Chapter 1

Protein–ligand interactions and their analysis

Babur Z. Chowdhry
School of Chemical and Life Sciences, University of Greenwich, Wellington Street, Woolwich, London SE18 6PF, UK.

Stephen E. Harding
NCMH Physical Biochemistry Laboratory, University of Nottingham, School of Biosciences, Sutton Bonington, Leicestershire LE12 5RD, UK.

1 Introduction

In each cell of an organism, a myriad of reactions, covalent and non-covalent, occur at any given point in time. These reactions are co-ordinated and regulated, both spatially and temporally. Each reaction has a specific purpose, occurs as a result of finely-tuned inter- and intramolecular recognition mechanisms, and forms part of an intricate network of interdependent multi-component linear/non-linear reactions in interconnected compartments (organelles etc.) of the cell. Moreover the frontiers of viability of such reactions—and the living organisms that depend on them—are marked by extreme conditions: 1–12 for pH, –5–110 °C for temperature, 0.1–120 MPa for hydrostatic pressure, and 0.6–1.0 for water activity. Amongst the many molecules that participate in such reactions in the complex—and, as yet hardly understood, milieu of the cell—are proteins.

Proteins play a pivotal, indeed essential role, in cellular (and in multicellular organisms, extracellular) activity. Their numerous biological functions together with the molecular basis of their biophysical properties/behaviour are therefore of multidisciplinary interest within the framework of what are generally called the biological (biochemistry, pharmacology, physiology, immunology, etc.) and physical (physics, chemistry, mathematics, computing, etc.) sciences.

Both our intrinsic curiosity-driven philosophical knowledge base and our need or desire to modify (or use), from an applied perspective (medical, agricultural, biotechnological), the properties or behaviour of proteins are increasingly dependent on the ability to modulate the physico-chemical, and hence biological behaviour of these molecules. Because *protein–ligand interactions play a key role in cellular metabolism*, detailed knowledge of such interactions, at both a microscopic and macroscopic level, is also required. Just two examples, which immediately stand out from the many, are antigen–antibody interactions and proteins which

act as receptors (membrane or non-membrane bound): neurotransmission depends on the ability of small and large molecular weight molecules to recognize and bind to specific sites on large membrane bound receptors. Another example: enzyme–ligand interactions form a very large group of important complexes, which have been investigated for many years. In these systems not only is the substrate(s) of obvious importance, but so are other molecules, which regulate enzyme activity such as coenzymes and positive and negative modulators in both allosteric and non-allosteric proteins. Other important systems, which are now being studied at a molecular level, include ligand binding to structural proteins, protein–DNA binding, protein–saccharide, protein–protein, and protein–peptide interactions. The formation—and maintenance—of the quaternary structure of multisubunit proteins, the self-assembly of large structures such as microtubules or chromatin and transcription factor–promoter interactions are other examples; although the list is actually almost endless.

The term 'ligand' in biological systems can have many different meanings. It is usually, in its broadest sense, used to mean any molecule which interacts with a given molecule (in this case a protein). The term 'ligand' thus includes other macromolecules (peptides/proteins/nucleic acids/lipids/carbohydrates or mixed molecular species thereof) as well as 'small' molecular mass (arbitrarily $< \sim 1$–2 kDa) molecules. Ligands therefore comprise a very large and structurally diverse group of molecules which, unsurprisingly also display a wide variety of physicochemical characteristics. This makes it difficult to understand, delineate, and draw generalized conclusions concerning their biophysical properties. Perhaps the major criterion, in the context of the discussion that follows, is that ligands can interact in a (potentially) reversible, non-covalent manner with a protein and thereby modulate its biological role in a controllable way (i.e. without the requirement to make or break covalent bonds).

To *fully* understand protein–ligand interactions requires that, at a minimum, the following criteria be met. First that the biophysical properties of both the protein and the ligand under investigation are examined independently and that their biophysical behaviour be fully understood: this requirement applies to both existing and/or newly designed and (bio)synthesized peptides/proteins and ligands. *Currently, and contrary to popular opinion, we are far from achieving such an aim, except at an extremely superficial level.* Knowledge of the structure/conformation, if possible, at the atomic level of the protein and the ligand in the unbound form is a prerequisite for protein–ligand interaction studies. Although it may appear trivial to the reader, the high level resolution structures of most proteins (either by X-ray crystallography or high resolution NMR) have not, to date, been established. It is currently estimated that such data exists for only about 30 peptides/proteins (1) but it is seriously arguable, in the sense of the generally accepted meaning of the term high (atomic) resolution (i.e. < 0.8 Å), that even this target has not been met. Additionally the process of forming a complex between a small ligand molecule and a protein is a complicated equilibrium process. Both the ligand and the protein in the solvated state probably exist as an equilibrium mixture of several conformers. Admittedly for many 'small mol-

