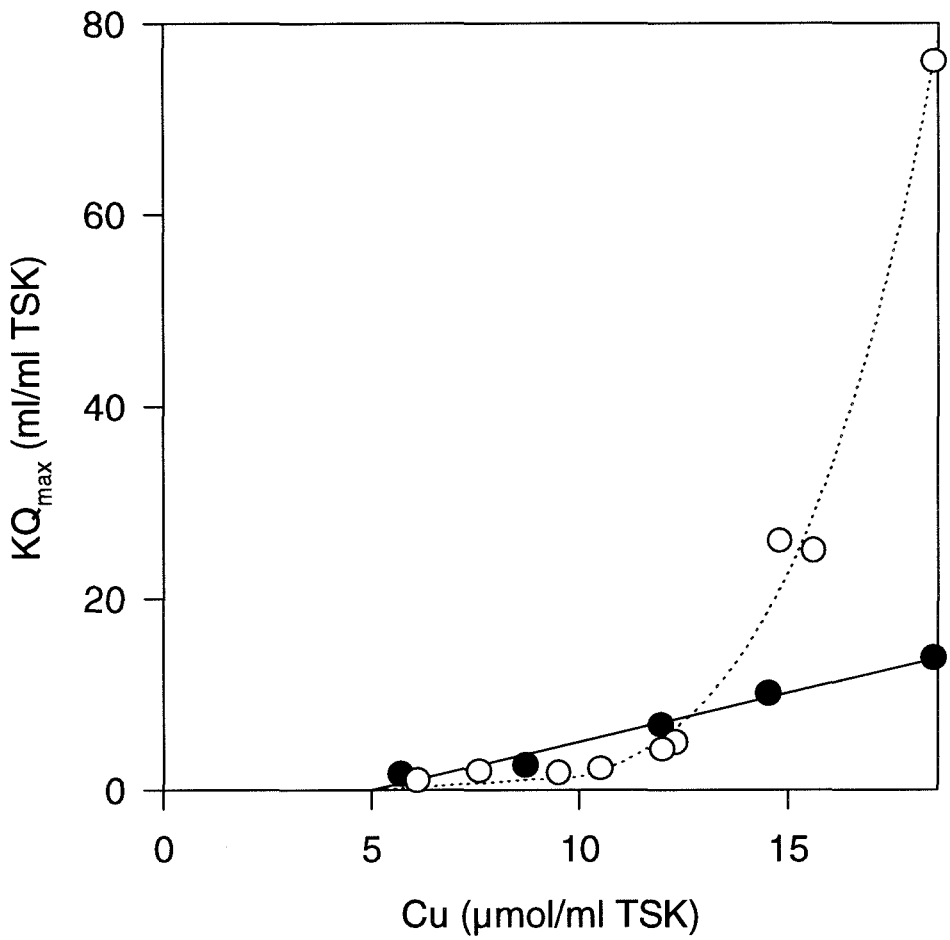


Figure 3.9. Effect of copper loading on acHIS and cytochrome *c* binding in IMAC. Initial slope of adsorption isotherms for (●) *N*-acetylhistidine, and (○) horse heart cytochrome *c*. The line of best fit for acHIS extrapolates to zero adsorption at Cu=5.0 μmol/ml TSK.

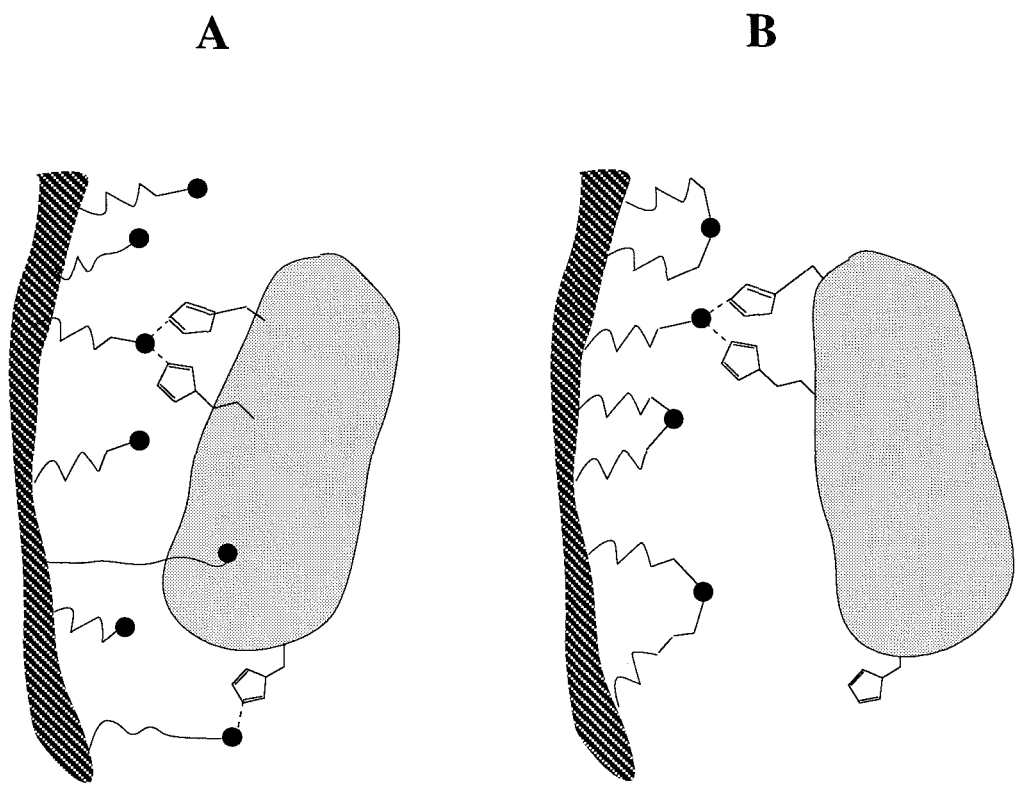


Figure 3.10. Effect of copper loading on IMAC of protein with di-histidine chelation site. **(A)** Multivalent binding of protein to IMAC surface. Di-histidine site results in high-affinity binding relative to protein with isolated histidines. **(B)** Protein with di-histidine chelation site retains high binding affinity to IMAC surface even at reduced copper loading.

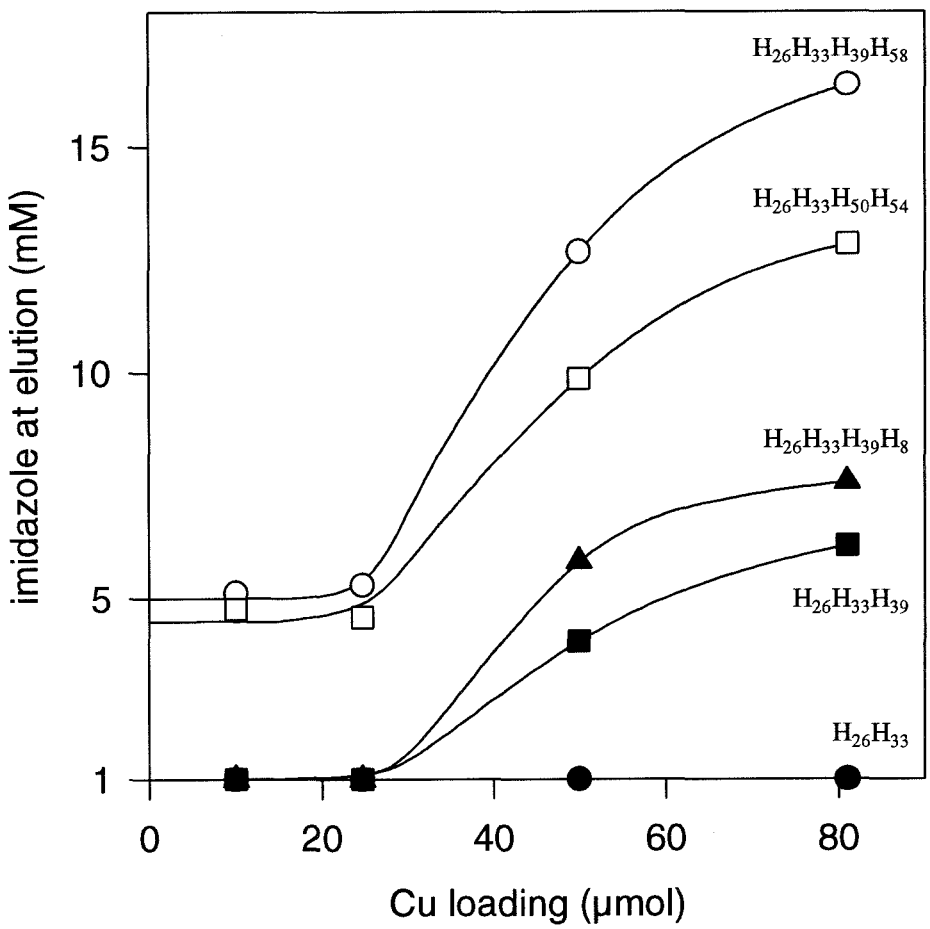


Figure 3.11. Imidazole gradient IMAC of cytochrome *c* at reduced copper loading. The amount of imidazole at elution of protein from IDA-TSK HPLC column prepared at decreased copper loading is measured at pH 7.0 for *S. cerevisiae* iso-1-cytochrome *c* histidine variants (●) $H_{26}H_{33}$, (■) $H_{26}H_{33}H_{39}$, (▲) $H_{26}H_{33}H_{39}H_8$, (□) $H_{26}H_{33}H_{50}H_{54}$ (metal chelation site at H_{50}-H_{54}), (○) $H_{26}H_{33}H_{39}H_{58}$ (metal chelation site at H_{39}-H_{58}).

APPENDIX

Table 3A.1. Equilibrium binding data for horse heart cytochrome *c* at reduced copper loading. The amount of adsorbed protein at equilibrium (Q) is measured at increasing protein solution concentration (c). Binding experiment is performed at decreased copper loading (18.6 μmol/ml is saturation), with mean copper content determined by EDTA extraction (standard deviation given in parenthesis).

Cu=18.6 (0.5) μmol/ml		*Cu*=15.6 (0.2) μmol/ml		*Cu*=14.8 (0.2) μmol/ml	
c (mM)	Q (μmol/ml)	c (mM)	Q (μmol/ml)	c (mM)	Q (μmol/ml)
0.263	0.903	0.261	0.930	0.268	0.767
0.193	0.998	0.202	0.838	0.199	0.720
0.118	0.689	0.148	0.715	0.122	0.524
0.053	0.529	0.088	0.541	0.078	0.444
0.047	0.458	0.050	0.515	0.032	0.286
0.037	0.435	0.027	0.380	0.020	0.209
0.027	0.374	0.016	0.283	0.014	0.162
0.017	0.285	0.012	0.239	0.0085	0.130
0.015	0.292	0.0078	0.200	0.0053	0.101
0.0061	0.183	0.0048	0.146	0.0022	0.045
0.0024	0.103	0.0026	0.103	0.0014	0.033
0.0018	0.087	0.0014	0.065		
0.0011	0.064	0.0009	0.044		

Table 3A.1. (continued)

Cu=12.33 (0.2) μmol/ml		*Cu*=12.0 (0.1) μmol/ml		*Cu*=10.5 (0.2) μmol/ml	
c (mM)	*Q* (μmol/ml)	*c* (mM)	*Q* (μmol/ml)	*c* (mM)	*Q* (μmol/ml)
0.280	0.514	0.276	0.484	0.253	0.314
0.136	0.366	0.202	0.455	0.132	0.173
0.060	0.230	0.139	0.280	0.063	0.104
0.047	0.198	0.067	0.223	0.024	0.045
0.037	0.156	0.047	0.160	0.014	0.033
0.024	0.097	0.026	0.095	0.0065	0.014
0.012	0.060	0.020	0.079	0.0025	0.005
0.0059	0.031	0.014	0.049		
0.0049	0.025	0.0064	0.030		
0.0038	0.016	0.0048	0.019		
0.0026	0.010	0.0027	0.011		

Cu=9.5 (0.3) μmol/ml		Cu=7.1 (0.4) μmol/ml		Cu=6.1 (0.4) μmol/ml	
c (mM)	*Q* (μmol/ml)	*c* (mM)	*Q* (μmol/ml)	*c* (mM)	*Q* (μmol/ml)
0.290	0.403	0.301	0.294	0.076	0.058
0.123	0.168	0.229	0.161	0.048	0.019
0.073	0.076	0.156	0.106	0.033	0.007
0.028	0.046	0.074	0.138	0.025	0.007
0.015	0.023	0.060	0.118	0.019	0.008
0.0071	0.010	0.045	0.059	0.013	0.006
0.0026	0.006				

Table 3A.2. Equilibrium binding data for *N*-acetylhistidine (acHIS) at reduced copper loading. The amount of adsorbed acHIS at equilibrium (Q) is measured at incrasing acHIS solution concentration (c). Binding experiment is performed at decreased copper loading (18.6 μmol/ml TSK is saturation), with mean copper content determined by EDTA extraction (standard deviation given in parenthesis).

Cu=18.5 (0.4) μmol/ml		*Cu*=14.5 (0.3) μmol/ml		*Cu*=12.3 (0.9) μmol/ml	
c (mM)	Q (μmol/ml)	c (mM)	Q (μmol/ml)	c (mM)	Q (μmol/ml)
19.6	17.702	14.6	10.219	14.5	11.281
11.1	16.658	14.5	10.844	7.85	11.764
7.35	16.103	5.24	8.467	4.99	10.860
3.36	11.685	5.11	9.730	2.34	7.695
2.77	12.390	1.21	5.444	1.32	4.377
1.80	9.835	1.20	5.550	0.88	3.715
1.20	9.070	0.33	2.739	0.57	2.727
0.77	6.748	0.32	2.771	0.38	2.259
0.48	5.871			0.16	1.298
0.21	3.518				

Cu=8.7 (0.5) μmol/ml		*Cu*=5.7 (0.3) μmol/ml	
c (mM)	Q (μmol/ml)	c (mM)	Q (μmol/ml)
15.08	5.209	1.64	1.292
8.75	3.180	1.16	0.993
5.58	5.219	0.77	0.868
2.79	3.397	0.54	0.655
1.56	2.062	0.26	0.349
1.07	1.813		
0.72	1.334		
0.51	0.973		
0.25	0.466		

REFERENCES

1. Arnold, F. H. (1991) *Bio/Technology 9*, 151-156.
2. Wong, J. W., Albright, R. L. and Wang, N. L. (1991) *Sep. Pur. Meth. 20*, 49-106.
3. Abelson, J. N., Simon, M. I. and Arnold, F. H., Editors (1992) *Methods: A Companion of Methods in Enzymology 4*, No. 1, 1992.
4. Hemdan, E. S., Zhao, Y., Sulkowski, E. and Porath, J. (1989) *Proc. Natl. Acad. Sci. USA 86*, 1811-1815.
5. Mrabet, N. T. (1992) *Biochemistry 31*, 2690-2702.
6. Hutchens, T. W., Yip, T. T. and Porath, J. (1988) *Anal. Biochem. 170*, 168-182.
7. Hutchens, T. W. and Yip, T. T. (1990) *Anal. Biochem. 191*, 160-168.
8. Mallik, S., Johnson, R. D. and Arnold, F. H. (1993) *J. Am. Chem. Soc. 115*, 2518-2520.
9. Mallik, S., Johnson, R. D. and Arnold, F. H. (1994) *J. Am. Chem. Soc. 116*, 8902-8911.
10. Dhal, P. K. and Arnold, F. H. (1991) *J. Am. Chem. Soc. 113*, 7417-7418.
11. Dhal, P. K. and Arnold, F. H. (1992) *Macromolecules 25*, 7051-7059.
12. Zhao, Y., Sulkowski, E. and Porath, J. (1991) *Eur. J. Biochem. 202,* 1115-1119.
13. Mayo, S. L., Ellis, W. R. Jr., Crutchley, R. J. and Gray, H. B. (1986) *Science 233*, 948-952.
14. Todd, R. J. (1993) *Ph.D. Thesis*, California Institute of Technology, Pasadena, CA.
15. Louie, G. V. and Brayer, G. D. (1990) *J. Mol. Biol. 214*, 527-555.
16. Sinha, P. C., Saxena, V. K., Nigam, N. B. and Srivastava, M. N. (1989) *Ind. J. Chem. 28A*, 335-336. The reported binding constant of 2.2×10^3 M^{-1}

(imidazole-Cu^{2+}IDA at 35 C) was extrapolated to 25 C using ΔH = -7.6 kcal/mol (imidazole-Cu^{2+}) [17].

17. Martell, A. E. and Smith, R. M. (1975) *Critical Stability Constants*, vol. 1-6, Plenum Press, New York.
18. Hutchens, T. W. and Yip, T. T. (1990) *J. Chromatogr. 500*, 531-542.
19. Sulkowski, E. (1989) *BioEssays 10*, 170-175.
20. Moore, C. D., Al-Misky, O. N. and Lecomte, J. T. J. (1991) *Biochemistry 30*, 8357-8365.
21. Belew, M., Yip, T. T., Andersson, L. and Porath, J. (1987) *J. Chromatogr. 403*, 197-206.
22. Arnold, F. H., Schofield, S. A. and Blanch, H. W. (1986) *J. Chromatogr. 335*, 1-12.
23. Todd, R. J., Van Dam, M., Casimiro, D., Haymore, B. L. and Arnold, F. H. (1991) *Proteins 10*, 156-161.
24. Kellis, J. T. Jr., Todd, R. J. and Arnold F. H. (1991) *Bio/Technology 9*, 994-995.
25. Muheim, A. and Arnold, F. H. (1993) unpublished results.
26. Mallik, S., Plunkett, S., Dhal, P. K., Johnson, R. D., Pack, D., Shnek, D. and Arnold, F. H. (1994) *New J. Chem. 18*, 299-304.
27. Chicz, R. M. and Regnier, F. E. (1988) *J. Chromatogr. 443*, 193-203.
28. Margoliash, E. and Frohwirt, N. (1959) *Biochem. J. 71*, 570-572.
29. Roosemont, J. L. (1978) *Anal. Biochem. 88*, 314-320.
30. Lundblad, R. L. and Noyes, C. M., Editors (1988) *Chemical Reagents for Protein Modification*, vol. 1, CRC Press, Boca Raton, FL.

CHAPTER 4

MULTIVALENT BINDING II: PH EFFECTS IN IMMOBILIZED METAL AFFINITY CHROMATOGRAPHY OF CYTOCHROME *C*

INTRODUCTION

The high cost of protein isolation and purification is an important obstacle in the development of protein products. Traditional separations techniques such as precipitation and ion exchange chromatography are often insufficient by themselves; a purification train of several different techniques in series is usually necessary to obtain a product of desired quality. On the other hand, the enhanced selectivity of biological affinity separations comes at great cost; the required materials are often fragile, prohibitively expensive, and limited in application. There is a call for novel separations materials which can recognize proteins with specificities approaching those of biological affinity chromatography, but providing the durability, low operating expense, and versatility of traditional chromatographic techniques.

Among the many protein purification approaches introduced in recent years, one stands out for its ease of use and wide applicability: immobilized metal affinity chromatography, or IMAC [1,2]. This highly versatile technique, based on the interactions between biological molecules and immobilized metal ions, is appropriate for both analytical and large-scale separations and is rapidly becoming a technique of choice for purification of recombinant proteins and peptides. Binding selectivity in IMAC is derived from the multiplicity and local environment of metal-coordinating residues [3,4,5], and as such can be tailored through the choice of metal ion, solvent conditions, or by modification of the target protein [6].

In Chapter 3, the role of surface histidines in IMAC was probed by equilibrium adsorption of a set of yeast cytochrome *c* variants differing only in surface histidine content [7]. A simple model, which assumed the copper sites were densely packed and that approximately 15 sites could be blocked by protein adsorption, yielded binding

constants for single-histidine proteins comparable to those for free imidazole and *N*-acetylhistidine. At low protein coverage, the binding affinity of proteins containing multiple accessible histidines increased by as much as a factor of 1000, apparently the result of simultaneous coordination to more than one surface copper site.

This simple model assumes that equilibrium protein adsorption in metal affinity chromatography is due primarily to surface histidines. Therefore, the pH dependence of protein binding in isocratic chromatography should reflect the effect of histidine imidazole protonation. The histidines of the yeast cytochrome *c* variants have pK's ranging from 5.5 to 6.5, as previously measured by ^{1}H NMR spectroscopy [8]. Because these histidines will be mostly deprotonated at pH 7, protein binding affinity and chromatographic retention should level off at basic conditions. However, the retention of a single-histidine protein on the IMAC support increases dramatically at pH > 7 [8]. Such behavior cannot be explained by the proton titration of the lone surface histidine.

In this investigation, chromatographic retention of *N*-acetylhistidine to a metal-complexing support (Cu^{2+} chelated by the iminodiacetate-derivatized resin TSK-IDA) under isocratic conditions demonstrates that metal-imidazole coordination is indeed inhibited at lower pH. The pH dependence of acHIS binding is described accurately by protonation of the imidazole nitrogens with the assumption of local equilibrium [8, 9]. The pH dependence of protein adsorption in IMAC is explored using of a set of three yeast (*S. cerevisiae*) iso-1-cytochrome *c* variants having no surface histidines or a single surface histidine. In contrast to the results for acHIS, simultaneous coordination to more than one surface copper site is necessary to account for observed chromatographic retention, particularly at elevated pH. The additional metal-coordinating groups on the protein are most likely primary amines (e.g., lysine and amino-terminus). Such amine groups have previously been implicated in the adsorption of small peptides to IMAC supports [10,11]. Thus, even for single-histidine proteins at neutral pH, protein

adsorption at low coverage appears to depend upon multiple interactions between the protein and immobilized copper ligands.

While the additional role of surface amine groups in IMAC complicates a molecular description of protein adsorption, it generalizes the phenomenon of multivalent binding to a wider class of proteins. Simultaneous coordination of multiple copper sites, particularly at low protein coverage and elevated pH, is no longer limited to those proteins with multiple surface histidines. It has previously been demonstrated (Chapter 3) that processes which manipulate the extent to which multivalent binding can occur (e.g., metal ligand loading) represent powerful operating parameters to tailor IMAC separations for particular target proteins. In particular, interactions at multiple sites can greatly enhance the binding affinity and dramatically alter specificity for certain classes of proteins. It should now be possible to target individual protein molecules by matching the distribution of metal ions to the spatial distribution of metal-coordinating imidazoles *and amines*. Efforts are underway to design new chromatographic supports capable of specific multiple-site binding to target molecules [12]. The additional role of deprotonated amine groups to multivalent binding in IMAC therefore has important implications for the design of efficient separations as well as for the design of new materials for IMAC supports.

RESULTS AND DISCUSSION

Effect of pH on N-acetylhistidine isocratic elution

To verify the effects of imidazole protonation in IMAC, chromatographic retention of *N*-acetylhistidine on the CuIDA-TSK column was measured for isocratic conditions over the pH range 6.0 to 9.0. Protein retention in chromatography is usually described by the capacity factor (k', ml solvent/ml column) which represents the

partitioning of the protein between the mobile phase (aqueous solution) and the stationary phase (column material, in this case TSK). The capacity factor, calculated from of the volume of solvent required to elute the protein from the chromatography column under isocratic conditions, is given by

$$k' = \frac{V_e - V_o}{\nu} \quad , \tag{1}$$

where V_e is mean elution volume of the protein (ml), V_o is the column void volume (ml), and ν is the total column volume (ml TSK). In the previous chapter, equilibrium adsorption of single-histidine proteins and *N*-acetylhistine could be described by the Langmuir isotherm,

$$Q(c) = \frac{Q_{max} K c}{1 + Kc} \quad , \tag{2}$$

where Q is the amount of protein adsorbed per ml of packed TSK gel and c is the liquid-phase protein concentration. The isotherm is fully described by two parameters, an equilibrium binding constant K (M^{-1}) and the limiting capacity for the adsorbed solute, Q_{max} (mol/ml TSK).

According to local equilibrium theory, chromatographic retention can be predicted directly from the Langmuir isotherm parameters K and Q_{max} [9] as

$$k' = KQ_{max} \quad . \tag{3}$$

This expression is strictly true only under analytical chromatographic conditions, that is, in the linear portion of the Langmuir adsorption isotherm ($Kc << 1$).

The isocratic zonal elution of *N*-acetylhistidine (acHIS) on CuIDA-TSK verifies the correspondence between equilibrium binding isotherms and chromatographic capacity factors as predicted by eqn. 3. To compare the capacity factor measured on the CuIDA-TSK HPLC column (24.5 μmol/ml TSK) with the equilibrium binding isotherm previously determined for CuIDA-TSK Guardgel (18.6 μmol/ml TSK, see Chapter 3), it is necessary to account for the different copper ligand densities. As shown in Table 4.1, the capacity factor of acHIS measured by chromatographic retention at pH 7.0 ($k' = 21$ ml/ml TSK, eqn. 1) is then in excellent agreement with the initial slope of the equilibrium binding isotherm ($KQ_{max} = 20.9$ ml/ml TSK, eqn. 2).

At lower pH, protonation of the imidazole group precludes interaction with the immobilized copper ions. As shown in Figure 4.1, acHIS retention increases with pH from 5.0 - 7.0 and levels off at pH > 7.0. It is straightforward to extend eqn. 3 to include the effects of protonation [8,9],

$$k' = \frac{k'_o}{1+10^{pK-pH}} \quad , \tag{4}$$

where k'_o is the chromatographic capacity factor for the deprotonated species and pK is the logarithm of the equilibrium constant for protonation (M^{-1}). The pH dependence of the acHIS capacity factor (k') can then be described by the protonation of the lone imidazole (Figure 4.1). In fact, as shown in Table 4.2, the pK determined from these data (6.98) is in excellent agreement with the reported pK of 6.96 for the imidazole nitrogen of *N*-acetylhistidine [13]. An analogous pH dependence has also been reported for the primary amine group of tryptamine; isocratic retention by a copper-chelating

chromatography support is described by a single protonation reaction with an apparent pK ~ 9 [10], in excellent agreement with the pK of primary amines [13].

In light of the greater than expected protein retention previously observed at elevated pH [8], the isocratic chromatography of *N*-acetylhistidine validates two important principles. First, the chromatographic capacity factor is closely related to the initial slope of the equilibrium binding isotherm (eqn. 2), as predicted by the local equilibrium assumption [9]. Second, copper-imidazole coordination in IMAC indeed is inhibited at low pH as described by the equilibrium protonation of imidazole nitrogens (eqn. 4) [8].

Effect of pH on cytochrome c isocratic elution

Isocratic chromatography of three *S. cerevisiae* iso-1-cytochrome *c* variants, two with a single surface histidine (H_8 and H_{26}, labeled as in Chapter 3) and one with none (H(-)), was also measured on the CuIDA-TSK column for isocratic conditions over the pH range 5.0 to 9.0. As shown in Table 4.1, the capacity factors measured at pH 7.0 by isocratic elution are also in good agreement with the initial slope of the adsorption isotherms from equilibrium binding experiments (Chapter 3 [7]). The chromatographic retention of these proteins shows that the H_8 variant, with a single, fully accessible surface histidine, partitions to the IMAC support with approximately an order of magnitude higher binding affinity than the H_{26} variant which has a single, partially accessible surface histidine. At this pH the partitioning of the H(-) variant which has no surface histidines, is negligible.

However, protein retention in immobilized metal affinity chromatography can be extremely sensitive to pH, as shown in Figure 4.2. The steep increase in chromatographic retention of all three variants under slightly basic conditions (pH 7 - 8)

is in sharp contrast to the pH dependence of *N*-acetylhistidine adsorption (Figure 4.1) which reaches a plateau as the lone imidazole group becomes completely deprotonated.

The pK of the imidazole nitrogen of the histidine at position 8 has previously been measured by NMR to be 5.7 [8]. If, as modeled in Chapter 3, adsorption of a single-histidine protein is due solely to metal coordination of the single surface histidine, then the pH profile of the H_8 single-histidine variant should be similar to that of acHIS. Instead, there is a continued increase in protein retention with increasing pH which suggests metal coordination by other protein surface groups in addition to surface histidines, particularly at elevated pH. Moreover, even the H(-) variant, with *no* surface histidines, is retained strongly by the IMAC column at moderately basic conditions (pH ~ 8).

Multivalent binding of single-histidine proteins

To account for the contribution of additional metal-coordinating groups to protein adsorption, the two-site binding model developed in Chapter 3 [7] can be easily extended. As in the two-site binding model, protein adsorption via a single surface histidine is described by

$$P + nCu \rightleftharpoons PCu + \{(n-1)\,Cu\} \qquad K_1 = \frac{[PCu]}{[P][Cu]},$$

where the protein (P) blocks n - 1 additional copper sites from further interactions (blocked sites designated by brackets) upon coordinating to a metal ion (Cu) on the chromatographic support. Protein binding to a nearby copper site by a second surface metal-coordinating group is described by

$$PCu + \{(n-1)\,Cu\} \rightleftharpoons PCu_2 + \{(n-2)\,Cu\} \qquad K_2 = \frac{[PCu_2]}{[PCu]},$$

where PCu_2 represents a protein bound to the IMAC support at two copper sites. The binding constant (K_2) represents the apparent binding affinity of the second side chain for a nearby copper site on the surface; this reaction is of zero order in protein concentration. For those surface sites which support protein binding to additional copper sites, the equilibria can be generalized to z+1 copper sites as

$$PCu_z + \{(n-z)\,Cu\} \rightleftharpoons PCu_{z+1} + \{(n-z-1)\,Cu\} \quad K_2 = \frac{[PCu_{z+1}]}{[PCu_z]},$$

where PCu_{z+1} represents a protein bound to the support by the single surface histidine and z additional protein surface groups. Again, the binding constant (K_2) represents the apparent binding affinity of each additional side chain for a nearby copper site on the surface. These equilibria yield the modified Langmuir isotherm described by eqn. 4,

$$Q = \frac{Q_{max} K_1 (1+K_2)^z c}{1 + K_1 (1+K_2)^z c}. \tag{5}$$

By analogy to eqn. 3, under analytical chromatographic conditions (in this case, $K_1(1+K_2)^z << 1$), isocratic protein retention can be expressed as

$$k' = Q_{max} K_1 (1+K_2)^z. \tag{6}$$

Finally, if the single surface histidine with binding constant K_1 titrates at one pK (e.g., pK_1~ 6) and the set of z additional coordinating groups titrate at a higher pK (e.g., amines with pK_z ~ 9), then from eqn. 6, the pH dependence of protein retention is

$$k' = \frac{k'_o}{1+10^{pK_1-pH}} \cdot \left(1+\frac{K_2}{\left(1+10^{pK_z-pH}\right)}\right)^z , \tag{7}$$

where k'_0 (ml/ml TSK) represents the capacity factor for protein adsorption by the single surface histidine ($k'_0=K_1Q_{max}$). Thus at moderate to low pH (pH < pK_1< pK_z) where all z groups are fully protonated, eqn. 7 collapses to the expression for a single metal-coordinating group (eqn. 4). At elevated pH (pH > pK_1), or if all surface coordinating groups are equivalent, then all z groups are titrating simultaneously and this expression is analogous to the Hill model for cooperative binding [14]. Moreover, if the pH is significantly less than the pK of the z titrating groups (pH < pK_z), and if the binding affinity of each additional surface group is favorable (K_2 >> 1), then the logarithm of the chromatographic capacity factor reduces to

$$\log k' = z \cdot pH + \left[\log k'_o + z \cdot \left(\log K_2 - pK_z\right)\right] . \tag{8}$$

Under these conditions, the logarithm of the capacity factor is linear in pH, with a slope equal to the number of metal-coordinating groups (z) titrating in that pH range. This relationship is analogous to the "stoichiometric displacement model" previously developed for ion exchange chromatography, in which the effect of equilibrium competitors is used to estimate the number of interactions between the protein and the surface [15].

As shown in Figure 4.3, chromatographic retention of the H_{26} and H(-) variants appears to follow the linear relationship between the logarithm of the capacity factor (log k') and pH predicted for pH < pK_z. In fact, the pH dependence of these proteins is described by eqn. 8 with z of 2.7 (H_{26}) and 3.4 (H(-)), as presented in Table 4.3. Thus, approximately three surface metal-coordinating groups each with pK > 8 contribute to protein adsorption at basic conditions. Clearly, these additional groups are not surface histidines: the H(-) variant, which has none, displays the same highly cooperative behavior, and the elevated pK of these groups is not consistent with that of imidazoles (pK= 6 - 7).

Tthe elevated pK of these groups is consistent with that of primary amines. For example, the pK of the primary amine group of glycine methyl ester is approximately 7.7, while that of the side chain amino group of lysine is approximately 9 [13]. Indeed, the potentiometric titration of wild type yeast iso-1-cytochrome *c* shows approximately five groups titrating in the pH range of 5.0 - 9.0 [16], only two of which correspond to surface histidines (positions 33 and 39, which are not present in the variants used in this study). The remaining three titrating groups are likely to consist of the amino terminus and/or some subset of the ~15 surface lysines of the protein. The equilibrium binding constants of primary amines for Cu^{2+} are comparable to those of imidazole (e.g., $10^{3.8}$ M^{-1} for glycine methyl ester versus $10^{4.2}$ M^{-1} for N-acetylhistidine) [13]. It is known that the amine group is required for tryptamine adsorption to IMAC supports [10]. Thus the additional metal-coordinating groups demonstrated by the H_{26} and H(-) variants are likely to be primary amines.

Over the entire range of measurable chromatographic retention ($1 < k' < 200$), the logarithm of the capacity factor of both the H_{26} and H(-) variants remains linear in pH. Under these conditions, it is not possible to independently determine any of the three parameters making up the pH-independent term of eqn. 8 (k'_o, K_2, pK_z). However,

the only difference between these two variants is a single, partially accessible histidine at position 26 (see Chapter 3 [7]). The result of this substitution is to increase protein retention by approximately a factor of 50 over the pH range 7.5 to 7.7 (the limited range over which chromatography of both proteins can be measured). In the equilibrium model developed above (eqns. 5 - 8), such a substitution would affect only the first equilibrium binding constant K_1. Presumably, the increase in retention of the H_{26} variant therefore reflects a 50-fold increase in the binding affinity of this single, partially accessible histidine over the background of surface amine groups.

