1. THEORETICAL PART

1.1 Iron-Sulfur Proteins: [2Fe-2S] Ferredoxin, Bovine Adrenodoxin

1.1.1 Classification of non-heme iron proteins

Non-heme iron proteins are ubiquitous proteins found in all life forms: bacteria, plants, and animals. The investigation of these proteins began in the 1960s. New classes of oxidation-reduction proteins were defined. Their primary function lies in mediating one-electron redox processes by molecular interactions with a variety of redox enzymes and proteins involved with carbon, oxygen, hydrogen, sulfur, and nitrogen metabolism.

Non-heme iron proteins were termed as ferredoxins, because of their iron-sulfur (Fe-S) centers. But, it was difficult to classify these proteins from others, containing iron-sulfur cluster(s) as well. Nevertheless, they were strictly differentiated from those non-heme proteins and polypeptides which contain additional prosthetic groups such as flavin, molybdenum centers or heme. Recommendations for the nomenclature of iron-sulfur proteins were first formulated in 1971 and revised in 1978. The most common types of an iron-sulfur cluster are [Fe_2S_2], [Fe_3S_4], and [Fe_4S_4] (Figure 1.1.1.1).

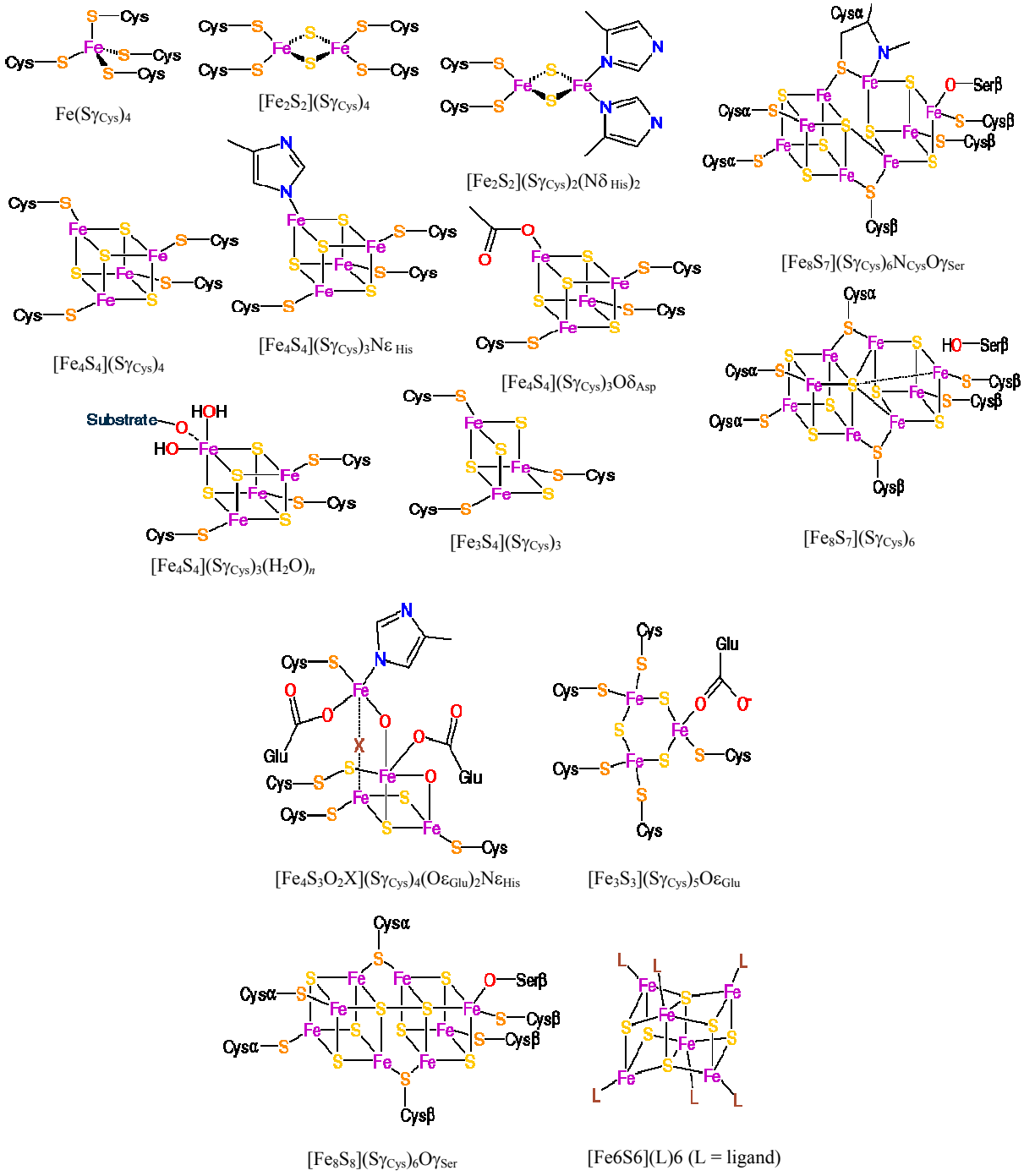

Figure 1.1.1.1 The geometry of typical Fe-S clusters. From: metallo.scripps.edu/PROMISE/FESMAIN.html

Structural analyses have shown that there is a little variation in the structure of these clusters of a given type in different proteins [Capozzi et al., 1998]. In most instances, the iron is

bound to sulfur(s) from cysteine residues in the polypeptide backbone and also to inorganic sulfurs in an iron-sulfur cluster.

This PhD thesis is focused on one particular ferredoxin family protein, the [2Fe-2S] ferredoxin, adrenodoxin (Adx). Before describing a typical family member, adrenodoxin, it will be of importance to give a short description of this family of proteins in general.

1.1.2 The [2Fe-2S] proteins

There are plant and vertebrate [2Fe-2S] ferredoxins, which are classified by structure and function [Müller et al., 1999]. Plant-type ferredoxins, originally found in chloroplast membranes, have been termed “chloroplast-type”. In chloroplasts, the [2Fe-2S] ferredoxins function as electron carriers in the photosynthetic electron-transport chain and as electron donors to various cellular proteins, such as glutamate synthase [Hirasawa et al., 1991], nitrate reductase [Privalle et al., 1995] and sulfite reductase [Krueger & Siegel, 1982]. Vertebrate-type iron-sulfur proteins are present in oxygenase systems of bacteria and vertebrates, where they transfer electrons from a reduced nicotinamide adenine dinucleotide phosphate- (NADPH) dependent ferredoxin reductase to different cytochrome P450 enzymes. In hydroxylating bacterial dioxygenase systems, they serve as intermediate electron-transfer carriers between reductase and oxygenase [Mason & Cammack, 1992]. In vertebrates, they are present in the adrenal cortex, placenta, liver, kidney and brain, where they participate in cytochrome P450-catalyzed hydroxylation reactions to produce steroid hormones, vitamin D metabolites, and bile acids [Hannemann et al., 2001].

Despite some structural differences, mostly in a variable flexible interaction domain, and low sequence identity (20%) between vertebrate and plant-type [2Fe-2S] proteins, the common feature of these proteins is that two irons in the active center are tetrahedrally coordinated, both by two inorganic sulfurs and by sulfurs provided by four conserved cysteine residues. This, together with a mostly conserved core domain, is a very important structural feature of these proteins that maintain their function - electron transport *via* [2Fe-2S] (Figure 1.1.2.1).

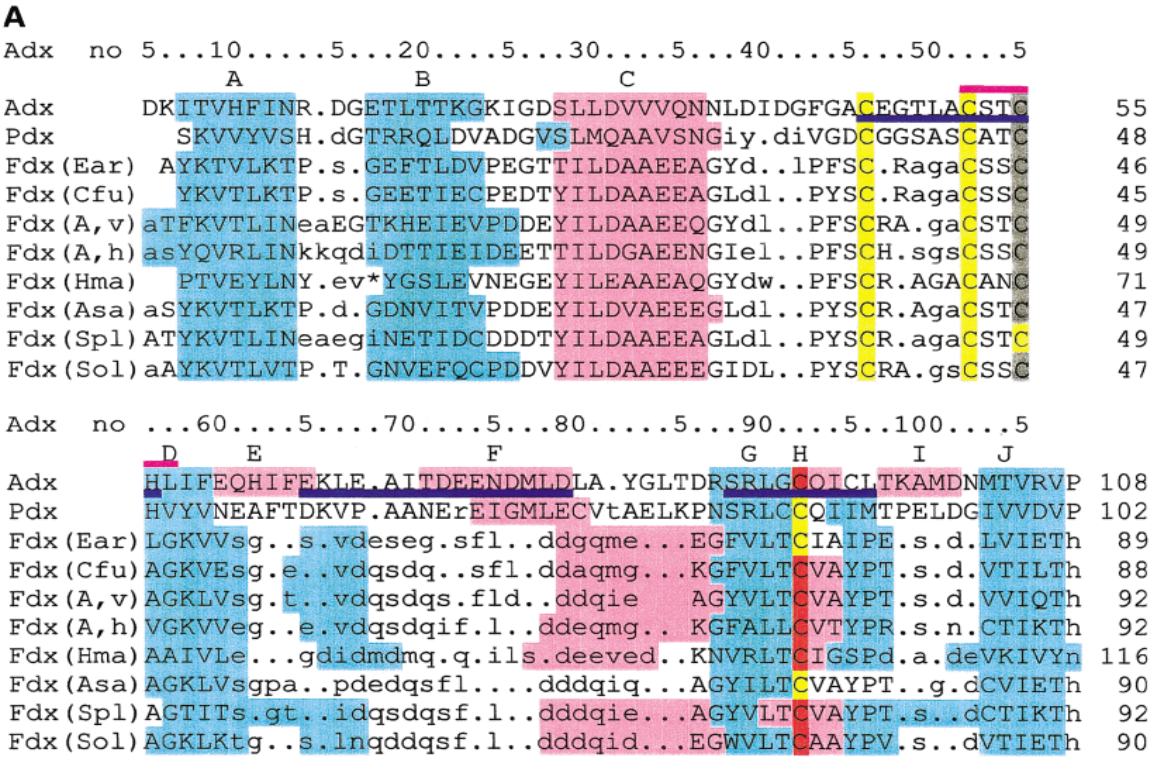

Figure 1.1.2.1 A. Structure-based alignment of vertebrate-type and plant-type ferredoxins. Residues at equivalent positions to adrenodoxin are represented by upper-case letters. Helical regions are red, β-strands green, and cluster-liganding cysteinyls yellow (appearing orange in helices, olive in sheets). The residues of Adx marked with blue bars belong to the adrenodoxin-specific motif, and the peptide strand marked magenta binds specifically to cytochrome c (defined in PROSITE [Bairoch et al., 1997]; Adx [Müller et al., 1998]; Pdx, putidaredoxin [Pochapsky et al., 1996]; Fdx(Ear), ferredoxin from *Equisetum arvense* [Ikemizu et al., 1994]; Fdx(Cfu), Fdx from *Chlorella fusca* [Bes et al., 1999]; Fdx(A,v), Fdx from *Anabaena*, vegetative form [Rypniewski et al., 1991]; Fdx(A,h), Fdx from *Anabaena*, heterocyst form [Jacobson et al., 1993]; Fdx(Hma), Fdx from *Haloarcula marismortui* [Sussman et al., 1989]; Fdx(Asa), Fdx from *Aphanothece sacrum* [Tsukihara et al., 1990]; Fdx(Spl), Fdx from *Spirulina platensis* [Tsukihara et al., 1981]; Fdx(Sol), Fdx from *Spinacia oleracea* [Binda et al., 1998]).