Table 1 Parameters which have to be defined for the protein, the ligand, and the protein–ligand complex

Chemical composition, concentration/purity (contaminants) /stability/specific activity or assay for, e.g. structural proteins and ligand/state of structural integrity. Ability to freeze or air dry and pre-history of protein and ligand samples may be important.
Solubility (in different solvents: aqueous, organic, and mixed phase); state of aggregation.
Structure: size, shape, geometry, topology; protein sequence, and post-translational modifications must be known; ability to use molecular biology techniques (strategies for over-expression and mutagenesis of recombinant protein are critical); combinatorial methods for over production of ligand.
Dynamics of protein/ligand/ protein–ligand 'complex'.

ecular weight ligands' high resolution X-ray and/or NMR structures do exist. The fact that our knowledge base is deficient in this area has significant implications for understanding protein–ligand interactions, which unfortunately will not be touched upon, in this brief overview. In addition a myriad of other properties of the reactants also require to be established (see *Table 1*).

Secondly the protein–ligand complex must also be fully characterized. Normally, at a simplistic level, the non-covalent interaction between a protein (P) and a ligand (L) is often represented as follows:

$$P + L \leftrightarrow PL \qquad [1]$$

However as Williams and Westwell have pointed out (2) the interaction should actually be written in the following form:

$$P + L \leftrightarrow P'L' \qquad [2]$$

Equation 2 may, at first glance appear identical to *Equation 1*. However the formulism of *Equation 2* recognizes, or is taken to mean, that once P and L have undergone an interaction or association, they no longer exist. They have, instead been replaced by the modified entities P′ and L′. (Tight complexes often result when protein–ligand interactions are significantly stronger than ligand–solvent and protein–solvent interactions.)

The well described equilibrium molar association or binding constant, K_a (other popular symbols are K_b or just K) corresponding to *Equation 2* is then described by:

$$K_a = [P'L'] / ([P].[L]) \qquad [3]$$

where the square brackets means the molar concentration (M, or $mol.l^{-1}$), which of course is not an SI unit. For the system described by *Equation 2* (or *Equation 1*) the units[1] of K_a will be M^{-1} ($l.mol^{-1}$). The more popularly used 'molar dissociation constant' K_d is simply the reciprocal of K_a, so for the system of *Equation 2*:

$$K_d = [P].[L] / [P'L'] \qquad [4]$$

[1]In strict thermodynamic terms K_a and K_d are both dimensionless. Dimensions are however normally added, but only to indicate the dimensions of the quantities used to calculate them. See Price, N. C., Dwek, R. A., Wormald, M. R. and Ratcliffe, R. G., *Principles and Problems in Physical Chemistry*, 3rd edition, Oxford University Press, 2001.

with the units of K_d, M or mol.l^{-1}. Traditionally, K_ds of < 5 μM are regarded as strong interactions, and > 50 μM as 'weak'. Any technique chosen to measure K_d (or K_a) has to cope with a concentration/concentration range where all species (P, L, and P′L′) are present: this means to probe the K_d for strong interactions, low concentrations are required and for weak interactions, high concentrations. Of course there are other reactions more complicated than that described by *Equations 1* or *2*. A more generalized form of *Equation 2* for a binary system is:

$$nP + mL \leftrightarrow P'_nL'_m \quad [5]$$

and the corresponding K_a will be:

$$K_a = [P'_nL'_m] / ([P]^n.[L]^m) \quad [6]$$

Other variants of Equations 3–6 include the use of weight concentrations C (g/l or g/ml: again, not SI units) rather than molar concentrations: (the K_a, K_d notation is then replaced by, e.g. X_a, X_d).

Establishing the stoichiometry and association/dissociation equilibrium constants (or the corresponding rate constants) is only one step: the goal of research into protein–ligand interactions is to understand, again it must be emphasized, in minute molecular detail the relationship between function and molecular recognition, structure, kinetics, energetics, and dynamics of as many defined systems as possible. Use can then be made of this plethora of knowledge such that either the behaviour of previously unknown systems can be hypothesized in advance of experimental information being gained or that completely new systems can be designed. The pursuit of this goal is a time-consuming and difficult task! Especially when it is realized that the inter-relationship between all of the foregoing parameters (intermolecular recognition–structure–dynamics–kinetics–thermodynamics) need to be assessed and need to be further ascertained over a wide variety of environmental solution conditions. *Table 2* cites the most commonly used experimental variables (but of course, there are countless others) and *Table 3* provides an overall summary of the factors that need to be taken into account when applying a technique (or, preferably, a collection of

Table 2 Experimental variables in the analysis of the protein, the ligand, and the protein–ligand complex