In contrast to the H_{26} and H(-) variants, the chromatographic retention of the H_8 variant shows two regimes in the pH titration curve corresponding to acidic and basic conditions (Figure 4.3). Under acidic conditions (pH 5 - 6), the increase in the logarithm of the capacity factor with pH has a slope very nearly equal to 1. This pH dependence is entirely consistent with protein adsorption by the single fully accessible histidine at position 8, as illustrated in Figure 4.4A. In fact, as shown in Figure 4.3 (dotted line), protein retention in this regime can be described by the single-site model of eqn. 3 using the previously determined pK of 5.7 [8]. As shown in Table 4.3, the predicted capacity factor for protein adsorption via the single surface histidine ($k'_0 = 10$) is a factor of four smaller than that measured for acHIS (Table 4.2), which reflects a decrease in the product $K_1 Q_{max}$ (eqn. 5). This decrease may indicate a decrease in limiting capacity of the large protein relative to acHIS, a decrease in the equilibrium binding constant of copper for a protein imidazole relative to an acHIS imidazole, or an *increase* in the energy penalty of immobilizing the large protein molecule on the IMAC surface relative to the small acHIS molecule. It is not possible to evaluate the relative contribution of these factors from the chromatographic experiments alone.

At neutral to slightly basic conditions (pH 6.5 - 7.5), the logarithm of the capacity factor increases much more rapidly with pH, approaching a slope of 3 for pH ~ 7.2. In

this regime, the single-histidine protein appears to bind the IMAC support via multiple copper sites, as illustrated in Figure 4.4B. This slope of the curve is consistent with the deprotonation of approximately three amine groups of pK > 8 displayed by the H_{26} and H(-) variants at slightly higher pH. In fact, the pH dependence of the H_8 variant over the entire pH range studied (5.2 to 7.2) is accurately portrayed by the multiple-site adsorption model considering the contribution of these three groups in addition to the lone surface histidine (eqn. 8). The binding affinity predicted for a single deprotonated amine group (K_2 = 22, Table 4.3), presuming an amine pK of 8.5 which corresponds to the basic transition of cytochrome c [16], is in excellent agreement with that previously determined for a single histidine imidazole in the two-site binding model of Chapter 3 (K_2 = 22 - 27) [7].

Although the amine groups clearly play a subsidiary role to a single imidazole under slightly acidic conditions (comparing solid line to dotted line of Figure 4.3), together they enhance the adsorption of the H_8 variant at moderate pH by more than an order of magnitude (e.g., pH 7.0 - 7.3). The contribution of additional surface groups explains the previously observed result that imidazole is a more effective inhibitor of H_8 chromatographic retention than predicted by its binding affinity for CuIDA-TSK [8]. If instead approximately four groups (e.g., histidine 8 and three primary amines) act in concert to display the observed binding affinity at neutral to slightly basic conditions, then the imidazole inhibition will be of higher order, analogous to the higher-order dependence on proton concentration observed in the pH dependence.

Implications to protein recognition

The binding mode of a single-histidine protein in IMAC can be modulated simply by changing the pH. As the pH is increased, the chromatographic retention of the H_8 variant changes from that typical of single-site binding under acidic conditions (Figure

4.4A) to that typical of multivalent binding under basic conditions (Figure 4.4B). As demonstrated with metal ligand loading in Chapter 3, control over the binding mode is a powerful means to fine-tune IMAC separations to the properties of individual proteins. Under acidic conditions, metal coordination would be limited exclusively to surface histidines (Figure 4.4A). In this regime, IMAC displays sequence specificity analogous to biological affinity separations. In contrast, under basic conditions, any of a large number deprotonated nitrogen side chains (e.g., histidines and lysine, Figure 4.4B) can contribute to protein adsorption via metal coordination.

Protein recognition by multivalent binding provides the opportunity to design materials to recognize specific patterns of metal-coordinating residues on the protein surface, primarily surface histidines at neutral pH, with complementary arrangements of metal ions. The secondary role played by surface lysines provides a mechanism to generalize protein recognition based on metal-to-ligand coordination to a much wider array of proteins, particularly at elevated pH. However, with such a gain of generality comes a loss of specificity: lysine is a much more common amino acid (~ 10% of globular proteins). In fact, the adsorption characteristics of cytochrome *c* at elevated pH in IMAC (Figure 4.3) are similar to those displayed by charged proteins in ion exchange chromatography [15]. However, even under these conditions IMAC retains two important advantages over ion exchange chromatography: the much higher binding constant of individual metal-ligand interactions relative to electrostatic interactions, and the ability to perform the separation at the high salt concentrations.

Finally, it must be stressed that chromatographic experiments focus upon protein binding at very low coverage, analogous to the initial slope of the adsorption isotherm. Protein adsorption in this regime will be dominated by those surface sites which optimize the number of interactions between the protein and the chromatographic support. As demonstrated in Chapter 2, it is possible to target individual molecules by matching the

distribution of metal ions to a spatial distribution of metal-coordinating ligands [17,18]. The additional contribution of a common amino acid such as lysine may therefore provide mechanism to distinguish among single-histidine proteins by the secondary effects of additional metal-coordinating groups. For example, template polymerization [12] at higher pH would correctly position the metal ions to complement the arrangement of protein surface imidazoles and amines. In this case, specific high-affinity binding sites would be created which match the arrangement of metal ligands on the surface to the distribution of histidines *and lysines* on a target protein. The ensuing protein recognition could be performed at a lower pH, making surface histidines a prerequisite to adsorption, but also placing a premium on those proteins with appropriate patterns of surface lysines.

CONCLUSIONS

At the low coverages typical of chromatographic separations, protein adsorption will be limited to those surface sites which accommodate simultaneous binding of multiple protein metal-coordinating groups. One consequence of this is a significant contribution of additional metal-coordinating residues such as lysines to protein binding, although surface histidines remain a prerequisite to binding at neutral pH. The wide range of pK's spanned by imidazoles and amines provides the opportunity to use pH to more specifically control the number of protein groups binding the IMAC support. Thus, as shown with the H_8 cytochrome *c* variant, adsorption of single-histidine proteins can be modulated from single-site binding under acidic conditions to multivalent binding under basic conditions. Such control over the number of the protein surface interactions, coupled with the relatively high binding affinity of a single protein-metal interaction,

uniquely positions IMAC at the interface between biological affinity separations and ion exchange chromatography.

EXPERIMENTAL METHOD

Variants of *Saccharomyces cerevisiae* iso-1-cytochrome *c* containing different distributions of histidine residues were expressed in yeast and purified as described previously [8,19] (see also Chapter 3). In each of the cytochrome *c* variants, the lone cysteine at position 102 was replaced by a serine to prevent oxidative dimerization at the surface sulfhydryl group [20]. Additional surface amino acids were replaced to alter surface histidine content as described previously [7]. The structural integrity of each mutant was confirmed by UV-visible spectroscopy, and the histidine content was confirmed by ^{1}H NMR spectroscopy. Expression of these variants in a strain of yeast lacking cytochrome *c* guarantees that each is biologically functional, which in turn ensures that a conformation very similar to the native one has been maintained.

Isocratic chromatography

A TSK Chelate-G6000XL (10 μm macroporous beads) HPLC column (7.5 cm × 7.5 mm ID, total volume 3.31 ml, column void volume 2.3 ml) from Tosohaas (#80645) was used in all experiments. Experiments were performed at a flow rate of 1.0 ml/min as described previously [8]. The column was regenerated by first washing with 15 ml of 50 mM EDTA pH 8.0, and then reloading with 15 ml of 50 mM $CuSO_4$ in water and finally washing with 15 ml of 0.1 M sodium acetate pH 4.0.

All isocratic elution experiments were carried out at maximum copper loading. The TSK Chelate column was regenerated as described above and then equilibrated with 10 ml of elution buffer (50 mM sodium acetate, 50 mM sodium phosphate, 50 mM boric

acid, 0.5 M NaCl), with pH adjusted in the range 6.0 - 9.0. For isocratic elution experiments with *N*-acetylhistidine (acHIS), 100 µl of 1 mM acHIS dissolved in elution buffer was injected and the eluent monitored by absorbance at 235 nm. For zonal isocratic elution experiments with cytochromes *c*, 25 µl of 3 µM cytochrome *c* dissolved in elution buffer was injected and the eluent monitored by absorbance at 410 nm. The elution time reported is that of the absorbance peak maximum.

To quantify copper loading, the column was washed with 15 ml of 50 mM EDTA pH 8.0. The effluent was pooled and the concentration of CuEDTA determined by absorbance at 800 nm (ε_{800} = 73.6 cm^{-1} M^{-1}). The copper loading was found to be 81 µmol of copper (24.5 µmol per ml of column) in agreement with values previously determined [8].

All nonlinear regression calculations were performed on SigmaPlot (v. 6.0, Jandel Scientific).

Table 4.1. Comparison of chromatographic retention with equilibrium adsorption. Chromatographic capacity factors (k') of *N*-acetylhistidine (acHIS) and *S. cerevisiae* iso-1-cytochrome *c* surface histidine variants (H_8, H_{26}, H(-)) measured under isocratic conditions for CuIDA-TSK HPLC column (24.5 μmol/ml TSK) at a flow rate of 1.0 ml/min, 298 K, pH 7.0 (50 mM sodium acetate, 50 mM sodium phosphate, 50 mM boric acid, 0.5 M NaCl). Equilibrium binding experiments (Chapter 3 [7]) measured for CuIDA-TSK Guardgel material (18.6 μmol/ml TSK) at 298 K, pH 7.0 (50 mM sodium phosphate, 0.5 M NaCl). Initial slope of adsorption isotherms (KQ_{max}) are scaled for increased copper loading of HPLC column.

	k' (ml/ml TSK)	KQ_{max} (ml/ml TSK)
acHIS	21	20.9
H_8	52	66
H_{26}	2.3	5.3
H(-)	<<1	0.3

Table 4.2. Imidazole protonation in acHIS chromatography on CuIDA-TSK. Chromatographic retention of *N*-acetylhistidine measured for CuIDA-TSK HPLC column (24.5 μmol/ml TSK) under isocratic conditions at a flow rate of 1.0 ml/min, 298 K, pH 6.0 - 9.0 (50 mM sodium acetate, 50 mM sodium phosphate, 50 mM boric acid, 0.5 M NaCl). Protonation parameters (with standard error) determined by nonlinear regression of observed capacity factors to equilibrium proton competition (eqn. 3).

pK	k_o' (ml/ml TSK)
6.98 (0.07)	40.4 (1.8)

Table 4.3. Protonation equilibria in cytochrome *c* chromatography on CuIDA-TSK. Chromatographic retention of *S. cerevisiae* iso-1-cytochrome *c* variants measured for CuIDA-TSK HPLC column (24.5 μmol/ml TSK) under isocratic conditions at a flow rate of 1.0 ml/min, 298 K, pH 6.2 - 8.25 (50 mM sodium acetate, 50 mM sodium phosphate, 50 mM boric acid, 0.5 M NaCl). Protonation parameters for H_{26} and H(-) variants determined by linear regression of logarithm of capacity factors to eqn. 8. Protonation parameters for H_8 variant (K and k'_o) determined by nonlinear regression of capacity factors (including data of Todd [8]) to eqn. 9, using z determined for H(-) variant, pK_1 for histidine 8 previously determined by NMR [8], and pK_z for additional (amine) groups determined for basic transition of wild type iso-1-cytochrome *c* [16].

	H_8	H_{26}	H(-)
z	3.4	2.7	3.4
pK_z	8.5	--	--
K_2	22	--	--
k'_o	10 ml/ml TSK	--	--
pK_1	5.7	--	--

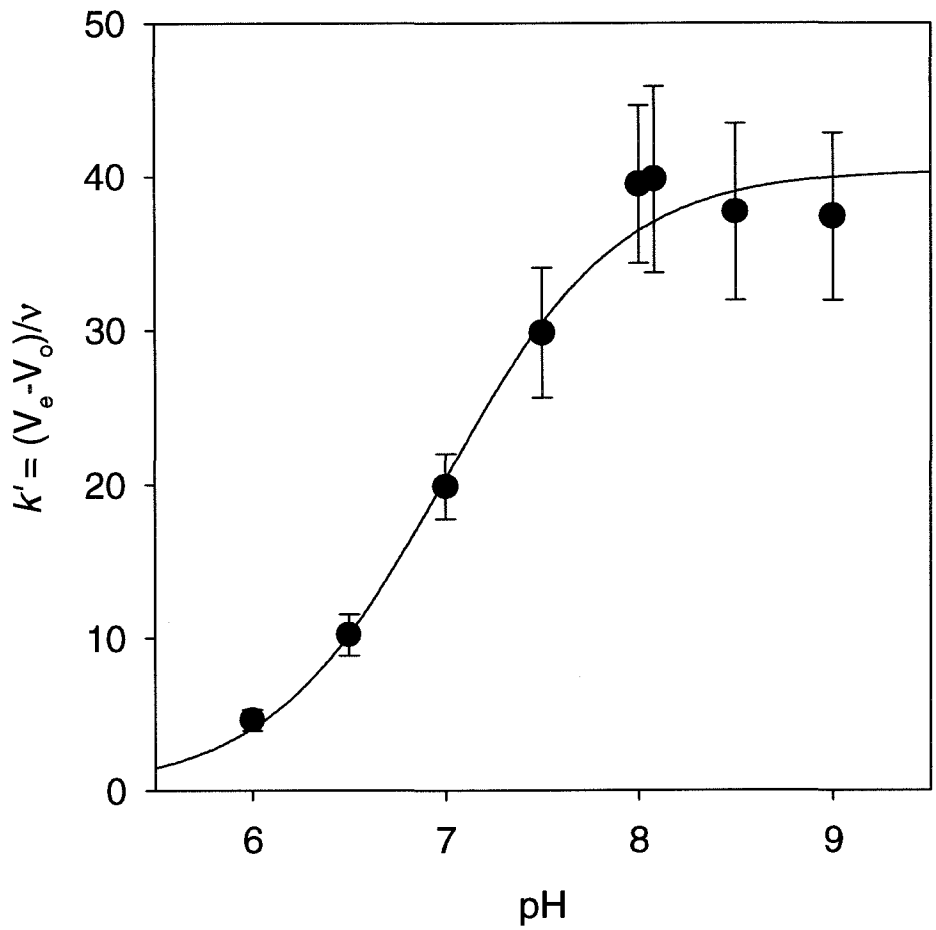


Figure 4.1. Effect of pH on isocratic zonal elution of acHIS in IMAC. The capacity factor (*k'*) of *N*-acetylhistidine for CuIDA-TSK HPLC column was measued under isocratic conditions. Error bars represent peak width at half-maximum absorbance. Curve represents best fit to equilibrium protonation of acHIS imidazole (eqn. 4) calculated from parameters of Table 4.2.

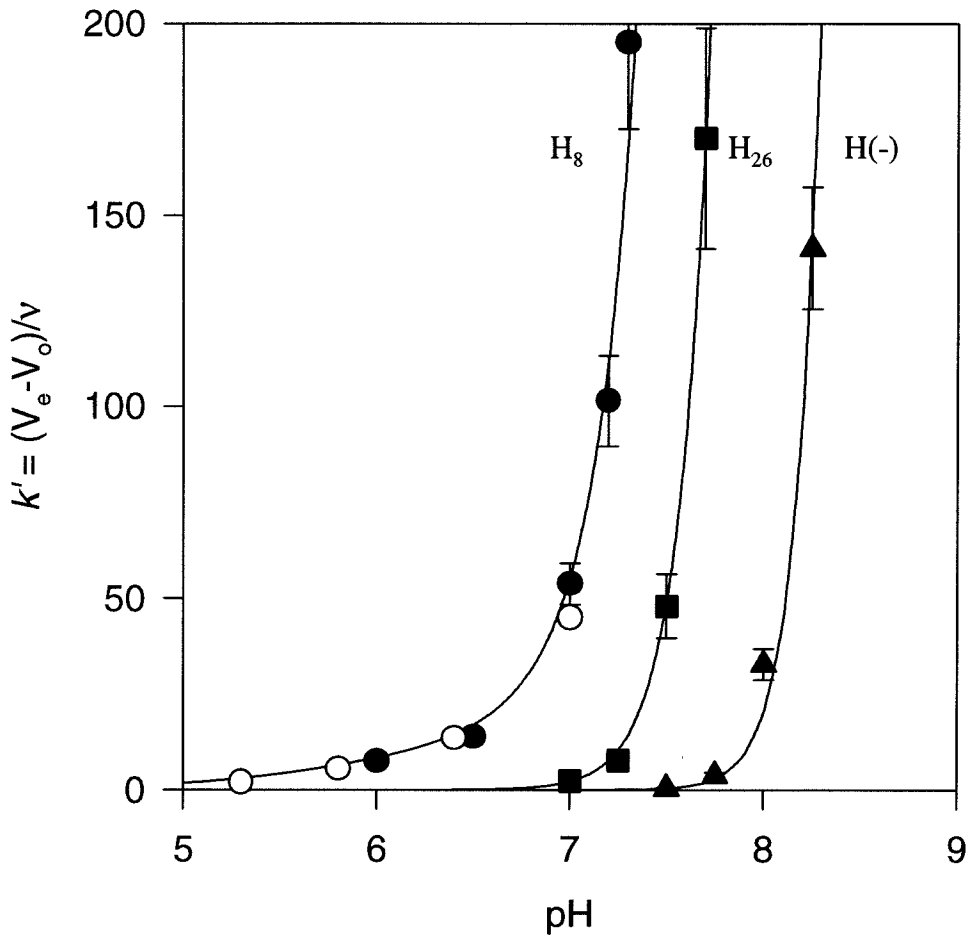


Figure 4.2. Effect of pH on isocratic zonal elution of cytochrome *c* in IMAC. The capacity factor (*k'*) for CuIDA-TSK HPLC column is measured at constant pH for *S. cerevisiae* iso-1-cytochrome *c* histidine variants (○) H_8 (adapted from Todd [8]), (●) H_8, (■) H_{26}, and (▲) H(-). Labels indicate surface histidines. Error bars represent peak width at half maximum absorbance. Curves represent best fit to multiple proton competition (see Figure 4.3).

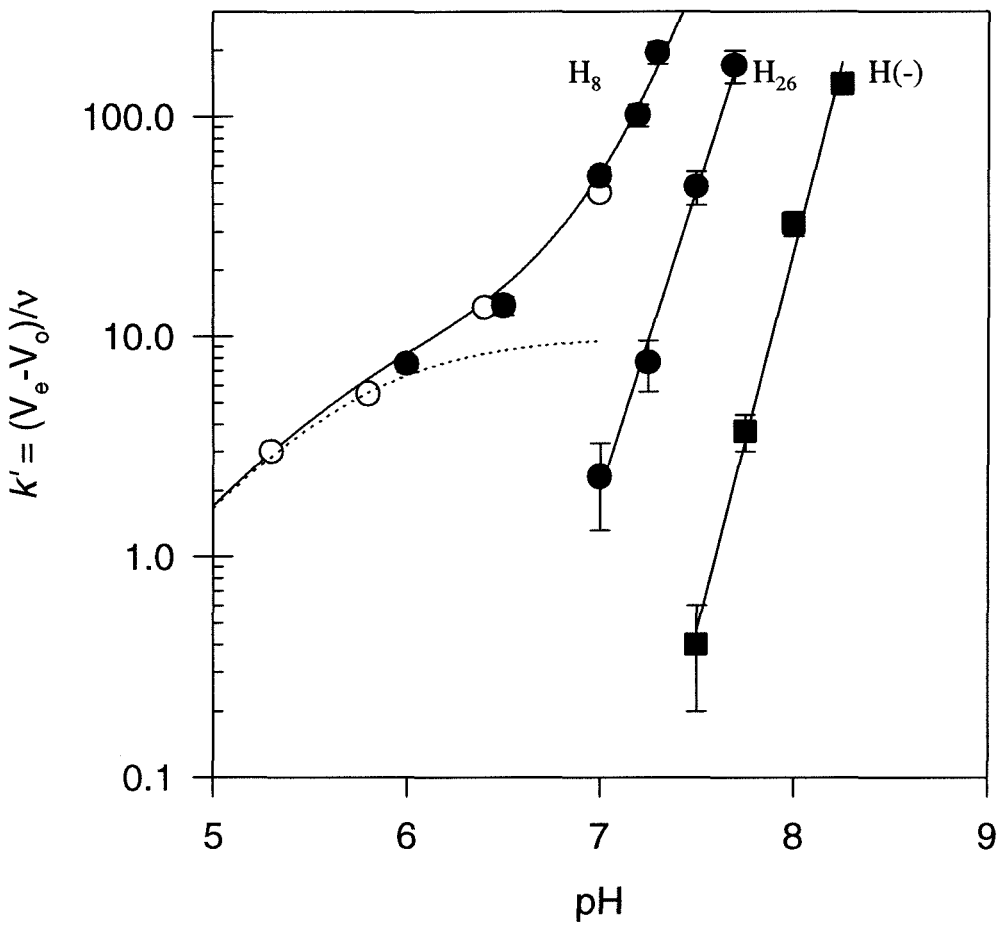


Figure 4.3. Effect of pH on logarithm of cytochrome *c* capacity factor in IMAC. The capacity factor (*k'*) is measured as described in Figure 4.2 for cytochrome *c* histidine variants (○) H_8 (adapted from Todd [8]), (●) H_8, (■) H_{26}, and (▲) H(-). Labels indicate surface histidines. Error bars represent peak width at half maximum absorbance. Solid lines are best fit to multiple proton competition of eqn. 8 (H_{26} and H(-)) and eqn. 7 (H_8) calculated from parameters of Table 4.3. Dotted line extrapolates H_8 data as predicted by single imidazole proton competition (eqn. 4).

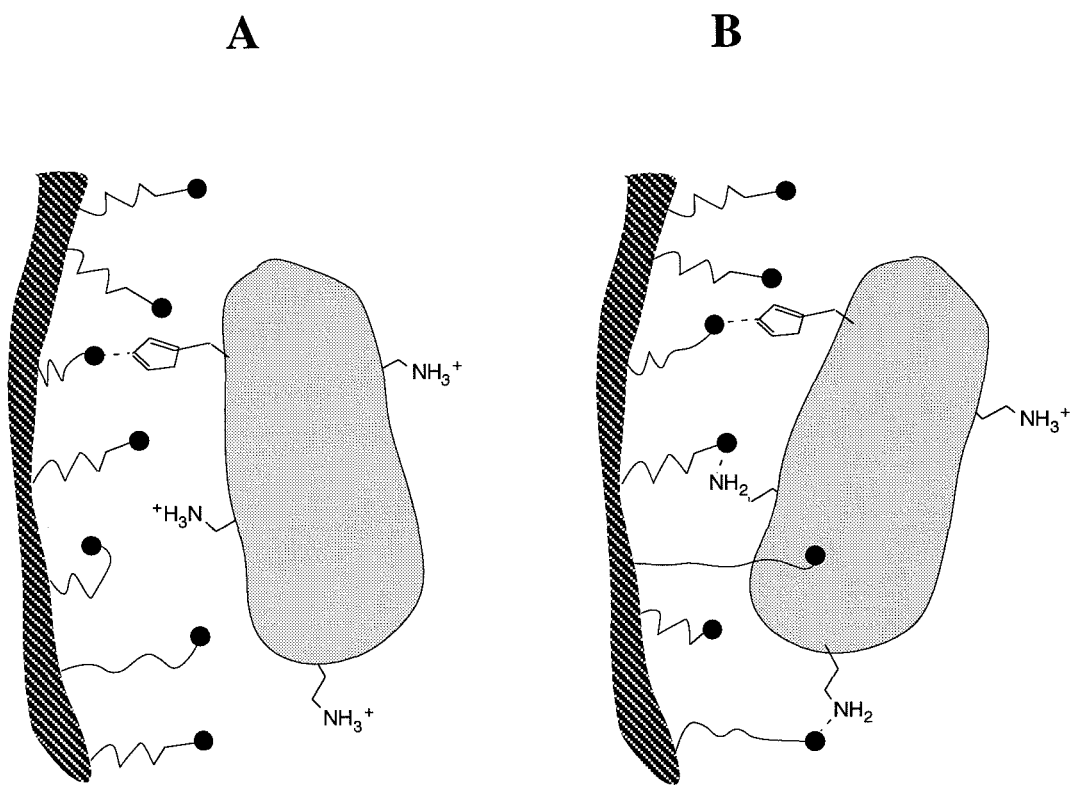


Figure 4.4. Proposed binding interactions in IMAC at elevated pH. **(A)** At low to moderate pH (pH < 7), single-histidine protein binds IMAC surface at a single copper site. Other surface coordinating groups are fully protonated. **(B)** At elevated pH (pH > 7), single-histidine protein exhibits multivalent binding to IMAC surface. Multiple deprotonated surface groups simultaneously coordinate individual copper sites.

REFERENCES

1. Arnold, F. H. (1991) *Bio/Technology 9*, 151-156.
2. Wong, J. W., Albright, R. L. and Wang, N. L. (1991) *Sep. Pur. Meth. 20*, 49-106.
3. Hemdan, E. S., Zhao, Y., Sulkowski, E. and Porath, J. (1989) *Proc. Natl. Acad. Sci. USA 86*, 1811-1815.
4. Mrabet, N. T. (1992) *Biochemistry 31*, 2690-2702.
5. Zhao, Y., Sulkowski, E. and Porath, J. (1991) *Eur. J. Biochem. 202,* 1115-1119.
6. Abelson, J. N., Simon, M. I., and Arnold, F. H., Editors (1992) *Methods: A Companion to Methods in Enzymology 4*, 1-108.
7. Todd, R. J., Johnson, R. D. and Arnold, F. H. (1994) *J. Chromatogr. A 662*, 13-26.
8. Todd, R. J. (1993) *Ph.D. Thesis*, California Institute of Technology, Pasadena, CA.
9. Arnold, F. H., Schofield, S. A. and Blanch, H. W. (1986) *J. Chromatogr. 355*, 1-12.
10. Hansen, P. and Lindeberg, G. (1994) *J. Chromatogr. A 662*, 235-241.
11. Yip, T., Nakagawa, Y. and Porath, J. (1989) *Anal. Biochem. 183*, 159-171.
12. Mallik, S., Plunkett, S., Dhal, P. K., Johnson, R. D., Pack, D., Shnek, D. and Arnold, F. H. (1994) *New J. Chem. 18*, 299-304.
13. Martell, A. E. and Smith, R. M. (1975) *Critical Stability Constants*, Vol. 1-6, Plenum Press, New York.
14. Hill, A. V. (1910) *J. Physiol. (London) 40*, iv.
15. Kopaciewicz, W., Rounds, M. A., Fausnaugh, J. and Regnier, F. E. (1983) *J. Chromatogr. 266*, 3-21.
16. Barker, P. L., Mauk, M. R. and Mauk, A. G. (1991) *Biochemistry 30*, 2377-2383.

17. Mallik, S., Johnson, R. D. and Arnold, F. H. (1993) *J. Am. Chem. Soc. 115*, 2518-2520.

18. Mallik, S., Johnson, R. D. and Arnold, F. H. (1994) *J. Am. Chem. Soc. 116*, 8902-8911.

19. Todd, R. J., Van Dam, M., Casimiro, D., Haymore, B. L. and Arnold, F. H. (1991) *Proteins 10*, 156-161.

20. Mayo, S. L., Ellis, W. R., Crutchley, R. J. and Gray H. B. (1986) *Science 233*, 948-952.

CHAPTER 5

MULTIVALENT BINDING III: MODELING BINDING HETEROGENEITY IN PROTEIN-SURFACE INTERACTIONS

INTRODUCTION

The adsorption of proteins to functional surfaces is often described by the Langmuir isotherm [1,2], originally derived for gas adsorption to a homogeneous surface. According to this model, adsorption is characterized by a single binding energy and a maximum adsorption capacity corresponding to monolayer surface coverage. Several researchers have noted that the Langmuir model does not adequately describe the heterogeneity of protein adsorption to materials commonly used in ion-exchange and affinity chromatography [3,4,5]. We have observed heterogeneous binding in the adsorption of cytochrome *c* variants to a metal-affinity chromatography support; these variants differed only in the number and accessibility of surface histidines ([6], see also Chapter 3). In the course of these studies, we noted that the extrapolated maximum capacity appeared to be correlated with the binding affinities of the cytochrome *c* variants. In addition, a correlation between limiting capacity and binding strength has been observed for adsorption of *the same protein* to ion exchange [4] and affinity [7] chromatography supports when protein adsorption is inhibited by soluble competitors.

The Langmuir model predicts a single maximum binding capacity for a given protein in the absence of significant conformational changes. This model therefore cannot explain the observed increase in limiting capacity with increasing surface histidine content for what are otherwise identical proteins. The capacity of these materials to bind protein is ultimately dictated by the degree of functionalization (i.e., the density of ligands) and steric factors (i.e., the size of the protein) and should not depend on binding affinity or competitor concentration. The current investigation was motivated by the presumption that a given protein, or a series of its variants, should always reach

the same limiting coverage at sufficiently high concentrations during reversible adsorption to a functional surface [8].

Binding heterogeneity is a recurrent theme among chromatographic systems in which adsorption involves multiple interactions between the protein and surface. For example, in affinity chromatography the adsorption of oligomeric proteins containing more than one ligand-binding site can result in extremely strong binding for those proteins adsorbed via two or more affinity ligands, compared to weaker binding for proteins adsorbed by only one affinity ligand [9,10,11]. Similarly, binding heterogeneity in metal affinity chromatography can arise from multiple interactions between metal-coordinating groups on a monomeric protein and complementary metal sites distributed on a seemingly homogeneous surface, as illustrated in Figure 5.1. On a densely-derivatized surface, a random distribution of functional sites or affinity ligands can be viewed as a set of individual "binding sites," each supporting a different number of specific protein-ligand interactions. A protein shows the highest affinity for the surface arrangements which best match its own pattern of functional groups, resulting in a distribution of binding energies. Depending on the strength of the individual interaction (a charge-charge interaction in ion-exchange chromatography, a metal-histidine coordination interaction in metal-affinity chromatography, or a specific ligand-binding interaction in affinity chromatography) and the number of interactions that can form simultaneously, adsorption can occur with a wide range of binding affinities. The range and distribution of binding energies should depend strongly on the density and distribution of functional groups, both on the protein and on the surface.

The Langmuir isotherm was originally derived for gas adsorption to a homogeneous surface. We have found that reversible protein adsorption to metal affinity and other chromatography supports is described well by another model also widely used in studies of gas adsorption, the Temkin isotherm. The Temkin isotherm

helps elucidate the effects that multivalent interactions have on the adsorption of proteins with multiple surface histidines to the metal-complexing surface. With this model, we show that changing the number of target groups on a protein (surface histidines) is equivalent to changing the density of binding sites on the surface (immobilized metal ions). This result underscores the role of multivalent interactions in protein adsorption and has important implications for the design of materials for separations as well as the interpretation of biological recognition phenomena.

Binding heterogeneity commonly observed in biological receptor systems is often described in terms of a probability distribution in binding energies. A proposed general model for binding energy distributions in biological systems is a bell-shaped curve with a maximum probability occurring at a relatively low binding energy [12]. Gaussian and Poisson energy distributions have been used to analyze antigen binding to polyclonal antibodies [13,14]. Bi-modal Langmuir adsorption has been used to interpret the adsorption of interleukins to cellular interleukin receptors in heterogeneous receptor populations [15,16]. The form of the binding energy distribution dictates predictions of the repertoire size, detection threshold, and population polymorphisms [13]. It is interesting to ask whether a model that describes heterogeneous protein adsorption to functionalized surfaces can also capture the relevant features of heterogeneous binding in such diverse biological systems. We therefore have compared the Temkin model to another, widely-used model for heterogeneous binding, the Hill model (or its two-parameter analog, the Freundlich model). This comparison illuminates a fundamental difference in the shape of energy distributions appropriate for functionalized surfaces versus those appropriate for biological receptors, a result which has repercussions in the design of synthetic materials to mimic biological systems.

CALCULATIONS

Langmuir isotherm

Under the assumption of local Langmuir adsorption, each individual surface "binding site" (illustrated in Figure 5.1C) independently obeys Langmuir adsorption with a particular equilibrium binding constant [17]. Binding heterogeneity is presumed to result from a distribution in binding constants over the ensemble of binding sites. The bi-Langmuir model used previously to describe protein binding to metal-affinity chromatographic supports [6] (see also Chapter 3) can be seen to be a special case of local Langmuir adsorption. Net protein adsorption Q (mol/ml material) is described by integrating over the distribution of binding constants,

$$Q(c) = Q_{\max} \int \theta(K) \frac{Kc}{1+Kc} dK \quad , \tag{1}$$

where θ(K) is the fraction of binding sites with binding constant K (M^{-1}), c is the protein solution concentration (M) and Q_{max} (mol/ml material) is the limiting protein adsorption capacity. The probability distribution over binding sites is more conveniently expressed as an integral distribution over the binding energies [17],

$$Q(c) = Q_{\max} \int \theta(x) \left[1+(c_o/c)e^{-x}\right]^{-1} dx \quad , \tag{2}$$

where $x = -\Delta G/RT = \ln Kc_o$, and c_o defines the solution concentration of unit activity (usually taken to be 1 M). For a finite number of adsorption sites, the probability distribution must integrate to unity,

$$1 = \int \theta(x) dx \quad . \tag{3}$$

This constraint on the distribution function has two important physical consequences for the adsorption isotherm: (1) adsorption approaches zero coverage at low solute concentrations, and (2) adsorption approaches a limiting coverage at high solute concentrations.