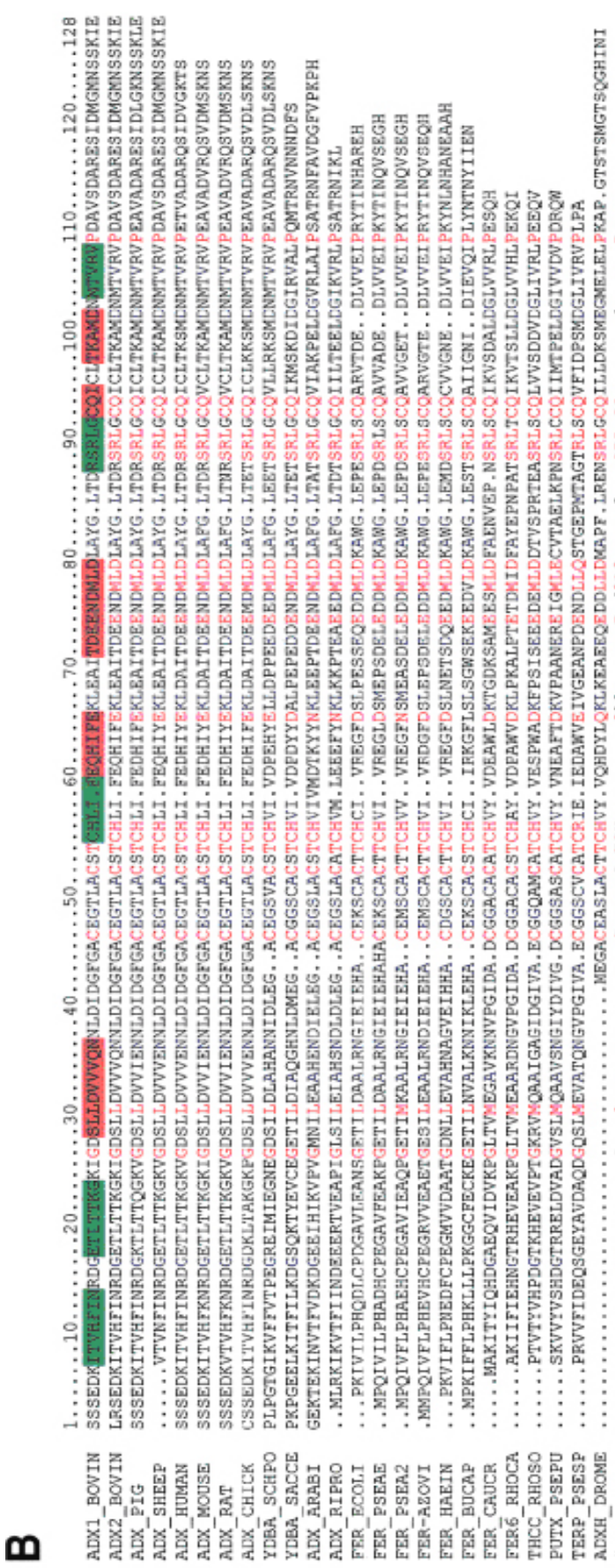

Figure 1.1.2.1 (Continued). B. Sequence alignment of vertebrate-type [2Fe-2S] ferredoxins that contain the adrenodoxin motif C-x(2)-[STAQ]-x-[STAMV]-C-[STA]-T-C-[HR] (PS00814, PROSITE database [Bairoch et al., 1997]). Database searches (GeneStream; http://wwwz.igh.cnrs.fr/) were done with FASTA [Pearson & Lipman, 1998] and the sequence alignment results are from MultAlin [Corpet 1988]. Consensus levels: high (colored red) ≥90%, intermediate (colored blue) ≥50%, low (colored gray) ≤50% residue identity. $ stands for L or M, # stands for N, D, Q, or E. Helical regions in Adx(4-108) are marked by a red background, β-strands in green. The abbreviations used for ferredoxins are self-explanatory for ADX1-BOVIN, ADX2-BOVIN, ADX-PIG, ADX-SHEEP, ADX-HUMAN, ADX-MOUSE, ADX-RAT, ADX-CHICK; YDBA-SCHPO *Schizosaccharomyces pombe*, YDBA-SACCE *Saccharomyces cerevisiae*, ADX-ARABI *Arabidopsis thaliana*, ADX-RIPRO *Rickettsia prowazekii*, FER-ECOLI *E. coli*, FER-PSEAE *Pseudomonas aeruginosa*, FER-PSEA2 *P. aeruginosa*, FER-AZOVI *Azotobacter vinelandii*, FER-HAEIN *Haemophilus influenzae*, FER-BUCAP *Buchnera aphidicola*, FER-CAUCR *Caulobacter crescentus*, FER6-RHOCA *Rhodobacter capsulatus*, THCC-RHOSO *Rhodococcus sp* (rhodocoxin), PUTX-PSEPU *Pseudomonas putida* (putidaredoxin), TERP-PSESP *Pseudomonas sp.* (terpredoxin), ADXH-DROME *Drosophila melanogaster*. The picture was borrowed from: [Grinberg et al., 2000].

1.1.2.1 Bovine mitochondrial adrenodoxin

The nature of the enzymes responsible for steroid biosynthesis remained unclear for a long time. It was suggested that the steroid hydroxylase system consists of at least three components, which can be separated by purification procedures. This electron-transfer system functions to provide active reduced oxygen, which can be incorporated into the steroid molecule [Kimura 1968]. The first successful purification and characterization of these components were achieved in the 1960s [Kimura 1966; Suzuki & Kimura, 1965; and Omura et al., 1965]. The discovery of adrenodoxin reductase (AR), Adx, and side-chain cleavage cytochrome P450 ($P450_{scc}$, CYP11A1) showed the whole complexity of this system and specific relationships between the redox partners (Figure 1.1.2.1.1).

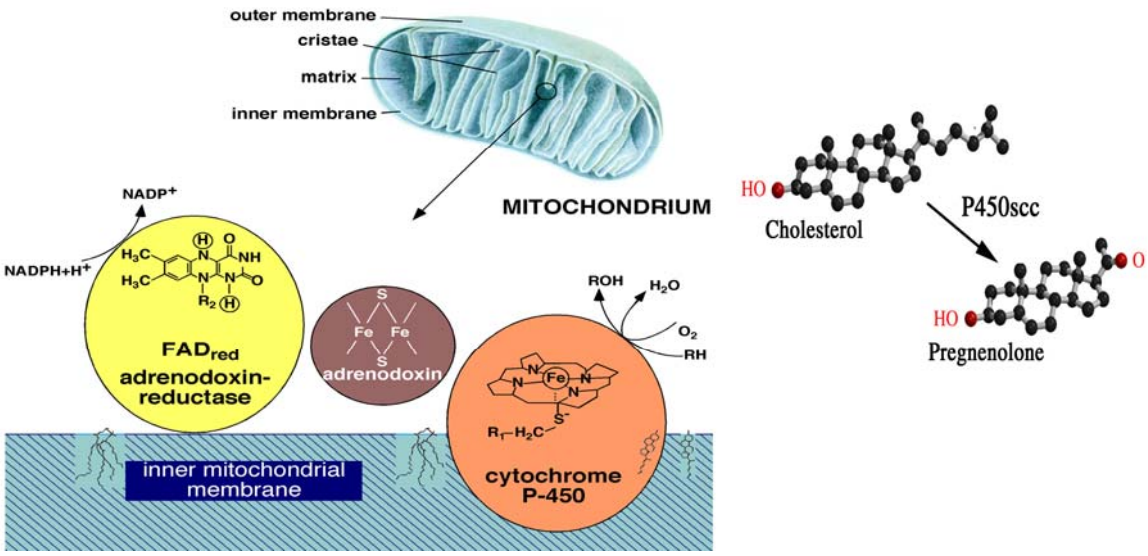

Figure 1.1.2.1.1 Organization of the steroid hydroxylase system in mitochondria of the adrenal cortex.

The mature bovine Adx is a low-molecular-mass (12-14 kDa) soluble protein, consisting of 128 amino-acid residues. The protein is negatively charged at neutral pH values and contains the [2Fe-2S] cluster as a redox-active group. Bovine Adx is encoded by a nuclear gene, synthesized in the cytoplasm and processed upon mitochondrial uptake. The mechanism of iron-sulfur cluster incorporation is still unclear, although during heterologous expression in *E. coli*, it can be assembled in the cytoplasm as well as in the periplasm [Hannemann et al., 2001]. Adx functions as an electron carrier in the steroid hydroxylase system where it transfers electrons from AR to cytochromes $P450_{scc}$ and bovine 11β-hydroxylation cytochrome P450 (P45011β, CYP11B1), thus, participating in the biosynthesis of all steroids.

Adx is the best characterized member of the family of the vertebrate [2Fe-2S]-type ferredoxins [Kimura & Suzuki, 1965; Suzuki & Kimura, 1965; Watari & Kimura, 1966; Mukai et al., 1973; Omura et al., 1965; Vallee & Ulmer, 1965; Müller et al., 1998; Pikuleva et al., 2000; Takeuchi et al., 2001; Grinberg & Bernhardt, 2001; Cammack et al., 1971; Fu et al., 1992; and Beilke et al., 2002]. Early studies dealt with attempts to classify this protein, so that Adx was often studied in comparison to other known plant- and vertebrate-type ferredoxins.

Before the first crystal structures of the proteins of the steroid hydroxylase system were known, the most powerful techniques to study the protein-protein interactions and the architecture of individual components of the system were chemical modification and mutagenesis studies. Structural, spectroscopic, redox and physiological consequences of site-directed mutations and specific chemical modifications were all explored for bovine Adx. Individual residues of Adx important for its electron-transfer reactivity were identified by this approach.

Site-directed mutagenesis and chemical modification studies. Tyr82, Pro108, Thr54, and His56 are conserved among ferredoxins (Figure 1.1.2.1.B and Figure 1.1.2.1.2) and are the most thoroughly investigated amino-acid residues of the core domain of bovine Adx. They were found to be responsible for tuning of the electron-transfer rates or transferring structural changes during the redox process [Beckert et al., 1994; Beckert et al., 1995; and Uhlmann & Bernhardt, 1995].

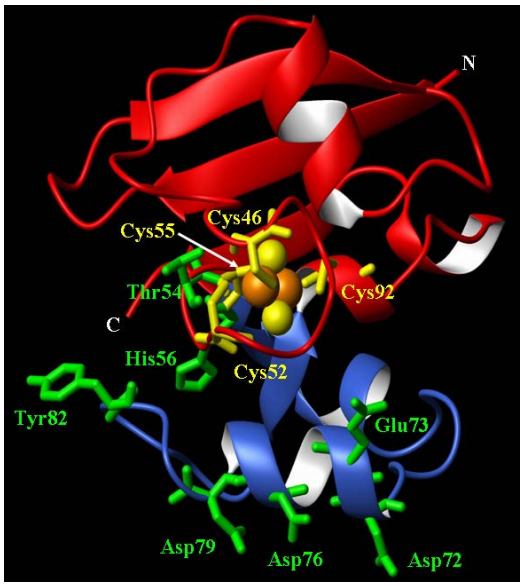

Figure 1.1.2.1.2 The chemical modification and site-directed mutagenesis residues of bovine adrenodoxin. Modified amino-acid residues of the core (red) and flexible interaction (royal blue) domains are shown. A white arrow points to Cys55. The iron and inorganic sulfur atoms of the [2Fe-2S] cluster are shown as dark orange and yellow spheres, respectively. The picture was prepared using the program MOLMOL [Koradi et al., 1996].

Tyr82 was proposed to play a role in interactions with AR and/or in electron transfer [Taniguchi & Kimura, 1975; Taniguchi & Kimura, 1976]. It is known that tyrosine residues may participate in electron-transfer processes in different proteins [Adman et al., 1973; Debus et al.,

1988; and Packer et al., 1972]. To investigate the role of Tyr82 of Adx, *Beckert et al.* [Beckert et al., 1994] carried out mutations of this residue. The calculated V_{max} values in cytochrome *c* reduction, CYP11A1- and CYP11B1-dependent activities remained unchanged when using Y82F, Y82S, and Y82L Adx mutants as electron donors, demonstrating that Tyr82 is not involved in the electron transfer. There were, however, changes in K_m values up to 4-fold when the enzymatic activities of mutant Adx with CYP11A1 and CYP11B1 were measured. The result suggested that mutation of Tyr82 either directly or indirectly (by inducing small conformational changes of the binding domain) affects the binding of P450 cytochromes.

The role of Thr54 was studied by analyzing T54S and T54A mutants of Adx [Uhlmann & Bernhardt, 1995]. It was found that the T54S substitution has no influence on the stability of the ferredoxin. This conclusion was made from the absorption, circular dichroism, fluorescence, and electron paramagnetic resonance spectra. The decreased stability of the T54A mutant as compared to the wild-type form and T54S mutant indicates that a hydrogen bond donor at this position stabilizes the protein. Despite the fact that both mutants were functionally active, the replacement of Thr54 led to rearrangements at the recognition sites for Adx redox partners.

His56 seems to play a key role in stabilization of the Adx structure. Replacement by Gln reduces the enthalpy change of unfolding by 100 kJ/mol and the Gibbs energy change by about 7 kJ/mol [Burova et al., 1996]. The effects of replacing His56 by Gln or Thr on cytochrome *c* reduction and cytochromes P45011β- and $P450_{scc}$-dependent substrate conversions were studied [Beckert et al., 1995]. As for Tyr82, it was shown that His56 does not play a role in electron transfer. Since His56 is located in the immediate environment of Tyr82, the possible involvement (directly or indirectly) of His56 in protein-protein interaction with redox partners is also not excluded.