Variables
(a) Reagent variables
Protein/ligand concentration
Ligand structural variants (analogues/homologues)
Protein structural variants (site/group specific, conservative/non-conservative mutants; denatured form/partially unfolded forms, presence or absence of cofactors)
(b) Environmental solution condition variables
Temperature, pH, buffers (ionic strength and nature of buffers)
Osmolytes, solvents including co-solvents, salt /ion concentration
Denaturants (urea /GdnCl (and other lyotropic /chaotropic agents)/detergents, surfactants), stabilizers (azide, glycerol etc.); mercaptoethanol, DTT; gaseous phase used for experiments, etc.
(c) Combinations of the above

Table 3 Factors to be taken into account in technique(s) used for investigating protein/ligand /protein–ligand complexes

Design of experimental protocol
Information content
Availability, cost
Complementary/alternative techniques[a]
Analytical parameters associated with the technique(s): *concentration of material required, compared to other techniques,* sensitivity, detection limit, accuracy, precision, reproducibility/repeatability, use or necessity for isotopic or non-isotopic labelling, ease of use, computing resources, methods/complexity and different methods of data analysis, errors/error propagation/statistical analysis of results (if possible); time scales of experimental (spectroscopic) techniques used.
Ability to use the technique in the presence of buffer additives and/or environmental solution conditions given in *Table 2*.
The experimentalist should insure that at every stage of the experiment and (particularly) in the data analysis, the correct (appropriate) theoretical models are used. In using instrumental techniques the 'black box data analysis' syndrome should be avoided. Critically compare and contrast with work on other related systems. Each method will have its own advantages and drawbacks for a given system. The ability to interpret experimental data both intrinsically and in terms of a theoretical basis is very important. The systematic/accumulated instrumental errors are often important, but usually ignored.

[a] Try, if possible to use more than one technique (method) to verify the overall results or the values for particular parameters.

techniques) to the analysis of protein–ligand interactions. *It is extremely important to cross-reference parameters in Tables 1, 2, and 3.*

Within the requirement that the changes in the conversion of P to P′ and L to L′ (i.e. changes in the reacting species in the 'complex' compared to their individual states) need to be ascertained as a step towards a full understanding of protein–ligand interactions, the following factors include some of the important parameters which require consideration. The changes in structure, in either species, may range from extremely subtle (i.e. an 'insignificant' structural reorganization; 'simply' a tightening or loosening of the internal structure) to large scale changes in both local and global conformational properties. The list of changes which occur not only include changes in structure (secondary/tertiary/quaternary), conformation, size, shape geometry, and topology but also include changes in the charge distribution, the state of hydration and protonation, and the partial molar volume as well as changes in the surface accessible surface area, polarity (hydrophobicity), and intra- and intermolecular entropy factors (e.g. rotational and translational motional properties of molecules: and the dynamics of water molecules). The molecular nature of the binding interface, the identification and quantification of the (individually) 'weak' cooperative molecular forces holding the protein–ligand complex together (such as hydrogen bonding, dipole–dipole interactions, van der Waals forces—induction and dispersion forces alone or combined—hydrophobic, electrostatic interactions, etc.) and the number/role, if any, of water molecules at the binding interface are other important factors which have to be considered. Association of other proteins or other

ligand species subsequent to the interaction of the initial species may also be involved; thus making the problem even more complicated! Furthermore if multisite ligand binding to either a monomeric or multimeric (homomeric/heteromeric subunit) protein is involved then cooperativity effects and linkage phenomena (allosteric effects) have also to be taken into account and examined by the appropriate experimental/theoretical techniques. Williams and Westwell (2) have also emphasized the important, but often overlooked, point that the binding energy between two reacting species is not only a property of the interface between them, but also depends on the modifications (as alluded to above) of the internal structures of one or both of the reacting partners.

Obviously in order to achieve the objectives of investigating protein–ligand interactions a multidisciplinary technique/methods approach has to be adopted as we stress in *Table 3*. Some of the methods which are used to examine protein and/or protein–ligand interactions have been compared and contrasted in *Table 4* which has been based to a certain extent on a recent review by Philo (ref. 3), and the contributions to this book (other useful sources can be found in refs. 4 and 5).

2 A brief comparison of the 'high resolution' methods

Since structural elucidation is so important in protein–ligand interactions it makes sense to compare and contrast the application of the two most powerful and commonly used methods for attaining this goal, namely X-ray crystallography and high resolution NMR. Both methods complement each other. There is no doubt, however that, currently at least, X-ray techniques can give a better-defined structure in less time compared to NMR techniques. However the two methods should be considered convergent (they do after all use similar software tools) in that for molecules (that are amenable to NMR analysis) the X-ray structure, if available, can be a substantial aid in the elucidation of the NMR structure. In addition both methods require, as a general rule of thumb, the same amount of material for analysis; approximately 10 mg per 10 kDa of protein. However the fact that NMR can be used for obtaining information on the dynamics of a system, the association constant for a protein–ligand interaction and, for small molecular weight proteins and/or protein–ligand complexes, the structure of different conformations in equilibrium should also be taken into account.