Temkin isotherm

The Temkin isotherm describes adsorption as a logarithmic function of the solute concentration. This isotherm appears as a straight line on a semi-logarithmic plot and corresponds to a uniform distribution in binding energies [17]. The Temkin isotherm can be generalized to low and high solute concentrations by limiting the distribution to a finite energy range [18],

$$\theta_T(x) = \begin{cases} 0, & x \le x_o \\ 1/(x_T - x_o), & x_o \le x \le x_T \\ 0, & x_T \le x \end{cases} , \tag{4}$$

where x_o is the minimum binding energy and x_T is the maximum binding energy. Applying the assumption of local Langmuir adsorption results in the following isotherm equation

$$Q(c) = \frac{Q_{\max}}{a} \ln\left(\frac{1 + K_T c}{1 + e^{-a} K_T c}\right) , \tag{5}$$

where K_T is the equilibrium binding constant corresponding to the maximum binding energy ($K_T = \exp(x_T)/c_0$), and a represents the range of binding energies ($a = x_T - x_0$).

It is often the case that adsorption isotherms cannot be measured accurately near the limiting coverage, that is, at relatively high solution concentrations. At low to moderate solute concentrations ($K_T c << e^a$), the denominator of the logarithm of eqn. 5 can be taken to be unity, giving

$$Q(c) = q_T \ln(1 + K_T c) \quad . \tag{6}$$

In this concentration regime, adsorption has only two undetermined parameters (q_T and K_T). The limiting capacity for solute adsorption (Q_{max}) and the range of binding energies (a) cannot be determined independently under these conditions. However, the two parameters q_T and K_T are related by the underlying parameters Q_{max} and a, as illustrated in the Appendix.

Hill isotherm

Heterogeneous adsorption often can be described by a power-law function of the solute concentration [17]. This isotherm appears as a straight line on a log-log plot and corresponds to an exponential distribution of binding energies. This form can be generalized to satisfy a limiting coverage at high solute concentrations using the following energy distribution [19],

$$\theta_H(x) = \frac{\sin \alpha\pi}{\pi} \cdot \frac{e^{-\alpha(x-x_o)}}{1 + 2e^{-\alpha(x-x_o)} \cos\alpha\pi + e^{-2\alpha(x-x_o)}} \quad , \tag{7}$$

where x_0 is the average binding energy and α defines the breadth of the energy distribution ($0 < \alpha < 1$). This function approximates a normal distribution of mean x_0 and standard deviation $1/\alpha$. Applying the assumption of local Langmuir adsorption results in the following isotherm,

$$Q(c) = Q_{\max} \frac{(K_o c)^{\alpha}}{1+(K_o c)^{\alpha}} \quad , \tag{8}$$

where K_0 is the binding constant corresponding to the mean binding energy ($K_0 = \exp(x_0)/c_0$). This isotherm is analogous to the Hill model derived for cooperative binding interactions ($\alpha > 1$) [20]. When the parameter $\alpha = 1$, this expression collapses to the Langmuir isotherm. In the limit of low solute concentrations ($K_0 c < 1$), this expression collapses to the Freundlich model,

$$Q(c) = b \cdot c^{\alpha}, \tag{9}$$

and, as such, is often referred to as the Langmuir-Freundlich isotherm [21]. At these concentrations the solute adsorbs to only the highest affinity binding sites ($x >> x_0$), and the energy distribution of eqn. 7 can be approximated by an exponential distribution,

$$\theta_F(x) = \frac{\sin \alpha\pi}{\pi} \cdot b \cdot e^{-\alpha x}. \tag{10}$$

In this concentration regime, adsorption again has two undetermined parameters (α and b). Under these conditions the limiting capacity Q_{max} and representative binding constant K_0 cannot be determined independently. However, as seen with the Temkin

isotherm, the two parameters α and b are related by the underlying parameters Q_{max} and K_o, as illustrated in the Appendix.

RESULTS AND DISCUSSION

Heterogeneity in metal-affinity adsorption

Equilibrium isotherms have been measured previously for a series of yeast cytochrome *c* surface histidine variants adsorbing to a copper-chelating support used for metal affinity chromatography (Cu^{2+}-IDA-TSK) [6]. Adsorption of those proteins containing a single surface histidine could be described by the Langmuir isotherm. A dramatic increase in the apparent binding affinity was observed for proteins with more than one accessible histidine. Binding heterogeneity was apparent from Scatchard plots of the isotherm data for the multiple-histidine variants; these data were analyzed using a bi-Langmuir model with adsorption to two populations of surface "binding sites." One population of (weak) sites was presumed to involve only single histidine-copper interactions, while the second population of (strong) sites apparently allowed two (or more) simultaneous metal-protein interactions [6] (see also Chapter 3).

Increasing surface histidine content also increased maximum binding capacity, according to fits to the data using the bi-modal Langmuir isotherm [6]. Even though the individual experimental isotherms could all be fit to a bi-Langmuir isotherm, changes in maximum binding capacity with surface histidine content for otherwise identical proteins could not be explained. As shown in Figure 5.2, semi-logarithmic plots of the data demonstrate that the Temkin model can also fit the individual isotherms. These data are not adequately fit by the Hill model (data not shown). The interesting fact here, however, is that the Temkin parameter q_T (slope of the plot in Figure 5.2) is approximately constant (~ 0.21 μmol/ml TSK) for all ten of the cytochrome *c* variants,

despite the fact that the maximum binding affinity K_T varies over five orders of magnitude, as shown in Table 5.1.

Adsorption isotherms of horse cytochrome *c* to the chelating support prepared with different densities of the copper have also been measured previously [6]. These adsorption isotherms are also described well by the Temkin model, *with the same value of the parameter* q_T, as shown in Table 5.2. When the concentration of each protein solution (c) is scaled by the equilibrium constant related to that protein's maximum binding energy (K_T), the isotherms of the ten yeast cytochrome *c* histidine variants, tuna cytochrome *c*, and horse cytochrome *c* at nine different copper loadings *all collapse to a single curve*, as shown in Figure 5.3. This result indicates a fundamental relationship between the target groups of the protein (surface histidines) and the density of binding sites on the surface (metal ions).

It is difficult or impossible to achieve the (>> mM) protein concentrations necessary to distinguish the apparent saturation of the two-parameter isotherm (eqn. 6) from the true saturation of the generalized isotherm (eqn. 5). Although the underlying parameters Q_{max} and x_o would validate the suggestion of consistent limiting coverage, it is difficult to determine these values independently from individual protein adsorption experiments. If, however, the cytochrome *c* variants are *presumed* to share the same limiting capacity (Q_{max}) and minimum binding energy (x_o), then q_T is predicted to depend weakly on the maximum binding constant K_T. There is, in fact, a trend for smaller q_T values at the larger binding affinities (see Tables 5.1 and 5.2). From this trend, consensus values of Q_{max} (3.5 μmol/ml TSK) and x_o (0) can be obtained. The consensus limiting capacity Q_{max} is physically realistic and consistent with our estimate of the cytochrome *c* monolayer coverage (3 - 4 μmol/ml TSK) based on the protein's crystal structure dimensions [22, 23].

All twenty cytochrome *c* adsorption isotherms now can be fit to the three-parameter Temkin model (eqn. 5), using common values of Q_{max} and x_o, and leaving only the maximum binding constant K_T as a free parameter. The resulting values of K_T are shown in Figure 5.4. As predicted by the Temkin model, the differences in equilibrium binding behavior in two disparate experiments, adsorption of yeast cytochrome *c* variants to a surface of constant metal ion density and adsorption of horse cytochrome *c* to a surface of varying metal ion density, are described completely by differences in binding energies.

The Temkin model provides a practical measure of the affinity displayed in a heterogeneous system. It has been observed that protein retention in metal-affinity chromatography depends strongly on the number of surface histidines and on the density of immobilized metal ion [6,24,25]. The molecular basis of this result becomes more apparent upon comparing the effect of adding histidines to the protein (Figure 5.4A) to that of adding copper ions to the metal-affinity chromatography support (Figure 5.4B). According to the Temkin model, adding a single histidine increases the maximum binding energy of yeast cytochrome *c* by ~ 4 kcal/mol (roughly a factor of 60 increase in the maximum binding constant). The same increase in binding can also be achieved by increasing the concentration of surface copper ions, from 10 to 18 μmol/ml TSK (for horse cytochrome c) [26]. In both cases, adsorption involves more copper-histidine coordination bonds between protein and surface, increasing apparent binding affinity. This functional equivalence between histidines on the protein and metal ions on the adsorbent is to be expected when the binding energy reflects the number of metal-ligand interactions formed between the protein and surface "binding sites" of multiple metal ions.

Heterogeneous protein binding to other chromatography supports

Binding heterogeneity, as demonstrated by nonlinear Scatchard plots, has been previously reported for other chromatographic systems, including ion exchange [27] and affinity adsorption to immobilized antibodies [14]. In both cases, protein adsorption was originally described using the Hill model (see Appendix). These data can also be adequately described by the Temkin model.

Adsorption isotherms for rat cytochrome b_5 to Mono Q anion exchange resin have been measured in the presence of equilibrium competitor (0.100-0.175 M NaCl) [27]. These isotherms were analyzed by Gill *et al.* [27] using the Hill model with an undetermined, but large, value of the limiting capacity Q_{max}; addition of NaCl increased the Hill coefficient α toward unity (more homogeneous binding). Their adsorption data can also be described by the Temkin isotherm, in which case addition of NaCl decreases the maximum (apparent) binding constant K_T and increases the Temkin parameter q_T, as shown in Table 5.3. Because these binding data do not reach saturation, it is not possible to determine consensus values of the limiting capacity Q_{max} for either the Hill or Temkin models. However, if the limiting capacity for rat cytochrome b_5 is presumed constant, then the Temkin parameter q_T is similar to the Hill coefficient α in that both parameters increase with decreasing binding heterogeneity. Indeed, both the Hill coefficient α and the Temkin parameter q_T increase by a factor of three as NaCl concentration is increased from 0.100 to 0.175 M (Table 5.3, see also Appendix).

The equilibrium adsorption of bovine serum albumin (BSA) to immobilized anti-BSA antibodies has been reported by Sada *et al.* [14]. Adsorption to immobilized monoclonal (homogeneous) antibodies results in a linear Scatchard plot typical of Langmuir adsorption with a single equilibrium binding constant ($K \sim 10^8\ M^{-1}$). Adsorption to immobilized polyclonal (heterogeneous) antibodies, however, yields nonlinear Scatchard plots. The binding data at four values of pH were analyzed by Sada

et al. using the local Langmuir assumption with gaussian binding energy distributions [28]. As shown in Table 5.4, these data can also be adequately described by the Temkin model (Table 5.4) presuming a single limiting capacity at all four pH values (see also Appendix) . The resulting value of Q_{max} (= 0.57 mg/ml) is in excellent agreement with the capacity determined by Sada *et al.* for the gaussian energy distribution [14]. The results given in Table 5.4 show that decreasing pH decreases the maximum binding constant K_T and slightly decreases the Temkin parameter q_T. This also agrees very well with the observation of Sada *et al.* that the mean binding energies and standard deviations decreased with decreasing pH. Thus, the Temkin model captures the relevant features of adsorption in this strongly heterogeneous system with a relatively simple isotherm expression.

Decreasing the pH (antibody-affinity materials) and increasing the salt concentration (ion exchange materials) serve to block surface binding sites and/or protein functional groups with equilibrium competitors (NaCl, H^+). Thus there is a parallel between the effect of increasing the competitor concentration for these chromatographic materials and that of decreasing the density of copper sites or the number of surface histidines for metal-affinity materials. All of these processes serve to reduce the number of the protein-surface interactions, and as a result decrease the maximum binding constant determined by fits the Temkin model.

Binding heterogeneity in biological receptors

Binding heterogeneity has also been previously reported in biological receptor systems, including immunoglobulins [13] and interleukin cellular receptors [15,16]. Equilibrium binding of iodovanillin to polyclonal anti-BSA immunoglobulin-G has been described using local Langmuir adsorption with a Poisson distribution in binding energies [13], presuming approximately ten vanillin binding sites per immunoglobulin

molecule [29]. We find that adsorption in this system can also be described by the Hill model, shown in Table 5.5, but *cannot* be adequately described by the Temkin model (data not shown). Although the Hill parameters K_o and Q_{max} cannot be determined from the published data, presuming a limiting capacity approximately ten times the antibody concentration results in a mean binding energy (K_o) similar to that reported previously (data not shown) [13]. Thus, the Hill model describes these adsorption data adequately in a simple isotherm that represents a binding energy distribution similar to more complicated expressions.

Interleukins display non-Langmuir adsorption to mouse fibroblasts expressing particular combinations of interleukin cellular receptor subunits [15,16]. Interleukin cellular receptors are proposed to consist of multiple subunits, resulting in a heterogeneous population of receptors, each consisting of a different combination of subunits. This system was analyzed originally by Sugamara and co-workers [15,16] using a bi-model Langmuir isotherm. Again, these data can be described adequately by the Hill model (Table 5.6), but not by the Temkin model (data not shown). The expression of additional receptor subunits decreases the Hill coefficient α, increasing the observed range of binding energies. Conversely, the presence of antibody against a particular subunit increases the Hill coefficient α, decreasing the range of binding energies.

Energy distributions

The Temkin model appears to better provide a general description of protein adsorption to chromatographic materials, while the Hill model appears to better describe receptor binding in biological systems. The difference between these two models of heterogeneous binding is a difference in the shape of the binding energy distributions.

However, both the Temkin and Hill models predict that the binding energy decreases as increasing amounts of adsorbate are bound to the surface.

A distribution of binding energies can be explained if we consider protein adsorption to surface sites involving multivalent contacts in terms of two opposing contributions [30]: a favorable energy from the specific metal-to-protein contacts ("intrinsic binding energy") and an unfavorable energy required to match each binding site and protein to make these contacts ("rearrangement energy"). On densely derivatized metal-affinity chromatography supports, the protein will adsorb with the highest binding energy to those arrangements of metal ions which most closely match its pattern of histidines (Figure 5.1B). Less optimal patterns of metal ions require some rearrangement to obtain the same number of interactions, resulting in a lower net binding energy. The Temkin model predicts a uniform distribution of binding energies over the population of surface binding sites. Theoretically, such a uniform distribution of binding energies would arise from a truly random arrangement of surface binding sites [31].

In contrast, the Hill model presumes that the net binding energy follows a quasi-gaussian distribution; that is to say, the binding energies are clustered around a mean value, as illustrated in Figure 5.5. Considering multivalent interactions between a protein and the surface, such an expression can be obtained from a probability distribution of the number of "successful" interactions between the target and the receptor relative to the total number of possible interactions (with each interaction contributing a constant factor to the overall binding energy) [11,13]. In this case, the most probable binding energy corresponds to the net binding energy of the most probable number of protein-surface interactions. Increased or decreased numbers of interactions (higher or lower binding affinities) occur with lower probability.

Surface heterogeneity: biological vs. nonbiological adsorption

The Langmuir, Temkin, and Hill models can be compared directly to one another by describing the degree of binding heterogeneity as a "characteristic temperature" of the surface [31]. Indeed, in this interpretation both the Hill coefficient α and Temkin parameter q_T are inversely proportional to the characteristic temperature. If the "temperature" is low ($\alpha \sim 1$), then the surface is approximately uniform and adsorption is adequately described by the Langmuir model for a homogeneous surface. If the "temperature" is very high ($\alpha \sim 0$), then the surface is strongly heterogeneous. All binding energies are equally likely and the Temkin model will result (e.g., chromatographic materials). For intermediate values ($0 < \alpha < 1$), the surface is moderately heterogeneous. Binding energies will be clustered around a most probable value, and the Hill model will result (e.g., biological receptors).

Thus the ability of the Temkin or Hill models to describe the binding to a material is an important indication of its heterogeneity. As demonstrated by the capability of some data to be described by both the Hill and Temkin models (e.g., Tables 5.3 and 5.4), accurate prediction of the shape of the energy distribution from equilibrium adsorption isotherms requires careful experimentation over a wide range of conditions (as done for the metal-affinity chromatography studies).

Titration microcalorimetry would provide a means to more directly measure the binding energy distribution [17]. In such experiments, the differential equilibrium heat of adsorption is measured for incremental increases in protein coverage. The Temkin model predicts a linear dependence of the differential heat of adsorption with increasing protein coverage; in fact, this linear dependence was the source of the original derivation of the Temkin isotherm [32]. The Hill model, on the other hand, predicts a logarithmic decrease in the differential heat of protein adsorption, diverging toward infinity at zero coverage [17]. Because of the relatively large enthalpy of metal-imidazole bond (~ 7

kcal/mol [33]), adsorption of multiple-histidine proteins to metal-affinity supports is well placed for such measurements.

Implications for chromatography

In order to accurately define an adsorption isotherm, equilibrium binding experiments are usually performed at low to moderate protein concentrations. Chromatographic applications, on the other hand, are often performed at much lower protein concentrations to optimize protein separation. For a model of heterogeneous binding in adsorption to be applied to chromatography, it must have predictive power at concentrations significantly lower than those generally used in equilibrium binding experiments. In terms of proposed binding energy distributions, this requires predictive power for binding energies *greater* than those which can be explored in equilibrium adsorption experiments.

For example, a common assumption in analytical affinity chromatography is that equilibration among all species is instantaneous. Under this local equilibrium assumption [34], the chromatographic capacity factor (see Chapter 4) is linearly related to the slope of the equilibrium adsorption isotherm. If the adsorption isotherm reaches a constant slope at low solute concentrations (Henry's Law), then according to the local equilibrium assumption, the elution time reaches a limiting value at small concentrations. If a constant slope is not obtained in the limit of low solute concentrations, then the elution time will always increase with decreasing solute concentration. The Henry's Law limit is therefore a requirement of the local equilibrium assumption in order to predict the bell-shaped elution profiles observed in analytical chromatography.

In the case of Hill model, the slope of the adsorption isotherm diverges toward infinity as solute concentration tends toward zero (see Appendix); thus it cannot satisfy Henry's Law. The incompatibility of this isotherm with the local equilibrium assumption

is a severe limitation of its practical application to chromatographic systems. Previous investigators have replaced the local equilibrium assumption with assumptions of a constant pattern of concentration and a linear driving force for intraparticle mass transport in modeling the chromatography of the Freundlich isotherm (the limit of the Hill model at low solute concentration) [35]. However, such interpretations are hampered by the need for apparent mass transfer coefficients in the particle pores, which may in fact simply describe the concentration regime in which adsorption in fact follows Henry's Law and therefore deviates from the Freundlich isotherm [31]. The bottom line of such requirements, however, are to prevent meaningful descriptions of chromatographic behavior at very low solute concentrations from model parameters determined by equilibrium adsorption at more moderate concentrations.

In contrast, The Temkin isotherm satisfies the Henry's Law limit (see Appendix). The initial slope of the Temkin adsorption isotherm (also the vertical intercept of the Scatchard plot) is the product $K_T q_T$. Thus the chromatographic capacity factor is predicted to be the product of an equilibrium binding constant (K_T) and an adsorption capacity (q_T), directly analogous to the Henry's Law limit of the Langmuir isotherm (see Chapter 4). In fact, zonal elution profiles predicted using the local equilibrium assumption for the Temkin isotherm have the same qualitative behavior as those predicted for the Langmuir isotherm [36]. Thus, based on parameters determined by equilibrium adsorption behavior at moderate concentrations, the Temkin isotherm can, with a simple expression, predict chromatographic behavior at low solute concentrations.

Implications for biological receptor repertoires

Predictions about biological receptor repertoires often depend on extending a binding model to concentrations lower than those used in equilibrium experiments; in this respect, they are analogous to predictions made in chromatographic systems. Once

again, these predictions require the extension of a model to binding energies *greater* than those explored in equilibrium binding experiments; thus predictions about the receptor repertoire are strongly dependent on the shape of the binding energy distribution [13].

In the case of distributions described by the Hill model, at higher binding energies the probability distribution decays exponentially with a decay constant α (eqn. 10). Although the average binding energies of the biological systems studied vary widely (from 10^3 to 10^9 M^{-1}), the Hill coefficients α are consistently close to ½, a value similar to that measured for dose-response relationships in the human olfactory system [37] and commonly seen in other biological receptor systems [38]. Expressing fewer receptor subunits [15] or disabling a particular one with a monoclonal antibody [16] increases the Hill coefficient. These changes decrease the "temperature" of the system resulting in a more homogenous receptor repertoire.

As the receptor repertoire becomes more homogeneous (α increases), the size of the repertoire necessary to achieve a particular binding affinity also increases. For example, the likelihood that a finite receptor repertoire can detect a target compound at a low concentration threshold is proportional to the probability of achieving sufficient binding energy [13]. As a result, the likelihood of detecting the target compound decreases as $n^{-\alpha}$ for an n-fold decrease in concentration. Alternatively, the repertoire size needed to detect a target compound is inversely proportional to the probability of achieving sufficient binding energy. Thus, the total number of receptors needed for successful detection increases as n^{α} for an n-fold decrease in target concentration.

Implications on design of materials for protein recognition

Comparisons of the degree of heterogeneity among the receptor systems studied show that only in the case of polyclonal antibodies binding a protein antigen [14] does the degree of adsorption heterogeneity surpass that of a three-histidine protein on the

metal affinity support ($\alpha \sim 0.1$). This suggests that the metal-affinity chromatography material provides a "receptor repertoire" for multiple-histidine proteins that is of comparable complexity to the repertoire provided by the immune system for complex protein antigens. That is to say, the random metal affinity support has a subset of binding sites that bind multiple-histidine proteins with selectivities approaching those of antibodies, and that the size of that subset is analogous to the size presented by the immune system.

The Temkin model predicts a uniform distribution in binding energy over the population of surface binding sites, which is presumed to arise from the random arrangement of affinity ligands on the chromatographic surface [31]. Therefore, the optimal binding sites for a multiple-histidine protein already exist in a random arrangement of metal-affinity ligands, and these sites display the practical limit of binding that can be achieved at a particular ligand density. Alternatively, the vast majority of binding sites have a somewhat decreased binding affinity, due to increased energy penalties to "arrange" the optimal interactions between the protein and the support. The Temkin model therefore provides a framework to separate a background of "nonspecific" high-affinity binding sites that naturally occur in random materials from those sites successfully manipulated to recognize specific protein targets.

This link between surface structure and binding energy is the motivation behind the design of novel materials for molecular recognition. For example, the CuIDA-TSK metal affinity support already has a small subset of sites (~ 5%) with binding affinities on the order of 10^9 M^{-1} for three-histidine yeast cytochrome *c* variants (Figure 5.4). To create a material with a high specificity for a particular target molecule, our laboratory has taken the approach of molecular imprinting to "pre-arrange" the *less optimal* sites to be more complementary to the target [39]. Alternatively, one could take a combinatorial approach and devise a means to "screen" for the *optimal* surface binding sites. By either

approach, a successful strategy will skew the probability distribution toward higher binding energies for the target molecules and toward lower binding energies for other molecules.

CONCLUSIONS

Both the Temkin and Hill models provide a practical measure of the binding affinity displayed in a variety of heterogeneous systems, and represent binding heterogeneity with simple expressions that have predictive power over a wide range of concentrations. Both models describe a heterogeneous population of noninteracting surface binding sites by a distribution in binding energies. These models are dependent on three physically realistic parameters: one expressing limiting binding capacity (Q_{max}), one describing the range of binding energies (a or α), and one expressing a characteristic binding energy (K_T or K_0). The choice between these two models is a choice between shapes of the binding energy distribution: either a uniform distribution over a limiting range of energies (Temkin), or a bell-shaped distribution with a most probable binding energy (Hill). Most importantly, the Temkin isotherm also satisfies Henry's Law, a necessary requirement to employ local equilibrium theory for chromatographic applications.

Moreover, these isotherms collapse to simpler expressions under conditions in which limiting capacity cannot be determined. Under these conditions, for both models one of the binding parameters retains physical relevance, either the maximum binding energy (Temkin) or the range of binding energies (Hill). In this respect, these two models have two important practical advantages over other models: (1) they contain as few as two undetermined parameters, and (2) they can be expressed in closed form. None of the other isotherms considered (Dubinin-Radushkevich and Toth isotherms [21], and

local Langmuir adsorption using binomial, Poisson [13] and gaussian [14] distributions in binding energy) satisfy these requirements.

In the case of metal affinity chromatography, the Temkin model provides a framework for protein adsorption with a consistent limiting capacity for essentially identical proteins. In this framework, it can be seen that adding histidines to the target protein and adding immobilized copper to the chromatographic support both increase the number of metal-to-ligand interactions dictating protein adsorption. These in turn increase the binding energy of the most accommodating surface sites and increase the binding heterogeneity. This model also describes adsorption to other materials in which Scatchard plots demonstrate heterogeneous protein-surface interactions [27].

Although the original motivation was to describe heterogeneous protein adsorption in chromatographic systems, binding heterogeneity is also common among biological receptors. In the systems studied, the bell-shaped energy distribution of the Hill isotherm more often captured the relevant features of biological binding data. It is difficult, if not impossible, to distinguish such a distribution from more complicated energy distributions (such as binomial, Poisson, and gaussian). This simple isotherm thus can thus be employed to describe the binding behavior of heterogeneous biological receptors in diverse systems such as polyclonal antibodies or interactions with complex cellular receptors.

Table 5.1. Temkin model for IMAC adsorption of yeast cytochrome c variants. Equilibrium adsorption data for *S. cerevisiae* iso-1-cytochrome *c* variants differing in surface histidine content to CuIDA-TSK (see Chapter 3 [6]). Parameters determined by nonlinear regression to Temkin isotherm in low concentration limit (eqn. 6). Error estimates are given in parentheses. Labels indicate the surface-accessible histidines. Regression did not converge for H(-) variant.

cytochrome *c* variant	K_T (M^{-1})	q_T (μmol/ml TSK)
$H_{26}H_{33}H_{39}$	1.8 (0.2) × 10^8	0.197 (0.01)
$H_{26}H_{33}H_8$	1.3 (0.4) × 10^8	0.198 (0.01)
$H_{26}H_{33}H_4$	9.8 (2.6) × 10^7	0.243 (0.01)
$H_{26}H_4$	1.5 (0.8) × 10^7	0.186 (0.02)
$H_{26}H_8$	9.8 (2.3) × 10^6	0.169 (0.01)
$H_{26}H_{33}$	1.9 (0.2) × 10^6	0.215(0.01)
H_4	4.5 (1.1) × 10^5	0.250 (0.02)
H_8	3.5 (0.8) × 10^5	0.242 (0.02)
H_{26}	9.5 (2.3) × 10^3	0.433 (0.07)
tuna (H_{26})	5.3 (0.9) × 10^3	0.562 (0.08)

Table 5.2. Temkin model for IMAC adsorption at decreased copper loading. Equilibrium adsorption data for horse cytochrome *c* to IDA-TSK prepared at decreased copper loading (Cu =18.6 μmol/ml TSK is maximum copper loading [6]). Parameters determined by nonlinear regression to Temkin isotherm in low concentration limit (eqn. 6). Error estimates are given in parentheses. Regression did not converge for copper loading < 10 μmol/ml TSK.

Cu (μmol/ml TSK)	K_T (M^{-1})	q_T (μmol/ml TSK)
18.6	$1.2\ (0.4) \times 10^5$	0.267(0.05)
15.6	$1.4\ (0.3) \times 10^5$	0.240 (0.02)
14.8	$6.0\ (1.0) \times 10^4$	0.268 (0.02)
12.3	$2.3\ (0.2) \times 10^4$	0.260 (0.01)
12.0	$1.3\ (0.4) \times 10^4$	0.315 (0.06)
10.5	$3.6\ (1.4) \times 10^3$	0.480 (0.15)

Table 5.3. Temkin model for ion exchange chromatography. Data adapted from Gill *et al.* [27]. Equilibrium adsorption of rat cytochrome b_5 to Mono Q ion exchange resin measured at increasing concentration of NaCl. Parameters are best fit to Temkin isotherm in low concentration limit (eqn. 6). Error estimates are given in parentheses.

[NaCl] (M)	K_T (M^{-1})	q_T (μmol/10^8 beads)
0.100	6.1 (1.9) × 10^6	0.076 (0.01)
0.125	4.0 (1.0) × 10^5	0.116 (0.01)
0.150	6.8 (1.4) × 10^4	0.155 (0.01)
0.175	2.0 (0.6) × 10^4	0.241 (0.04)

Table 5.4. Temkin model for antibody affinity chromatography. Data adapted from Sada *et al.* [14]. Equilibrium adsorption of bovine serum albumin (BSA) to polyclonal anti-BSA covalently immobilized to Sepharose-4B measured over pH range of 3.6 to 7.2. Parameters determined by nonlinear regression of Scatchard plots to Temkin model (see Appendix) with a common value of the limiting adsorption capacity (Q_{max} = 0.57 (0.03) mg/ml) for all four experiments. Error estimates are given in parentheses.

pH	K_T (M^{-1})	a
7.6	$5.1\ (1.3) \times 10^{10}$	12.2 (1.0)
4.2	$8.6\ (1.6) \times 10^{9}$	14.4 (0.9)
3.9	$4.3\ (0.6) \times 10^{8}$	12.0 (0.9)
3.6	$5.1\ (0.9) \times 10^{7}$	10.1 (1.1)

Table 5.5. Hill model for solution binding to polyclonal antibodies. Data adapted from Lancet *et al.* [13]. Equilibrium binding of iodovanillin to polyclonal anti-BSA immunoglobulin-G (0.5 mM total concentration) measured by equilibrium dialysis. Parameters determined by nonlinear regression of Hill model in low concentration limit (Freundlich approximation, eqn. 9). Error estimates are given in parentheses.

α	b ($M^{1-\alpha}$)
0.535 (0.031)	1.77 (0.2) $\times 10^{-2}$

Table 5.6. Hill model for solution binding to cellular receptors. **(A)** Data adapted from Kumaki *et al.* [15]. Equilibrium adsorption of human Interleukin-2 to mouse fibroblasts expressing particular mouse interleukin receptor subunits measured by centrifugation. **(B)** Data adapted from Kondo *et al.* [16]. Equilibrium adsorption of Interleukin-7 to mouse fibroblasts in presence of antibody against γ subunit of interleukin receptor measured by centrifugation. Parameters determined by nonlinear regression of Scatchard plots to Hill model in low concentration limit (Freundlich approximation, see Appendix). Error estimates are given in parentheses.

(A) Adding γ receptor subunit

Receptor subunits	α	***b*** (molecules cell^{-1} M$^{-\alpha}$)
α,β	0.695 (0.018)	$2.6\ (1.1) \times 10^{10}$
α,β,γ	0.616 (0.013)	$4.8\ (1.5) \times 10^{9}$

(B) Inhibiting γ receptor subunit

[anti-γ Ab] (μg/ml)	α	***b*** (molecules cell^{-1} M$^{-\alpha}$)
0	0.547 (0.013)	$4.7\ (1.4) \times 10^{8}$
100	0.750 (0.030)	$2.0\ (1.4) \times 10^{10}$

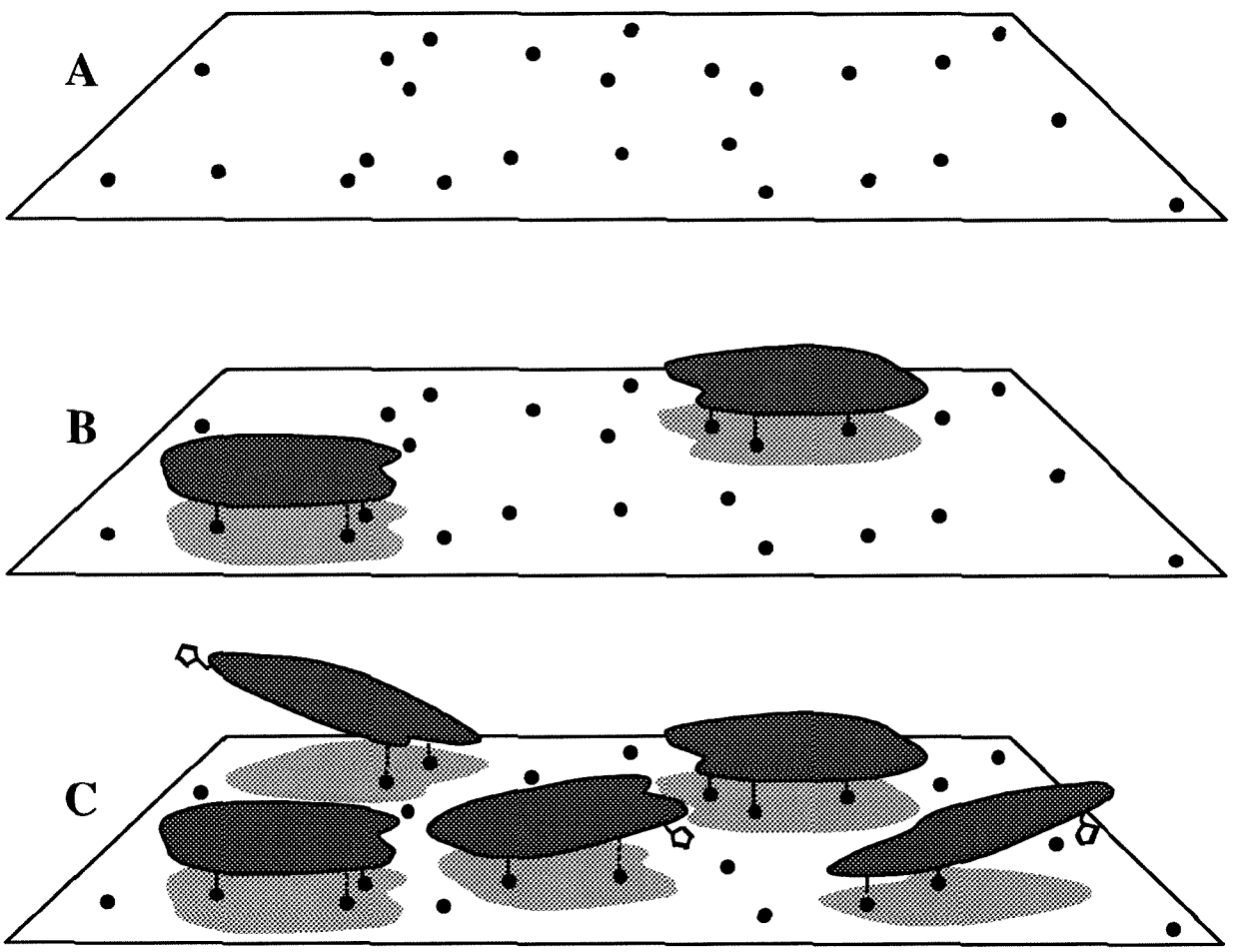


Figure 5.1. Herogeneous protein adsorption to a random arrangement of binding sites. **(A)** Surface has a nonuniform arrangement of ligands for protein binding (●, e.g., copper sites). **(B)** Protein adsorbs with highest affinity to surface "binding sites" in which ligand patterns complement arrangement of protein functional groups (○, e.g., histidines). **(C)** Protein adsorbs with lower affinity to surface "binding sites" with less optimal ligand patterns, resulting in a range of protein binding energies.