The C-terminal tail of Adx may play a role in stabilizing an interaction between ferredoxin and cytochromes P450. *Schiffler et al.* [Schiffler et al., 2001] and *Burova et al.* [Burova et al., 1996] investigated the influence of the C-terminal truncation of Adx on the kinetics of protein-protein interaction and heme reduction of $P450_{scc}$ and P45011β. Two different truncated mutants were characterized, Adx(4-108) and Adx(4-114). *Burova et al.* [Burova et al., 1996] could show using calorimetric and limited proteolysis experiments that Adx(4-108) had a more compact overall structure indicating that differences between the wild-type Adx (Adx(WT)) and the truncated Adx exist in solution. This deletion mutant is apparently more stable than the wild-type one, as it was judged by higher specific denaturation enthalpy and resistance toward proteolytic

degradation. On the contrary, the deletion mutant Adx(4-114) was less stable. Thus, the hydrophilic C-terminal sequence 108-128, which could not be allocated in model building of Adx, produces unfavorable interactions rather than contributing to the overall protein stability. On the other hand, further cleavage of C-terminal residues, in particular removal of Pro108, led to the loss of the [2Fe-2S] cluster in Adx [Uhlmann et al., 1994]. Pro108 has been shown to be important for correct folding of the protein and subsequent incorporation of the [2Fe-2S] cluster [Uhlmann et al., 1994; Grinberg & Bernhardt, 1998]. It was found that these mutants exhibit lowered redox potentials, which improved the ability to reduce $P450_{scc}$. Another function of the C-tail was thought to reside in dimerization of Adx molecules [Pikuleva et al., 2000]. Evidence for this was provided by the crystal structure analysis of full-length Adx, as will be discussed further.

An enormous number of cross-linking studies were carried out, involving all three components of the P450 system. In all these investigation, the main role belongs to Adx that makes a large contribution to the protein-protein interactions and to the electron-transfer process in general. These studies can be divided into three parts: (i) cross-linking of AR and Adx [Lapko et al., 1997; Geren et al., 1984; Usanov et al., 1985; and Hara & Kimura, 1989]; (ii) cross-linking of Adx molecules [Beilke et al., 2002]; (iii) cross-linking of Adx and $P450_{scc}$ [Müller, E.-Ch. et al., 2001].

Müller et al. [Müller, E.-Ch. et al., 2001] were able to show that there is no steroidogenic hydroxylase activity of a cross-linked Adx-$P450_{scc}$ complex which could be detected in the reconstituted test system. An explanation was that there is a crosslink between Asp79 of Adx and Lys403 of $P450_{scc}$. Because, the acidic region between Asp72 and Asp79 belongs to the primary binding site of Adx to AR, as well as to $P450_{scc}$, the covalent bond within the Adx-$P450_{scc}$ complex prevents electron transfer by the proposed "shuttle" mechanism. This assumption is also supported by the fact that, in general, oxidized Adx inhibits cholesterol side-chain cleavage activity [Hanukoglu & Jefcoate, 1980]. However, the cross-linked complex between oxidized AR and Adx was shown to be functional in the side-chain cleavage reaction, but with one assumption that an excess of Adx will be present in solution [Lapko et al., 1997]. Covalent AR-Adx complex could also catalyze the reduction of cytochrome *c*, but at a three- to four-fold slower rate as compared to free Adx. Thus, cross-linking studies between AR, Adx, and $P450_{scc}$ provide evidence favoring the "shuttle" model over the cluster model, when a ternary 1:1:1 complex between proteins is formed or even a quaternary complex with a 1:2:1 stoichiometry.

Indeed, experiments with Adx, only, indicate that the iron-sulfur protein may form dimers in the oxidized state, whereas monomers exist predominantly in the reduced form [Beilke et al., 2002].

To summarize, it should be said that the chemical and mutagenesis studies have prepared a good basis for understanding the mechanism of action of Adx on the molecular level, which can now be investigated using X-ray crystallography.

X-ray crystallographic studies: Mechanism of electron transfer by adrenodoxin. The crystal structure of the truncated form of bovine Adx was solved using the multiple-wavelength anomalous diffraction technique (MAD) at 1.85-Å resolution [Müller et al., 1998]. The structure provided the first detailed description of a vertebrate [2Fe-2S] ferredoxin, displaying the typical compact *αβ*-fold. The polypeptide chain is organized into a large core domain and a smaller interaction domain (Figure 1.1.2.1.2).

Based on the high-resolution crystal structure of Adx(4-108), the detailed geometry of the [2Fe-2S] cluster can now be described. The redox center is localized in the core domain and it is accessible to solvent from six sides. In the [2Fe-2S] prosthetic group two iron atoms are coordinated to four cysteines (46, 52, 55, and 92) and two labile sulfur atoms (Figure 1.1.2.1.3). The cluster is in a plane, which is perpendicular to the plane formed by the Sγ atoms of the cysteine residues. The active center is stabilized by hydrogen bonds between inorganic sulfurs, cysteines, and surrounding donor atoms.

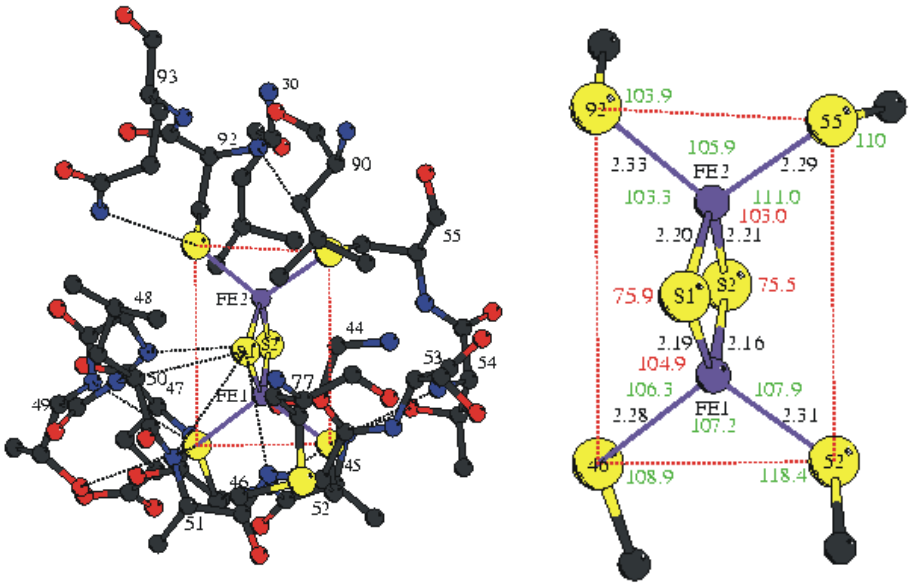

Figure 1.1.2.1.3 View of the [2Fe-2S] cluster binding region of Adx(4–108). The plane defined by the Sγ atoms of the cluster-binding cysteines is marked by red dotted lines; the plane spanned by the iron and sulphur atoms of the cluster is perpendicular to it. The hydrogen-bonding pattern is indicated by black dotted lines. The insert on the right side of the figure contains bond distances (in Å; black) and angles (in degrees; green) for the cluster and covalently bound cysteinyl sulphurs. Angles within the [2Fe-2S] cluster are labeled in red. Fe atoms are shown in purple, sulfur atoms in yellow and other atoms are in standard colours. The picture is borrowed from: [Müller et al., 1998].

The crystal structure also suggests how a change in the redox state of the cluster may be coupled to a motion of the flexible interaction domain of Adx. The residues that are involved in interactions with redox partners, AR and $P450_{scc}$, are located on the molecular surface and coupled to the [2Fe-2S] cluster *via* structurally equivalent hydrogen bonds. The clearly asymmetric charge distribution on the protein surface and the resulting strong molecular dipole are involved in electrostatic steering of the interactions with AR and $P450_{scc}$ (Figure 1.1.2.1.4) [Müller et al., 1998].

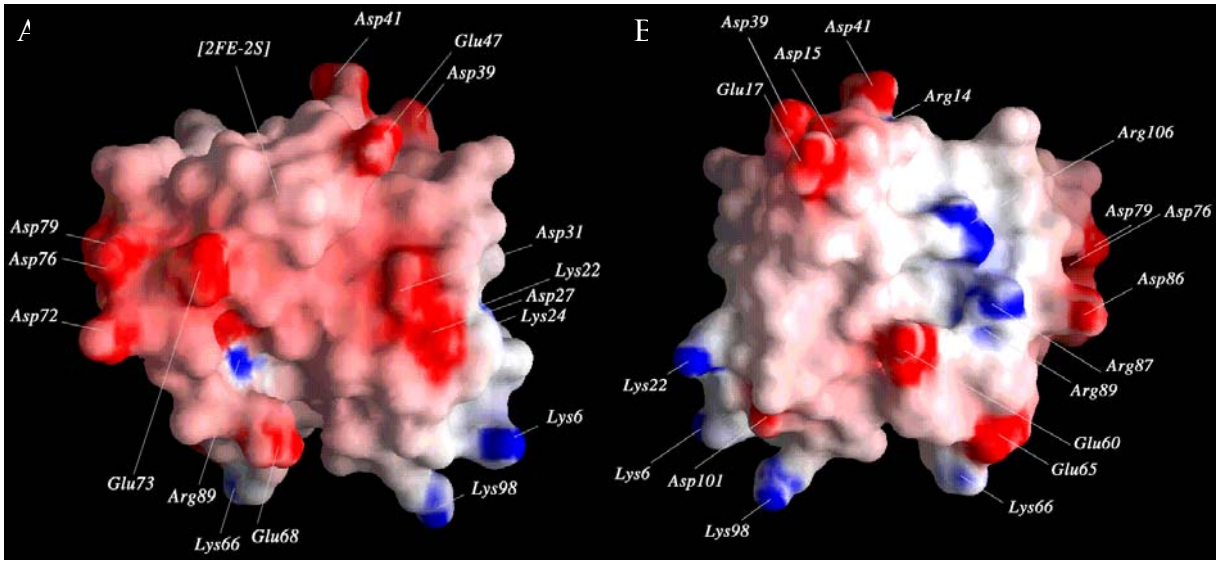

Figure 1.1.2.1.4 Molecular surface of Adx(4-108) colored according to electrostatic potential (blue, positive; red, negative). A. Front view of Adx(4-108). The [2Fe-2S] cluster is located within the protuberance at the top of the figure; the positions of charged residues are indicated. B. View of Adx(4-108) after rotation by 180° around the vertical axis. The picture was borrowed from: [Müller et al., 1998].

The mechanism of transfer of six electrons from NADPH to $P450_{scc}$, *via* AR, and Adx has not been finally explored. Three alternative models have been suggested in the literature (Figure 1.1.2.1.5): (i) a "shuttle" model in which Adx carries electrons from AR to $P450_{scc}$ by sequential binding to these proteins [Lambeth et al., 1982; Hanukoglu et al., 1981], (ii) an electron transfer through an organized ternary complex of AR, Adx, and $P450_{scc}$ [Kido & Kimura, 1979], and (iii) a quaternary complex of AR, Adx dimer, and $P450_{scc}$ [Hara & Takeshima, 1994].

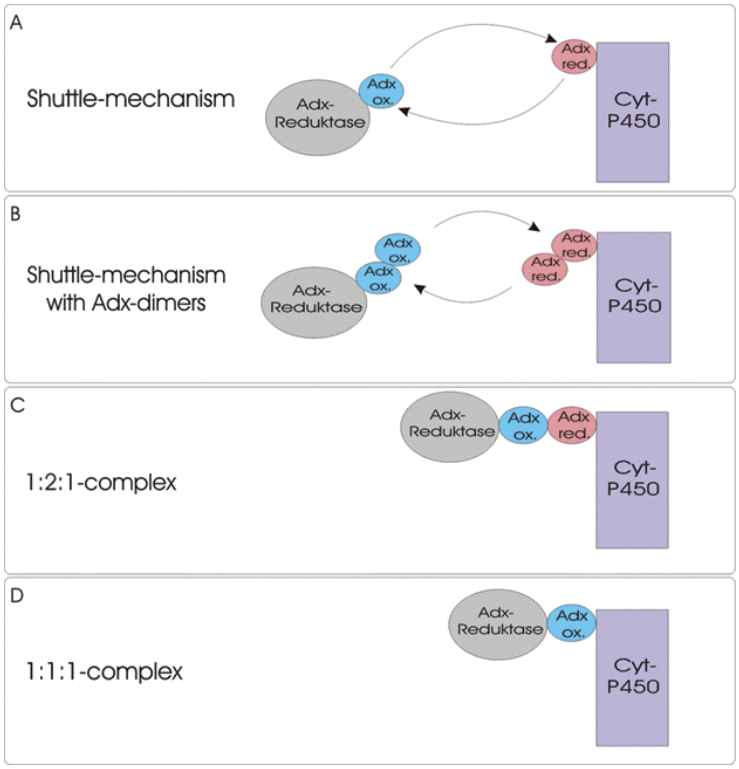

Figure 1.1.2.1.5 The proposed mechanisms of electron transfer in the steroid hydroxylase system. A. The "shuttle" model where a Adx monomer serves as mobile carrier between AR and P450. B. Modified "shuttle" model in which Adx dimers serve as electron carrier. C. A quaternary complex. D. A ternary complex. The picture was borrowed from: [Beilke et al., 2002].