The advantages of X-ray crystallography, for structure determination of proteins, ligands, and their complexes are as follows:

(a) It is a well-established technique.

(b) More mathematically direct image construction is required, compared to NMR.

(c) Objective interpretation of data (usually) easier than NMR.

(d) Raw data processing highly automated.

(e) Quality indicators available (resolution, *R*-factor).

Table 4 Techniques for characterizing protein–ligand interactions (adapted and extended from ref. 3; see also refs 4, 5)[a]

Technique	Basis for detection	Interaction information	Minimum protein required[b]	Analysis time	Comments
Hydrodynamic and calorimetric (Volume 1)					
Equilibrium dialysis and rate dialysis; affinity (zonal and frontal) gel chromatography	Partitioning of free and bound ligand	Stoichiometry, equilibrium binding strength (K_a or K_d) Range of $K_d \sim 10^{-1}$ to 10^{-13} M	10 μl	15 min to 24 h	Chapters 2 and 3. Relatively straightforward compared to other techniques, but still very useful and adaptable.
Analytical size exclusion chromatography	Change in hydrodynamic size	Stoichiometry, strength ($K_d \sim 10^{-4}$ to 10^{-13} M)	1000 μl	30–60 min	Chapter 3. Wide range of potential on-line or off-line detection methods (including multi-angle laser light scattering for absolute M_r); may be labour-intensive depending on degree of automation.
Sedimentation velocity (in analytical ultracentrifuge)	Change in sedimentation coefficient	Stoichiometry, strength ($K_d \sim 10^{-3}$ to 10^{-9} M), conformation changes	10 μg; 400 μl	2–5 h (3–7 samples)	Chapter 4. Newer software makes data analysis fast and user-friendly; generally not as good as sedimentation equilibrium for K_d.
Sedimentation equilibrium (in analytical ultracentrifuge)	Change in solution mass	Stoichiometry, strength ($K_d \sim 10^{-3}$ to 10^{-9} M)	10 μg; 100 μl	12–24 h per rotor speed (9–21 samples)	Chapter 5. Powerful method when there is a significant change in mass; much more accessible with modern hardware and software.
Surface plasmon resonance	Change in mass bound to surface	Stoichiometry, strength ($K_d \sim 10^{-3}$ to 10^{-13} M), kinetics	0.01 μg; 100 μl	15–30 min (1 analyte, up to 4 surfaces)	Chapter 6. High throughput; fairly insensitive to contaminants; potential problems from surface coupling.
Capillary electrophoresis	Change in size, shape, or charge	Equilibrium binding strength ($K_d \sim 10^{-3}$ to 10^{-7} M), kinetics	μM; nl	Minutes	Chapter 7. Particularly useful if only tiny volumes are available, or material available at low purity.

Table 4 *Continued*

Technique	Basis for detection	Interaction information	Minimum protein required[b]	Analysis time	Comments
Hydrodynamic and calorimetric (Volume 1) (*continued*)					
Electro-optics	Change in rotational diffusion and/or electrical properties	Conformation change on binding. Kinetics of slow (~ minutes) reactions	~ mg/ml, 100–200 μl	Minutes (per concentration)	Chapter 8. Electric dichroism (and decay) or electric birefringence (and decay). Risks of sample damage through field effects now minimized by better shielding.
X-ray and neutron scattering	Change in solution mass and conformation	Conformation change on binding	~ mg/ml	Minutes to hours	Chapter 9. Some risk of radiation damage to sample during measurement.
Isothermal titration microcalorimetry, and Differential scanning calorimetry	Heat release or uptake; affect on thermal transitions of ligand binding	Stoichiometry, strength ($K_d \sim 10^{-3}$ to 10^{-9} M), ΔH; ΔC_p	0.1–1 mg/ml; 0.5–2 ml	3 h	Chapters 10 and 11. Universal signal; works well for small molecules; useful additional thermodynamic information.
Structural and spectroscopic (Volume 2)					
X-ray crystallography	Change in structure	Conformation change on binding	~ 100 mg for crystals	Weeks to months	Chapter 1. Protein and protein–ligand complex must be crystallizable
Molecular modelling	Change in structure	Conformation change on binding	Nothing	Days to weeks	Chapter 2. X-ray or NMR structure needed to begin with; in cases where there is no structure available, one can resort to the techniques of protein structure prediction.
Circular dichroism	Change in optical activity	Secondary structure change. Interaction constants ($K_d \sim 10^{-3}$ to 10^{-7} M)	< 1 mg/ml (~ 10 μM)	~ 2 h	Chapter 3. Optimum $A_{max} \sim 0.8$.
Fluorescence methods	Change in fluorescence property. Rotational diffusion (for polarization anisotropy decay measurements)	Change in surface electrical potential or solution conformation. Kinetics from quenching studies	~ 1 mg/ml (~ 10 μM) Depends on amount of chromophore	Hours	Chapter 4. Applicability depends on nature of protein and/or ligand; chromophore label may be required.