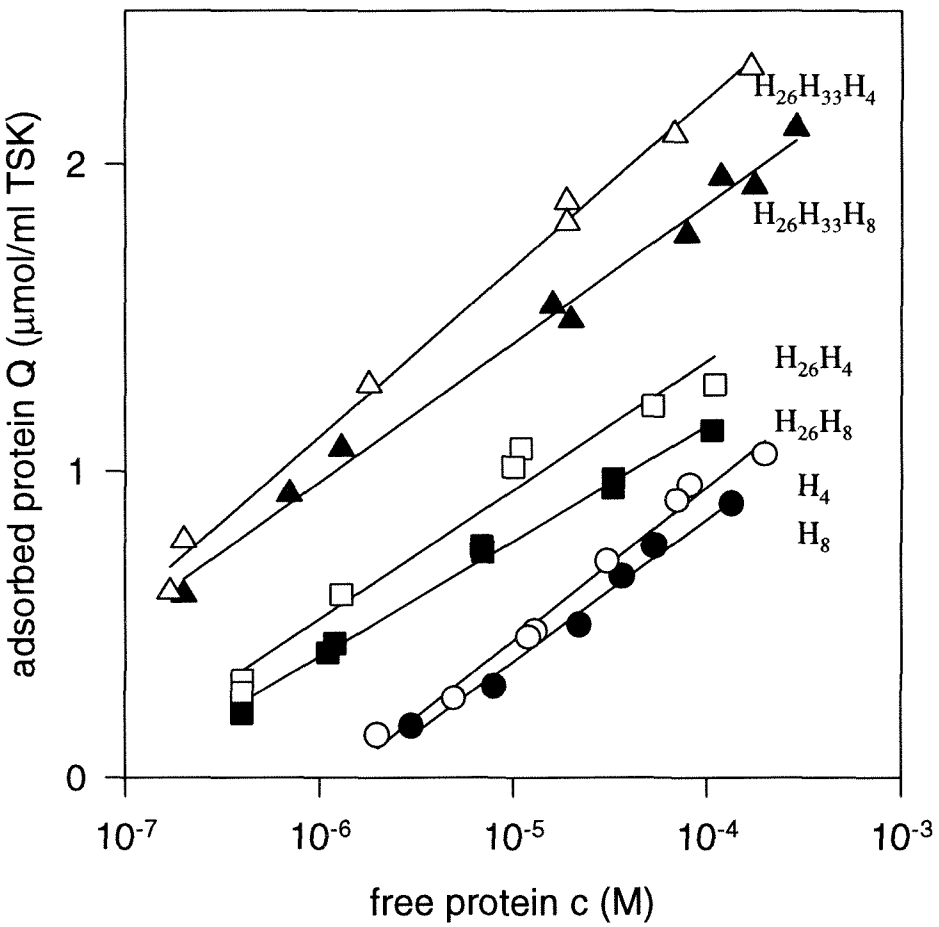


Figure 5.2. Equilibrium adsorption isotherms of yeast cytochromes *c*. Adsorption data is presented for yeast cytochrome *c* histidine variants to CuIDA-TSK (see Chapter 3). Adsorbed protein is plotted against the logarithm of equilibrium protein solution concentration for cytochrome *c* histidine variants (●)H_8, (○) H_4, (■) $H_{26}H_8$, (□)$H_{26}H_4$, (▲) $H_{26}H_{33}H_8$, (Δ) $H_{26}H_{33}H_4$. Labels indicate the surface-accessible histidines. Lines represent fit to Temkin model (eqn. 6), calculated from parameters in Table 5.1

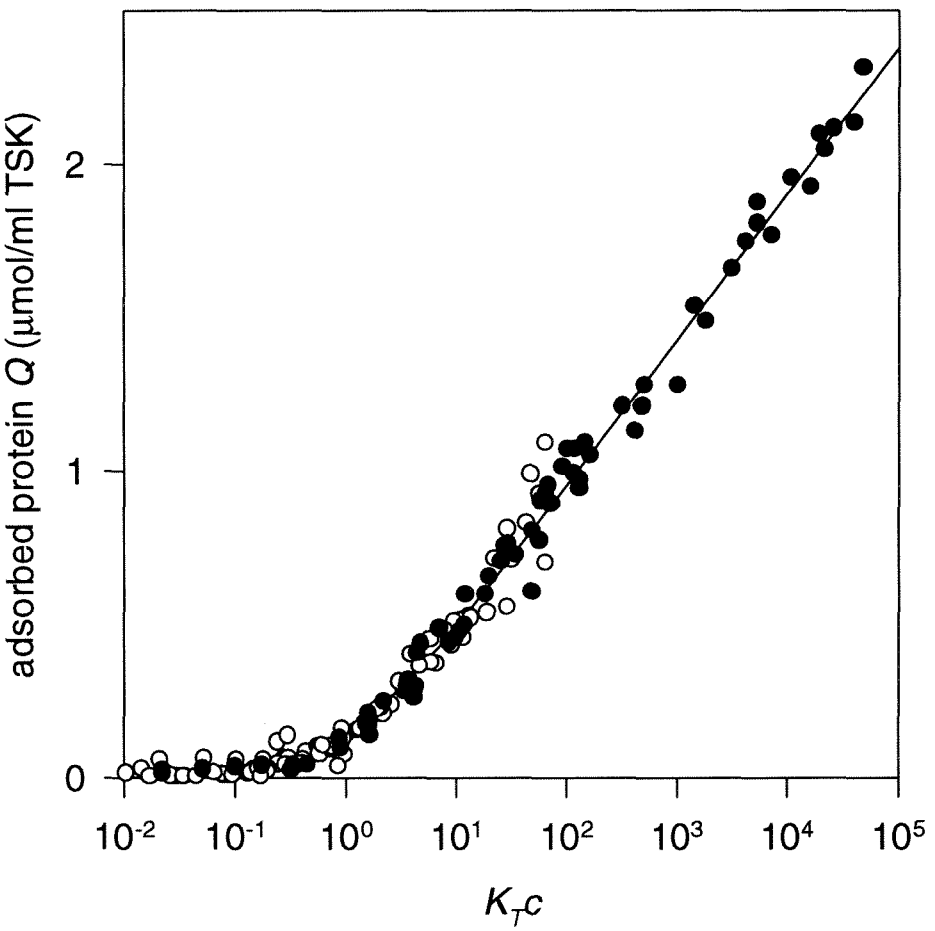


Figure 5.3. Temkin model for equilibrium adsorption of cytochromes *c*. Free protein concentration (*c*) scaled by Temkin isotherm maximum binding constant (K_T) using parameters of Table 5.1. Adsorption data are presented for cytochromes *c* to CuIDA-TSK (see Chapter 3), (●) yeast cytochrome *c* histidine variants and tuna cytochrome c at maximum copper loading, (O) horse cytochrome *c* at reduced copper loading. Curve represents Temkin model (eqn. 6) calculated using $q_T = 0.21$ μmol/ml TSK.

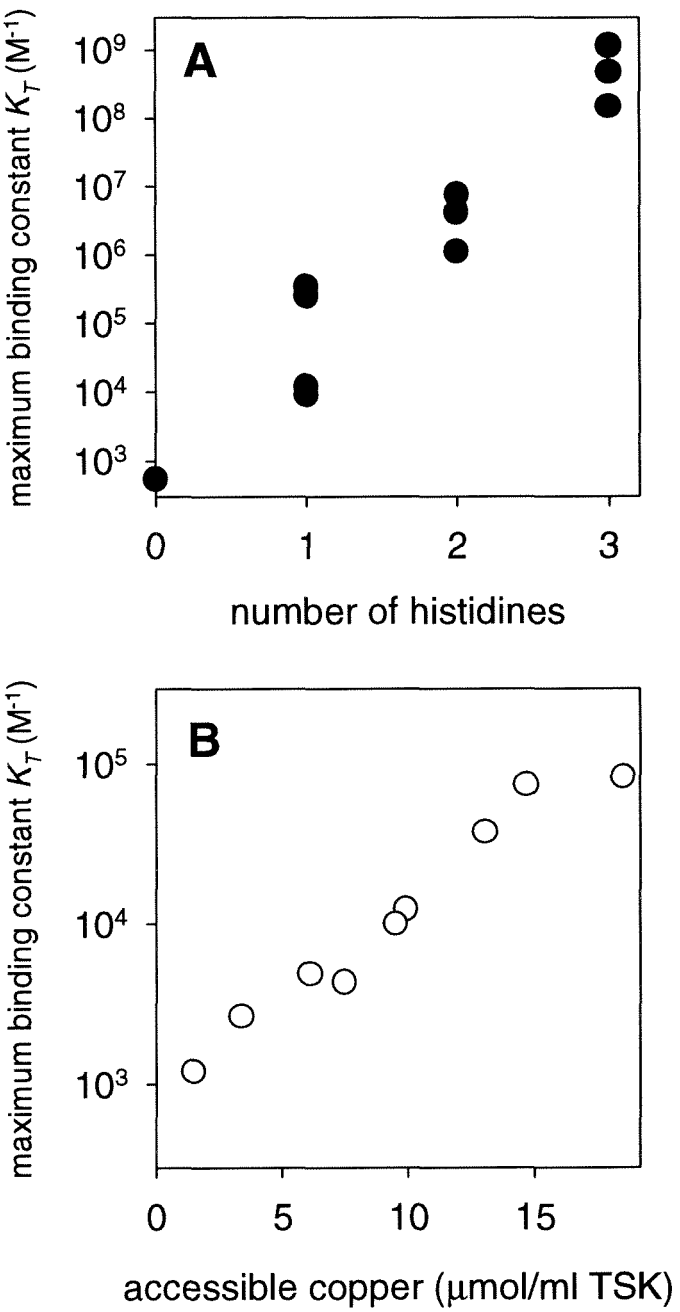


Figure 5.4. Maximum Temkin binding constants for cytochromes *c* in IMAC. **(A)** Adsorption of yeast cytochrome *c* surface histidines variants to IDA-TSK at maximum copper loading (see Chapter 3). **(B)** Adsorption of horse cytochrome *c* to IDA-TSK at reduced copper loading (see Chapter 3). Maximum binding constants K_T calculated from best fit to Temkin isotherm (eqn. 5) using Q_{max}= 3.5 µmol/ml TSK and x_o=0.

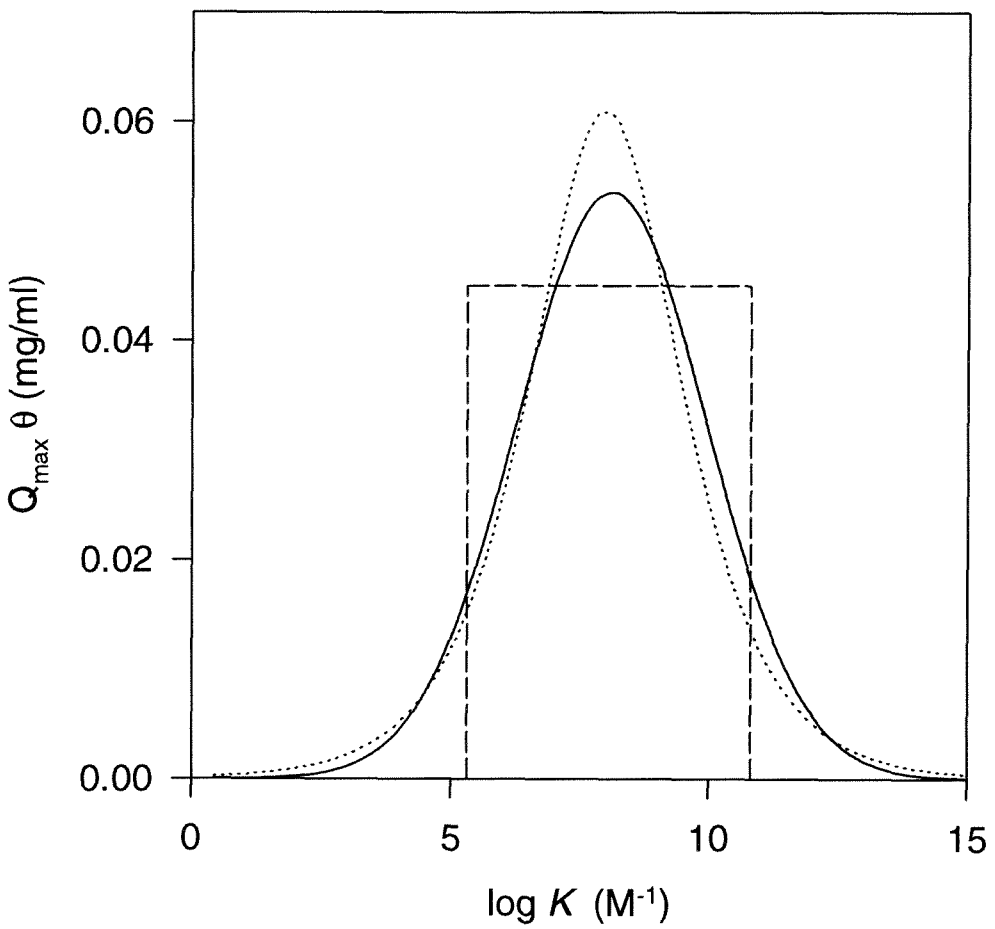


Figure 5.5. Energy distributions of Temkin, Hill, and Gaussian models. Energy distributions calculated from Temkin, Hill, and the gaussian models for BSA adsorption on polyclonal anti-BSA immobilized to Sepharose-4B at pH 7.6 (data adapted from Sada *et al.* [14]). Gaussian distribution [14] (—), uniform distribution of Temkin model (- - -) calulated using parameters of Table 5.3, and bell-shaped distribution of Hill isotherm (·····) (see Appendix).

APPENDIX

Henry's Law limit of Local Langmuir isotherms

The Henry's Law limit is an important criteria for practical application of an adsorption isotherm to chromatographic experiments. Henry's Law states that, in the limit of infinite dilution of solute, the adsorption isotherm approaches a straight line. This limit can be defined as

$$\lim\ Q(c)\Big|_{c\to 0} = Hc\ ,$$

where H is the Henry's Law constant, analogous to the chromatographic capacity factor (Chapter 4). For local Langmuir adsorption, the Henry's Law constant can be expressed as an integral over the probability distribution in energy (θ) as

$$H = Q_{max}\int K\theta\,(K)dK = \frac{Q_{max}}{c_o}\int e^{x}\theta\,(x)dx\ .$$

This integral is in fact the moment generating function of the probability distribution [40]. From this expression, it can be seen that any distribution which at large energies corresponds to an exponential decay with decay constant less than one (e.g., Hill or Freundlich isotherms) will fail to satisfy Henry's Law.

For the uniform distribution in binding energy (Temkin isotherm) described by a maximum capacity Q_{max}, a maximum binding energy x_t, and a range of binding energy a, the Henry's Law limit is given by:

$$H = \frac{Q_{max} K_T (1 - e^{-a})}{a} \approx K_T q_T \quad ,$$

where K_T and q_T are defined previously.

For local Langmuir adsorption of a binomial energy distribution characterized by a maximum capacity Q_{max}, a total number of interactions n, a probability of each interaction p (mean number of interactions np), and an energy per interaction of x_o [13], the Henry's Law limit is given by

$$H = \left[p e^{x_o} + (1 - p) \right]^n \approx p e^{n x_o} \quad .$$

For local Langmuir adsorption of a Poisson energy distribution characterized by a maximum capacity Q_{max}, a mean number of interactions λ, and an energy per interaction of x_o [13], the Henry's Law limit is given by

$$H = \exp\left(\lambda \left(e^{x_o} - 1 \right) \right) \quad .$$

For local Langmuir adsorption of a gaussian energy distribution characterized by a maximum capacity Q_{max}, a mean energy x_o, and a standard deviation σ_x, [14] the Henry's Law limit is given by

$$H = \exp\left(x_o + \frac{(\sigma_x)^2}{2} \right) \quad .$$

Strictly speaking, local Langmuir adsorption isotherms derived from Poisson, binomial, and gaussian energy distributions satisfy Henry's Law. However, their behavior is similar to that of the Freundlich model, in that the resultant Henry's Law constants are

impractically high and the concentrations required to achieve Henry's Law are impractically low.

Scatchard Plots of Temkin and Hill isotherms

The Scatchard plot of the Temkin isotherm can be expressed analytically for all coverages,

$$\frac{Q}{c} = K_T Q\left[1-\left(\frac{K_T}{K_o}\right)^{Q/Q_{max}-1}\right]\Big/\left[\left(\frac{K_T}{K_o}\right)^{Q/Q_{max}}-1\right] \approx K_T Q/\left[e^{Q/q_T}-1\right] ,$$

where all parameters are as described previously. The Scatchard plot of the Temkin isotherm approaches a constant ($K_T q_T$ = Henry's Law constant) as Q approaches 0.

The Scatchard plot of the Hill isotherm can also be expressed analytically for all coverages,

$$\frac{Q}{c} = K_o\left[Q_{max}-Q\right]^{1/\alpha} Q^{(\alpha-1)/\alpha} \approx b^{1/\alpha} Q^{(\alpha-1)/\alpha} ,$$

where all parameters are a described in Chapter 5. For $\alpha < 1$, the Scatchard plot diverges as Q approaches 0, again indicating that the Hill isotherm does not satisfy Henry's Law.

Table 5A.1. Hill model for ion exchange chromatography. Equilibrium adsorption of rat cytochrome b_5 to Mono Q ion exchange resin at increasing concentration of NaCl (adapted from Gill *et al.* [27]). Parameters are best fit to Hill model in low concentration limit (Freundlich approximation, eqn. 9). Error estimates are given in parentheses.

[NaCl] (M)	α	b (μmol/10^8 beads $M^{-\alpha}$)
0.100	0.209 (0.012)	3.3 (0.4)
0.125	0.326 (0.011)	8.4 (0.9)
0.150	0.460 (0.033)	21 (6)
0.175	0.609 (0.054)	67 (32)

Table 5A.2. Hill model for antibody affinity chromatography. Data for equilibrium adsorption of bovine serum albumin (BSA) to polyclonal anti-BSA covalently immobilized to Sepharose-4B from Sada *et al.* [14]. **(A)** Parameters determined by nonlinear regression of Scatchard plots to Hill model in limit of low concentration (Freundlich approximation, see above). **(B)** Parameters determined by nonlinear regression of Scatchard plots to Hill model (see above) with a common value of the parameter Q_{max} for all four experiments. Error estimates are given in parentheses.

A. Low concentration limit (Freundlich model)

pH	α	b (mg/ml gel $M^{-\alpha}$)
7.6	0.114 (0.006)	2.5 (0.3)
4.2	0.152 (0.002)	3.1 (0.1)
3.9	0.270 (0.007)	13.4 (1.6)
3.6	0.458 (0.033)	147 (83)

B. Hill model: Q_{max} = 0.68 (0.059) mg/ml gel

pH	α	K_o (M^{-1})
7.6	0.274 (0.033)	$3.4\ (2.2) \times 10^7$
4.2	0.248 (0.015)	$1.7\ (1.1) \times 10^6$
3.9	0.366 (0.016)	$5.5\ (2.4) \times 10^5$
3.6	0.546 (0.025)	$3.5\ (1.1) \times 10^5$

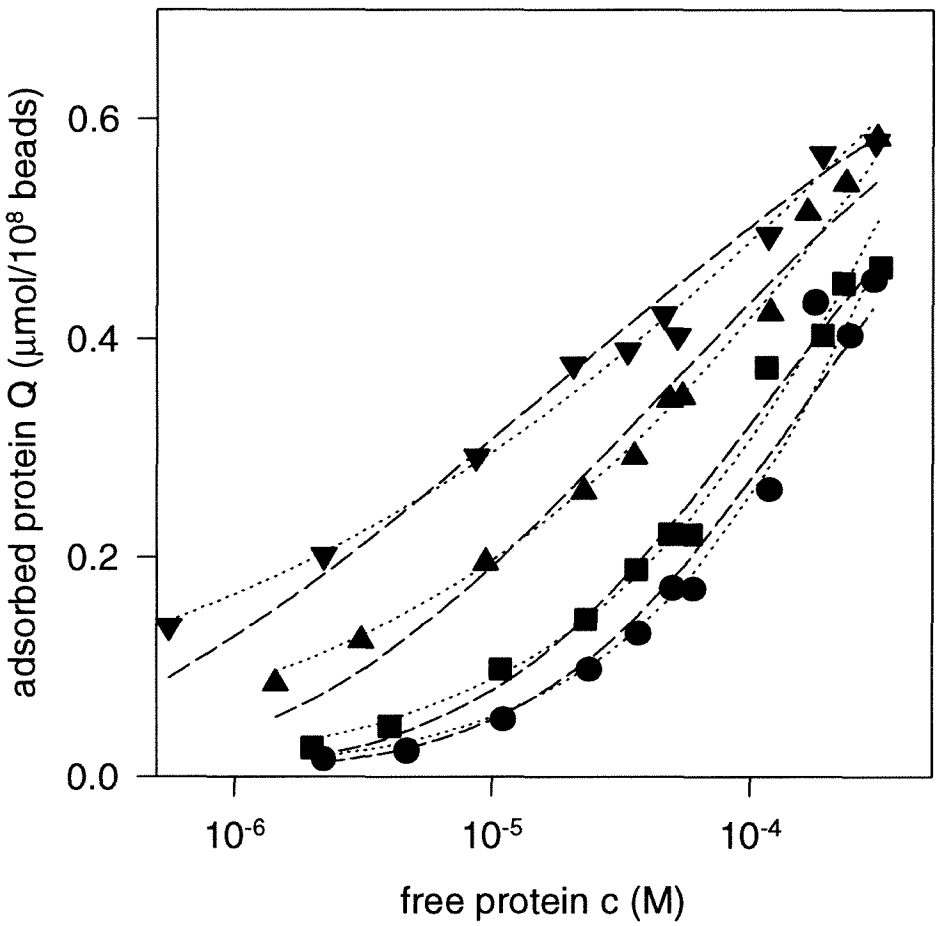


Figure 5A.1. Equilibrium adsorption of cytochrome b_5 in ion exchange chromatography. Equilibrium adsorption of rat cytochrome b_5 to Mono Q ion exchange resin at increasing concentration of NaCl (data adapted from Gill *et al.* [27]): (▼) 0.100 M NaCl, (▲) 0.125 M NaCl, (■) 0.150 M NaCl, (●) 0.175 M NaCl. Parameters determined by nonlinear regression of adsorption isotherms to Temkin (- - -) and Hill (····) models (Tables 5.3 and 5A.1).

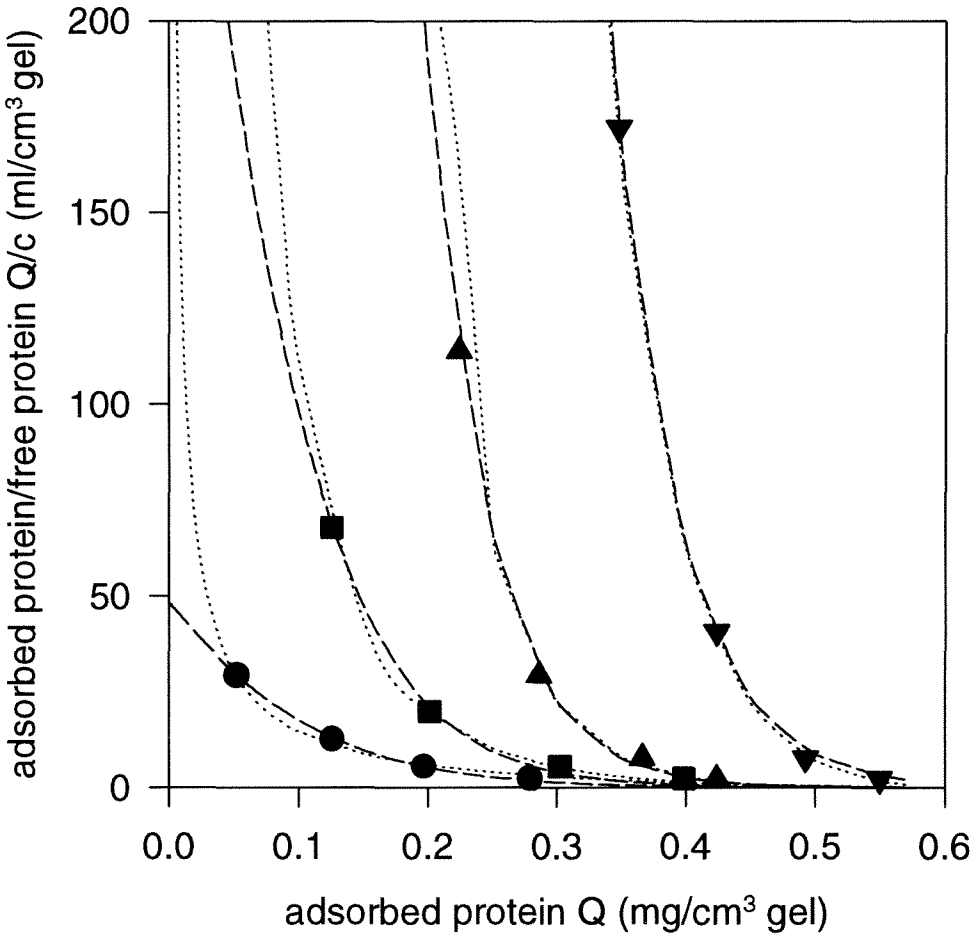


Figure 5A.2. Scatchard plots of adsorption of BSA by immobilized ploYclonal anti-BSA antibodies. Data for equilibrium adsorption of bovine serum albumin (BSA) to polyclonal anti-BSA covalently immobilized to Sepharose-4B, adapted from Sada *et al.* [14]: pH (▼) 7.2, (▲) 4.2, (■) 3.9, and (●) 3.6. Parameters determined by nonlinear regression of Scatchard plots to Temkin (- - -) and Hill (···) models (Tables 5.4 and 5A.2).

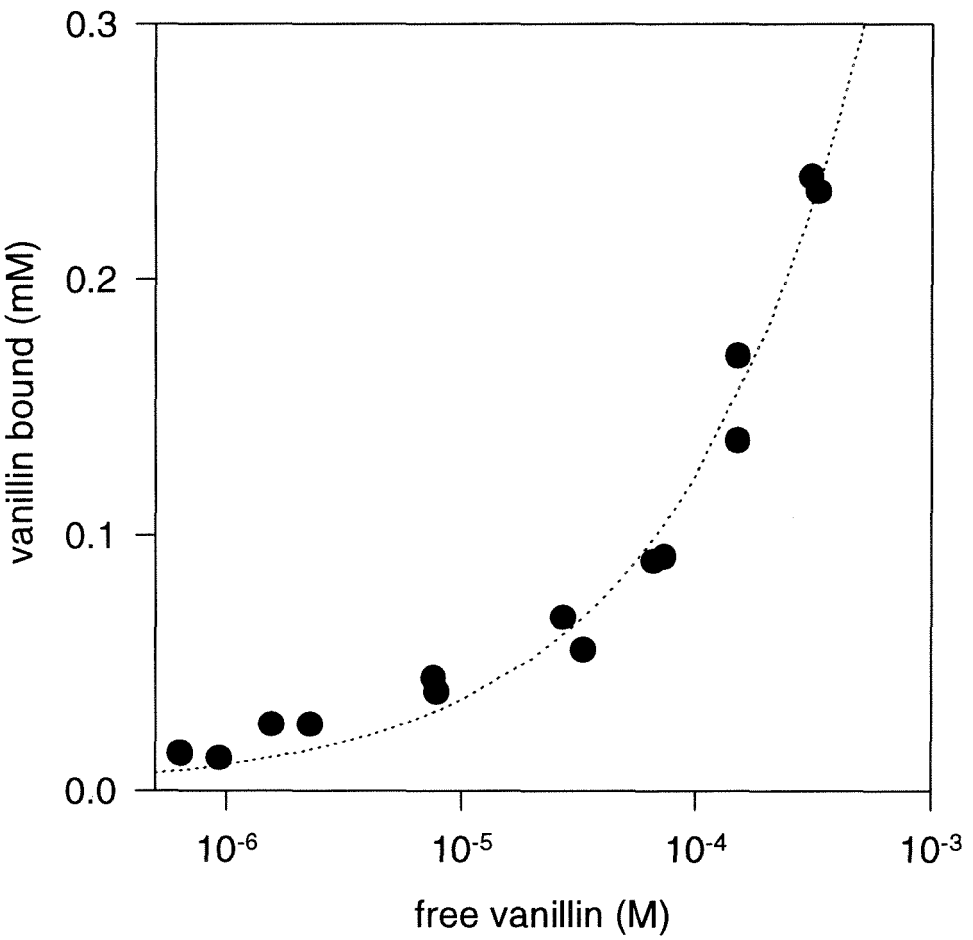


Figure 5A.3. Solution binding of vanillin by polyclonal anti-BSA antibodies. Equilibrium binding of iodovanillin to polyclonal anti-BSA (0.5 mM total concentration) measured by equilibrium dialysis (data adapted from Lancet *et al.* [13]). Parameters determined by nonlinear regression to Hill isotherm (Table 5.5).

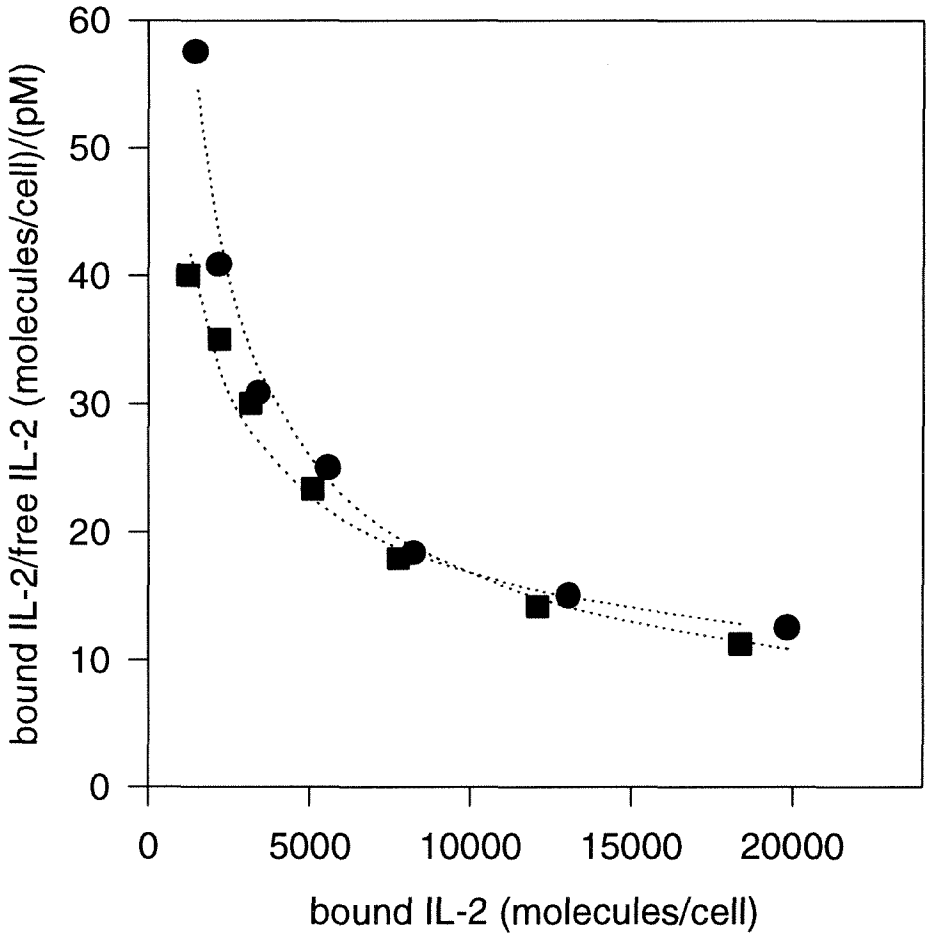


Figure 5A.4. Scatchard plots of interleukin adsorption:effect of cellular receptor subunits. Equilibrium adsorption of human Interleukin-2 to mouse fibroblasts expressing particular mouse interleukin receptor subunits (data adapted from Kumaki *et al.* [15]): (■) α and β subunits, (●) α, β, and γ subunits. Parameters determined by nonlinear regression of Scatchard plots to Hill model (Table 5.6).