Direct evidence on protein-protein interactions relevant to these models can be provided by crystallization of biologically active protein-protein complexes. The recently solved crystal structure of the cross-linked complex between AR and Adx provides first detailed description of their structural elements involved in intermolecular interaction and favors the "shuttle" hypothesis (Figure 1.1.2.1.6) [Müller, J. J. et al., 2001]. In the 2.3-Å resolution crystal structure of the AR-Adx complex, the main interaction sites are around Asp79, Asp76, Asp72, and Asp39 of Adx and around Arg211, Arg240, Arg244, and Lys27 of AR. Thus, negatively charged amino-acid residues from the recognition-interaction domain of Adx participate in the binding with positively charged amino acids of AR.

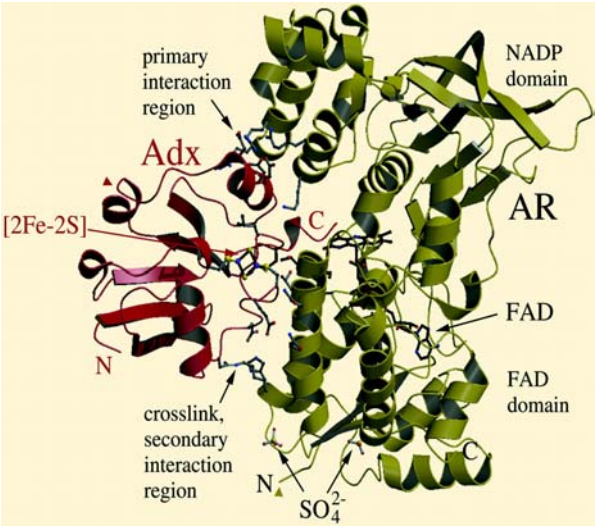

Figure 1.1.2.1.6 The crystal structure of the AR-Adx complex [Müller, J. J. et al., 2001]. Contacts occur at the primary and secondary interaction regions and the region between the [2Fe-2S] cluster of Adx and the isoalloxazine ring of the FAD group of AR. C-terminal residues of Adx are also in contact with AR. The side chains of some residues involved in polar AR-Adx interactions are displayed. The brown triangle marks the position of Adx Lys66 and the green marks AR Glu4, both residues involved in another cross-link as reported recently [Hara & Kimura, 1989; Hara & Miyata, 1991].

Müller et al. [Müller, J. J. et al., 2001] were able to compute possible electron-transfer routes between AR and Adx. The electrons would travel from the flavin-adenine dinucleotide (FAD) group of AR to the [2Fe-2S] of Adx along covalent bonds, requiring of two through-space jumps within the pathway.

The hypothesis, invoking a quaternary protein complex with Adx dimers [Hara & Takeshima, 1994] has also been considered [Pikuleva et al., 2000; Beilke et al., 2002]. In the work of *Pikuleva et al.* [Pikuleva et al., 2000] two models of such an Adx dimer are discussed. The first is based on the assumption that the iron-sulfur clusters of both Adx molecules are involved in electron transfer; the other assumes that the iron-sulfur cluster of only one Adx molecule in the dimer participates in this process. Since it was shown that an electron could be transferred from AR to $P450_{scc}$ by an Adx dimer [Hara & Kimura, 1989], the dimer might be functional. *Pikuleva et al.* [Pikuleva et al., 2000] have predicted possible electron-transfer pathways from one Adx molecule to another within a dimer. Based on NMR and cross-linking studies, *Beilke et al.* [Beilke et al., 2002] have proposed a mechanism of electron transfer *via* Adx dimers that supplements the "shuttle" mechanism suggested by *Lambeth et al.* [Lambeth et al., 1982]. The *Beilke* model combines the two models of *Pikuleva*, in way of a suggesting a possible final model, if a dimer would really exist *in vivo*. In the *Beilke* model, both Adx molecules in a dimer are reduced in a consecutive manner. First, Adx molecule in a dimer that does not interact with AR will be reduced and will dissociate from the second Adx molecule still contacting the reductase. Second, Adx still bound will be reduced by the second electron transfer and will also dissociate, but from AR. From the physiological point of view this model might have merit. In general, it confirms that Adx functions as a shuttle, either as a monomer or dimer.

It is probable that if the Adx dimers exist, they will speed up the electron transfer, interacting with adrenodoxin reductase in competitive manner, side by side with monomers of Adx. Moreover, the question why the truncated adrenodoxin affects the binding to P450 [Burova et al., 1996; Schiffler et al., 2001; and Uhlmann et al., 1994] could also be answered by suggesting that Adx may dimerize [Pikuleva et al., 2000; Xia et al., 1998].

In order to complete the discussion of the mechanism of electron transport in the P450 system, the crystal structure of $P450_{scc}$ will be required, and ideally even a crystal structure of a complex between Adx and the cytochrome. Electron-transfer pathways may be then computed and interaction sites can be identified in comparison with those obtained from the cross-linking studies. However, it is common knowledge that membrane proteins are difficult to crystallize at all, and in fact, there was no real progress in obtaining suitable for diffraction crystals of $P450_{scc}$ in the last 20-30 years. While there is a lack of structural studies on $P450_{scc}$, investigators are trying some other techniques to clarify the mechanism of electron transfer in the mitochondrial steroid hydroxylase system of adrenal cortex. In this PhD thesis an alternative approach was used to examine whether a single Adx could reduce $P450_{scc}$ in the absence of AR, and, hence, to probe the "shuttle" model for Adx. A photo-induced electron transfer from a photo-sensitive ruthenium(II) complex covalently attached on Adx surface near the [2Fe-2S] cluster was investigated. Photoreduction electron-transfer rates were calculated. EPR measurements were carried out to confirm photoreduction of the iron-sulfur cluster of Adx.

1.2 Electron transfer in chemistry and biology

Since the late 1940s, electron-transfer reactions between ions and molecules in solution have been the subject of considerable experimental study, because of a big interest to calculate the rate of chemical reaction. The idea of Willard Libby to apply the Franck-Condon principle to the movement of an electron between two molecules was further investigated and developed by Rudolf Marcus. He was the first who was able to calculate a size of the energy barrier, which was know as a factor that slows down the rate between bivalent and trivalent iron ions in aqueous solution. This gave him the stimulus to extend his mathematical models to cover electron transfer between different kinds of molecules and to formulate his theory. The contribution of the Marcus theory to our understanding of such widely varying phenomena as the capture of light energy in green plants, electron transfer in biological systems, inorganic and organic oxidation and reduction processes and photochemical electron transfer led to the development of many new research programs, thereby demonstrating the lasting impact of his work on science as a whole. In October 1992, Dr. Rudolf Marcus was awarded the Nobel Prize in Chemistry for his work describing definable differences in the transfer rates of electrons from one atom to another, and from one molecule to another.

1.2.1 Electron transfer in chemistry: The classical theory

How can the rate constant of an electron-transfer reaction between two reactants be calculated if they do not share a common atom or group? For a bimolecular out-sphere electron-transfer reaction to occur, the reactants must first approach each other to enhance the coupling of their orbitals. As before and after the electron transfer such a system proceeds in the vicinity of the equilibrium, from the vibrational coordinates of the reactants (R) to those of the products (P) with an influence of orientational coordinates of the surrounding solvent molecules. This defines a many-dimensional potential-energy surface between reactants, products, and surrounding medium and could be schematically represented by a one-dimensional profile as in Figure 1.2.1.1.A.

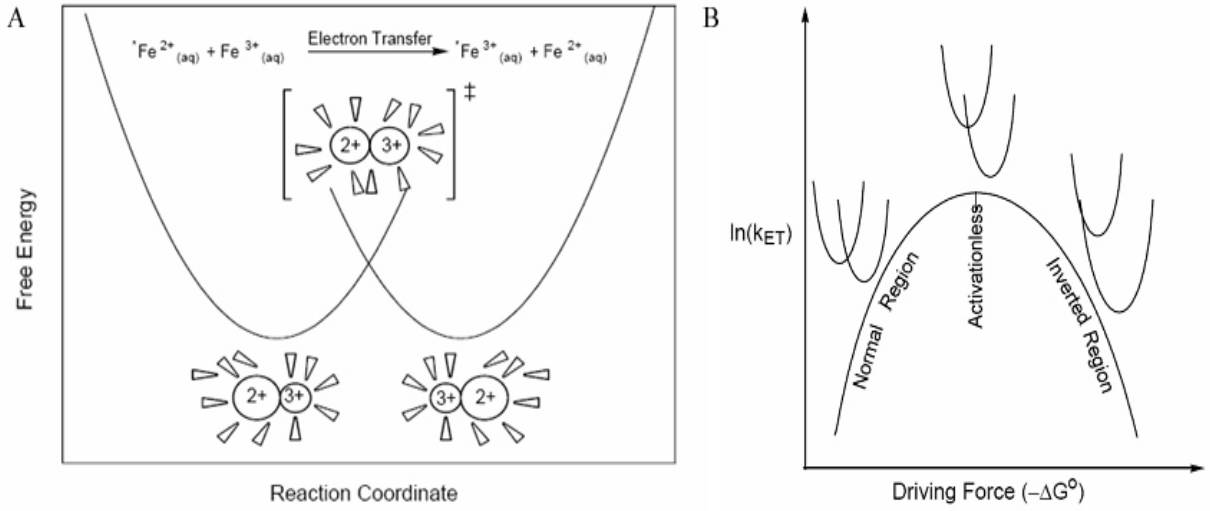

Figure 1.2.1.1 A. Plot of the free energy of the reactants and products as a function of nuclear configuration (reaction coordinate) for ferric-ferrous self-exchange, as an example. Thermal electron transfer occurs at the transition-state configuration. The change in nuclear configuration of reactants (where the circles represent inner-sphere coordinates, and the arrows represent solvent dipoles) to form products is illustrated by the difference in the radii of the oxidized and reduced forms of the redox couple and the change in the average orientations of the solvent dipoles [Sutin et al., 1988]. B. Logarithm (base e) of electron-transfer rate *versus* free energy plot. From: [Sutin et al., 1988].

The transfer will occur only at or near nuclear configurations for which the total potential energy of the reactants and surrounding medium is equal to that of the products and surrounding medium, i.e., at the intersection of curves in Figure 1.2.1.1.A. To reach this electron-transfer region, the nuclei must undergo fluctuations of their positions. Once the system reaches the transfer region, the probability of going from the *R* to the *P* surface (Figure 1.2.1.1.A) depends on the extent of coupling of the electronic orbitals, which in turn depends on the separation distance of the reactants. One has to denote κ(*r*), the transmission coefficient or averaged transition probability for electron transfer per passage of the system through the intersection region at a fixed separation distance *r* of the two reactants. The latter will contribute to the rate constant. To make κ(*r*) larger, *r* should be small; on the other hand, any short-range of Coulombic repulsion between the reactants will favor large *r*. Taking into account all noted factors one can calculate the rate constant of a bimolecular electron-transfer reaction which will be given by the equation:

$$k = \kappa \cdot A \cdot \sigma^2 \exp\left(-\Delta G^* / RT\right) \qquad (1.2.1.1)$$

where $A\sigma^2$ has dimensions of collision frequency, σ is the average center-to-center distance in the reacting pair during the electron transfer, ΔG^* is the free energy of activation. The latter is related to λ, reorganization energy, to ΔG^0, the 'standard' free energy of reaction in the prevailing medium, and to the work of bringing the reactants (w^r) or products (w^p) to the mean separation distance σ[Marcus 1965; Marcus 1982]:

$$\Delta G^* = w^r + \frac{\lambda}{4}\left(1 + \Delta G^{0'}/\lambda\right)^2 \qquad (1.2.1.2)$$

with λ given by:

$$\lambda = \lambda_i + \lambda_o \qquad (1.2.1.3)$$

where λ_i reflects changes in the bond lengths of the reactants and λ_o reflects changes in solvent orientation coordinates [Marcus & Sutin, 1985].

$$\Delta G^{0'} = \Delta G^0 + w^p - w^r \qquad (1.2.1.4)$$

where $\Delta G^{0'}$ in Eqn. 1.2.1.4 equals the free energy of reaction when the reactants are a distance r apart in the prevailing medium. For example, by suitable variation of a ligand in a reactant, $\Delta G^{0'}$ is made increasingly negative at constant λ so that ΔG^* initially decreases and that, when $-\Delta G^0$ exceeds λ, ΔG^* begins to increase. This region where $-\Delta G^{0'} > \lambda$ has been termed as the inverted region [Marcus 1965]. In Figure 1.2.1.1.B it could be explained by lowering the P surface vertically (or raising the R surface) that intersection of both surfaces occurs at the minimum of the R surface and that there is no barrier. Thus, electron-transfer reactions occur when $-\Delta G^0 < \lambda$, $-\Delta G^0 = \lambda$, $-\Delta G^0 > \lambda$. Further lowering the P surface raises the intersection, and hence, it raises the barrier (ΔG^* increases).