Stopped flow	Rapid mixing of reactants	Association and dissociation rate constants, k_{on} and k_{off} respectively	~ 1 mg/ml (~ 10 μM) Depends on amount of chromophore	Hours	Chapter 5. Alternative to rapid-reaction methods are relaxation methods in which an equilibrium mixture is perturbed by rapid changes in temperature (T-jump) or pressure (P-jump).
Fourier transform infra-red spectroscopy	Changes in the amide I absorption band	Secondary structure change	10 μl at 5–10 mg/ml	12–24 h	Chapter 6. Especially suited for membrane as well as water soluble proteins. Suitable for small and large proteins (and complexes thereof).
Raman spectroscopy	Change in inelastic light scattering or 'Raman spectrum'	Secondary structure, k (rate constants) and equilibrium constants ($K_d > 10^{-3}$ M)	Higher concentrations needed (several mg/ml) 30 μl	Minutes	Chapter 7. Various forms, such as difference, resonance, time resolved, surface enhanced, coherent anti-Stokes.
Mass spectrometry	Change in molecular mass	Stoichiometry	< 0.001 μg; <1 μl	Minutes to hours	Chapter 8. May not detect weaker interactions. Electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). Maximum M_r of complex ~ 1–2 × 10^5 Da.
Electron paramagnetic spin resonance	Changes in electronic environment, rotational diffusion	Changes in structure. Dynamics (rate of change of spectral line shape)	50–100 μg aliquots (1–2 mg/ml)	~ 2 h	Chapter 9. Particularly relevant to membrane proteins. Chemical labelling often necessary.
Nuclear magnetic resonance	Change in chemical shift, diffusion rates (rotational and translational T_2 relaxation)	Changes in structure. Dynamics (k_{on}, k_{off}), equilibria ($K_d > 10^{-3}$ M– ~ 10^{-9} M)	50–100 μM for 1D; 3 mM for 2D; 0.5 ml	Hours (weeks/months for high resolution structure)	Chapter 10. Sample requirements depend on resolution required. 2D high resolution work much more demanding and restricted to small proteins.
Atomic force microscopy	Surface imaging of the force between molecules and the changes on ligand binding	Change in structure	~ 1 mg/ml	Minutes	Chapter 11. Can visualize single molecule: particularly useful for membrane proteins inaccessible by X-ray crystallography.

[a] Abbreviations: ΔH enthalpy change; ΔC_p heat capacity increment; K_a association constant; K_d dissociation constant; k_{on} rate of association; k_{off}, rate of dissociation.

[b] Concentration depends on the strength of the interaction being analysed: weaker interactions require higher concentrations.

(f) Mutant proteins, different ligands, and homologues structures (as low as 25% sequence identity) may be extracted by using molecular replacement and subsequent use of electron density/electron density difference methods as an aid for comparison.

(g) Large molecules and assemblies can be determined, e.g. virus particles (however ordering between such assemblies is not usually as good as—i.e. they often diffract less well than—small molecular weight molecules).

(h) Surface water molecules relatively well defined.

(i) Often produces a single structural model that is easy to visualize and interpret.

(j) Use of synchrotron radiation combined with cryo-conditions usually speeds up data acquisition, improves resolution and stability of crystals.

The disadvantages of X-ray crystallographic methods include the following:

(a) Protein/protein–ligand complex has to form stable crystals that diffract well and X-ray crystallography does not directly yield hydrogen atom positions. (Neutron diffraction techniques make this possible but usually only for small molecules and large size crystals are required.)

(b) Need heavy atom derivatives that form isomorphous crystals, unless molecular replacement methods are used.

(c) Crystal production can be difficult and time-consuming and often impossible.

(d) Unnatural, non-physiological environment, i.e. may not wholly represent structure as it exists in solution.

(e) Difficulty in apportioning uncertainty between static and dynamic disorder.

(f) Surface residues may be influenced by crystal packing.

(g) Large molecular weight flexible modular proteins can be problematic and the final structural model represents a time-averaged structure, where details of molecular mobility remain unresolved.

The advantages of NMR for structural elucidation are as follows:

(a) Experiments are conducted under solution conditions which are 'closer to biological conditions' than X-ray methods (i.e. free from artefacts resulting from crystallization).