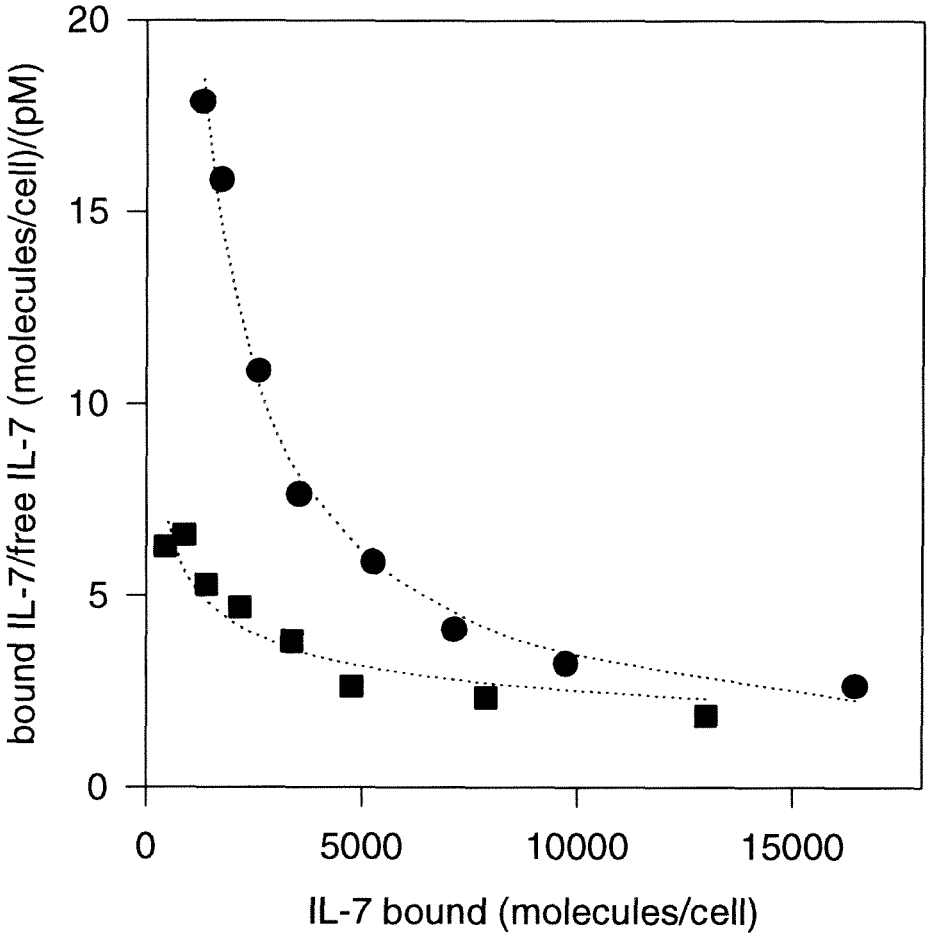


Figure 5A.5. Scatchard plots of interleukin adsorption: inhibition by anti-receptor antibody. Equilibrium adsorption of Interleukin-7 to mouse fibroblasts in presence of antibody against γ subunit of interleukin receptor (data adapted from Kondo *et al.* [16]): (●) no anti-γ, (■) 50 μg/ml anti-γ. Parameters determined by nonlinear regression of Scatchard plots to Hill model (Table 5.6).

REFERENCES

1. Horstmann, B. J. and Chase, H. A. (1989) *Chem. Eng. Res. 67*, 243-254.
2. Belew, M. B., Yip, T. T., Andersson, L. and Porath, J. (1987) *J. Chromatogr. 403*, 197-206.
3. Kopaciewicz, W., Rounds, M. A., Fausnaugh, J. and Regnier, F. (1983) *J. Chromatogr. 266*, 3-21.
4. Whitley, R. D., Brown, J. M., Karajgikar, N. D. and Wang, N. H. L. (1989) *J. Chromatogr. 483*, 137-155.
5. Livingston, A. G. and Chase, H. A. (1989) *J. Chromatogr. 481*, 159-174.
6. Todd, R. J., Johnson, R. D. and Arnold, F. H. (1994) *J. Chromatogr. A 662*, 13-26.
7. Yon, R. J. (1988) *J. Chromatogr. 457*, 13-23.
8. This argument presumes that the protein's adsorption cross section (the space it occupies on the surface) does not change. We could envision situations in which it would change, for example when the protein becomes oriented in a certain direction that depends on the specific interactions with the surface.
9. Hogg, P. J. and Winzor, D. J. (1984) *Arch. Biochem. Biopys. 234*, 55-60.
10. Kyprianou, P. and Yon, R. J. (1982) *Biochem. J. 207*, 549-556.
11. Dowd, V. and Yon, R. J. (1992) *J. Chromatogr. 627*, 145-151.
12. Burnet, F. M. (1963) *Cold Spring Harbor Symp. Quant. Biol. 32*, 1-8.
13. Lancet, D., Sadovsky, E. and Seidelman, E. (1993) *Proc. Natl. Acad. Sci. USA 90*, 3715-3719.
14. Sada, E., Katoh, S., Sukai, K., Tohma, M. and Kondo, A. (1986) *Biotech. Bioeng. 28*, 1497-1502.

15. Kumaki, S., Kondo, M., Takeshita, T., Asao, H., Nakamurea, M. and Sugamura, K. (1993) *Biochem. Biophys. Res. Comm. 193*, 356-363.

16. Kondo, M., Takeshita, T., Higuchi, M., Nakamura, M., Sudo, T., Nishikawa, S. and Sugamura, K. (1994) *Science 263*, 1453-1454.

17. Tompkins, F. C. (1978) *Chemisorption of Gases on Metals*, Academic Press, New York.

18. Sips, R. (1950) *J. Chem. Phys. 18*, 1024-1026.

19. Sips, R. (1948) *J. Chem. Phys. 16*, 490-495.

20. Hill, A. V. (1910) *J. Physiol. (London) 40*, iv.

21. Jaroniec, M., Dabrowski, A. and Toth, J. (1984) *Chem. Eng. Sci. 39*, 65-70.

22. Louie, G. V. and Brayer, G. D. (1990) *J. Mol. Biol. 214*, 527-555.

23. We estimate a TSK surface area of 16 m^2/ml (see Chapter 3). The average radius of yeast cytochrome *c* is 15.2 Å, that of the asymmetric unit (protein plus water) is 17.6 Å [22]. Distributing circular shadows of these radii over the TSK surface area results in a monolayer coverage of 2.7 - 3.7 μmol/ml TSK.

24. Hemdan, E. S., Zhao, Y., Sulkowski, E. and Porath, J. (1989) *Proc. Natl. Acad. Sci. USA 86*, 1811-1815.

25. Mrabet, N. T. (1992) *Biochemistry 31*, 2690-2702.

26. This energy gain is consistent with the selectivity displayed (~ 100-fold) by rationally designed bis-mercury receptors for a target bis-imidazole (a two-histidine "protein analog") over the single imidazole control (Chapter 2).

27. Gill, D. S., Roush, D. J. and Willson, R. C. (1994) *J. Coll. Int. Sci.*, in press.

28. Pauling, L., Pressman, D. and Grossbery (1944) *J. Am. Chem. Soc. 66*, 784-792.

29. Lancet, D. (1994) personal communication.

30. Jencks, W. P. (1981) *Proc. Natl. Acad. Sci. USA 78*, 4046-4050.

31. Cerofolini, G. (1982) *J. Coll. Int. Sci. 86*, 204-212.

32. Temkin, M. and Pyshev, V. (1940) *Acta Physicochim. U. R. S. S. 12*, 327.

33. Martell, A. E. and Smith, R. M. (1975) *Critical Stability Constants*, vol. 1-6, Plenum Press, New York.

34. Arnold, F. H., Schofield, S. A. and Blanch, H. W. (1986) *J. Chromatogr. 335*, 1-12.

35. Hashimoto, K. and Miura, K. (1976) *J. Chem. Eng. Japan 9*, 388-392.

36. Lau, J., Johnson, R. D. and Arnold, F. H. (1994) data not published.

37. Gross-Iseroff, R. and Lancet, D. (1988) *Chem. Senses 13*, 191-204.

38. Levitzki, A. (1984) *Receptors: A Quantitative Approach*, Benjamin/Cummings Inc., Menlo Park, CA.

39. Dhal, P. K. and Arnold, F. H. (1992) *Macromolecules 25*, 7051-7059.

40. Beyer, W. H., Editor (1987) *CRC Standard Mathematical Tables*, Boca Raton, FL.

CHAPTER 6

MAPPING PROTEIN METAL BINDING SITES BY 2D PARAMAGNETIC DIFFERENCE NMR SPECTROSCOPY

INTRODUCTION

With the ultimate goal of designing metal-complexing materials that are capable of recognizing individual proteins [1], it is necessary to identify the key chemical properties that dictate the selectivity and strength of interactions between metal complexes and individual protein surface histidines. A model system of bis-metal ion receptors has shown relatively high selectivity (>10) among complementary bis-imidazole "protein analogs" in which the distance between imidazole functionalities differed by as little as 4 Å [2,3] (see also Chapter 2). In this investigation, paramagnetic effects on 2D NMR spectra are used to describe quantitatively the binding of metal complexes to specific histidines on two model proteins, cytochrome *c* and myoglobin.

By altering specific properties of metal complexes, it is possible to modulate their interactions with protein surfaces. For example, the choice of metal ion dictates which protein functional groups are targeted. In the case of Cu^{2+} at neutral pH, the relevant groups are primarily the imidazole groups of histidine side chains and, to a lesser degree primary amine groups of lysine side chains and the amino-terminus [4]. Indeed, the selectivity of separations that use copper (II) iminodiacetate (CuIDA) derivatized materials is based primarily on the histidines accessible on the protein surface [5,6] (see also Chapters 3 - 5). The choice of metal-chelating ligand can also alter the selectivity for individual proteins [7]. The most common chelating ligand used for divalent metal ions is iminodiacetate, or IDA, which binds three metal coordination sites via an amine and two carboxylate groups. This results in a complex with no net charge, as illustrated in Figure 6.1. Metal complexes derived from diethylenetriamine, or DIEN, have been used to target imidazole groups (see Chapter 2 [3]). In addition to IDA, the two DIEN-based chelating ligands shown in Figure 6.1B and 6.1C are used in this investigation.

The unpaired electronic spin of paramagnetic species, such as copper complexes, dramatically shortens the lifetime of nuclear magnetic resonance (NMR) signals of nearby nuclei. In the case of copper complexes, the result is an increase in the relaxation rate of the NMR signal for protons near the specific protein binding sites of the paramagnetic metal complex [8]. Because paramagnetic relaxation is a transient process, the lifetime of the protein-metal complex dictates which type of information can be derived from the increased relaxation rates of protein NMR signals [9]. Under certain conditions, these relaxation rates are sensitive to the rate of chemical exchange, allowing the interpretation of binding kinetics from NMR spectra [10]. Under other conditions, paramagnetic relaxation effects are instead sensitive to the distance between the observed nucleus and the metal center, making them an excellent tool to explore macromolecular structure [11, 12].

A quantitative analysis of the effect of copper complexes on NMR of protein surface histidines requires a general description of the relaxation processes of divalent copper on protons of coordinated imidazole. For small copper complexes, paramagnetic effects on the transverse (T_2, or spin-spin) relaxation rate are dominated by through-bond interactions [13]. Effects on longitudinal (T_1, or spin-lattice) relaxation rates are dominated by through-space interactions. In this investigation, the relaxation of *N*-acetylhistidine in the presence of copper iminodiacetate is compared to parameters reported for other copper complexes [14], taking into account the simultaneous multiple equilibria predicted from known equilibrium constants. These parameters indicate that paramagnetic effects of copper on directly coordinated histidines will be dominated by through-bond effects, while through-space effects will dominate for nearby residues. In order to interpret these relaxation phenomena in proteins, it is necessary to have estimates of three parameters: (1) the electronic relaxation rate of the copper ion, (2) the

rotational correlation time of the copper-imidazole complex, and (3) the scalar coupling constants for copper-coordinated imidazole.

Paramagnetic relaxation can often be distinguished in one-dimensional NMR spectra. As the size and complexity of proteins increase, however, it becomes necessary to resolve spectra in two or more dimensions. In 2D NMR, the resonance signatures of pairs of nuclei combine to form cross peaks; the resulting "map" of the relationships between the individual protons isolates individual protons that could not be distinguished in one-dimensional spectra. In 1D NMR, the effect of increasing relaxation rates is to broaden the linewidth of the NMR signal of protons near the metal center. In 2D NMR, the effect is to reduce the intensity of cross peak derived from pairs of protons near the metal center. Thus, a 2D NMR difference spectrum can be formed by subtracting the spectrum recorded in the presence of metal complex from a suitable control spectrum; those protons nearest the binding site appear as the largest peaks in the 2D difference spectrum [15]. This approach has been used to identify amino acids which are surface accessible [16] or near the binding sites of spin-labeled substrate analogs from cross peaks in the 2D difference spectrum [15,17].

Using these techniques, the relaxation effects of several copper complexes for individual surface amino acids of cytochrome *c* (isolated from yeast, horse, and tuna) and myoglobin (horse heart) are explored using 2D NMR spectroscopy. These highly conserved, monomeric proteins have been studied extensively by NMR [18,19,20] and X-ray crystallography [21,22,23,24]. Three different types of 2D NMR proton correlation experiments are used in this investigation: TOCSY [25], COSY [26], and 2Q [27]. The relative advantages of each of these experiments will be discussed in context (see Results and Discussion).

Although the three-dimensional structures and amino acid sequences are highly conserved among the three cytochromes *c*, as shown in Figure 6.2 and Table 6.1, they

differ in their surface histidine content and in their ability to bind immobilized CuIDA, as demonstrated by immobilized metal affinity chromatography experiments (see Chapter 3). In this investigation, paramagnetic difference 2D NMR spectra of these proteins are calculated from 2D spectra recorded in the presence of copper complexes. Only the cross peaks with a greater relaxation rate in the presence of copper complex relative to a control spectrum appear in the 2D difference spectrum, limiting effects to those residues near surface histidines. These effects are used to identify and compare specific copper binding sites conserved among cytochromes *c*.

Horse heart myoglobin has a total of eleven histidines (Table 6.2), but according to the crystal structure [24], only four of them (H48, H81, H113, H116, see Figure 6.3) have greater than 30% of the surface area at the imidazole nitrogens accessible to a probe the size of CuIDA, as reported in Table 6.3. Copper complexes have previously been shown to interact with these four histidines [28]. However, attempts to quantify metal complex binding to surface histidines have been limited to the effect of histidine substitutions [28] or choice of metal ion and chelating ligand [29] on the *overall* rate of heme oxidation by the metal ions. Relative binding affinities of *individual* surface histidines for copper complexes can instead be measured directly by paramagnetic relaxation effects in 2D NMR. A quantitative relationship between the 2D cross peak intensity of a yeast cytochrome *c* histidine and its binding constant for a copper complex is derived by observing how 2D cross peak intensity decreases with increasing concentration of CuIDA. This relationship is also applied to measure the relative affinity of two different copper complexes (CuDIEN and bis(CuDIEN)) for four individual surface histidines of myoglobin.

These results taken together will demonstrate that individual surface histidines differ in their abilities to bind certain copper complexes, suggesting that it may be possible to target specific protein surface histidines with particular metal complexes.

These capabilities will be invaluable in exploring the molecular mechanisms behind metal affinity separations and protein-metal interactions in general.

CALCULATIONS

Paramagnetic relaxation of nuclei by copper(II)

The relaxation rate of the unpaired electrons (T_{1e}^{-1}, s^{-1}) determines how a paramagnetic metal affects the chemical shift and relaxation rate of observed nuclei. In general, paramagnetic shifts will dominate for those species with faster electronic relaxation rates (e.g., $T_{1e}^{-1} > 10^{10}$ s^{-1} for Fe^{3+}), and paramagnetic relaxation will dominate for those with slower electronic relaxation rates (e.g., $T_{1e}^{-1} < 10^{9}$ s^{-1} for Cu^{2+}) [8]. In the case of divalent copper, two relaxation processes, spin-rotation [30] and hyperfine-anisotropy [31], contribute to the net electronic relaxation rate. The contribution of spin-rotation (SR) to electronic relaxation is given by

$$\left.\frac{1}{T_{1e}}\right|_{SR} = \frac{1}{9\tau_R}\left[\left(g_{\parallel}-g_e\right)^2 + 2\left(g_{\perp}-g_e\right)^2\right], \qquad (1)$$

where τ_R is the rotational correlation time (s), and $g_{\parallel}$, $g_{\perp}$, and g_e ($g_e = 2.0023$) are the electronic g-factors of the copper complex. The contribution of hyperfine anisotropy (HA) to electronic relaxation is given by

$$\left.\frac{1}{T_{1e}}\right|_{HA} = \frac{8\pi^2}{15}\left[\frac{\left(g_{\parallel}-g_{\perp}\right)^2 \mu_B^2 H_o^2}{h^2} + \frac{\left(A_{\parallel}-A_{\perp}\right)^2 S(S+1)}{3}\right]\left(\frac{\tau_R}{1+\omega_s^2\tau_R^2}\right), \qquad (2)$$

where μ_B is the Bohr magneton (9.27×10^{-24} J T^{-1}), h is Planck's constant (6.626×10^{-34} J s), H_0 is the magnetic field strength (11.75 T), ω_S is the electronic Larmor frequency (2.1×10^{12} rad s^{-1}), S is the unpaired electronic spin (½ for Cu^{2+}) and $A_{\|}$, $A_{\perp}$ are the electronic hyperfine coupling constants of the copper complex (s^{-1}). The rotational correlation time (τ_R) is presumed to follow the Stokes-Einstein relationship [8],

$$\tau_R = \frac{4}{3} \cdot \frac{\pi a^3 \eta}{k_B T} \quad , \tag{3}$$

where a is the radius of the molecule (m), η is the viscosity of D_2O (Pa s) [32], k_B is Boltzmann's constant (1.38×10^{-23} J K^{-1}) and T is the absolute temperature (K). In this investigation, it is presumed that the volume and therefore rotational correlation time of copper complexes scales as the molecular weight, including coordinated water.

In the high-field limit, where the Larmor frequency of the electron is fast compared to the rotation rate of the copper complex ($\omega_S \tau_R >> 1$), both spin-rotation and hyperfine anisotropy relaxation are proportional to the rotation rate. In this limit, the longitudinal electronic relaxation rate (T_{1e}^{-1}) is proportional to the rotation rate and can be expressed as [33]

$$\frac{1}{T_{1e}} = \left.\frac{1}{T_{1e}}\right|_{SR} + \left.\frac{1}{T_{1e}}\right|_{HA} = c_{1e} \frac{1}{\tau_R} \quad , \tag{4}$$

where the proportionality constant c_{1e} is given by

$$c_{1e} = \frac{1}{9}\left[\left(g_{\parallel} - g_e\right)^2 + 2\left(g_{\perp} - g_e\right)^2\right] + \frac{8\pi^2}{15\omega_s^2}\left[\frac{\left(g_{\parallel} - g_{\perp}\right)^2 \mu_B^2 H_o^2}{h^2} + \frac{\left(A_{\parallel} - A_{\perp}\right)^2 S(S+1)}{3}\right] . \quad (5)$$

In this investigation, the parameter c_{1e} (9×10^{-3}) is calculated from electronic g and A values of copper in bis-(cyclo-*l*-histidyl-*l*-histidyl)copper(II) ($g_{\parallel}$ =2.253, $g_{\perp}$ =2.050, $A_{\parallel}$ =5.7×10^8 s^{-1}, and $A_{\perp} = 9.0 \times 10^7$ s^{-1}) [34]. This value is in close agreement with that previously reported for copper complexes with histidine ($c_{1e} = 7 \times 10^{-3}$) [14].

Paramagnetic electrons relax nearby nuclear spins by scalar (through-bond) and dipolar (through-space) mechanisms. The *paramagnetic relaxation rate* (T_{2M}^{-1}) of an observed nucleus, the relaxation rate while in contact with the coordinated metal ion, is the sum of these two relaxation processes,

$$\frac{1}{T_{2M}} = \left.\frac{1}{T_{2M}}\right|_{scalar} + \left.\frac{1}{T_{2M}}\right|_{dipolar} . \quad (6)$$

In copper complexes, scalar relaxation processes contribute significantly only to the transverse (T_2) relaxation. Under the conditions of these experiments ($\omega_s T_{1e} >> 1$), the contribution is given by [8]

$$\left.\frac{1}{T_{2M}}\right|_{scalar} = \frac{4\pi^2}{3} S(S+1) A_c^2 T_{1e} , \quad (7)$$

where S is the unpaired electronic spin (½ for Cu^{2+}) and A_c is the scalar coupling constant (s^{-1}), which describes the delocalization of unpaired electronic spin on the

observed nucleus. If the complex is in fast exchange among different conformations, then scalar relaxation effects are weighted among the different coupling constants resulting in an effective scalar coupling constant $\langle A_c \rangle$ given by

$$\langle A_c \rangle = \left[\sum A_{c,i}^2 \right]^{1/2}. \tag{8}$$

Dipolar relaxation processes contribute significantly to both longitudinal (T_1) and transverse (T_2) relaxation by copper complexes. Under the conditions of these experiments ($\omega_s \tau_R >> 1$), these contributions become [8]

$$\left. \frac{1}{T_{1M}} \right|_{dipolar} = \frac{2}{15} \left(\frac{\mu_o}{4\pi} \right)^2 \frac{\gamma_I^2 g_e^2 \mu_B^2 S(S+1)}{r^6} \left(\frac{3\tau_R}{1+\omega_I^2 \tau_R^2} \right), \tag{9}$$

$$\left. \frac{1}{T_{2M}} \right|_{dipolar} = \frac{1}{15} \left(\frac{\mu_o}{4\pi} \right)^2 \frac{\gamma_I^2 g_e^2 \mu_B^2 S(S+1)}{r^6} \left(4\tau_R + \frac{3\tau_R}{1+\omega_I^2 \tau_R^2} \right), \tag{10}$$

where ω_I is the nuclear Larmor frequency (3.1×10^9 rad s^{-1} for ^1H, 7.9×10^8 rad s^{-1} for ^{13}C), μ_0 is the permittivity of free space ($4\pi \times 10^{-7}$ H m^{-1}), μ_B is the Bohr magneton, g_e is the electronic g-factor (2.0023), γ_I is the nuclear magnetogyric ratio (2.67×10^8 rad T^{-1} s^{-1} for ^1H, 6.73×10^7 rad T^{-1} s^{-1} for ^{13}C), and r is the distance between the observed nucleus and the metal center. Again, if the complex is in fast exchange among different conformations [35], then dipolar relaxation effects are weighted among the difference distances resulting in an effective distance <r> given by

$$\langle r \rangle = \left[\sum \frac{1}{r_i^6} \right]^{-1/6}. \tag{11}$$

This expression is analogous to distance constraints derived from NOESY experiments [36], and biases the observed distance toward the distance of closest approach.

The efficiency of relaxation depends on the ability of the paramagnetic species to sample all the observed nuclei. The observed relaxation rate therefore depends on the fraction of observed nuclei bound to the copper complex at equilibrium (f_M) and on the lifetime of bound copper complex (τ_m) [8]

$$\frac{1}{T_{1P}} = f_M \frac{1}{T_{1M} + \tau_m} \quad , \tag{12}$$

$$\frac{1}{T_{2P}} = f_M \frac{1}{T_{2M} + \tau_m} \quad , \tag{13}$$

where f_M is the equilibrium mole fraction of bound metal complex, and T_{2P}^{-1} and T_{1P}^{-1} are the *net* increase in relaxation rate due to the metal complex. Under conditions of fast exchange ($\tau_m^{-1} >> T_{2M}^{-1}$), paramagnetic relaxation dominates the net relaxation process. However, under conditions of slow exchange (high affinity binding, $\tau_m^{-1} >> T_{2M}^{-1}$), the rate of chemical exchange dominates the net relaxation process.

Measuring binding constants from relaxation rates

In these experiments, the concentration of paramagnetic species is very much less than that of the observed nucleus (mole fraction $c_M << 1$). Under these conditions, the mole fraction of species bound (f_M) will be approximately linear in the molar ratio of metal complex compared to observed nucleus (c_M). The net increase in transverse relaxation rate (T_{2P}^{-1}) will therefore be linear in the total mole fraction of metal (c_M),

$$\frac{1}{T_{2P}} = Kc_M \frac{1}{T_{2M}} \quad , \tag{14}$$

where K is an apparent equilibrium binding constant [37,38]. Thus, given a reliable estimate of the paramagnetic relaxation rate (T_{2M}^{-1}), paramagnetic relaxation effects at a particular histidine provide a direct measure of its equilibrium binding constant (K) for the metal complex.

In a 1D NMR experiment, T_1 and T_2 relaxation rates can be measured directly. In a 2D experiment, the NMR signal amplitude decays exponentially during acquisition with a rate equal to the transverse relaxation rate (T_2^{-1}). Thus there will be some characteristic decay time for the experiment (τ_{exp}) such that the volume of a 2D cross peak (V) can be expressed by

$$V \propto \exp\left(-\tau_{exp}/T_2\right) \quad . \tag{15}$$

The characteristic decay time may represent the evolution time of the two-dimensional pulse sequence [15] or the acquisition time of the free induction decay [39]. In either case, this characteristic time will be constant for given set of data acquisition and spectral processing parameters. If transverse relaxation is due to a paramagnetic species, then the ratio of the cross peak intensity in the presence of paramagnetic species (given by V) to the intensity in its absence (given by V_0) will be given by

$$V/V_o = \exp\left(-\tau_{\exp}/T_{2P}\right) \quad , \tag{16}$$

where T_{2P}^{-1} is the net increase in relaxation rate due to the paramagnetic species as given by eqn. 13. By analogy to eqn. 14, under the conditions of these experiments the cross peak intensity will decay exponentially with increasing concentration of paramagnetic species,

$$V/V_o = \exp\left(-\tau_{exp} K c_M / T_{2M}\right) , \tag{17}$$

where K is the apparent equilibrium binding constant for the copper complex at that particular site, and T_{2M} is the paramagnetic relaxation rate (eqn. 6). Relaxation effects in 2D spectra thus provide the means to directly measure the equilibrium binding constant (K) of the metal complex for particular protein surface groups, given a reliable estimate of the paramagnetic relaxation rate (T_{2M}^{-1}).

RESULTS AND DISCUSSION

Paramagnetic relaxation of acHIS

The paramagnetic relaxation of *N*-acetylhistidine protons, labeled in Figure 6.4, was measured at fixed molar ratio of CuIDA to acHIS ($c_M = 3 \times 10^{-4}$) for acHIS concentration ranging from 0.5 to 100 mM. As shown in Figure 6.5, the increase in the T_1 relaxation rate of the acHIS protons is accurately described by calculating the concentrations of copper-containing species from known formation constants [40,41,42], as presented in Table 6.4. According to these calculations, the acHIS concentrations used in this investigation are sufficient to remove a majority of the copper ion from the IDA chelate. Although the formation constant for CuIDA is very large ($K = 3 \times 10^{10}$ M^{-1}), the reported formation constants for $Cu(acHIS)_3$ and $Cu(acHIS)_4$ are also large enough ($K_3 = 2 \times 10^{10}$ M^{-3}, $K_4 = 6 \times 10^{11}$ M^{-4}, Table 6.4) that acHIS, when present in

large excess, can effectively compete with IDA for free copper ion. The proton linewidths of copper-coordinate imidazole ($T_2^{-1} \sim 10^5$ Hz) necessitates large excesses of acHIS (3000 fold) to maintain observable NMR signals (linewidth < 100 Hz). In fact, at the maximum concentration used in this study (92 mM acHIS), approximately 90% of the copper ion is calculated to be in the form of $Cu(acHIS)_3$ and $Cu(acHIS)_4$.

The rotational correlation time is a critical parameter in determining both the T_1 and T_2 relaxation rates of the copper complexes. As presented in Table 6.4, the rotational correlation time was presumed to scale with the molecular weight of the complex including coordinated water. The rotational correlation times subsequently predicted for the ternary copper complexes, in particular $Cu(acHIS)(IDA)(H_2O)$ and $Cu(acHIS)_2(H_2O)_4$ (1.18×10^{-10} s and 1.46×10^{-10} s respectively) are in excellent agreement with that previously reported for Cu(*l*-HIS)$_2$ (1.25×10^{-10} s) [14].

The T_1 relaxation rate is dominated by dipolar effects and is predicted to fall off as the sixth power of distance of the proton from the copper center (eqn. 9). From this dependence an effective distance (<*r*>) of acHIS protons from the copper center can be calculated. As shown in Table 6.5, this distance increases for protons more remote from the imidazole nitrogens. The presence of multiple copper-containing species which change with increasing acHIS concentration prevents any meaningful discussion of the "structure" of the acHIS-CuIDA complex. Nevertheless, the calculated distances are more consistent with those observed in crystal structures of copper-histidine complexes representing N_1 coordination [43] than for those representing N_3 coordination [44,45,46], as illustrated in Figure 6.6. In fact, the observed distances are consistent with N_1 coordination 80% of the time and N_3 coordination 20% of the time, as calculated by eqn. 11.

While longitudinal (T_1) relaxation is dominated by dipolar effects, transverse (T_2) relaxation is instead dominated by scalar effects. The observed T_2 relaxation rates shown

in Figure 6.7 can also be predicted by known formation constants. The resultant effective scalar coupling constant ($<A_c>$) of the C_2 proton (Table 6.5) is in excellent agreement with that reported for the C_2 proton of copper complexes with histidine [14]. The observed coupling constant of the C_4 proton (0.62 MHz) is between that reported for N_3 coordination (0.68 MHz) and that estimated for N_1 coordination (0.57 MHz). The ^{13}C linewidths of acHIS imidazole carbons (data not shown) were also consistent with N_1 coordination a majority of the time.

Of course, the above description refers predominately to the structures of the complexes $Cu(acHIS)_3$ and $Cu(acHIS)_4$, and not that of the complex Cu(acHIS)(IDA). To accurately evaluate paramagnetic relaxation in such a mixed complex, a more appropriate chelating ligand would maintain complexes with only one equivalent of acHIS, even in the presence of large molar excesses of acHIS. One such chelating ligand is diethylenetriamine (DIEN, $K = 10^{16}$ M^{-1}). According to published equilibrium constants [40], greater than 99% of the copper ion would remain chelated by DIEN, even at 92 mM acHIS.

Although detailed structural information could not be obtained for acHIS-CuIDA, the relaxation parameters obtained for copper paramagnetic relaxation of acHIS lead to some important generalizations for paramagnetic effects in protein NMR. First, macromolecules such as proteins can be expected to have rotation correlation times at least an order of magnitude longer ($\tau_R > 10^{-8}$ s [8]) than these small molecules. Thus, as illustrated in Figure 6.8, a protein such as cytochrome *c* ($\tau_R \sim 10^8$ s^{-1}) may have an order of magnitude higher paramagnetic relaxation rate ($T_{2M}^{-1} \sim 10^6$ s^{-1}) than the acetyl histidine complexes. At such rotation rates, longitudinal (T_1) relaxation effects will not be observable due to line broadening. All relevant information must therefore be gleaned from T_2 relaxation effects. Second, as seen with acetylhistidine, T_2 relaxation rates for imidazole protons of histidines directly coordinated by copper will be dominated by

scalar paramagnetic effects. Because the predicted effects at the C_2 and C_4 protons are very similar (see Figure 6.8), it is inconsequential which imidazole nitrogen (N_1 or N_3, see Figure 6.7) coordinates the copper ion. Third, T_2 relaxation rates for surrounding amino acids not directly coordinated by copper will instead be dominated by dipolar relaxation effects, and these are strictly distance dependent. Thus any distance constraints of the metal binding site must be derived from paramagnetic effects at nearby amino acids and not from copper-coordinating histidines.

Paramagnetic relaxation and chemical exchange

The paramagnetic effect of a fixed concentration of CuIDA (28 μM) on the T_1 relaxation rate of the acHIS imidazole protons (92 mM) shown in Figure 6.9 decreases with increasing temperature over the range studied (25 C to 60 C). Such a decrease is consistent with the temperature dependence predicted by taking into account changes in the viscosity of D_2O and the formation constants of copper-containing species (see Figure 6.4) and *assuming a fast exchange limit* ($\tau_m^{-1} > T_{1M}^{-1}$). The chemical exchange rate of the acHIS (τ_m^{-1}) must therefore be significantly greater than 10^5 s^{-1} (eqn. 12), which is consistent with a value of 1.3×10^6 s^{-1} previously reported for Cu(*l*-HIS)$_2$ [14].

The reported chemical exchange rate of Cu(*l*-HIS)$_2$ [14] illustrates an important distinction between small molecules and proteins. The rate of chemical exchange is in fact that of the *self-exchange* reaction,

$$\text{A*-Cu} + \text{A} \rightleftharpoons \text{A-Cu} + \text{A*} \quad ,$$

where A* is the observed nucleus. In the case of Cu(*l*-HIS)$_2$, the observed exchange rate ($\tau_m^{-1} = 1.3 \times 10^6$ s^{-1}) was much faster than that of the estimated dissociation rate (~10^{-1} s^{-1}) calculated from the ratio of the bimolecular collision rate due (~ 10^8 M^{-1} s^{-1}) and the equilibrium association constant (~10^9 M^{-1}). As a result, the observed exchange rate was

presumed to result from the first-order dissociation of a single imidazole moiety of histidine, followed by rapid replacement with an imidazole of another histidine. However, if A* is a large protein rather than a small molecule, then it is unlikely that imidazole moieties from different proteins could simultaneously coordinate the same metal ion. Under such conditions, the self-exchange reaction portrayed above would require the ligand to dissociate completely from the protein binding site. Thus the exchange rate of a protein-ligand interaction is likely to be dominated by the *dissociation rate*, as observed by Thieuralt *et al.* [9] for antibody-antigen interactions.