1.2.2 The quantum-mechanical aspects of the electron-transfer theory

In this section, extensions to the classical theory described above will be discussed. The first-order rate constant will be now written for very low electronic coupling, when $\kappa << 1$, as

$$k = \frac{2\pi}{h} H_{AB}^2 (FC) \qquad (1.2.2.1)$$

where H_{AB} is the electronic matrix element describing the electronic coupling of the reactants' electronic state with the products' and is equal to one-half the separation of the dotted curves at the intersection of the *R* and *P* surfaces in Figure 1.2.2.1.A.

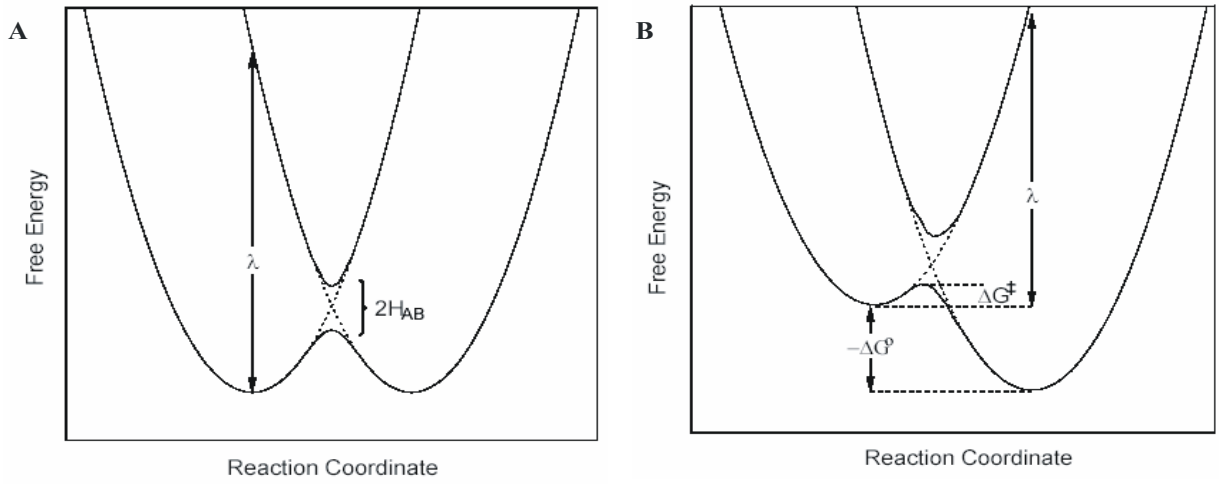

Figure 1.2.2.1 A. Plot of the free energy *versus* reaction coordinate for an electron-transfer reaction. The splitting at the intersection of the curves is defined as $2H_{AB}$. B. The activation barrier of the reaction is designated $\Delta G^{\ddagger}$ (ΔG^{*}). From: [Hopfield 1982].

The quantity (*FC*) is the Franck-Condon factor. It is a sum of products overlap integrals of the vibrational and solvational wave functions of the reactants with those of the products, suitably weighted by Boltzmann factors. The quantum-mechanical theories for the nuclear motion, which have treated reaction coordinates differently, are reviewed in [Warshel 1982; Marcus & Siders, 1982; and Siders & Marcus, 1981]. In the classical treatment, one writes:

$$k = \frac{2\pi}{h} H_{AB}^{2} \sum_{v_r, v_p} (FC)_v \, p(v_r) \qquad (1.2.2.2)$$

where $(FC)_v$ is the Franck-Condon (vibrational overlap) factor for any given set of vibrational quantum numbers v_r of the reactant and v_p of the product system, $p(v_r)$ is the equilibrium (Boltzmann) probability of finding the system in the vibrational state v_r, and the sum is over all v_r and v_p. If (*FC*) could be written as:

$$FC = \frac{1}{\sqrt{4\pi\lambda RT}} \cdot \exp\left[-\frac{\left(\Delta E^{0'} + \lambda\right)^2}{4\lambda RT} \right] \qquad (1.2.2.3)$$

then Eqn. 1.2.2.2 could be rewritten as:

$$k = \frac{2\pi}{h} H_{AB}^2 \frac{1}{(4\pi\lambda RT)^{1/2}} e^{-(\Delta E^{0'}+\lambda)^2/4\lambda RT} \qquad (1.2.2.4)$$

where $\Delta E^{0'}$ is the energy (enthalpy) of the reaction at the separation distance r in the prevailing medium. It is negative for an exothermic reaction.[1] But, when only some of the vibration frequencies are treated as relatively small und unchanged by the reaction, with a contribution to λ denoted by λ_o, one obtains [Kestner et al., 1974; Ulstrup & Jortner, 1975] from Eqn. 1.2.2.2:

$$k = \frac{2\pi}{h} H_{AB}^2 \sum_{v_r, v_p} \frac{1}{(4\pi\lambda_o RT)^{1/2}} \times e^{-(\Delta E^{0'}+\Delta\in_v+\lambda_o)^2/4\lambda_o RT} (FC)_v \, p(v_r) \qquad (1.2.2.5)$$

where $\Delta\in_v$ is the vibrational energy of the products minus that of the reactants for the set of quantum numbers (v_r, v_p).

If the nuclear motion is treated quantum-mechanically, the inner-shell coordinates behave as harmonic (or Morse) oscillators, but the motion (orientational, librational) of the solvent is treated outside the inner-coordination shell as classical, using the free energy of solvent reorganization, then one writes for k [Warshel 1982; Marcus & Siders, 1982]:

$$k = \frac{2\pi}{h} H_{AB}^2 \sum_{v_r, v_p} \frac{1}{(4\pi\lambda_o RT)^{1/2}} \times e^{-(\Delta G^{0'}+\Delta\in_v+\lambda_o)^2/4\lambda_o RT} (FC)_v \, p(v_r) \qquad (1.2.2.6)$$

which is now contains $\Delta G^{0'}$ instead of the $\Delta E^{0'}$ in the exponent. At the high-temperature limit, equation 1.2.2.6 could be reduced to:

$$k = \frac{2\pi}{h} H_{AB}^2 \frac{1}{(4\pi\lambda RT)^{1/2}} \times e^{-(\Delta G^{0'}+\lambda)^2/4\lambda RT} \qquad (1.2.2.7)$$

with λ given by Eqn. 1.2.1.3. Equation 1.2.2.7 has the same form as Eqn. 1.2.1.1 and 1.2.1.2 ($w^p = w^r = 0$ in a fixed site intramolecular reaction) when $\kappa << 1$.

Equations 1.2.2.5 and 1.2.2.6 have a major defect when there is significant entropy of reaction ΔS^0, e.g. when a polar environment is present, which finally can influence the reaction rate. These entropic effects result mainly from changes in degrees of freedom of rotations and librations of the solvent molecules. Equations 1.2.2.1-1.2.2.7 were obtained for reactions with $\kappa << 1$, as was noted in the beginning of this section. However, the main quantum correction in Eqn. 1.2.2.6 to the classical equation, Eqn. 1.2.2.7, proves to be one for nuclear tunneling, and

[1] If the chemical reaction releases or absorbs heat, it is called exothermic or endothermic, respectively.

the latter can occur regardless of whether κ is small or close to unity. Thus, one quantum correction to Eqn. 1.2.1.1, which has been used for κ≈1, is given by [Marcus & Siders, 1982; Siders & Marcus, 1981]:

$$k_{quantum} = k_{classical}\left(FC\right)_q / \left(FC\right)_c \qquad (1.2.2.8)$$

where $(FC)_q/(FC)_c$ is the ratio of the right-hand sides of Eqn. 1.2.2.6 and 1.2.2.7, where the vibrational motion is treated in a quantum manner, but the solvent motion is treated classically.

1.2.3 Nuclear and electron tunneling, electronic and steric aspects, relaxation-limited rates

In most chemical reactions, nuclear tunneling is important only for protons. However, if the potential-energy barrier for changes in, e.g. metal-ligand bond length is very narrow, then even ligands can tunnel at room temperature. However, nuclear tunneling tends to occur at low temperature, because then the probability that the reacting system has enough energy to surmount the barrier is small. At sufficiently low temperatures, in fact, all reactions occur by nuclear tunneling. The rate of the reaction is then independent of temperature, since the reaction is now occurring at the energy equal to that of the zero-point vibrational energy of the pair of reactants and surrounding system. Electron tunneling occurs when the potential energy acting on the electron in the space between the reacting pair is higher than the energy of the electron in its initial localized site. Although electron tunneling is a convenient term for describing the process of electron transfer, it is, at best, only a physically intuitive one-electron description of the process.

The nature of the intervening medium and transfers through a series of conjugated bonds (σ-bonds and π-bonds) will influence the rate of electron transfer [Larsson 1983; Beratan & Hopfield 1984; and Larsson 1981]. For example, orientation effects for π-electron systems are expected, because of both the non-spherical nature of the electronic orbitals and the non-spherical shape of the molecules themselves. Steric effects in reactions can also occur [Chang & Weismann, 1967; Tembe et al., and Newton 1982].

Throughout this section the rate expressions have been those, which are appropriate to 'activation control' in which the probability of forming reactive configurations can be calculated from an equilibrium-type distribution function. It is assumed that the probability of finding the system in the vicinity of the intersection region in Figure 1.2.1.1.A is a near-equilibrium one.

However, under certain conditions the motion, leading to the reactive configurations may partly involve very slow diffusive steps, either in the relative translation motion of the reactants or in fluctuations of the environment itself, e.g. a slow diffusive-type dielectric relaxation.

For more complex, biological systems, e.g. proteins, the calculation of an electron-transfer rate will have to account for the fact that the electron has to travel over large distances and through the protein matrices. This usually happens even without direct contact between the respective redox centers of proteins. Although this feature may at first seem to constitute a constraint, insights gained in recent years into long-rage electron-transfer processes suggest that the essential specificity and exquisite tuning of these processes are achieved to a considerable extent by using the protein matrix as the reaction medium. Thus, additional approximations should be taken into account for electron-transfer rate calculation. The next section will give an overview on electron transfer in proteins with a discussion of possible electron-transfer models.

1.2.4 Electron transfer in biological systems

Intermolecular electron transfer from one biological molecule to another is the fundamental reaction for energy conversion in the processes of respiration and photosynthesis [Stryer 1995]. Many biomolecules in these processes are proteins that contain redox sites buried within the protein matrix. The protein matrix prevents buried active sites from direct collision, and forces electron tunneling to occur over large distances. This leads to weakly coupled electron-transfer reactions, which are frequently described by the semi-classical electron transfer theory [Marcus & Sutin, 1985].

The biological electron transfers have certain features in common with the small-molecule electron transfers, as well as features which are different. As for the simple reactions, the reactants in biological systems can undergo vibrational and solvational changes, and so the reaction rate should depend on λ and ΔG^0. Despite many differences between these two systems, two important should be noted here. Whereas most electron transfers in solution are expected to occur at or near contact between the reactants, with a reaction probability, which depends on λ and ΔG^0, reactants (active or prosthetic groups) in biological systems, such as proteins, are frequently fixed in positions and, thus, prevented from coming in contact. The rate of biological reactions is expected to vary as $\exp(-\beta r)$, i.e. to decrease exponentially with separation distance r of the reactants, and the rate now depends not only on intrinsic (λ) and thermodynamic (ΔG^0)

factors, but also on β *and r*. Here, β describes a mean effect of the surrounded medium on the electron transfer. Moreover, protein-conformational changes may precede or follow the electron transfer. For example, such conformational changes may modify the relative binding free energy of the oxidized and reduced forms of a substrate to the protein.