(b) Can provide information on dynamics (relaxation, rotational, and translational diffusion measurements, proton exchange rates, and the vibrational motion of atoms) and identify individual side chain motions.

(c) Secondary structure can often be derived from limited experimental data.

(d) Increasingly used to monitor conformational change on ligand binding to protein or to fragments thereof.

(e) Good for comparing and checking the correct fold of mutants (useful for protein folding/time resolved studies).

(f) Solution conditions can be explicitly chosen and readily changed, e.g. pH, temperature, etc. (but high concentrations of buffers may be a problem).

(g) 'Internal' water molecules can be detected (as in X-ray), but surface water molecules can be problematic because of fast exchange (lifetimes can still be ascertained).

(h) H-bonds can now be detected (rather than inferred from distance constraints as in X-ray).

(i) The increasing availability of high resolution instruments (e.g. 700 MHz and above) combined with new data acquisition/analysis techniques and greater computing power is and will continue to allow significant improvements in the application of NMR techniques.

The disadvantages of NMR for structural elucidation include the following:

(a) Usually require concentrated solutions, which may present a danger of aggregation: this problem is manifested in e.g. many antibody systems.

(b) Currently limited to determination of relatively small proteins ($<$ 20 kDa), but there are exceptions.

(c) Surface residues *generally* less well defined than in X-ray crystallography because of mobility of surface residues and experimentally fewer interactions (NOEs).

(d) Often produces an ensemble of possible structures rather than one model (this can be advantageous) but conformational variability can make data interpretation difficult and complete structure determination required if protein sequence homology is less than $\sim$ 60%.

(e) Labelling often used, e.g. ^{2}H, ^{13}C, and ^{15}N and may cause problems in terms of protein yield and expense.

It has to be emphasized that true high resolution structures of the reacting species and the complex are sorely needed in order for example to:

(a) Obtain statistically unbiased data on both protein stereochemistry/precise geometry and the validity of the parameters used to obtain the data (i.e. their refinement).

(b) Check the application of 'normal mode' calculations.

(c) Calculate charge density distributions.

(d) Analyse hydration shells around protein molecules and also for a more accurate view of hydrogen bonds.

(e) Obtain precise information on the direction of amino acid side chain/domain movement.

(f) Model alternative conformations/multi-conformational species.

(g) Help to quantitate the physico-chemical parameters involved in the interaction, especially weak non-covalent interactions.

(h) Obtain precise surface areas.

Not only are these parameters required for intrinsic reasons but also for helping to analyse data obtained from other techniques and thereby improve the theoretical underpinning of experimental observations.

3 Other spectroscopic techniques

One of the virtues of spectroscopic techniques used in examining proteins, ligands, and protein–ligand interactions is that they are usually non-destructive and sometimes non-invasive, however some procedures do require the attachment of labels at specific sites on one or both of the interacting species.

Relatively soft ionization methods introduced in *mass spectrometry* (MS) allow a gentle phase transfer from solution into the gas phase of the mass spectrometer such that even weakly bound non-covalent complexes can be detected intact and their mass analysed. MS methods such as ESI-MS have begun to be recently used for the determination of association constants of non-covalent interactions. One of the advantages of these methods is that signals due to protein, ligand, and protein–ligand ion signals can be detected separately. In addition higher order structure formation, e.g. dimerization may also be detected by MS methods, although analytical ultracentrifugation is probably better suited for this purpose.

Fluorescence spectroscopy is a technique that is very sensitive (picogram quantities of material can be detected) and well-suited for measurement in the real time domain. Its principal shortcomings, however, are that only the structure of the fluorescent probe and the immediate environment is reported and the data obtained is not easy to interpret. Nonetheless, fluorescence spectroscopy has found wide use in studying the physico-chemical properties of proteins, protein–ligand interactions, and protein dynamics. This is because almost all proteins contain naturally fluorescent amino acid residues such as tyrosine and tryptophan. In addition, a large number of fluorescent dyes have been developed that can be used to specifically probe the function and/or structure of macromolecules. In cases where only limited quantities of proteins are available (e.g. a recently expressed recombinant protein or a precious protein pharmaceutical), fluorescence spectroscopy is often the method of choice for studying properties such as stability, hydrodynamics, kinetics, or ligand binding, because of its exquisite sensitivity. Even in cases where the structure of proteins is 'well-known' (either from X-ray diffraction or NMR), *electron paramagnetic resonance* (EPR), *time-resolved fluorescence spectroscopy*, and *stopped-flow methods* can be particularly useful for investigating their dynamic behaviour.