The procedure developed to measure binding constants directly from relaxation rates (eqn. 14) presumes that the complexes are in the fast exchange limit. To satisfy such conditions, the metal complex must therefore dissociate from the protein at a rate faster than relaxation rate of histidine imidazole protons ($\tau_m^{-1} > T_{2M}^{-1} \sim 10^6\ s^{-1}$). While such rates have been realized with small molecules, they may not always be accessible with proteins. For monodentate interactions, such as copper coordinating individual surface histidines ($K \sim 10^3\ M^{-1}$), the dissociation rate should be comparable to the exchange rate observed for acHIS, and fast enough to meet the requirements of the proposed mechanism. However, for multidentate interactions typical of high-affinity surface metal chelation sites ($K \sim 10^6\ M^{-1}$ [47]), the dissociation rate would likely be orders of magnitude slower, and therefore would not satisfy the fast exchange limit. Under these conditions eqns. 14 and 17 are no longer valid, and it would not be possible to measure the fraction of metal ion bound from observed relaxation rates without prior knowledge of the exchange rate. This complication limits the quantitative application of paramagnetic relaxation in copper-histidine interactions to measuring binding constants of order $10^3\ M^{-1}$.

Paramagnetic relaxation of cytochrome c surface histidines

The NMR of protein surface histidines, like that of acHIS, is dramatically affected by coordinating copper complexes. For example, adding small amounts of CuIDA to yeast cytochrome c (*S. cerevisiae* C102S variant) specifically broadens the NMR signals of the three surface histidines, as shown in Figure 6.10. It is clear that these three histidines (H26, H33, and H39) are affected to differing degrees by the same concentration of CuIDA. The crystal structure of the yeast protein indicates the imidazole ring of histidine 39 to be freely accessible to the solvent (see Chapter 3). Histidine 33 is also accessible, but somewhat constrained by the carboxyl-terminal alpha helix. Finally, the imidazole nitrogens of histidine 26 are involved in hydrogen bonds to backbone heavy atoms and thus are only sparingly accessible. The line broadening at histidine 26 is much less than at the other histidines (Figure 6.12), reflecting its reduced accessibility.

As predicted by eqn. 14 and shown in Figure 6.11, the T_2 relaxation rate of surface histidines increases linearly with the mole fraction of added CuIDA. As demonstrated with *N*-acetylhistidine, paramagnetic relaxation of imidazole ring protons is dominated by scalar interactions. The scalar coupling constants of all three protein surface histidines should be comparable to acHIS, so that transverse relaxation rates T_{2M}^{-1} would be approximately equal. Thus the different slopes of the titration curve shown in Table 6.6 (KT_{2M}^{-1}, see eqn. 14) reflect the different equilibrium binding constants of these individual histidines for CuIDA. In this context, the histidines at positions 33 and 39 have approximately 20-fold higher binding affinity for CuIDA than does the histidine at position 26. This is in excellent agreement with CuIDA metal affinity chromatography of single-histidine cytochrome *c* variants (see Chapter 3), where it was seen that variants with histidines 33 and 39 were retained much more strongly than those with histidine 26.

Cytochrome c COSY cross peak intensities

For relatively unique NMR signals (e.g., imidazole protons) of small proteins (e.g., cytochrome *c*), it is often possible to distinguish the paramagnetic effects in one-dimensional spectra. However, for aliphatic protons and for larger proteins, the one-dimensional NMR spectrum is so cluttered as to make distinction among individual amino acids impossible. In these cases, it is necessary to resort to two-dimensional NMR spectroscopy to isolate the resonance signals of individual protons. In 2D NMR, the increased relaxation rate of protons near the bound metal complex has the additional effect of reducing their 2D cross peak intensity.

Magnitude mode COSY spectra were employed to compare the effect of CuIDA on histidine proton relaxation rate (1D) to its effect on histidine proton cross peak intensity (2D). To quantify the effect of relaxation rates on cross peak intensity, it is imperative that the cross peak multiplet be of one phase. For example, in the phase-sensitive COSY experiment, cross peaks have a dispersive lineshape; a slight increase in relaxation rate results in a cancellation of the positive and negative components of the peak [48]. In contrast, for magnitude mode COSY spectra, all cross peaks are of the same sign.

Magnitude mode COSY C_2H-C_4H cross peak intensities at histidines 39 and 26 of yeast (*C. krusei*) cytochrome *c* measured at increasing concentrations of CuIDA are reported in Figure 6.12. The cross peak intensity decays exponentially with increasing T_2 relaxation rate, and therefore with increasing concentration of copper complex, in agreement with the proposed mechanism (eqn. 17). In fact, assuming an experimental decay time (τ_{exp}) equal to the free induction decay (FID) acquisition time as shown in Table 6.6, T_2 relaxation rates calculated from the decay of COSY cross peak intensities are in excellent agreement with those measured from linewidths in 1D spectra. The

relative rates of decay of the COSY cross peaks therefore indicate the relative binding constants of the histidines for the copper complex. Thus histidine 39 of yeast cytochrome c displays approximately a 20-fold higher affinity for CuIDA than does histidine 26. Using the paramagnetic effect of CuIDA upon both one-dimensional and two-dimensional spectra of cytochrome *c*, it is possible to measure directly the relative affinity of *individual* surface histidines for the metal complex.

Previous investigators have calculated 2D difference spectra by subtracting spectra recorded in the presence of relaxation reagent from those recorded in its absence [15]. This is possible only when site-specific binding is very strong, so that the concentration of free paramagnetic species is negligible under experimental conditions. Free paramagnetic species have been shown to relax nonspecifically protons on the protein surface [16]. Because the metal coordination interactions studied here are weak ($\sim 10^3\ M^{-1}$) compared to biological binding interactions ($>10^6\ M^{-1}$), in some cases relatively high concentrations of copper complex (> 0.1 mM) were used, which in turn resulted in substantial amounts of free paramagnetic species. Thus in some cases it is necessary to account for the *nonspecific* relaxation of surface amino acids. To accomplish this, difference spectra were calculated by subtracting spectra recorded in the presence of Cu-IDA from spectra recorded in the presence of the same concentration of CuEDTA (which does not interact specifically with the protein). This is the first time that nonspecific paramagnetic effects have been corrected for in this manner.

The paramagnetic difference 2D NMR TOCSY has made it possible to map the surface sites at which Cu-IDA binds to cytochromes *c*. In pure adsorption mode TOCSY spectra, all cross peaks due to scalar coupling are of the same phase; thus the TOCSY experiments are particularly appropriate for mapping the binding sites of paramagnetic species [49]. For example, it is not possible to isolate individual α and β proton resonances in the one-dimensional spectrum of cytochrome *c*. However, the α-β cross

peaks of these histitidines appear clearly in the 2D TOCSY difference spectrum of yeast (*C. krusei*) cytochrome *c* (Figure 6.13A) in the presence of CuEDTA. In the presence of the same concentration of CuIDA, the 2D cross peaks of H39 are attenuated strongly, while those of H26 are attenuated only slightly (Figure 6.13B). In the resultant 2D difference spectrum, only those cross peaks with a greater relaxation rate in the presence of CuIDA (compared to the same concentration of CuEDTA) can be distinguished. Thus the cross peaks of H39 appear strongly in the 2D difference spectrum (Figure 6.13C), while those of H26 appear only very weakly. Indeed, the signal attenuation by CuIDA, as measured by the ratio of cross peak volumes shown in Table 6.7, is much stronger for α-β protons of H33 and H39 than it is for those of H26. This is in excellent agreement with both 1D linebroadening and 2D COSY cross peak intensity measurements (Table 6.3).

Mapping CuIDA binding sites

Paramagnetic effects are not limited to the histidines directly coordinating the copper complex. The through-space effects enable the identification of amino acids in close proximity to CuIDA binding sites in cytochrome *c*. This is demonstrated by the amino acids surrounding residue 33, which are clearly manifested in TOCSY difference spectrum shown in Fig. 6.14. For example, histidine 33 is near the carboxyl-terminus of cytochrome *c* (lysine 103 in *C. krusei*). The complete spin system of lysine 103 is clearly visible in the difference TOCSY spectrum of this *C. krusei* cytochrome *c* (Figure 6.14C), while it is extremely difficult to distinguish it in the parent TOCSY spectrum (Figure 6.14A). This is a dramatic illustration of the selectivity of paramagnetic relaxation.

Paramagnetic relaxation at amino acid side chains not directly coordinated to the copper ion are limited to dipolar effects, which are strictly distance dependent (eqn. 10). Thus the TOCSY cross peaks of side chains in close proximity to bound CuIDA are more

strongly attenuated than those more distant, allowing a qualitative description of the local structure of the CuIDA binding site. For example, cytochromes *c* from yeast (*C. krusei*) and horse differ slightly in amino acid sequence (Table 6.1), particularly at the carboxyl terminus. As discussed above, histidine 33 is in close proximity to the c-terminus of cytochromes *c*. Comparing α-β cross peak volume ratios (Table 6.4), there are differences between the two proteins in the paramagnetic relaxation of amino acids near histidine 33 (tuna cytochrome *c* lacks histidine 33 and thus lacks observable effects in this region).

In the case of the yeast protein, TOCSY cross peaks from A101 and K103 (carboxyl terminus) are strongly attenuated by 0.15 mM CuIDA and those of G34 are less so. However, in the horse heart protein, cross peaks from residues S103 and E104 (carboxyl terminus) are strongly affected by the same concentration of CuIDA, and those of residues G34 and A101, less strongly. The strong effects at residues 101 and 34 in yeast cytochrome *c* are not duplicated in the horse protein, thus CuIDA bound at histidine 33 is farther removed from these residues in the horse protein than in the yeast isolate. This observation is consistent with differences in the local environment histidine 33 in the crystal structures of these two proteins [22,23] due to the extra residue at the carboxyl terminus of the horse protein (E104).

Myoglobin 2Q cross peak intensities

In the case of the cytochromes *c*, it was possible to assign the histidine cross peaks affected by copper complexes in one-dimensional spectra (Figure 6.12). In the case of myoglobin, however, individual histidine signals are not as well resolved from those of other aromatic protons. Approximately 6 - 8 peaks in the aromatic region of myoglobin 1D spectra are broadened by copper complexes such as CuDIEN. These

peaks were tentatively assigned as histidines based on their pH titration behavior (data not shown).

To assign these histidines individually, 2D double quantum (2Q) NMR spectra were acquired for horse myoglobin [50]. The 2Q spectrum has three advantages over COSY and TOCSY experiments for assigning myoglobin histidines. First, in the double quantum experiment, NMR signals resulting from zero-quantum (e.g., diagonal peaks) and multiple-quantum transitions are removed by phase cycling, optimizing the double-quantum signal from transitions between two individual protons (e.g., C_2H and C_4H of imidazole). Second, the longer double-quantum evolution time can be optimized to enhance nuclear transitions due to long-range coupling (e.g., 80 ms for imidazole protons, see Appendix). Finally, the chemical shift along the ω_1 axis corresponds to the *sum* of the chemical shifts of the pair of protons forming a double-quantum transition, distributing the cross peak map over twice the spectral width in this dimension.

In the myoglobin double quantum spectrum shown in Figure 6.15, cross peaks corresponding to the C_2H and C_4H resonances of the four surface histidines are clearly visible and distinct from other aromatic peaks. The 2Q experiment results in an antiphase cross peak multiplet structure (like the phase-sensitive COSY spectra) and as such is particularly sensitive to paramagnetic broadening. In the presence of copper complex (e.g., CuDIEN, molar ratio $c_M = 0.016$), the cross peaks of the four histidines are attenuated (H48, H81, H113, H116), while nearby cross peaks from aromatic residues are not significantly affected (Figure 6.16). These four histidines were previously identified by Hegetschweiler *et al.* to be specifically broadened by CuNTA [29]. According to the crystal structure [24], these are the only myoglobin histidines with greater than 30% of the area at the imidazole nitrogens accessible to a probe the size of CuIDA (see Table 6.3).

Although the antiphase multiplets of 2Q cross peaks are advantageous in assigning histidines specifically broadened by copper complexes, they complicate the calculation of 2D difference spectra and prevent the measurement of signal intensity by cross peak volumes (antiphase multiplets should integrate to zero volume). For these reasons, a 2D processing procedure was developed (see Appendix) to convert phase-sensitive 2Q data to magnitude-mode 2Q spectra more appropriate for difference spectra or cross peak volume measurements. In the magnitude-mode 2Q spectra, histidine imidazole cross peaks are still well separated from other aromatic signals (Figure 6.17A). In the presence of copper complex (Fig 6.17B), the cross peaks of the four histidines are still attenuated, while nearby cross peaks are not significantly affected; thus the signals from the four histidines appear clearly in the double quantum difference spectrum (Figure 6.17C).

Binding of copper complexes by myoglobin surface histidines.

According to the myoglobin crystal structure [24], two histidines at positions 113 and 116 are separated by approximately 8 Å (Figure 6.4). It has previously been demonstrated (Chapter 2) that a bis-metal complex can selectively recognize bis-imidazole species in which the imidazoles are approximately 8 Å apart. If this principle extends to protein binding, then these two surface histidines should distinguish between two analogous copper complexes, one containing a single metal ion (CuDIEN) and the other containing two metal ions separated by approximately 8 Å (bis(CuDIEN), see Figure 6.1).

To compare the binding affinity of these two complexes for myoglobin surface histidines, 2Q cross peak intensities were measured at the same total *copper* concentration (32 μM CuDIEN, 16 μM bis(CuDIEN)). For both copper complexes, histidine cross peak volumes were attenuated relative to those volumes measured in the

absence of copper complex, as reported in Table 6.5. Comparing the degree of cross peak attenuation at the same total copper concentration, there is no evidence that histidines 113 and 116 are specifically targeted by the bis-metal complex. There is at least one reason why this receptor may have failed to target this pair of surface histidines. In the model system the analogous bis-mercury ion receptor ($\mathbf{R}_{dxd}$, see Chapter 2) showed only an order of magnitude selectivity in binding bis-imidazoles over 1-benzylimidazole. It can be seen from the crystal structure of horse myoglobin (Figure 6.3) that histidines 113 and 116 do not face in the same direction. In molecular modeling studies, it was necessary to rotate both of these histidines to simultaneously coordinate the two metal ions of bis(CuDIEN). If these rotations were sterically forbidden, or even marginally unfavorable (> 1 kcal/mol), they could abolish any advantage provided by simultaneous coordination of the two copper ions.

However, there does appear to be a difference in the ability of the individual surface histidines of horse heart myoglobin to distinguish between CuDIEN and bis(CuDIEN). Histidines 113 and 81 are relaxed to the same degree by CuDIEN and bis(CuDIEN) at the same total copper concentration. These two histidines show approximately twice the affinity for the bis-copper complex as they do for the single-copper complex; such behavior is expected on statistical grounds. In contrast, histidines 48 and 116 are relaxed to approximately half as much by bis(CuDIEN) as by CuDIEN at the same total copper concentration. These histidines display approximately the same affinity for the bis-copper complex as they do for the single copper complex. Therefore, histidines 113 and 81 show approximately a factor of two selectivity for the bis-copper complex over the single-copper complex, while histidines 48 and 116 show no selectivity.

Comparing the surface accessibility of these four histidines, those at positions 113 and 81 have greater than 50% of the accessibility of an unhindered imidazole, while

those at positions 48 and 116 have less than 50%. Thus it is not surprising that the relatively large bis-copper complex would have a reduced affinity for histidines 48 and 116 relative to the two more accessible histidines. In fact, the degree of cross peak attenuation due to bis(CuDIEN)$_2$ follows the order of increasing solvent accessibility, H113 > H81 > H48 > H116 (Table 6.3). The observation that some surface histidines are more sensitive to the size of the copper complex than others suggests that steric interactions from nearby amino acids could be used to target individual surface histidines of a particular protein.

CONCLUSIONS

Paramagnetic difference 2D spectroscopy has been shown to be a useful tool for examining interactions between protein surfaces and metal complexes. This technique can quantify the relative binding strength that metal complexes show for individual surface histidines. Approximate distance constraints obtained from relative relaxation rates can be used to describe the local structure of protein metal binding sites. In the case of both cytochrome *c* and myoglobin, the affinity of copper complexes for surface histidines increases with increasing surface accessibility as calculated from the crystal structures. Moreover, in the case of myoglobin, individual surface histidines show differing preferences for a bis-copper complex over a smaller complex with a single copper ion. Such interactions can be modulated either by the local environment of the protein (as demonstrated by the different cytochromes *c*) or by the structure of the copper ligand (as demonstrated using myoglobin).

These results taken together demonstrate that on the molecular level there is selectivity to individual copper-imidazole interactions at a protein surface. Individual surface histidines differ in their abilities to bind certain copper complexes, suggesting

that it may be possible to target specific protein surface histidines with particular metal complexes. These capabilities will be invaluable in exploring the molecular mechanisms behind protein-metal interactions, leading to the ultimate goal of designing materials that exhibit multiple, specific interactions with individual surface groups of a particular protein.

EXPERIMENTAL METHOD

Relaxation of acHIS by CuIDA

To evaluate the paramagnetic relaxation of CuIDA on *N*-acetylhistidine, the ratio of CuIDA to acHIS (3.0×10^{-4}) was fixed, and the total concentration of acHIS was varied from 0.5 mM to 100 mM by diluting a stock solution (28 μM CuIDA, 92 mM acHIS in D_2O plus 100 mM NaCl) in D_2O plus 100 mM NaCl. NMR experiments (^1H and ^{13}C) were obtained on an AMX 500 (Bruker) spectrometer operating at a ^1H frequency of 500.13 Mhz and a ^{13}C frequency of 125.73 MHz. Proton longitudinal (T_1) relaxation rates were measured using "fast" inversion recovery with a recycle delay time of 200 ms [51] and nonlinear regression of the resultant peak integrals with an exponential decay function (SigmaPlot). Transverse (T_2) relaxation rates were calculated directly from measured linewidths determined by fitting to a Lorentzian lineshape (UXNMR). Linewidth measurements were obtained only for imidazole protons because of the multiplet structure of C_α and C_β protons. For temperature effects on paramagnetic relation, T_1 relaxation rates were measured at fixed concentration (28 μM CuIDA, 92 mM acHIS, 100 mM NaCl in D_2O) for temperatures ranging from 25 to 60 C.

Multiple equilibria of acHIS-CuIDA

The concentration of copper-containing species in the multiple equilibria between Cu, IDA, and acHIS were calculated using the program SPE.FOR [41]. The formation constants included the following: (1) protonation of IDA and acHIS, (2) formation of CuIDA, $Cu(IDA)_2$, and $Cu(IDA)(OH^-)$, and (3) formation of Cu(acHIS), $Cu(acHIS)_2$, $Cu(acHIS)_3$, and $Cu(acHIS)_4$ and using known formation constants [40, 52]. The formation constant for Cu(acHIS)(IDA) was kindly provided by Dr. Barry Haymore [53]. Formation constants were adjusted for temperature effects by the Clausius-Claypeyron relationship. The enthalpies of formation of Cu(IDA) and $Cu(IDA)_2$ are reported in the literature [40]. The enthalpies of formation of $Cu(acHIS)_n$ were approximated by those of $Cu(Im)_n$, where Im is imidazole [40]. The enthalpy of formation of Cu(acHIS)(CuIDA) was estimated as that of Cu(IDA) plus that of Cu(Im).

NMR of cytochromes c

The CuIDA and CuEDTA solutions were prepared by adding 5 mM cupric sulfate to a 5 mM solution of either iminodiacetic acid disodium salt (IDA) or ethylenediamine-tetraacetic acid disodium salt (EDTA) and titrating with sodium hydroxide to pH 7.0.

Cytochromes *c* from yeast (*Candida krusei*), horse heart, and tuna heart (Sigma Chemical Co.) were used without further purification. Recombinant *Saccharomyces cerevisiae* iso-1-cytochrome c (C102S variant) was produced and purified as described previously [4]. Protein samples (0.5 mM) were oxidized in a solution of phosphate buffer (50 mM potassium phosphate, 200 mM NaCl, pH 7.8) and 1.5 mM potassium ferricyanide. The protein was then concentrated by ultrafiltration in an Amicon 8050 stirred ultrafiltration cell and rediluted with phosphate buffer containing 1.0 mM EDTA to remove free metal ions. This solution was dialyzed (1:1000 dilution) three times against phosphate buffer.

Cytochrome *c* samples for NMR were concentrated to 1.0 mM by ultrafiltration. These samples were then lyophilized, redissolved in 99% D_2O and exchangeable protons removed by heating at 37 C for six hours. The sample was lyophilized again, and finally redissolved in 99.96% D_2O under argon. At this time, copper complex (CuIDA or CuEDTA) stock solutions diluted in phosphate buffer (similarly lyophilized and dissolved in 99.96% D_2O) were added. The final cytochrome *c* concentration was 0.80 mM; the final copper complex concentrations ranged from 0.0005 to 0.20 mM.

Cyctochrome *c* ^1H 1D NMR were obtained on a Bruker AMX 500 spectrometer operating at 500.13 MHz. NMR spectra were recorded at 303K, and chemical shifts are referenced to an internal standard of 3-(trimethylsilyl)propane sulfonic acid (TSPSA). Protein samples (0.5 ml of 0.5 mM) were titrated with increasing amounts of CuIDA (0.004 - 0.02 mM). Transverse T_2 relaxation times were determined by Lorenztian deconvolution of the fourier-transformed spectrum (UXNMR).

Cytochrome *c* ^1H 2D NMR spectra were obtained on a Bruker AM 500 spectrometer operating at 500.13 MHz, controlled by an Aspect 3000 pulse programmer. NMR spectra were recorded at 303 K, and chemical shifts are referenced to TSPSA. The pulse programs were standard sequences from the Bruker automation program library (COSYHG and MLEV17PC). Magnitude mode COSY [26] spectra were recorded with 512 t_1 transients (48 scans) with a relaxation delay time of 20 ms. Phase sensitive TOCSY [25] spectra were recorded with 512 t_1 transients (60-80 scans) with a mixing time of 30 ms.

The 2D spectra were Fourier transformed and processed on a Sun 3/160 workstation (Sun Microsystems), using the FTNMR spectral analysis program (Hare Research). Magnitude mode COSY spectra were multiplied by a sine bell window in both dimensions before Fourier transformation, and cross-peak intensities were measured as the sum of cross peak volumes both above and below the diagonal. Phase sensitive

TOCSY spectra were multiplied by a $\pi/4$-shifted, squared sinebell window in both dimensions before Fourier transformation. Paramagnetic difference spectra [54] were calculated by subtracting the Fourier-transformed TOCSY spectrum recorded in the presence of CuIDA from the transformed TOCSY spectrum recorded in the presence of the same concentration of CuEDTA. Negative peaks (ROESY) were zeroed before subtraction. Tentative assignments for proton resonances of *C. krusei* ferricytochrome c [37] were made by comparing the TOCSY and COSY spectra with the published assignments for those residues in homologous ferricytochromes *c* [18,19].

2D NMR of myoglobin

Bis(CuDIEN) was formed by adding 1,4-bis(bis(2-aminoethyl)aminoethyl) benzene free base (synthesized as described previously [3]) to a solution of $CuCl_2$ in methanol. A pale green precipitate appeared and was filtered, washed with ether, and vacuum dried. The precipitate was found to be pure bis(CuDIEN) (1,4-bis(bis(2-aminoethyl)aminoethyl) benzene bis copper(II)) by elemental analysis [55]. An analogous procedure was used to form CuDIEN by adding $CuCl_2$ to diethylenetriamine.

Horse heart myoglobin (Sigma Chemical Co.) was dissolved in 100 mM NaCl and adjusted to pH 6.8. The solution was eluted through a 10 ml column of metal-free chelating Sepharose 4B to remove contaminating metal ions. The resulting solution was concentrated by ultrafiltration and rediluted with 100 mM NaCl. This sample was concentrated by ultrafiltration to 2 mM and lyophilized, and exchangeable protons were removed as described above.

Proton NMR spectra were obtained on a Bruker AMX 500 spectrometer operating at 500.13 MHz. After acquiring a 2D spectrum in absence of metal ion, 10 µl of a solution of copper complex (0.10 to 1.6 mM in D_2O) was added to 500 µl of protein solution and a second 2D spectrum was acquired. NMR spectra were recorded at 300 K,

and chemical shifts are referenced to a TSPSA internal standard. The double quantum (2Q) pulse program [27] was modified (see Appendix) to include a composite 180 degree refocusing pulse [50,56]. Phase sensitive 2Q spectra were recorded with 512 t_1 transients (64-96 scans) with a total mixing time of 80 ms.

The 2D spectra were Fourier transformed and processed on an Indigo workstation (Silicon Graphics), using the FELIX spectral analysis program (v. 2.1, Biosym Inc.). Phase sensitive 2Q spectra were multiplied by a $\pi/4$-shifted squared sinebell window in the t_2 dimension and a $\pi/2$-shifted squared sinebell window in the t_1 dimension. Magnitude mode 2Q spectra were created from phase sensitive aquisition data by rotating each t_1 transient to reverse the TPPI phase cycling [57] before multiplying by squared sinebell window in both dimensions and Fourier transforming (see Appendix). Paramagnetic difference 2Q spectra were calculated by subtracting the Fourier-transformed 2Q spectrum recorded in the presence of Cu-complex from the transformed spectrum recorded in its absence, accounting for dilution effects.

Molecular Modeling

Molecular modeling studies on yeast, horse heart, and tuna heart cytochromes *c* and horse heart myoglobin were performed using InsightII (v. 2.2, Biosym Inc.). The atomic coordinates of tuna heart ferricytochrome *c* were obtained from Brookhaven data bank. The estimated structure of *C. krusei* ferricytochrome *c* was created by direct substitution of unconserved amino acids to the atomic coordinates of the *S. cerevisiae* iso-1-ferricytochrome *c* crystal structure, obtained from the Brookhaven protein data bank. The atomic coordinates of horse heart metaquomyoglobin and horse heart ferricytochrome *c* were kindly provided by Dr. G. Brayer [23, 24]. Surface accessibility is measured as the total surface area at deprotonated N_1 and N_3 accessible to probe the size of CuIDA (1.93 Å [58]).

Table 6.1. Amino acid sequence of eukaryotic cytochromes *c*. Sequences of yeast (*Saccharomyces cerevisiae* iso-1 and *Candida krusei*), horse heart, and tuna heart cytochromes *c* are numbered relative to the horse heart isolate.

	-1	10	20	30	40
C. krusei:	PAPFEQ	GSAKKGATLF	KTRCAQCHTI	EAGGPHKVGP	NLHGIFSRHS
S. cerevisiae:	TEFKA	GSAKKGATLF	KTRCLQCHTV	EKGGPHKVGP	NLHGIFGRHS
horse heart:		GDVEKGKKLF	VQKCAQCHTV	EKGGKHKTGP	NLHGLFGRKT
tuna heart:		GDVAKGKKTF	VQKCAQCHTV	ENGGKHKVGP	NLWGLFGRKT

	50	60	70	80
C. krusei:	GQAEGYSYTD	ANKRAGVEWA	EPTMSDYLEN	PKKYIPGTKM
S. cerevisiae:	GQAEGYSYTD	ANIKKNVLWD	ENNMSEYLTN	PKKYIPGTKM
horse heart:	GQAPGFTYTD	ANKNKGITWK	EETLMEYLEN	PKKYIPGTKM
tuna heart:	GQAEGYSYTD	ANKSKGIVWN	NDTLMEYLEN	PKKYIPGTKM

	90	100	
C. krusei:	AFGGLKKAKD	RNDLVTYMLE	ASK
S. cerevisiae:	AFGGLKKEKD	RNDLITYLKK	ACE
horse heart:	IFAGIKKKTE	REDLIAYLKK	ARNE
tuna heart:	IFAGIKKKGE	RQDLVAYLKS	ATS

Table 6.2. Amino acid sequence of horse myoglobin.

10	20	30	40	50	60
GLSDGEWQQV	LNVWGKVEAD	IAGHGQEVLI	RLFTGHPETL	EKFDKFKHLK	TEAEMKASED
70	80	90	100	110	120
LKKHGTVVLT	ALGGILKKKG	HHEAELKPLA	QSHATKHKIP	IKYLEFISDA	IIHVLHSKHP
130	140	150			
GNFGADAQGA	MTKALELFRN	DIAAKYKELG	FQG		

Table 6.3. Calculated accessible surface areas of myoglobin surface histidines. Calculations were performed using coordinates from the high-resolution crystal structure of horse heart cytochrome c [24]. The accessible surface area reported is the total Connoly area at N_1 and N_3 accessible to a probe the size of CuIDA (1.93 Å radius [58]). All other histidines of horse heart myoglobin (see Table 6.2) have zero accessible surface area. The percent accessibility is calculated relative to the accessible surface area of an unhindered imidazole. The pK_a of the histidines is reported by Cocco *et al.* [20].

histidine	**pK_a**	**accessible surface area** ($Å^2$)	**percent accessibility**
36	7.8	7.9	27
48	5.46	14.5	49
81	6.65	17.1	58
97	N. A.	6.1	20
113	5.37	17.6	60
116	6.70	12.0	41
119	6.39	6.6	23

Table 6.4. Equilibrium formation constants for calculation of acHIS-Cu-IDA equilibria. The enthalpies of formation of $Cu(acHIS)_n$ were approximated by those of $Cu(Im)_n$, where Im is imidazole. The enthalpy of formation of Cu(acHIS)(CuIDA) was estimated as that of Cu(IDA) plus that of Cu(Im). Rotational correlation times (τ_R) are presumed to scale with molecular weight of complex including coordinated water (given in parentheses).

log *K* [ref.]	**ΔH** (kcal mol^{-1})	**IDA**	**Cu**	**acHIS**	**H^+**	**τ_R** (10^{-10} s)
9.34 [40]	--	1	0	0	1	--
11.89 [40]	--	1	0	0	2	--
10.57 [40]	-4.5	1	1	0	0	--
16.54 [40]	-10.9	2	1	0	0	--
3.00 [40]	--	1	1	0	-1	--
14.37 [53]	-12.1	1	1	1	0	1.18 (1 H_20)
7.01 [52]	--	0	0	1	1	--
4.35 [52]	-7.6	0	1	1	0	0.97 (5 H_20)
7.75 [52]	-14.0	0	1	2	0	1.46 (4 H_20)
10.30 [52]	-19.0	0	1	3	0	1.96 (3 H_20)
11.80 [52]	-23.0	0	1	4	0	2.45 (2 H_20)

Table 6.5. Paramagnetic relaxation parameters for acHIS-CuIDA complexes. **(A)** Effective copper-proton distance (<*r*>) compared to those calculated from crystal structures representing N_1 and N_3 coordination [43,44]. **(B)** Effective scalar coupling constant (<A_c>) compared to those reported for N_3 coordination [14].

(A) dipolar relaxation parameters

	Cu-H distance *r* (Å)		
proton	<*r*>	N_1[a]	N_3[b]
C_2H	3.20	3.2	3.2
C_4H	3.29 (0.01)	3.2	5.1
$C_\beta H2$	4.63 (0.04)	5.9-6.2	3.3-4.4
$C_\alpha H$	4.10 (0.04)	7.1	3.8
$C_{me}H_3$	6.75 (0.23)	5.6-7.2	3.6-3.8

(B) scalar relaxation parameters

	scalar coupling constant A_c (MHz)		
proton	<A_c>	N_1 (est.)	N_3[c]
C_2H	0.57 (0.02)	0.57	0.57
C_4H	0.62 (0.02)	0.57	0.68

[a] histidine 2 of bis(cyclo-(*l*-histidyl-*l*-histidyl))copper(II); C_{me} is C_α of histidine 1 [43].

[b] histidine from (glycyl-*l*-histidinato)copper(II); C_{me} is C_α of glycine [44].

[c] reported for Cu(*l*-HIS)$_2$ [14].

Table 6.6. Paramagnetic effect of CuIDA on histidine protons of yeast cytochrome c. **(A)** Calculated from nonlinear regression of histidine COSY diagonal- and cross-peak volumes of *C. krusei* cytochrome to eqn. 17 using τ_{exp}= 0.135 s (acquisition time). N.D. = not determined (H33 imidazole COSY cross peaks were of negligible intensity). **(B)** Calculated from linear regression of histidine C_2H transverse (T_2) relaxation rate to eqn. 14 from Lorentzian deconvolution of 1D NMR spectra of *S. cerevisiae* iso-1-cytochrome *c* C102S variant.

	(A) 2D NMR COSY cross peak volumes $K\,T_{2M}^{-1}$			**(B) 1D NMR** $K\,T_{2M}^{-1}$
histidine	**C_2H-C_2H**	**C_2H-C_4H**	**C_4H-C_4H**	**C_2H**
H_{26}	41 (7)	24 (2)	17 (4)	38 (6)
H_{39}	> 400	> 400	> 400	590 (80)
H_{33}	N.D.	N.D.	N.D.	720 (80)

Table 6.7. Paramagnetic effect of CuIDA on cytochrome *c* TOCSY cross peak intensity. Ratio of TOCSY cross peak volumes for 0.8 mM cytochrome *c* with 0.15 mM CuIDA (V) or 0.15 mM CuEDTA (V_0). **(A)** Surface histidines of cytochrome *c*. **(B)** Amino acids near residue 33 of cytochrome *c*. N.A. = cross peak not assigned.

(A) surface histidines

		TOCSY V/V_o		
histidine	**cross peak**	**yeast**	**horse**	**tuna**
H_{39}[a]	α-β	0.10	--	--
H_{33}[b]	α-β	0.11	0.10	--
H_{26}	α-β	0.96	0.79	0.98
H_{26}	α-β′	1.17	0.77	1.00

(B) amino acids near histidine 33

		TOCSY V/V_o		
amino acid	**cross peak**	**yeast**	**horse**	**tuna**
G_{34}	α-α′	0.42	0.96	N. A.
A_{101}	α-β	<0.10	0.60	0.95
K_{103}[c]	α-β	<0.10	<0.10	1.05
E_{104}[d]	α-β	--	<0.10	--

[a] K39 in horse and tuna. [b] W33 in tuna. [c] N103 in horse, S103 in tuna. [d] yeast and tuna sequences terminate at position 103 (Table 6.1).