As mentioned above, electron-transfer reactions between reactants in biological systems occur between spatially fixed and oriented sites. In this case, the reaction is formally of first order. The first-order rate constant can then be written as:

$$k = \kappa(r)\nu \exp\left(-\Delta G_r^* / RT\right) \tag{1.2.4.1}$$

$$\Delta G_r^* = \frac{\lambda}{4}\left[1 + \frac{\Delta G^{0'}}{\lambda}\right]^2 \tag{1.2.4.2}$$

where λ is given by Eqn. 1.2.1.3, ΔG_r^* and $\Delta G^{0'}$ are the free energy barrier and the free energy of reaction, respectively, when the reactants are a distance *r* apart in the prevailing medium, and *v* has dimensions of a frequency (s^{-1}). For non-adiabatic[2] reactions (κ<<1), κ*v* is actually independent of the frequency of nuclear motion and is given by:

$$\kappa(r)\nu = 2\pi \frac{H_{AB}^2}{h(4\pi\lambda RT)^{1/2}} \tag{1.2.4.3}$$

H^2_{AB} decreases exponentially with separation distance, so one should write κ(*r*)*v* as:

$$\kappa(r)\nu = 1\cdot 10^{13} \exp\left[-\beta(r - r_0)\right] s^{-1} \tag{1.2.4.4}$$

where r_0 is the value of *r* at which κ*v* equals some pre-assigned value, $1\cdot 10^{13}$ s^{-1}. The values of β may vary from system to system, depending on the vertical ionization potential of the redox site and its determination for various model systems, for reactions of ground states, electronically excited states, anions, and cations. Its dependence on the intervening material presents a central experimental problem in this field.

Progress in protein electron transfer has been expedited by achievements in areas such as metalloprotein isolation and purification, site-directed mutagenesis, high-resolution X-ray and NMR structure determination, and time-resolved kinetics. However, up to now there is no a single suitable model that can describe electron transfer between and within proteins, taking into

[2] It is expected that for large H_{AB}, or when the solvent motions are sufficiently slow, the rate will become independent on H_{AB}, and non-adiabatic reaction will become adiabatic.

account all possible factors, which can influence the rate of the process. Three models have been proposed and discussed in the literature so far.

1.2.4.1 Models for studying the electron transfer within and between proteins

Despite many studies over the last 30 years, there is no common theory about how electron transfer occurs in proteins. There are many physics- and mathematics-based models, which were developed for better understanding of the whole complexity of the electron transfer in proteins. Contemporary models for electron transfer in proteins assume that the protein matrix mediates electron tunneling. In these models, the 3D protein structure provides specific pathways for electron tunneling, and the structural details of these pathways govern the efficiency in which electron transfer is mediated by the protein matrix. However, all of these models are different in a way of interpretation of the rate constant, namely in approximations, which should be taken into account by calculation of H_{AB}^2 and definition of the distance dependence. The necessity to understand these and other factors that control the rates of electron transfer in biological systems has stimulated much work on the study of electron tunneling between redox centers within and between protein molecules.

Straight-line model. According to this model proposed by *Dutton and Page* [Page et al., 1999], the rate of electron tunneling will be calculated with the single approximation that the protein environment is homogeneous, and the straight distance between donor and acceptor is considered. The authors have investigated some oxidoreductase proteins in which electrons can travel up to 14 Å between redox centers through the protein medium. They could show for those proteins that there is a common relationship between the rate and the edge-to-edge distance. Reorganization energy and the driving force of reaction are given by the following equation:

$$\log_{10} k_{et} = 15 - 0.6R - 3.1(\Delta G + \lambda)^2 / \lambda \qquad (1.2.4.1.1)$$

This equation is useful at the first approximation when the detailed structure is not taken into account [Page et al., 1999]. However, proteins are not homogeneous structures, thus additional factors should be included in Eqn. 1.2.4.1.1 for the rate calculation. Eqn. 1.2.4.1.2 includes the packing density (ρ) of protein atoms in the volume between redox centers:

$$\log_{10} k_{et}^{ex} = 13.0 - (1.2 - 0.8\rho)(R - 3.6) - 3.1(\Delta G + \lambda)^2 / \lambda \qquad (1.2.4.1.2)$$

where ρ is within the united van der Waals radius of intervening atoms. k_{et}^{ex} is an exergonic[3] electron tunneling rate. *Dutton and Page* [Page et al., 1999] have compared k_{et}^{ex} obtained from Eqn. 1.2.4.1.2 with the experimentally calculated free-energy-optimized electron-transfer rates for the photosynthetic reaction centers of *Rhodobacter sphaeroides* and *Rhodopseudomonas viridis* (Figure 1.2.4.1.1.a).

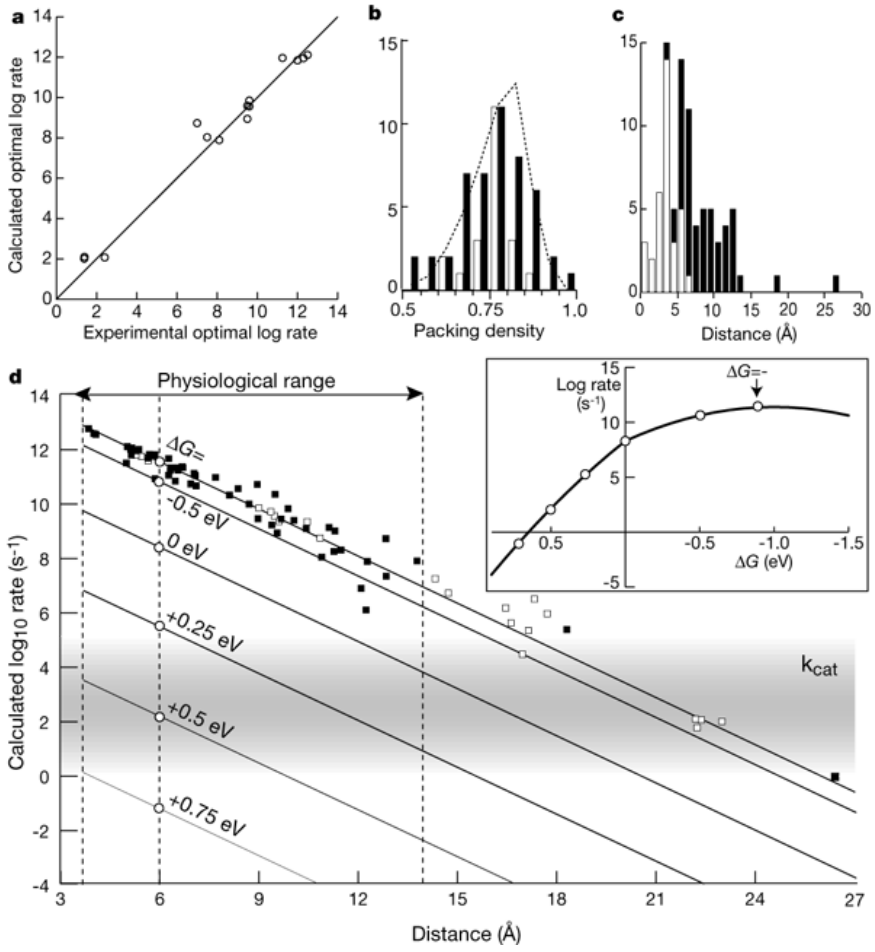

Figure 1.2.4.1.1 a. Comparison of calculated with experimental free-energy-optimized electron-transfer rates. **b.** Distributions of packing density ρ for physiologically productive (solid bars) and non-physiological (open bars) electron-transfer reactions for proteins taken by *Dutton and Page* [Page et al., 1999] from the PDB database for analysis. The dashed line shows a normalized distribution. **c.** A histogram of edge-to-edge distances of long-range physiological electron transfer. Solid bars, distances between cofactors within chains; open bars, distances between closely spaced substrates and redox centers within catalytic clusters. **d.** Free-energy-optimized rate of electron tunneling, predicted by equation 1.2.4.1.2, for natural multiredox center oxidoreductase structures in the Protein Data Bank (PDB) as of February 1999. Physiologically productive reactions (solid squares) and unproductive reactions (open squares)

[3] The products formed during the reaction have lower free energy than the reactants (energy is released during the course of the reaction) –exergonic reactions. The products have more free energy than the reactants, because energy has been incorporated during the reaction – endergonic reactions.

fall evenly on either side of an average log rate *versus* distance relationship ($\Delta G = -\lambda$). The calculated rates for non-optimal ΔG can be gauged from the parallel lines, which use an unremarkable $\lambda = 1$ eV. Inset, the expected free-energy dependence of the rate for both exergonic and endergonic tunneling reactions for a fixed distance of 6 Å and the same generic λ. Circles correspond to free energies chosen for the diagonal lines. The grey zone gives the typical range of turnover rates (k_{cat}) of electron-transfer protein. The figure is from: [Page et al., 1999].

Page and colleagues [Page et al., 1999] found a good correlation between the theoretically calculated and experimentally observed electron-transfer rates. In the analysis of the effect of protein structural heterogeneity and packing density (Figure 1.2.4.1.1.b), it was found that protein regions with well-bonded pathways tend to be well packed. There did not appear to be any general correlation between the quality of the path and the physiological benefit of the reaction. The well-bonded pathways radiated from a given redox center in many different directions with an average distance of ~12-14 Å between redox centers within the electron-transfer chain and a distance of ~2-3 Å between a substrate and redox center within the catalytic cluster (Figure 1.2.4.1.1.c). Both classical and quantum versions of the Marcus electron-transfer theory (see above sections 1.2.1 and 1.2.2) suggest a roughly parabolic dependence of log rate on ΔG. The distances of 14 Å being in the physiological range foster robust electron-transfer design. However, the enormous latitude in values of ρ, λ, and ΔG, which sustain tunneling rates faster than typical catalytic rates (Figure 1.2.4.1.1.d), shrinks sharply at distances greater than 14 Å where designs become less robust. From the Figure 1.2.4.1.1.d it can be concluded that proximity of redox centers in chains provides rugged, highly directional electron transfer, which slows roughly linearly with distance along the chain, in contrast to the sharply exponential slowing in all other directions.

PATHWAY model. Beratan and *Onuchic* [Onuchic et al., 1992] have discussed in their work the electron-transfer rates in proteins within the non-adiabatic limit and therefore proportional to H_{AB}^2. Their work compares the theoretically calculated H_{AB}^2 with experimentally obtained values [DeVault 1984; Hopfield 1974; Jortner 1980; and Newton & Sutin, 1984]. Based on these and earlier studies [Beratan & Onuchic, 1991; Onuchic & Beratan, 1987], the pathway model of electronic coupling in proteins [Beratan & Onuchic, 1991; Beratan et al., 1991; and Beratan et al., 1990] was developed. Generally, the pathway method is based on a search for the combination of bonded and non-bonded interactions that maximizes the total donor-to-acceptor (*D*-*A*) interaction mediated by a combination of through-bond and through-space coupling

through the protein. Here, *D* and *A* are donor and acceptor, the definitions used instead of reactant(s) and product(s).