Circular dichroism (CD) is an important, widely used and commonly available technique for secondary structure determination. It requires expertise in data collection/analysis to be used to its full potential, which is not always the case. Nevertheless it has the following advantages in that relatively small amounts of material are required (0.1–1 mg/ml can often suffice) and results can be obtained quickly (in hours rather than weeks or months as in the case of NMR and X-ray). The amino acid sequence of the protein need not be known—although obviously it helps. CD can be used to examine the effects of environmental conditions such as pH, temperature, etc., on the overall protein conformation, in the presence and absence of a ligand. It is in fact a very useful technique prior to X-ray crystallography and/or NMR for screening mutant proteins, obtained by molecular

cloning methods, in relation to their secondary structure characteristics. CD, used in the stop-flow mode is particularly useful for studying protein–ligand interactions in real time. In order to minimize the signal to noise ratio, in CD studies, solution components should, ideally, be UV transparent; this limits the use of certain buffers, salts, and solvents. Unfortunately because of theoretical problems, the technique cannot be easily/reliably used to determine and interpret tertiary structure or changes therein.

Fourier transform infrared spectroscopy (FT-IR) is also very useful for secondary structure determination of proteins, where the information content is very similar to CD techniques. However there is greater flexibility in FT-IR, at least from the viewpoint of the type of samples that can be investigated: tissue slices, cells, solid state (powder; freeze dried) samples, crystals, thin films, and aqueous proteins and protein–ligand samples. FT-IR can be used to investigate the structure of proteins in the presence and absence of ligands, solubilized in D_2O by using the protein amide 1 band for analysis, provided that the absorption bands of the ligand do not interfere. Use of H_2O is a problem. Other pros and cons of this technique are similar to those for CD and Raman spectroscopy. It is, for example, useful for examining the effect of variations in solution conditions on proteins and protein–ligand interactions. In addition the presence and absence of hydrogen bonds can also be investigated. FT-IR is limited to the use of short path length cells and research grade instruments are not commonly available.

Raman spectroscopy has been used for many years to study structural and enzymatic proteins as well as protein–ligand (particularly enzyme–substrate/inhibitor) interactions, albeit in laboratories specializing in the technique. Again it is an excellent method for the study of variable solution state conditions. The use of resonance Raman is advantageous, since specificity and sensitivity are improved relative to off-resonance Raman. Information about electronic states can also be gained from resonance Raman spectra. Fluorescence effects *may* pose a problem but the FT-Raman technique can very often eliminate these. Its use is not limited by cell path lengths and as in FT-IR different sample states, e.g. cellular tissue can be examined.

4 Non-spectral methods

Equilibrium dialysis and pH titration represents the traditional approach—which dates back to the classical Scatchard analysis—to the study of ligand binding. Another titration probe, *ITC (isothermal titration calorimetry)* together with the related *DSC (differential scanning calorimetry)* technique are now widely and increasingly, used to examine proteins and protein–ligand interactions. The two techniques are complementary and are the only methods currently available which allow the direct determination of the enthalpy of an intra- or intermolecular reaction. Other thermodynamic parameters can also be determined directly, indirectly, or by using model-based assumptions. Both techniques

provide invaluable information in relation to the energetics of a system and significant advances have been made, especially over the last two decades, in the theoretical basis of interpretation of experimental results. It has to be emphasized however that they are macroscopic techniques with attendant advantages and disadvantages which, should not be overlooked in terms of experimental design and interpretation of results. A similar complementary pair of techniques are *sedimentation velocity* and *sedimentation equilibrium analytical ultracentrifugation*. The latter, like calorimetry, provides an absolute thermodynamic probe of interactions (in terms of the stoichiometry and association or dissociation constants, K_d, and the related Gibbs free energy $\Delta G = RT.\ln K_d$) and molecular mass (oligomeric structure). The former (sedimentation velocity) is a good tool for the separation and analysis of heterogeneous mixtures of the various reaction components, as well as a highly useful solution conformation probe of the reactants and products. Another useful procedure—also based on a separation principle (but requiring a separation medium as opposed to a centrifugal field) is *affinity chromatography* (frontal and zonal procedures). Interaction stoichiometries can also be obtained by the technique of coupling size exclusion chromatography to multi-angle laser light scattering. This *SEC/MALLS* procedure (*Table 4*) is very much complementary to analytical ultracentrifugation and is considered in the earlier book by Creighton in the Practical Approach series (see Chapter 9 of ref. 6). Analytical ultracentrifugation and SEC are just two examples of hydrodynamic procedures. Others, based on the electrical properties of proteins and ligands, are *electo-optics* (which can provide valuable information on the effect of ligand binding on solution conformation of a protein) based—like fluorescence depolarization—on the measurement of rotational diffusion behaviour, and the rapidly emerging probe (with its great resolving power on very small quantities) of *capillary electrophoresis*. *Solution X-ray* and *neutron scattering*, also provide a very sensitive handle on the effects of ligand binding on the solution conformation of a protein, and the surface probe of *atomic force microscopy* provides the potential for individual complexes to be visualized (although very difficult to use and data interpretation can be problematic).