Table 6.8. Paramagnetic effect of CuDIEN complexes on myoglobin surface histidine 2Q cross peak intensity. Magnitude-mode 2Q cross peak volumes were determined for aromatic cross peaks of horse heart myoglobin (2 mM), in presence (V) or absence (V_o) of CuDIEN complexes.

		2Q *V/Vo*		
		CuDIEN	**CuDIEN**	**bis(CuDIEN)$_2$**
amino acid	**cross peak**	3.2 μM	32.0 μM	16.0 μM
W_{14}	η_2-ζ_2	1.00	1.12	0.94
H_{81}	C_2H-C_4H	0.93	0.53	0.50
H_{116}	C_2H-C_4H	0.96	0.43	0.66
H_{48}	C_2H-C_4H	0.86	0.39	0.53
H_{113}	C_2H-C_4H	0.96	0.35	0.30

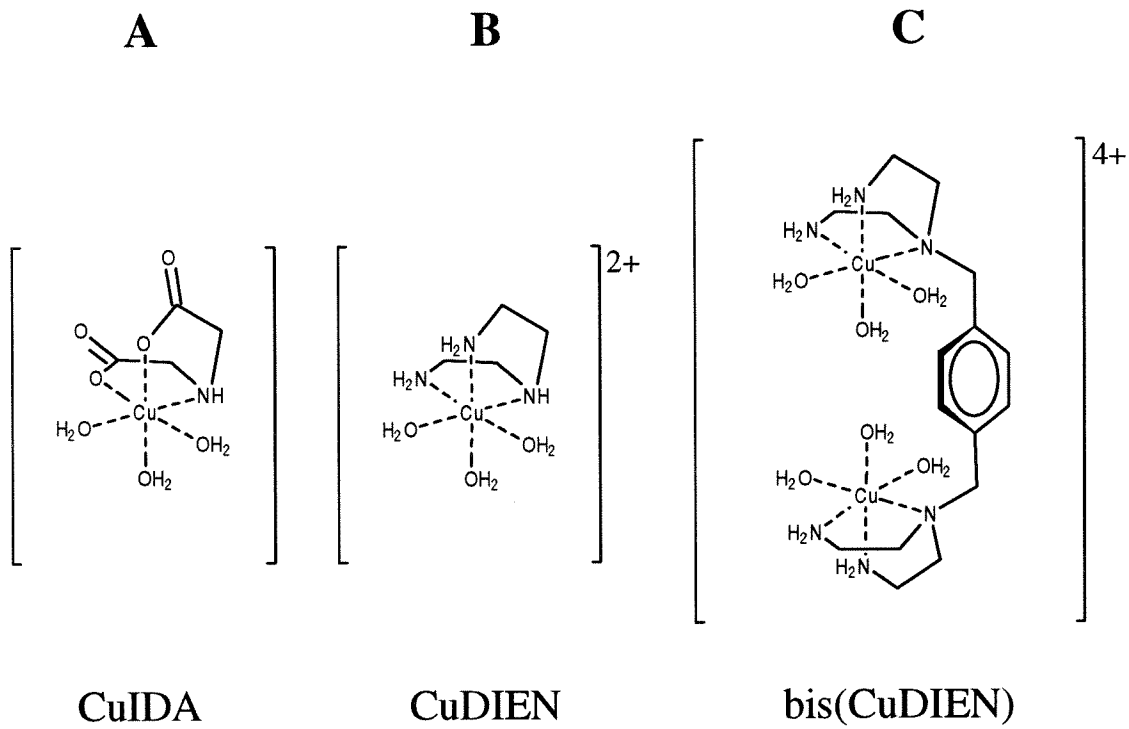


Figure 6.1. Copper ligands used to target protein surface histidines. **(A)** CuIDA (copper(II)iminodiacetate), **(B)** CuDIEN (copper(II)diethylentriamine), **(C)** bis(CuDIEN) (1,4-bis(bis(2-aminoethyl)aminoethyl)benzene bis copper(II)).

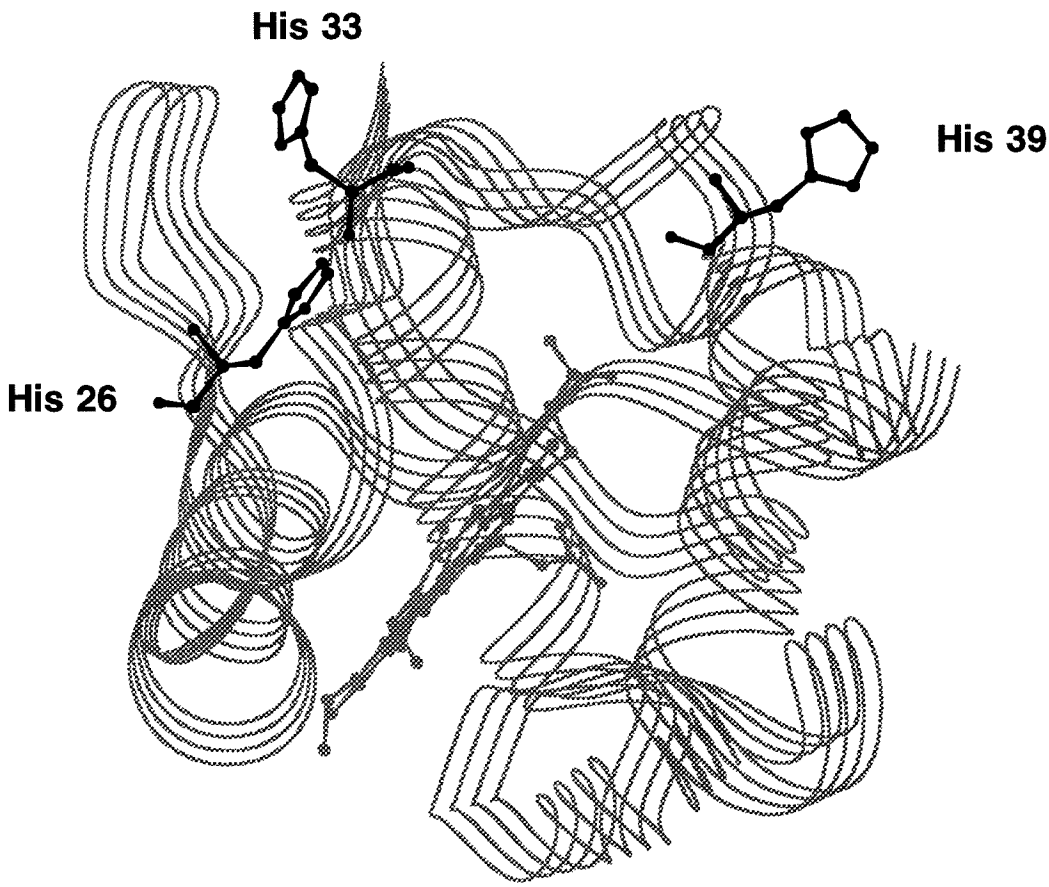


Figure 6.2. Surface histidines of yeast cytochrome *c*. Three-dimensional structure of *Saccharomyces cerevisiae* iso-1-ferricytochrome *c* determined by x-ray crystallography [36]. Surface-accessible histidines at positions 33 and 39 are indicated. The histidine at position 26 is only partially accessible to a probe the size of CuIDA (see Chapter 3).

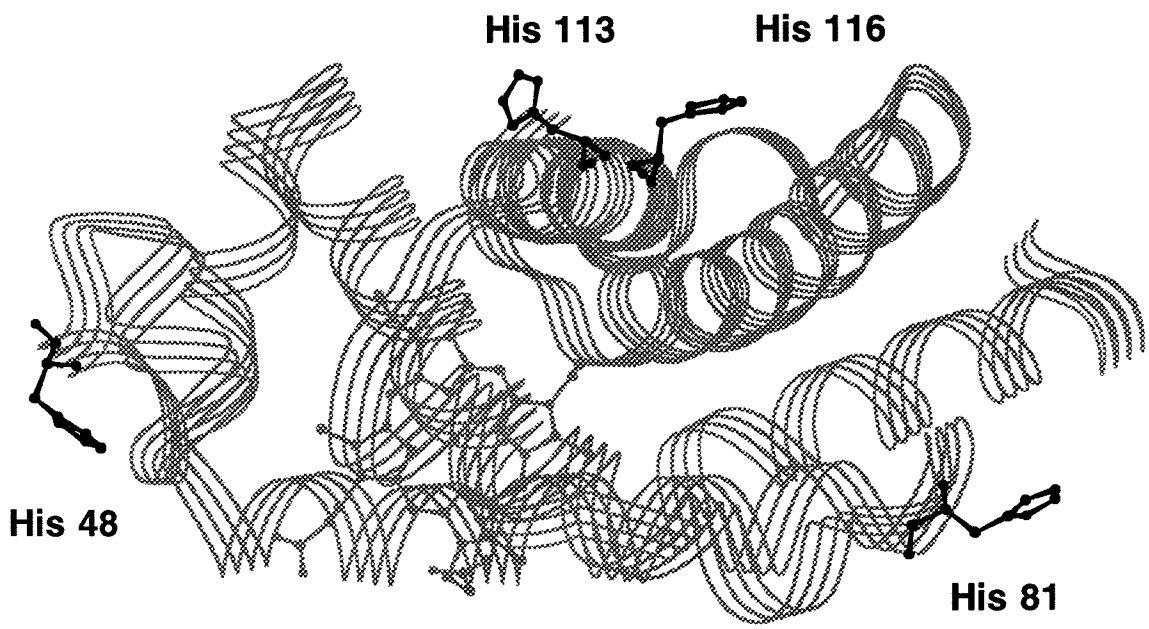


Figure 6.3. Surface histidines of horse heart myoglobin. Three-dimensional structure of horse heart metaquomyoglobin determined by x-ray crystallography [37]. Surface-accessible histidines (> 30% of the Connoly surface at the imidazole nitrogens accessible to a probe the size of CuIDA, see Table 6.3) at positions 48, 81, 113, and 116 are indicated.

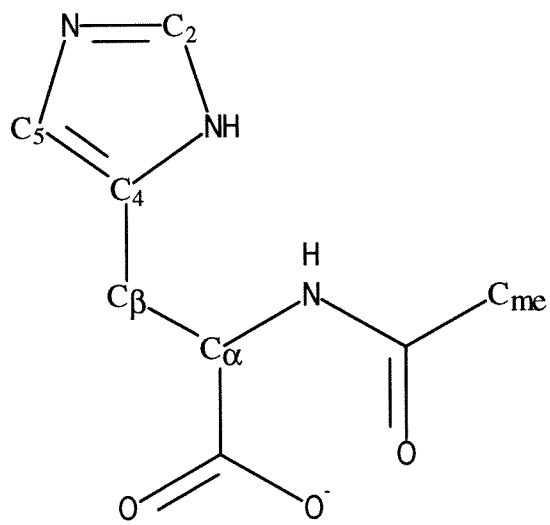


Figure 6.4. Nomenclature used in this study for *N*-acetylhistidine carbons. Protons are identified by the adjacent carbon atom.

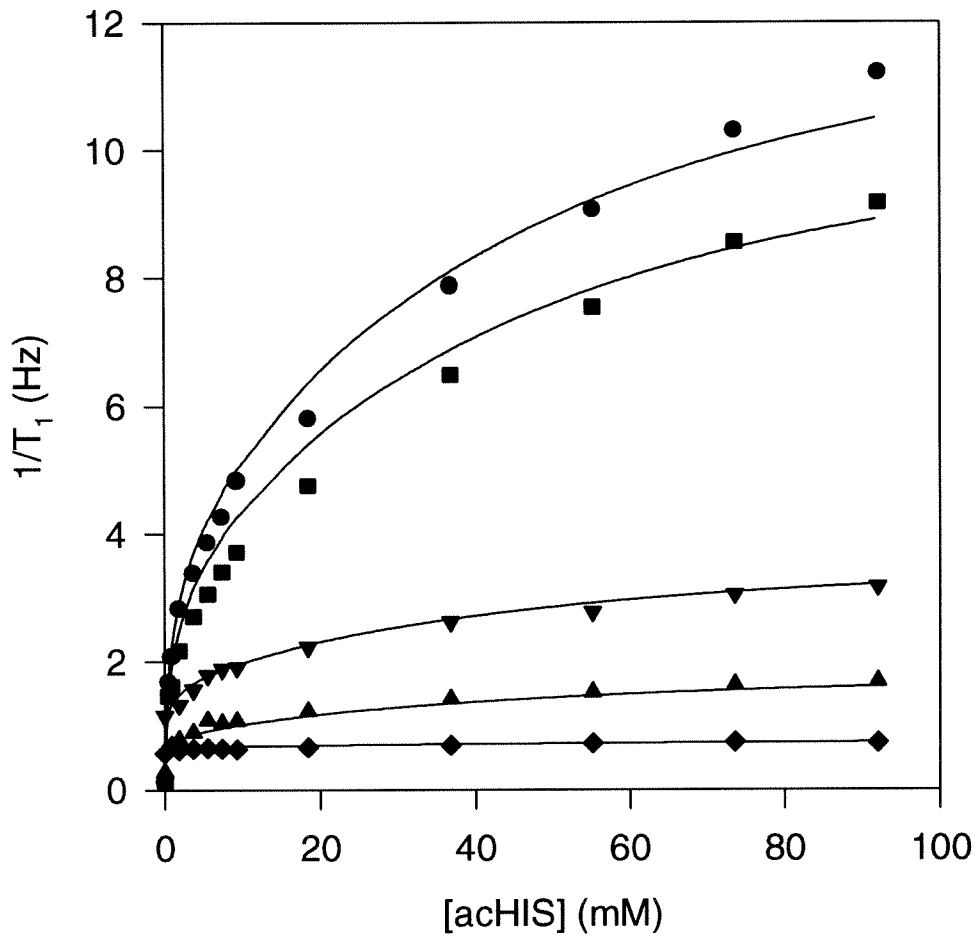
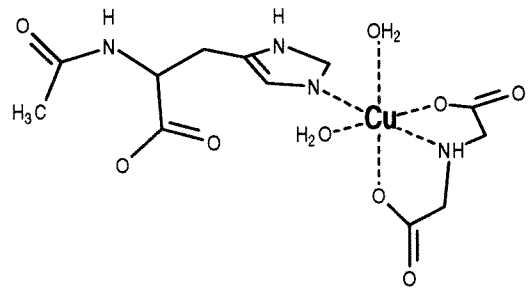
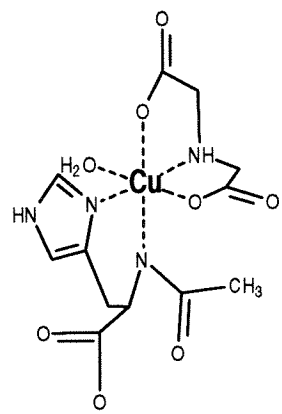


Figure 6.5. Longitudinal (T_1) relaxation rate of acHIS protons. Longitudinal relaxation rate measured by inversion recovery for increasing concentration of *N*-acetylhistidine at fixed molar ratio of CuIDA to acHIS ($c_M = 3.0 \times 10^{-4}$) for acHIS protons (●) C_2H, (■) C_4H, (▼) $C_\beta H_2$, (▲) $C_\alpha H$, (◆)$C_{me}H_3$. Curves calculated from parameters of Table 6.2.



(A) N_1 coordination



(B) N_3 coordination

Figure 6.6. Proposed CuIDA coordination at N_1 and N_3 imidazole nitrogens of acHIS. **(A)** Copper coordination at N_1 imidazole nitorgen, analogous to that of cyclo-(*l*-histidyl-*l*-histidyl) [39]. **(B)** Copper coordination at N_3 imidazole nitrogen and deprotonated peptide nitrogen, analogous to that of glycyl-*l*-histidine [40].

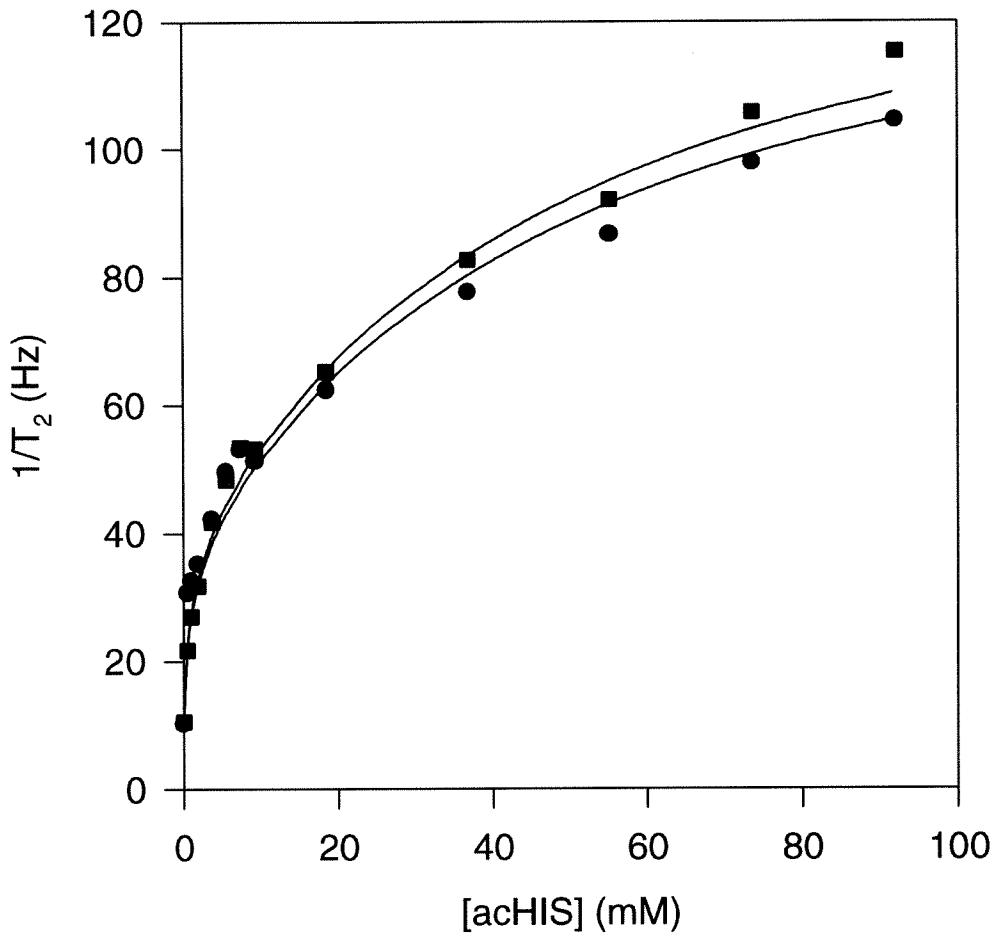


Figure 6.7. Transverse (T_2) relaxation rate of acHIS imidazole protons. Transverse relaxation calculated from linewidth measured by lorentzian deconvolution for increasing concentration of *N*-acetylhistidine at fixed molar ratio of CuIDA to acHIS ($c_M = 3.0 \times 10^{-4}$) for acHIS protons (●) C_2H, (■) C_4H. Curves calculated from parameters of Table 6.2.

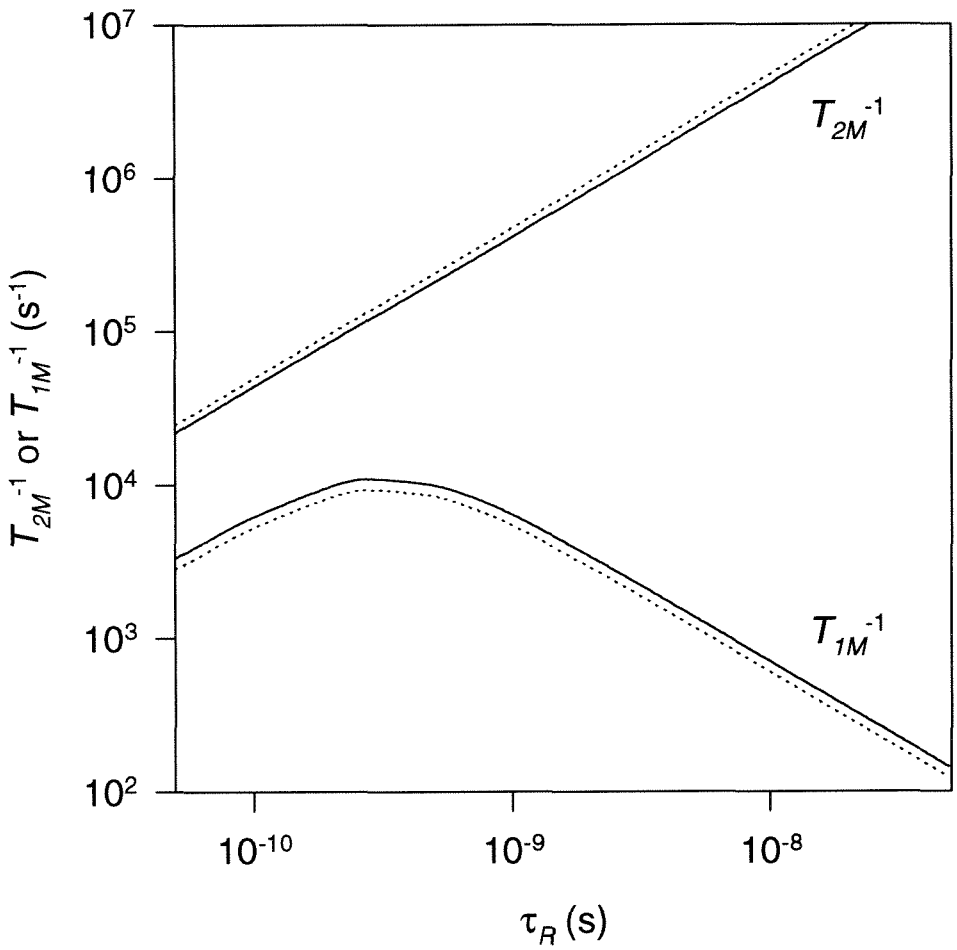


Figure 6.8. Calculated paramagnetic T_{1M} and T_{2M} relaxation rates of imidazole protons. Transverse (T_{2M}^{-1}) and longitudinal (T_{1M}^{-1}) relaxation rates calculated for increasing rotational correlation time (τ_R). Curves calculated for imidazole C_2H (—) and C_4H (····) protons of acHIS from parameters of Table 6.4.

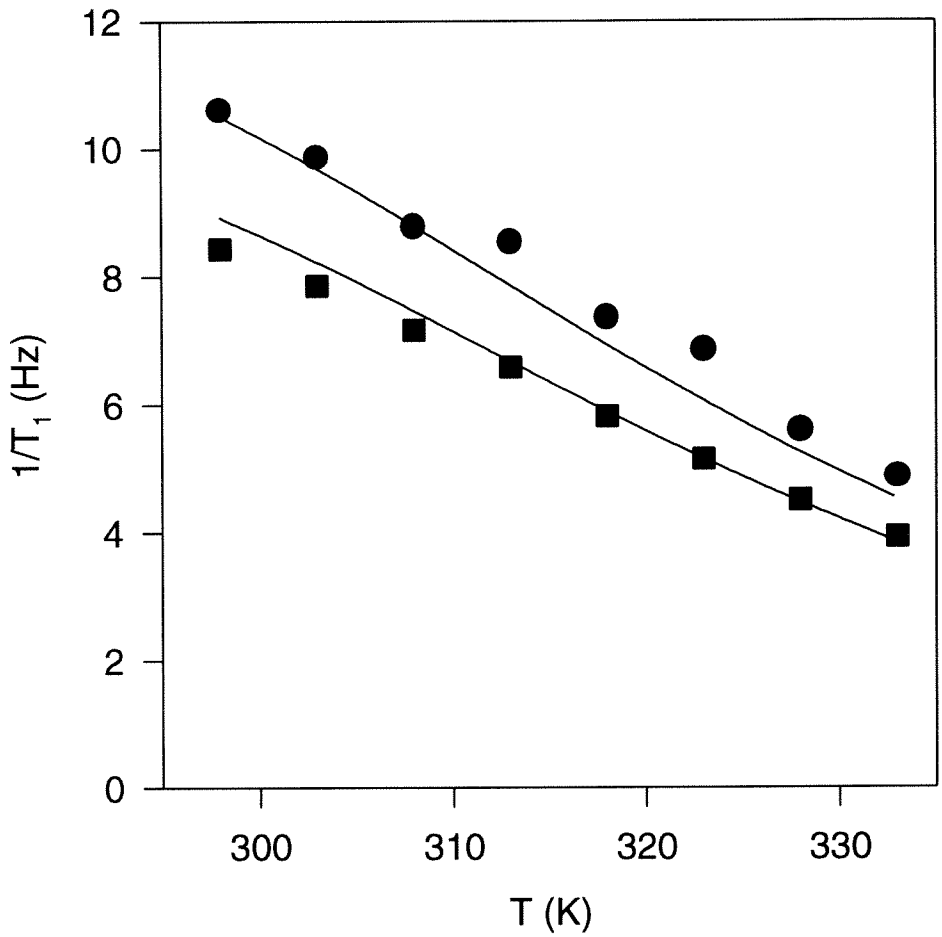


Figure 6.9. Effect of temperature on T_1 relaxation rate of acHIS imidazole protons. Longitudinal relaxation rate measured by inversion recovery for fixed concentration of *N*-acetylhistidine (92 mM) and CuIDA (28 μM, $c_M = 3.0 \times 10^{-4}$) at increasing temperature for acHIS protons (●) C_2H, (■) C_4H. Curves calculated from parameters of Table 6.4.

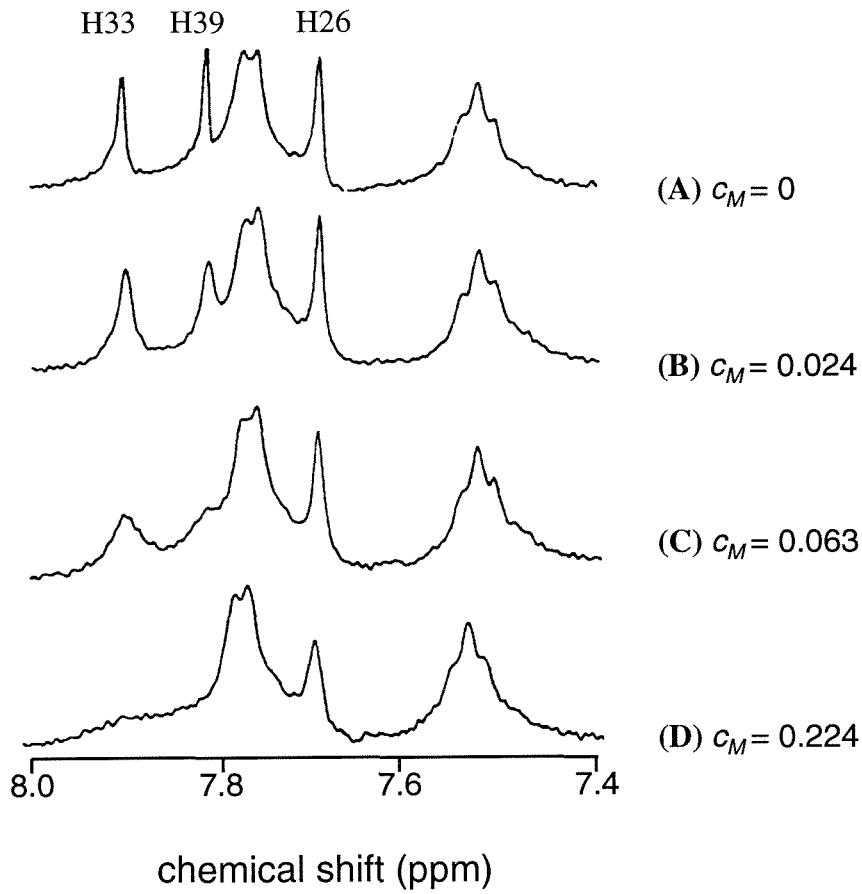


Figure 6.10. Paramagnetic relaxation of surface histidines of yeast cytochrome *c*. 1D NMR spectra of *S. cerevisiae* iso-1-cytochrome *c* C102S variant were recorded at 310 K, pH 7.0 for (A) 0.5 mM cytochrome *c*, (B) 0.45 mM cytochrome *c*, 11 μM CuIDA, (C) 0.43 mM cytochrome *c*, 27 μM CUIDA, (d) 0.38 mM cytochrome *c*, 85 μM CuIDA. NMR signals for histidines 26, 33, and 39 are indicated, along with molar ratio of CuIDA to protein (c_M).

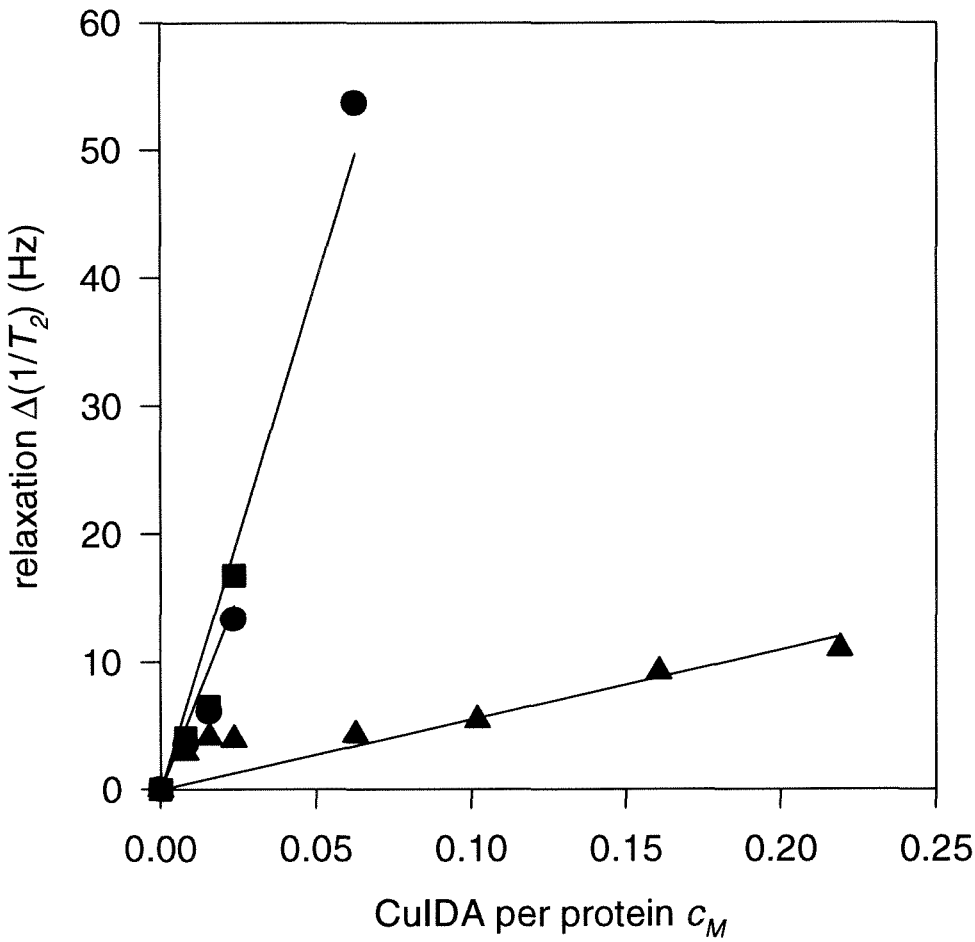


Figure 6.11. Transverse (T_2) relaxation rates of surface histidines of yeast cytochrome *c*. 1D NMR of S. cerevisiae iso-1-cytochrome *c* C102S variants were recorded at 310 K, pH 7.0 for 0.37 - 0.5 mM cytochrome *c*, 0 - 85 μM CuIDA. Increase in transverse relaxation rates determined by fitting Lorentzian lineshape to C_2H signals of (●) H33, (■)H39, and (▲) H26. Lines calculated from parameters of Table 6.5.

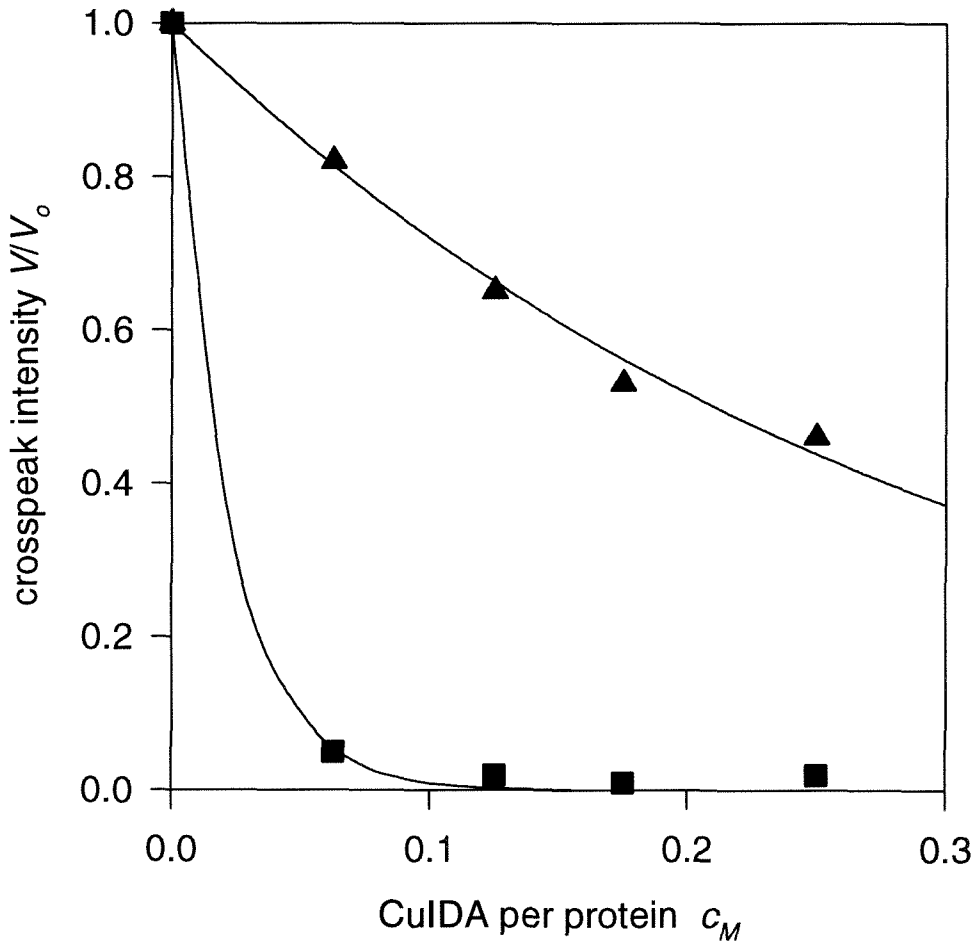


Figure 6.12. Intensity of histidine 2D COSY cross peaks of yeast cytochrome *c*. 2D magnitude mode COSY spectra of *C. krusei* cytochrome c (0.80 mM) were recorded at 303K, pH 7.85. Cross peak volumes in the presence of CuIDA (V) or CuEDTA (V_o) were measured at increasing concentration of copper complex for C_2H-C_4H cross peaks of (▲)H26, (■)H39. Curves calculated from parameters of Table 6.5.