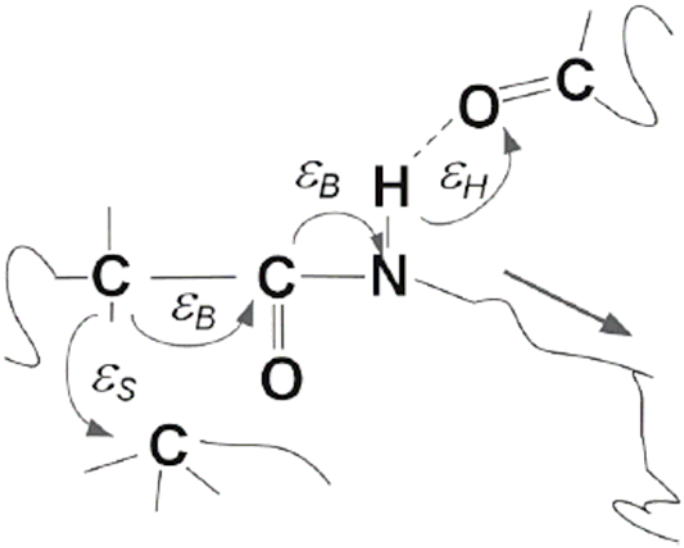

Figure 1.2.4.1.2 Coupling decay factors describing the computer-search strategy PATHWAY. Figure is borrowed from Dr. C. Jung's lecture “Theoretische Biochemie”.

$$H_{AB} = H_C \cdot H_H \cdot H_S = c \cdot \varepsilon_C \cdot \varepsilon_H \cdot \varepsilon_S \qquad (1.2.4.1.3)$$

where *c* is a constant, and ε_C, ε_H, and ε_S are covalent, hydrogen-bond and through-space coupling decay factors, respectively (Figure 1.2.4.1.2). ε_C was calculated to be in a range of ~0.7-0.4 [Mikkelsen & Ratner, 1988]. *Beratan* and *Onuchic* [Onuchic et al., 1992] choose 0.6 as an average value for the decay per bond. The hydrogen-bond decay is treated as two ε_C from hetero atom to hetero atom. The through-space interactions are treated as stretched bonds and an additional factor, usually 1/2, is accounting for generally unfavorable orientation effects associated with these interactions. Thus, one can write for ε_C, ε_H, and ε_S the following equations [Onuchic et al., 1992]:

$$\varepsilon_C = 0.6 \qquad (1.2.4.1.4)$$

$$\varepsilon_H = \varepsilon_C^2 \exp\left[-1.7(R-2.8)\right] \qquad (1.2.4.1.5)$$

$$\varepsilon_S = \left(1/2\right)\varepsilon_C \exp\left[-1.7(R-1.4)\right] \qquad (1.2.4.1.6)$$

where *R* is distance in Å and coupling decay factors are unitless. The reference covalent bond distance is 1.4 Å, and for two bonds it is 2.8 Å. Each decay factor is associated with an effective distance d_{eff} given by:

$$d_{eff}(i) = -\log \varepsilon(i) \qquad (1.2.4.1.7)$$

The strength of the coupling arising from single pathway is proportional to the product of decay factors for each step on the path. Thus, to compute the path one has to analyze the interconnected network of bonded and non-bonded contacts. However, because Eqn. 1.2.4.1.7 associates all decay factors with the effective distance, one can restate the search for the maximum pathway couplings as a search for the shortest effective distance between donor and acceptor. Graph theory [Buckley & Harary, 1990] is used to solve this minimum-distance problem, originating in the PATHWAY model. The lengths of edges, i.e. decays, that will joint the atoms, are determined by the distances between them and the nature of their interaction. Such atom information can be obtained from the PDB files, which are used as inputs to the PATHWAY software.

Thus, the obtained tunneling pathways contain mostly bonded interactions with occasional through-space connections. However, tunneling is much more efficient through bonded orbitals than through space; because the potential barrier is effectively lower [Onuchic et al., 1992]. *Beratan* and *Onuchic* [Onuchic et al., 1992] have also described the pathway search strategies, which can be shortly described here as follows: (i) depth-first and (ii) breadth-first. The difference between them is in the way of path searching. The breadth-first approach uses a simultaneous search in all direction form the origin site. The depth-first strategy, which is actually is used in the PATHWAY program, steps always in “one direction” along the allowed connections from origin to the target site.

1.3 Photo-induced electron transfer in proteins

The protein environment is composed of various arrangements of secondary structures such as α helixes, β strands, and random coils, as well as hydrogen-bond and van der Waals contacts, surrounded by lipids or water molecules. Therefore, determining which part of the protein environment is most significantly used in the electron tunneling between a redox pair (donor and acceptor) is a challenge. Recently developed photo-chemical techniques of studying the kinetics of intramolecular electron transfer in ruthenium(II)bipyridyl-modified proteins helped to answer the question of how exactly electron tunneling in proteins occurs. Since these ruthenium complexes can be covalently bound to the protein surface, the photo-induced electron transfer can be studied. That, together with a possibility of crystal structure determination of such modified protein complexes allows an electron transfer prediction.

Photo-induced electron transfer results from an electronic state, which is produced by the resonant interaction of electromagnetic radiation with matter. The first act of a photo-chemical or photo-physical process is the absorption of a photon by a molecule A, transforming it to an electronically excited state, A^*. The latter is an unstable species with high energy, which must undergo any of the following types of deactivation: (i) emission of light (luminescence) or (ii) different non-radiative transitions, where e.g. the excess energy is transferred to the environment as heat, quenching reaction in the presence of a quencher, or through a photo-chemical reaction generating another chemical species. These are shown in Figure 1.3.1 with the different energy levels involved and the different transitions, which are indicated by arrows.

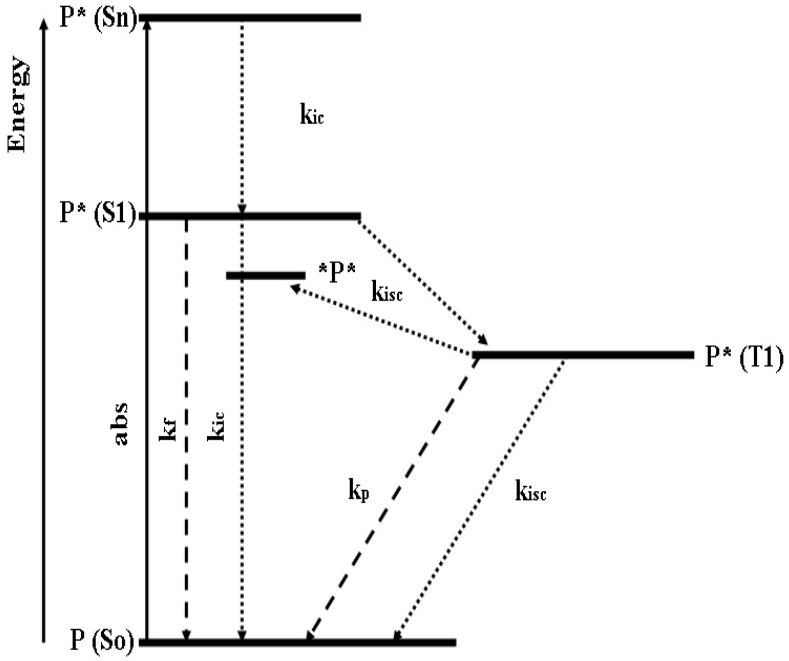

Figure 1.3.1 A schematic Jablonski diagram. Different types of transitions can occur with energy levels, where So is the singlet ground state, Sn and S1 are singlet excited states and T1 is the lowest triplet excited state. Radiative transitions are indicated by dashed arrows and non-radiative transitions by dotted ones. k_f, k_p, k_{ic}, and k_{isc} are fluorescence, phosphorescence, internal conversion, and intersystem crossing.

The following reactions additionally illustrate the nature of the possible transitions:

$$A + hv \xrightarrow{k_{ex}} A^* \qquad \text{(excitation process)}$$

$$A^* \xrightarrow{k_r} A + hv \qquad \text{(radiative process)}$$

$$A^* \xrightarrow{k_{nr}} A + heat \qquad \text{(non-radiative process)}$$

$$A^* + Q \xrightarrow{k_q} A + Q^* \qquad \text{(quenching)}$$

Each decay step is characterized by its own rate constant and each excited state is characterized by its life time τ, given by Eqn. 1.3.1, where $\sum_i k_i$ the summation of the first order rate constants is for an unimolecular process that causes the disappearance of the excited state.

$$\tau = \frac{1}{\sum_i k_i} \qquad (1.3.1)$$

When the lifetime of the excited state is sufficiently long, the excited molecule may have time to approach a molecule of another solute, i.e. the two reactants, here donor and acceptor, have to diffuse together to form an outer-sphere precursor complex *D*--*A*, where k_a is usually diffusion controlled. The precursor complex undergoes reorganization towards the successor complex (D^+--A^-); a transition state in which the electron transfer takes place. The latter dissociates to form the product ions, the D^+ and A^- [Marcus 1956; Marcus & Sutin, 1985]:

$$D + A \xleftrightarrow{k_a, k_{-a}} D--A \xleftrightarrow{k_{et}, k_{-et}} D^+ --A^- \rightarrow D^+ + A^- \qquad (1.3.2)$$

where k_a, k_{-a}, k_{et}, and k_{-et} are rate constants for forward and backward reactions. It was described in the previous section that the rate of an electron-transfer reaction decreases with increasing the distance between the reacting molecules. However, when *D* and *A* are covalently linked to each other (*DA*), an intramolecular[4] electron transfer can occur, using the orbitals of the coupling bridge that is unlimited by diffusion and therefore can be very rapid:

$$DA \xleftrightarrow{k_{et}, k_{-et}} D^+ A^- \qquad (1.3.3)$$

A semi-classical model (see section 1.2) for electron transfer will describe the first-order rate constant in Eqn. 1.3.3 from a *D* to *A* held at fixed distance.

A bimolecular process is possible for those excited states whose lifetimes are longer than ~10^{-9} s. For transition metal complexes, only the lowest spin-forbidden excited states fulfill this requirement, i.e. phosphorescence and intersystem crossing. Thus, the transition metal complexes should be stable, able to absorb visible light and have an excited state lifetime that is long enough to form the charge-separated state before it decays to the ground state.

1.3.1 Ruthenium(II) bipyridyl complex

Photo-chemical and photo-physical properties. The ruthenium(II) bipyridyl complex such as (2,2′-bipyridine)$_3$ ruthenium(II) (Ru(bpy)$_3$) is a d_6 transition metal complex with octahedral geometry (Figure 1.3.1.1).

[4] A bimolecular electron-transfer process involves inter- and intra-components. For the ruthenated proteins, intermolecular electron transfer takes place between neighboring *DA* complexes, and intramolecular electron transfer takes place within one DA complex, i.e. from a ruthenium complex to a prosthetic group of a protein and *vice versa*.

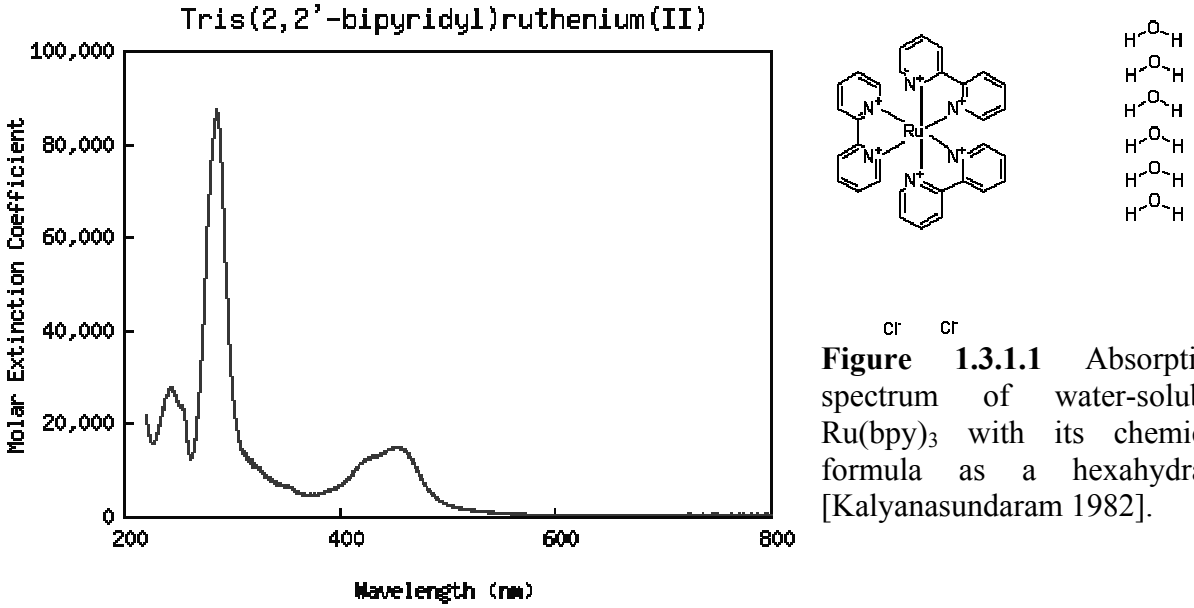

Figure 1.3.1.1 Absorption spectrum of water-soluble $Ru(bpy)_3$ with its chemical formula as a hexahydrate [Kalyanasundaram 1982].

Two very intense bands at 240 nm and 450 nm are caused by transition of an electron from a π_M metal orbital to the π_L^* ligand orbitals, and are therefore named metal-to-ligand charge transfer (MLCT) bands. Promotion of an electron from π_L to π_L^* results in the bands at 185 nm and 285 nm, which are called ligand-centered (LC) bands. The weak shoulders at 322 nm and 344 nm are due to the metal-centered (MC) transitions, i.e. promotion of an electron from π_M to σ_M^* [Kalyanasundaram 1982; Juris et al., 1988; Kalyanasundaram 1992; and Crosby 1975].