5 Computational methods/molecular modelling

Contemporary computing facilities and capacity have revolutionized virtually every aspect of scientific investigations related to protein–ligand interactions. This includes experimental protocol design as well as collection and analysis of data. Most people, unfortunately, associate molecular modelling with the now commonplace colourful complex two- and three-dimensional protein or protein–ligand docking images produced via graphic workstations. However this gives a totally erroneous picture of the importance of molecular modelling techniques for analysing and predicting the physical properties of molecules. It is perhaps, impossible to underestimate the usefulness of such techniques both currently and in the future.

Table 5 List of some of the protein servers on the World Wide Web

Server/software	Description	Web address (URL)
PDB	Database of experimentally determined protein and nucleic acid structures compiled at Brookhaven National Laboratory	http://www.pdb.bnl.gov/
SWIS-PROT	Database of protein sequences	http://expasy.hcuge.ch/www/tools.html
PIR	Database of protein sequences	http://www.gdb.org/Dan/proteins/pir.html
BioSCAN	A rapid search and sequence analysis	http://genome.cs.unc.edu/bioscan.html
BLITZ	Provides specific sequences most similar to or containing the most similar regions to a query sequence	http://www.ebi.ac.uk/searches/blitz_imput.html
NNPREDICT	Predicts secondary structure	http://www.cmpharm.ucsf.educ/~nomi/nnpredict.html
PHD	Automatic service for predicting a protein structure	http://embl-heidelberg.de/predictprotein/predictprotein.html
SWISS-MODEL (PoMod)	Experimental protein modelling server at the Glaxo Institute	http://expasy.hcuge.ch/swissmod/SWISS-MODEL.html
RASMOL (PC) and RASMAC (Mac)	Molecular graphics program intended for visualization of proteins and related molecules	http://www.unmass.edu/microbio/rasmol/

6 Information sources

Useful sources of information include the protein structural databases given *in Table 5*. In addition many scientific literature reference databases now exist. These include, but are not limited to, the following:

(a) Mimas (ISI®, Citation Indexes & ISTP® http://wos.mimas.ac.uk/). Access requires institutional affiliation.

(b) Chemical abstracts.

(c) MedLine.

(d) PubMed (www.ncbi.nlm.nih.gov/). A new facility called PubMed Central is due at the time of going to press.

7 Additional remarks

To date we have obtained, what may appear at first sight, an incredible amount of data and theoretical insight into the properties and behaviour underpinning the function of proteins in biological systems. However in reality it could, nay should, be argued that we have merely scratched the surface in our quest to understand and predict the intricate and interwoven complexities of protein–ligand interactions. There is no doubt that a multidisciplinary approach is required which entails intra- and/or inter-laboratory collaborations which in turn necessitate the collaborator or non-expert to have some level of understanding of the techniques/methods in which they themselves are not experts. It is to be hoped that the subject matter included in this volume will help, at least in part, to achieve this aim.

However the discerning reader will note that the foregoing overview has not dealt with any topic in detail but has merely attempted to wet their appetite for questioning current achievements and future exploration. There are an infinite number of unanswered questions compared to answers. For example, non-covalently controlled phenomena are still very poorly understood. It has not yet been possible to design, *from first principles*, even a small molecular weight molecule that binds with a 'designed' affinity and specificity to a binding site in a protein of known macroscopic structure, despite widespread interest in, e.g. rational drug design. Again at the expense of repetition, questions should be asked, above all by novice researchers as to the applicability of 'test-tube' experiments to biological systems. What are the effects for example, of the fact that biological systems do not operate under equilibrium thermodynamic conditions? Further what are the implications of the concentrations at which molecules occur, the packing density (as well as membrane association) of reactions and the activity of water, in cells compared to the manner and conditions under which *in vitro* experiments are conducted? In truth we are still at the first stage of a long quest. Nonetheless the techniques described in this volume are helping us at least make some inroads in the right direction.

References

1. Longhi, S., Czjzeck, M., and Cambillau, C. (1998). *Curr. Opin. Struct. Biol.*, **8**, 730.
2. Williams, D. H. and Westwell, M. S. (1998). *Chem. Soc. Rev.*, **27**, 57.
3. Philo, J. S. (1999). In *Current Protocols in Protein Science* (ed. P. T. Wingfield), pp. 20.1.1.–20.1.13. J. Wiley & Sons, NY.
4. Hensley, P. (1996). *Structure*, **4**, 367.
5. Harding, S. E. (1993). *Biotech. Gen. Eng. Rev.*, **11**, 317.
6. Creighton, T. E. (1997). *Protein structure: a Practical Approach*. Oxford University Press.