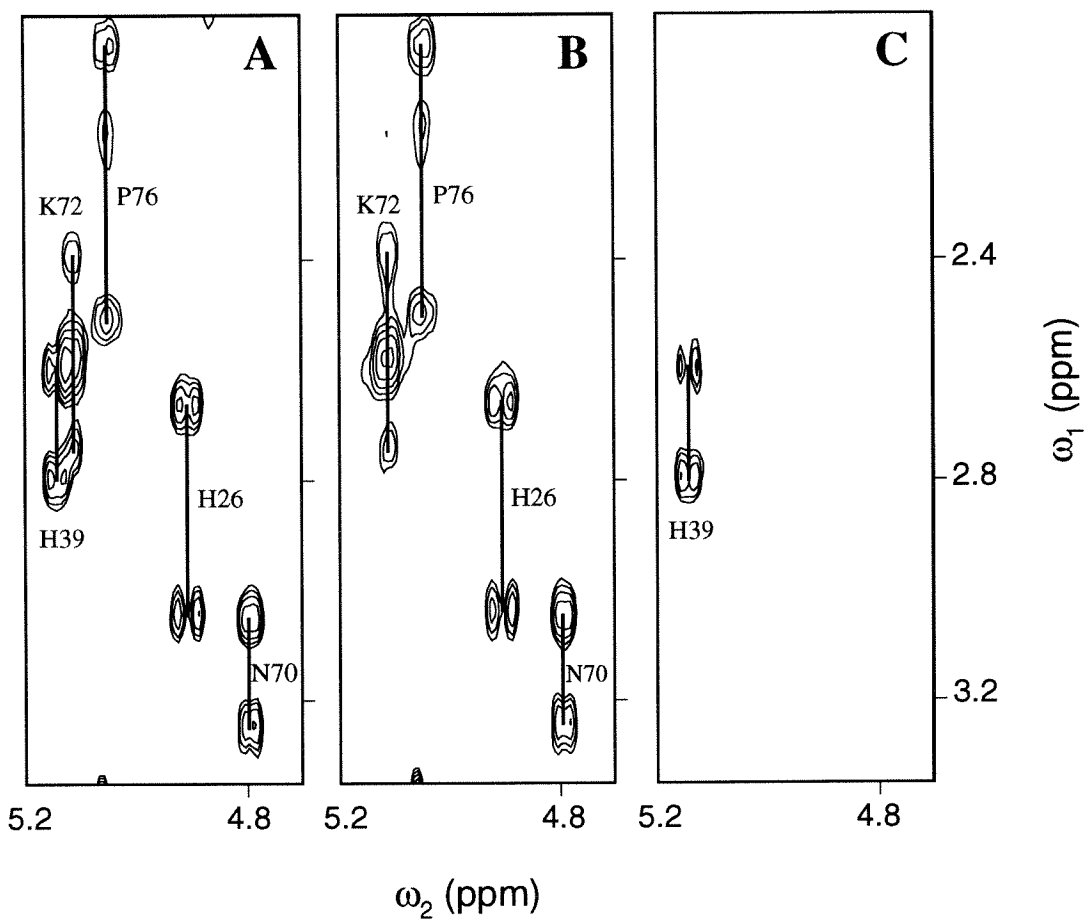


Figure 6.13. 2D paramagnetic difference TOCSY spectrum of yeast cytochrome *c*. TOCSY spectra of *C. krusei* cytochrome c were recorded at 303K, pH 7.85, for **(A)** 0.8 mM cytochrome c with 0.15 mM CuEDTA, **(B)** 0.8 mM cytochrome c with 0.15 mM CuIDA (c_M= 0.188), **(C)** difference spectrum obtained by subtracting (B) from (A). Assignments of α-β cross peaks of H26, H39, N70, K72, P76 are indicated.

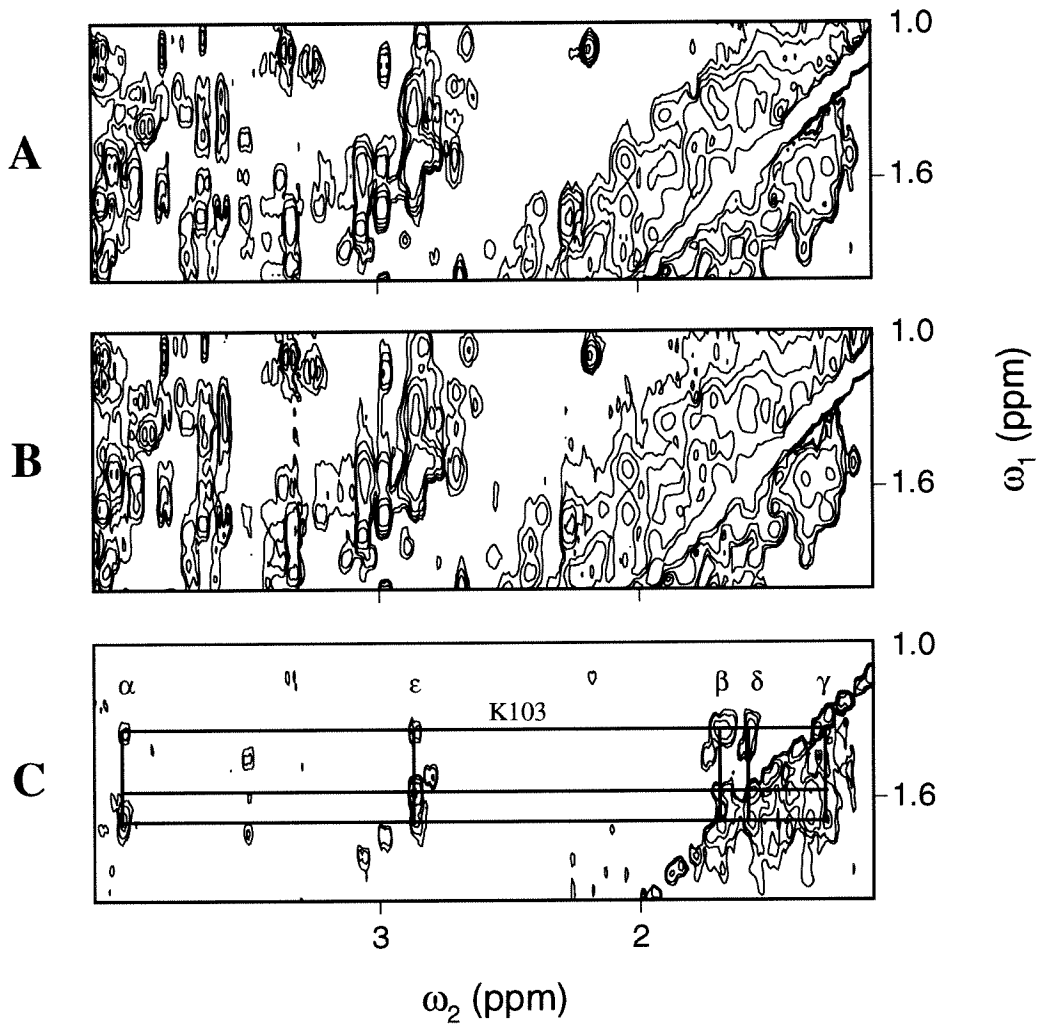


Figure 6.14. 2D paramagnetic difference TOCSY spectrum of yeast cytochrome *c*. TOCSY spectra of *C. krusei* cytochrome c were recorded at 303K, pH 7.85, **(A)** 0.8 mM cytochrome *c* with 0.15 mM CuEDTA, **(B)** 0.8 mM cytochrome *c* with 0.15 mM CuIDA (c_M= 0.188), **(C)** difference spectrum obtained by subtracting (B) from (A). Assignments of the complete spin system of K103 are labeled.

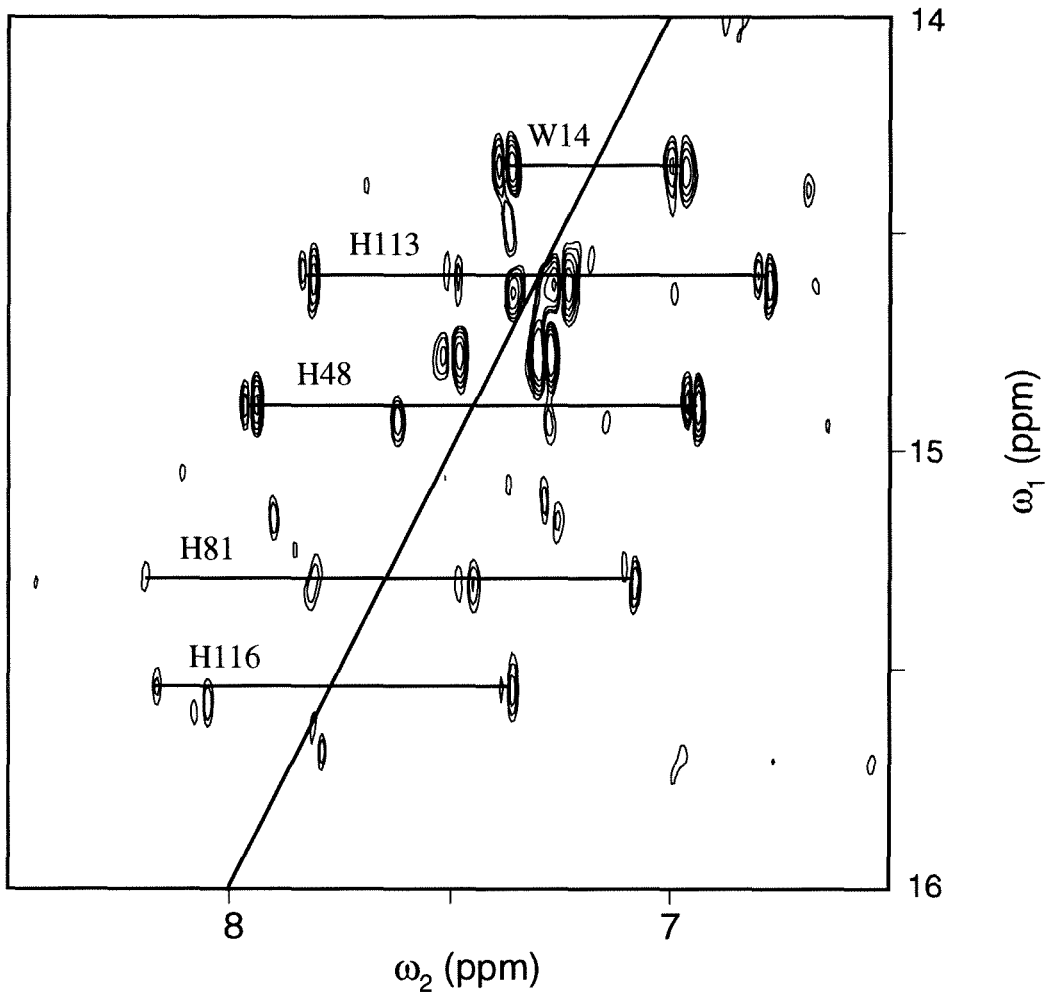


Figure 6.15. Double quantum (2Q) spectrum of horse myoglobin. Phase-sensitive double quantum spectra of horse myoglobin were recorded at 300K, pH 6.85, for 2.0 mM myoglobin. Both positive and negative contour levels are plotted. Assignments for C_2H-C_4H double-quantum cross peaks of H113, H116, H48, and H81 are indicated, along with assignments for W14 η_2-ζ_2 cross peaks.

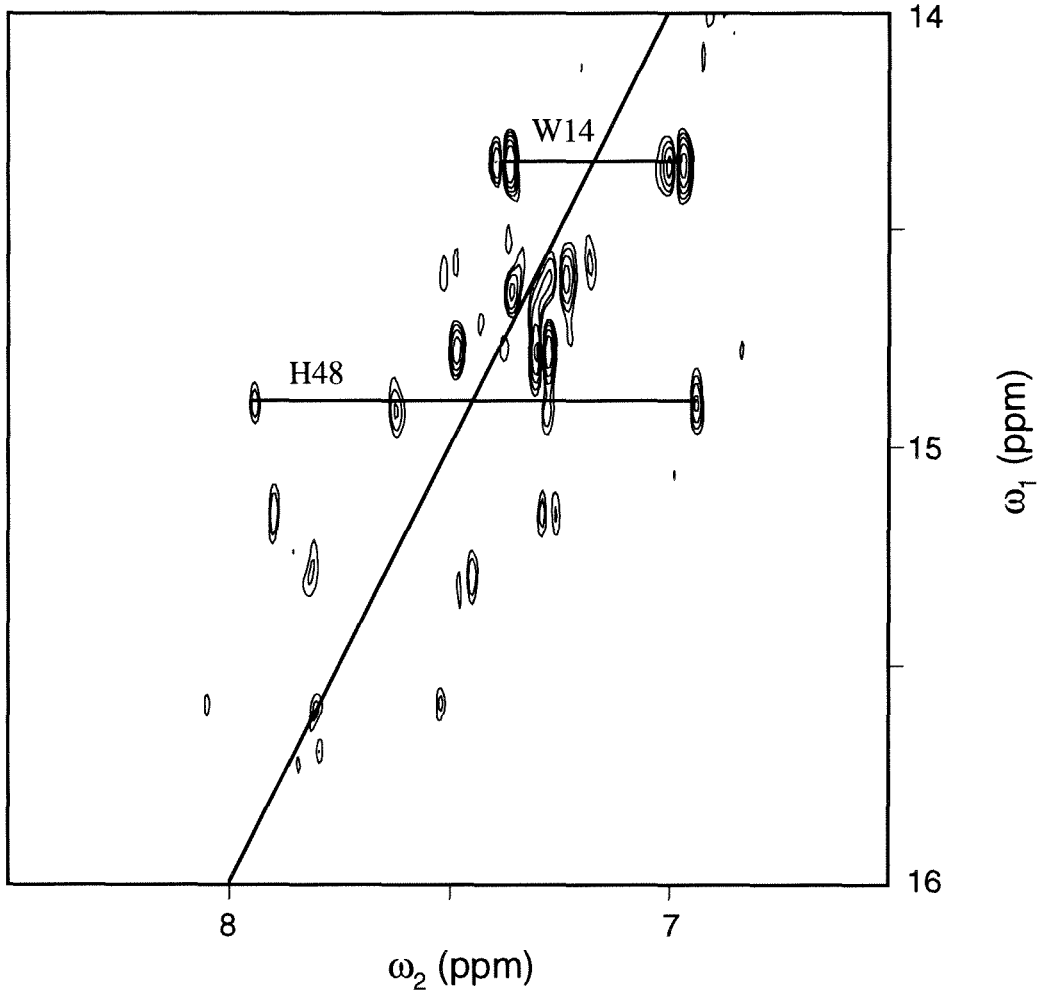


Figure 6.16. Double quantum (2Q) spectrum of horse myoglobin with CuDIEN. Phase-sensitive double quantum spectra of horse myoglobin was recorded at 300K, pH 6.85, for 2.0 mM myoglobin with 32 μM CuDIEN ($c_M = 0.016$). Both positive and negative contour levels are plotted. Assignments for C_2H-C_4H double-quantum cross peaks of H48 and η_2-ζ_2 cross peaks of W14 are indicated.

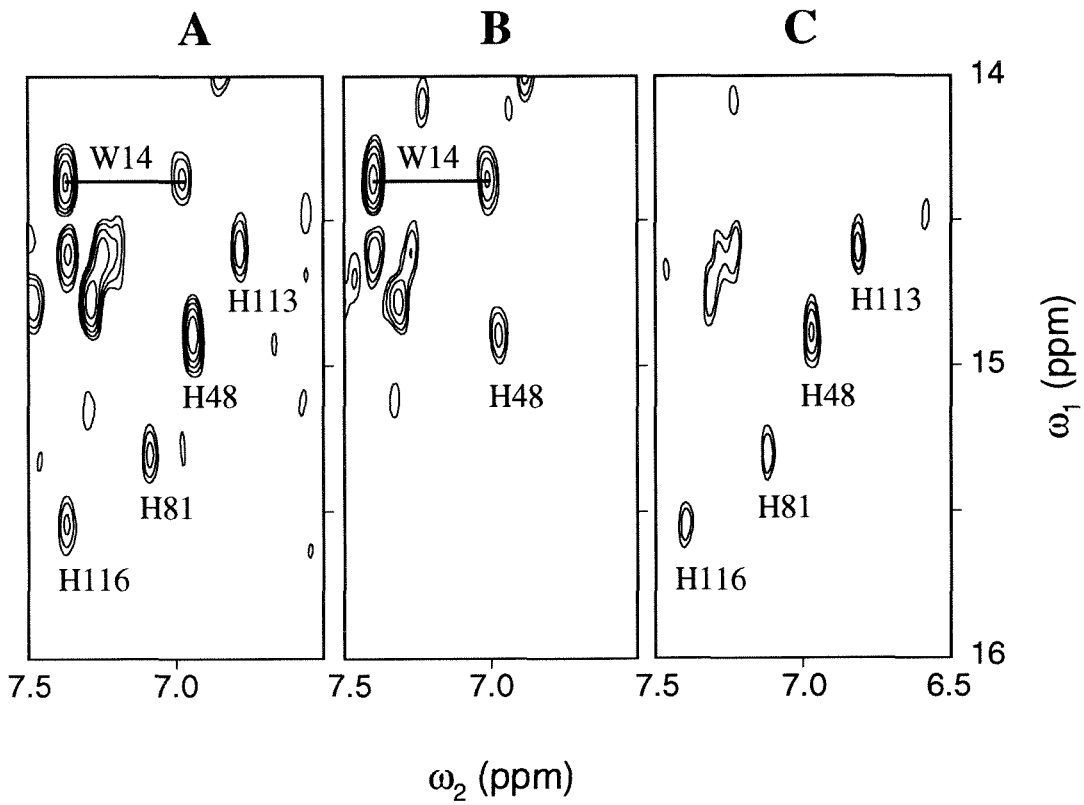


Figure 6.17. 2D paramagnetic difference 2Q spectrum of horse myoglobin. Magnitude mode double quantum spectra of horse myoglobin were recorded at 300K, pH 6.85, for **(A)** 2.0 mM myoglobin, **(B)** 2.0 mM myoglobin with 32 μM CuDIEN (c_M = 0.016), **(C)** difference spectrum obtained by subtracting (B) from (A). Assignments for C_2H-C_4H double-quantum cross peaks of H113, H116, H48, and H81 are indicated, along with assignment for W14 η_2-ζ_2 cross peaks.

APPENDIX

Double-Quantum experiment

The following AMX500 (UXNMR) pulse program performs a modified phase-sensitive (TPPI) double quantum experiment with composite 180 mixing pulse and solvent presaturation during relaxation delay.

```
;dqscrdj

1   ze
2   d11
    d11
3   d12 hl2
     p18 ph29
    d13
    d12 hl1
    p1 ph1
    d4
    p1 ph1
    d13
    p2 ph2
    d13
    p1 ph1
    d4
    p1 ph1
    d0
    p1 ph3
    go=2 ph31
    d11 wr #0 if #0 id0 ip1 zd
    d11 ip2
    lo to 3 times td1
exit

ph1=        (8)     0 2 4 6 2 4 6 0 4 6 0 2 6 0 2 4
ph2=        (8)     2 4 6 0 4 6 0 2 6 0 2 4 0 2 4 6
                    6 0 2 4 0 2 4 6 2 4 6 0 4 6 0 2
ph3=                0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
ph29=       0
ph31=       2 0 2 0 3 1 3 1 0 2 0 2 1 3 1 3
;hl1: ecoupler high power level, hl2: ecoupler power level for presaturation
;p1 :  90 degree transmitter high power pulse, p2 : 180 degree transmitter high power pulse
;p18: presaturation during relaxation delay
;d0 : incremented delay (2D) [3 usec], d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J), d11: delay for disk I/O  [30 msec], d12: delay for power switching  [20 usec]
;d13: short delay (e.g. to compensate delay line) [3 usec]
;in0: 1/(4 * SW) = (1/2) DW, nd0: 2
;NS: 16 * n, DS: 4, td1: number of experiments
;MC2: TPPI
```

Magnitude-mode processsing of TPPI data

The following FELIX (v.2.10) program processes TPPI data as a magnitude-mode spectrum suitable for 2D difference spectrum or for measurement of crosspeak volume intensities.

```
c**tppimag.mac
; conducts magnitude-mode transform of TPPI data
cl
cmx
get '2D filename = ' datfil
get 'number of fids = ' nfids
get 'matrix = ' mat
get 'D1 size = ' d1len
get 'D2 size ( > 2*# fids )= ' d2len
get 'D1 window function = ' wind1
get 'D2 window function = ' wind2
ty &datfil &nfids
ty &mat &d1len &d2len
ty Building matrix...
bld &mat 2 &d1len &d2len 0 y
mat &mat w
ty D1 transform, rotating for magnitude mode ...
  def nangle 4
  eva 2d1len (2*&d1len)
  eva 2d2len (2*&d2len)
  eva hfd2le (&d2len/2)
  eva hfd2le (int(&hfd2le))
  def iangle 0
  def row 1
  def rrow 1
loop:
  esc out
  if &out ne 0 quit
  for iangle 1 &nangle
  if &row gt &nfids d1done
  rn &datfil
  if &status ne 0 quit
  def datype 1
  bc 0.05
  &wind1
  zf &d1len
  bft
  gto &iangle quit ang1 ang2 ang3 ang4
ang1:
    mul  1  0   ; multiply by 1
    go back
ang2:
    mul  0 -1   ; multiply by -i
```

```
   go back
ang3:
   mul -1  0   ; multiply by -1
   go back
ang4:
   mul  0  1    ; multiply by i
   go back
back:
  sep  ; store real and imaginary parts in alternating rows
  def datsiz &d1len
  sto 0 &rrow
  if &status ne 0 quit
  eva rrow (&rrow+1)
  def datsiz &2d1len
  shl &d1len
  def datsiz &d1len
  sto 0 &rrow
  if &status ne 0 quit
  eva rrow (&rrow+1)
  ty row= &row$
  eva row (&row+1)
  next     ; increment phase angle for next fid
  go loop
d1done:
ty D2 transform ...
  for col 1 &d1len
  loa &col 0
  def datype 1
  &wind2
  zf &2d2len
  cnj
  ft
  ms
  red
  shl &hfd2le  ; keep only the middle half of the spectrum
  alt
  red
  sto &col 0
  if &status ne 0 quit
  ty col=&col$
  esc out
  if &out ne 0 quit
  next
done:
cmx
ex return
end
quit:
ty Error.
ex return
end
```

Table 6A.1. Longitudinal (T_1) relaxation of acHIS protons by CuIDA. Relaxation rates measured by inversion recovery for fixed concentraion of acHIS (92 mM) and CuIDA (28 μM, $c_M = 3.0 \times 10^{-4}$).

	C_2H δ=8.17		C_4H δ=7.13		$C_\beta H_2$ δ=3.23	
T (C)	CuIDA	no CuIDA	CuIDA	no CuIDA	CuIDA	no CuIDA
25	10.6	--	8.4	--	2.54	--
30	9.9	0.13	7.9	0.28	2.34	1.12
35	8.8	--	7.2	--	2.19	--
40	8.5	--	6.6	--	1.96	--
45	7.4	0.10	5.8	0.18	1.76	0.74
50	6.9	--	5.1	--	1.53	--
55	5.6	--	4.5	--	1.38	--
60	4.9	--	3.9	0.08	1.24	0.58

	$C_\alpha H$ δ=4.50		$C_{me}H_3$ δ=2.03		TSPSA δ=0.03	
T (C)	CuIDA	no CuIDA	CuIDA	no CuIDA	CuIDA	no CuIDA
25	1.41	--	0.82	--	0.28	--
30	1.26	1.15	0.72	0.57	0.25	0.31
35	1.11	--	0.66	--	0.22	--
40	1.02	--	0.59	--	0.20	--
45	0.85	0.81	0.53	0.41	0.19	0.18
50	0.73	--	0.48	--	0.18	--
55	0.54	--	0.44	--	0.16	--
60	--	0.58	0.40	0.31	0.15	0.16

Table 6A.2. Paramagnetic effect of CuIDA (compared to CuEDTA) on histidine imidazole ring COSY crosspeak volumes at 303 K for *Candidia krusei* cytochrome *c* (0.8 mM).

		V/V_o			
histidine	**crosspeak**	0.05 mM Cu	0.10 mM Cu	0.14 mM Cu	0.20 mM Cu
H_{26}	2-2*	0.73	0.63	0.73	0.63
H_{26}	4-4*	0.89	0.42	0.36	0.20
H_{26}	2-4	0.82	0.65	0.53	0.46
H_{39}	2-2*	0.05	0.03	0.03	0.03
H_{39}	4-4*	0.11	0.08	0.06	0.10
H_{39}	2-4	0.05	0.02	0.01	0.02

* diagonal peak

REFERENCES

1. Mallik, S., Plunkett, S., Dhal, P. K., Johnson, R. D., Pack, D., Shnek, D. and Arnold, F. H. (1994) *New J. Chem. 18,* 299-304.
2. Mallik, S., Johnson, R. D. and Arnold, F. H. (1993) *J. Am. Chem. Soc. 115*, 2518-2520.
3. Mallik, S., Johnson, R. D. and Arnold, F. H. (1994) *J. Am. Chem. Soc. 116*, 8902-8911.
4. Todd, R. J. (1993) *Ph.D. Thesis*, California Institute of Technology, Pasadena, CA.
5. Arnold, F. H. and Haymore, B. L. (1991) *Science 252*, 1796-1797.
6. Yip, T., Nakagawa, Y. and Porath, J. (1989) *Anal. Biochem. 183*, 159-171.
7. Wuenschell, G. E., Wen, E., Todd, R. J., Shnek, D. R. and Arnold, F. H. (1991) *J. Chromatogr. 543*, 345-354.
8. Bertini, I. and Luchinat, C. (1986) *NMR of Paramagnetic Molecules in Biological Systems*, Benjamin/Cummings, Menlo Park, CA.
9. Theriault, T. P., Leahy, D. L., Levitt, M. and McConnell, H. M. (1991) *J. Mol. Biol. 221*, 257-270.
10. Anglister, J., Frey, T. and McConnell, H. M. (1984) *Biochemistry 23*, 5372-5375.
11. Rehman, J. P. and Barton, J. K. (1990) *Biochemistry 29*, 1710-1717 .
12. Bocarsly, J. R. and Barton, J. K. (1992) *Inorg. Chem. 31*, 2827-2834.
13. Esperson, W. G. and Martin, R. B. (1976) *J. Am. Chem. Soc. 98*, 40-44.
14. Henry, B., Boubel, J. C. and Delpuech, J. J. (1986) *Inorg. Chem. 25*, 623-631.
15. deJong, E., Claesen, C., Daeman, C., Harmsen, B., Konings, R., Tesser, G. and Hilbers, C. (1988) *J. Mag. Reson. 80*, 197-213.
16. Petros, A., Mueller, L., Kopple, K. (1990) *Biochemistry 29*, 10041-10048.

17. Yu, L. Y., Meadows, R. P., Wagner, R. and Fesik, S. W. (1994) *J. Mag. Reson. B 104*, 77-80.

18. Gao, Y., Boyd, J., Williams, R. and Pielak, G. (1990) *Biochemistry 29*, 6994-7003.

19. Feng, Y., Roder, H., Englander, S. W., Wand, A. J. and Di Stefano, D. L. (1989) *Biochemistry 28*, 195-203.

20. Cocco, M. J., Kao, Y. H., Phillips, A. T. and Lecomte, J. T. J. (1992) *Biochemistry 31*, 6481-6491.

21. Takano, T. and Dickerson, R. (1981) *J. Mol. Biol. 153*, 79-105.

22. Bushnell, G., Louie, G. and Brayer G. (1990) *J. Mol. Biol. 214*, 585-595.

23. Louie, G. V. and Brayer, G. D. (1990) *J. Mol. Biol. 214*, 527-555.

24. Evans, S. V. and Brayer, G. D. (1990) *J. Mol. Biol. 213*, 885-897.

25. Braunschweiler, L. and Ernst, R. R. (1983) *J. Mag. Reson. 53*, 521-528.

26. Shaka, A. J. and Freeman, R. (1983) *J. Mag. Reson. 51*, 169-173.

27. Braunschweiler, L., Bodenhausen, G. and Ernst, R. R. (1983) *Mol. Phys. 48*, 535-560.

28. Van Dyke, B. R., Bakan, D. A., Glover, K. A. M., Hegenauer, J. C., Saltman, P., Springer, B. A. and Sligar, S. G. (1992) *Proc. Natl. Acad. Sci. USA 89*, 8018-8019.

29. Hegetschweiler, K., Saltman, P., Davlit, C. and Wright, P. E. (1987) *Biochim. Biophys. Acta 912*, 384-397.

30. Atkins, P. W. and Kivelson, D. (1966) *J. Chem. Phys. 44*, 169-174.

31. McConnell, H. M. (1956) *J. Chem. Phys. 25*, 709-711.

32. Kirillin, V. A., Editor (1971) *Heavy Water: Thermophysical Properties*, Israel Program for Scientific Translations, Jerusalem.

33. Henry, B. and Delpuech (1989) *J. Chimie Physique 86*, 2067-2079.

34. Arena, G., Bonomo, R., Impellizzeri, G., Izatt, R. M., Lamb, J. D. and Rizzarelli, E. (1987) *Inorg. Chem. 26*, 795-800.

35. Kuroda, Y. and Aiba, H. (1979) *J. Am. Chem. Soc. 101*, 6837-6842.

36. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley and Sons, New York.

37. Johnson, R. D. (1991) *M.S. Thesis*, California Institute of Technology, Pasadena, CA.

38. For example, if P is the concentration of protein (M) and K_P is the equilibrium association constant (M^{-1}) for protein binding the metal complex by a particular surface site, then K is approximately equal to the product K_PP [37].

39. Niccolai, N., Bonci, A., Rustici, M., Scarselli, M., Neri, P., Esposito, G., Mascagni, P., Motta, A. and Molinari, H. (1991) *J. Chem. Soc. Perkin II 1991*, 1453-1457.

40. Smith, R. M. and Martell, A. E. (1975) *Critical Stability Constants,* vol. 1-6, Plenum Press, New York, NY.

41. Martell, A. E. and Motekaitis, R. J. (1988) *The Determination and Use of Stability Constants*, VCH Pub. Inc., New York, NY.

42. Sinha, P. C., Saxena, Y. K., Nigam, N. B. and Srivastava, M. N. (1989) *Indian J. Chem. 28A*, 335-336.

43. Hori, F., Kojima, Y., Matsumoto, K., Oot, S. and Kuroya, H. (1979) *Bull. Chem. Soc. Japan 52*, 1076-1079.

44. Camerman, N., Fawcett, J. K., Kruck, T. P. A., Sarkar, B. and Camerman, A. (1978) *J. Am. Chem. Soc. 100*, 2690-2693.

45. Blount, J. F., Fraser, K. A., Freeman, H. C., Szymanski, J. T. and Wang C. H. (1966) *Acta. Cryst. 22*, 396-405.

46. Kojima, Y., Hirotsu, K. and Matsumoto, K. (1977) *Bull. Chem. Soc. Japan 50*, 3222-3231.

47. Todd, R. J., Van Dam, M., Casimiro, D., Haymore, B. L. and Arnold, F. H. (1991) *Proteins 10*, 156-161.

48. Weiss, M. A., Eliason, J. L. and States, D. J. (1984) *Proc. Natl. Acad. Sci. USA 81*, 6019-6023.

49. Luchinat, C., Stuernagel, S. and Turano, P. (1990) *Inorg. Chem. 29*, 4351-4353.

50. Dalvit, C. and Wright, P. E. (1987) *J. Mol. Biol. 213*, 885-897.

51. Canet, D., Levy, G. C. and Peat, I. R. (1975) *J. Mag. Reson. 18*, 199-204.

52. Martin, R. B. and Edsall, J. T. (1960) *J. Am. Chem. Soc. 82*, 1107-1111.

53. Haymore, B. L., personal communication.

54. Moonen, C. T. W., Scheek, R. M., Boelens, R. and Muller, F. (1984) *Eur. J. Biochem. 141*, 323-330.

55. Mallik, S. (1994) unpublished results.

56. Levitt, M. H. and Freeman, R. (1979) *J. Mag. Reson. 33*, 477-480.

57. Marion, D. and Wuthrich, K. (1983) *Biochem. Biophys. Res. Comm. 113*, 967-974.

58. Mrabet, N. T. (1992) *Biochemistry 31*, 2690-2702.