For most ruthenium(II)-polypyridine complexes, the lowest excited state, responsible for luminescence and bimolecular excited state reactions, is a ^{3}MLCT[5] state (in Figure 1.3.1 it is P* (T1)). Thus, excitation with visible light creates the lowest singlet excited state ^{1}MLCT (in Figure 1.3.1 it is P* (S1)) [Damrauer et al., 1997], which within a few hundred fs is converted into the lowest triplet state, ^{3}MLCT, *via* intersystem crossing (k_{isc}) [Juris et al., 1988; Yeh et al., 2000; and Demas & Taylor, 1979]. The quantum yield of the formation of the lowest excited states is unity, showing that intersystem crossing from the upper singlet excited states obtained by excitation to the lowest triplet is both fast and very efficient [Demas & Taylor, 1979; Demas & Crosby, 1971]. The ^{3}MLCT excited state decays to the ground state *via* three major pathways. Two of the pathways involve k_p and a crossover into the nearby ^{3}MC ($^*P^*$) excited state takes place (Figure 1.3.1), followed by the radiationless decay to the ground state. The lifetime of the lowest ^{3}MLCT excited state of $Ru(bpy)_3$ is in the order of 850 ns [Marcus & Sutin, 1985; Young et al., 1976]. With decreasing temperature, the luminescence intensity and lifetime will increase,

[5] Experimentally it has been shown that this state consists of three closely spaced energy levels [Hager & Crosby, 1975; Kober & Meyer, 1984].

and at 77 K the lifetime is ~5 μs and the luminescence is highly structured with a prominent vibrational progression [Crosby 1975; Demas & Addington, 1976].

Redox properties. One oxidation and three reduction processes, all one-electronic and reversible, can be observed [Tokel-Takvoryan et al., 1973]. The oxidation of $Ru(bpy)_3^{2+}$ occurs at fairly positive potential (+1.3 eV) and involves the removal of one electron from a metal-centered orbital. This results in the formation of the ruthenium(III) complexes according to Eqn. 1.3.1.1:

$$[Ru^{2+}(LL)_3]^{2+} \leftrightarrow [Ru^{3+}(LL)_3]^{3+} + e^- \qquad (1.3.1.1)$$

where *LL* is a bipyridine ligand. Three reductions occur at fairly negative potentials and are all ligand-centered. The added electron appears to be localized on a single ligand:

$$[Ru^{2+}(LL)_3]^{2+} + e^- \leftrightarrow [Ru^{2+}(LL)_2(LL^-)]^+ \qquad (1.3.1.2)$$

$$[Ru^{2+}(LL)_2(LL^-)]^+ + e^- \leftrightarrow [Ru^{2+}(LL)(LL^-)_2]^0 \qquad (1.3.1.3)$$

$$[Ru^{2+}(LL)(LL^-)_2]^0 + e^- \leftrightarrow [Ru^{2+}(LL^-)_3]^- \qquad (1.3.1.4)$$

Since the amount of electric charge localized on the metal (and thus, the tendency to lose an electron) is governed by the σ and π properties of the ligands, the nature of the ligands will affect the Ru(III/II) potential [Masui & Lever, 1993; Dodsworth et al., 1994]. Moreover, substitution of one or more polypyridine ligands can drastically change the redox potentials of a ruthenium(II) complex.

In the excited state of $Ru(bpy)_3^{2+}$, which is a ^{3}MLCT state, the ruthenium is oxidized and one of the ligands is reduced according to Eqn. 1.3.1.5.

$$[Ru^{II}(bpy)_3]^{2+} \xrightarrow{h\nu} [Ru^{III}(bpy)_2(bpy^-)]^{2+} \qquad (1.3.1.5)$$

The redox potential for reduction and oxidation of the excited state of $Ru(bpy)_3^{2+}$, $^*Ru(bpy)_3^{2+}$, are +0.84 and -0.86 eV (in water), respectively [Juris et al., 1988]. Thus, $^*Ru(bpy)_3^{2+}$ possesses suitable properties to work as a good energy donor, electron acceptor, and electron donor at the same time. By changing the ground-state redox potentials and/or the excited-state energy, the excited-state potentials can be tuned.

1.3.2 Proteins with linked ruthenium(II) bipyridyl complexes

Despite the importance of biological electron-transfer reactions, relatively few techniques are available to measure the rate of the actual electron-transfer step within a complex between two proteins. This is because the reactions are usually too rapid for stopped-flow techniques, and protein binding and dissociation steps are frequently rate-limiting [Pan et al., 1990]. The purpose of ruthenium(II) bipyridyl complexes is to mimic the electron transfer that can occur in many biological systems. This new strategy for the design of proteins and enzymes carrying non-natural functional ruthenium(II) groups includes their covalent attachment at specific amino-acid residues on the protein surface. Labeling of protein matrices with photosensitizer-electron donor or acceptor units has important advantage as compared to the simple molecular photosensitizer-electron assemblies or just simple proteins in biochemistry. This approach allows the placement of artificial groups in close vicinity to the natural protein cofactors providing new protein properties and possibilities to use them for the individual research goals. The potential modification sites are cysteine [Geren et al., 1991], lysine [Anderson et al., 1964], and histidine residues [Karpishin et al., 1994]. The strong reducing properties of the artificial ruthenium bipyridyl complexes in the excited state [Sutin & Creutz, 1978] can demonstrate that photoreduction of proteins *via* these complexes could be a good approach to investigate the mechanism of electron transfer in the assemblies of protein oxidation-reduction chains.

Investigations on the long-range electron transfer using proteins covalently modified with different ruthenium bipyridine complexes enhanced our understanding of factors involved in the biological electron transfer. The work of *Winkler et al.* [Winkler et al., 1982] and *Isied et al.* [Isied et al., 1982] with $Ru^{II}(NH_3)_5$-labeled cytochrome *c* provides information of the intramolecular rate constants. Additionally, such investigations were carried out on ruthenium-labeled azurin and myoglobin derivatives [Kostič et al., 1983; Axup et al., 1988; and Crutchley et al., 1986], mixed-metal hybrid hemoglobins [McGourty et al., 1987], and organic complexes [Closs et al., 1986] with the goal to determine the dependence of electron transfer on the driving force of the reaction, the distance between donor and acceptor, and the protein medium through which electron transfer occurs. The donor-acceptor distances of ruthenium-modified proteins can be estimated from crystal structure analysis. In the best of cases, the distance is reasonably well fixed during the course of electron transfer, but in other instances there is some uncertainty about the distance at the time the electron travels from the donor to the acceptor. As an example, the

work of *Isied et al.* [Isied & Vassilian, 1984] demonstrated that a polypeptide bridge is capable of mediating the passage of an electron between two redox centers at relatively short range, but the types of molecules, which were used were not rigid enough to allow the determination of the distance dependence of the transfer rate at long range. In further work [Winkler et al., 1982; Isied et al., 1982], when the long-range electron-transfer reaction had a driving force of 0.11 eV and the separation distance was about 12 Å between the ruthenium and the heme group of cytochrome *c*, the intramolecular rate constant was determined to be about 30 s^{-1}. Paradoxically, *Bechtold et al.* [Bechtold et al., 1986] found that the rate for the coupling of the ferrous heme group of cytochrome *c* and $Ru^{III}(NH_3)_4$(isonicotinamide)(histidine 33) was over 10^5 times slower than the above rate, even though the driving force was larger (0.18 eV).

In order to study the mechanism of electron transfer in the steroid hydroxylase system, in this PhD thesis bovine adrenodoxin was covalently modified by a ruthenium(II) bipyridyl complex. The modification site was proposed and then found from the crystal structure of the $Ru(bpy)_2$(mbpy)-adrenodoxin complex. The crystals structure was also important for the prediction of possible electron-transfer pathways within the $Ru(bpy)_2$(mbpy)-adrenodoxin complex.

1.4 Protein structure determination by X-ray crystallography

X-ray crystallography is an experimental technique that exploits the fact that X-rays are diffracted by crystals. It is not an imaging technique. X-rays have the proper wavelength (in the Ångstrom range, 10^{-8} cm) to be scattered by the electron cloud of an atom of comparable size. Based on the diffraction pattern obtained from X-ray scattering off the periodic assembly of molecules or atoms in the crystal, the electron density can be reconstructed. Additional phase information must be extracted either from the diffraction data or from supplementing diffraction experiments to complete the reconstruction (the phase problem in crystallography, see below). A model is then progressively built into the experimental electron density, refined against the data, and the result is mostly an accurate atomic structure.

Crystallography can reliably provide the answer to many structure related questions, from global folds to atomic details of bonding. In contrast to NMR, which is an indirect spectroscopic method, no size limitation exists for the molecule or complex to be studied. The price for the high accuracy of crystallographic structures is that a good crystal must be used, and that only limited information about a molecule's dynamic behavior is available from one single diffraction experiment. To carry out this experiment, one needs to grow a single crystal from a solution containing the macromolecule whose structure is going to be determined.

1.4.1 3D crystals of biomolecules

There are a number of potential bottlenecks in determining a crystal structure, but growing a useful crystal can be the most serious one. *McPherson’s* [McPherson 1990] review “Current approaches to macromolecular crystallization” describes most of the basics of the crystallization of proteins, nucleic acids, and their complexes with one another and with small molecules. Here, there are some of extracts from that review concerning protein crystallization only, because in this thesis a protein, bovine Adx, was crystallized.

A crystal is homogenous solid, exhibiting a high degree of internal order and a definite, although not necessarily stoichiometric overall chemical composition. The protein molecules have irregular shapes and are not ideally suited to be stacked into a periodic lattice, i.e., a crystal. Protein crystals are, thus, very fragile, soft and sensitive to all kind of environmental variations.

Protein crystals contain on average 50% solvent, mostly in large channels between the stacked molecules on the crystal. The interactions, which hold the molecules together, are usually weak hydrogen bonds, salt bridges, and hydrophobic interactions, compared to strong covalent or ionic interactions in mineral crystals. This explains the fragility of the crystals, but allows the possibility of soaking metal solutions (important for phasing) or even large enzyme substrates or inhibitors, into the crystals.

In order to obtain a crystal, the protein molecules must assemble into a periodic lattice, as mentioned above. One starts with a solution of a protein of the highest purity with a fairly high concentration (2-50 mg/ml) and adds reagents (to a total average volume of 3 μl) that reduce the solubility close to spontaneous precipitation. Precipitants like polyethylene glycol (PEG) or organic solvents are often used. pH, temperature, and salt concentration are other crucial parameters to take into account. By slow further concentration in a droplet, and under conditions suitable for the formation of a few nucleation sites (endothermic process), small crystals may start to grow (exothermic process) (Figure 1.4.1.1). Often very many conditions (initial screening) have to be tried to succeed.

The most common setup to grow protein crystals is by the vapor-diffusion technique (hanging- or sitting-drop) (Figure 1.4.1.2). A few microliters of protein solution are mixed with an about equal amount of reservoir solution containing the precipitants. As the precipitant in the drop is less concentrated than in the reservoir solution, water evaporates from the drop into the reservoir. As a result, the concentration of protein and precipitant in the drop slowly increases, and crystals may form. There is a variety of other techniques available such as dialysis buttons, gel, macro- and microbatch techniques. Nowadays, robots are also very useful, but much more expensive than techniques for manual screening and optimization of crystallization conditions.

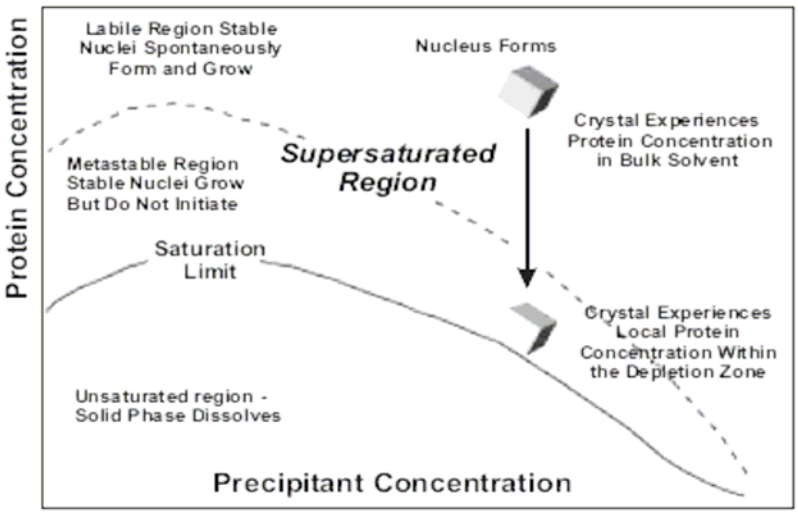

Figure 1.4.1.1 The phase diagram describing the process of the protein crystallization. The figure is borrowed from: [McPherson 1999].

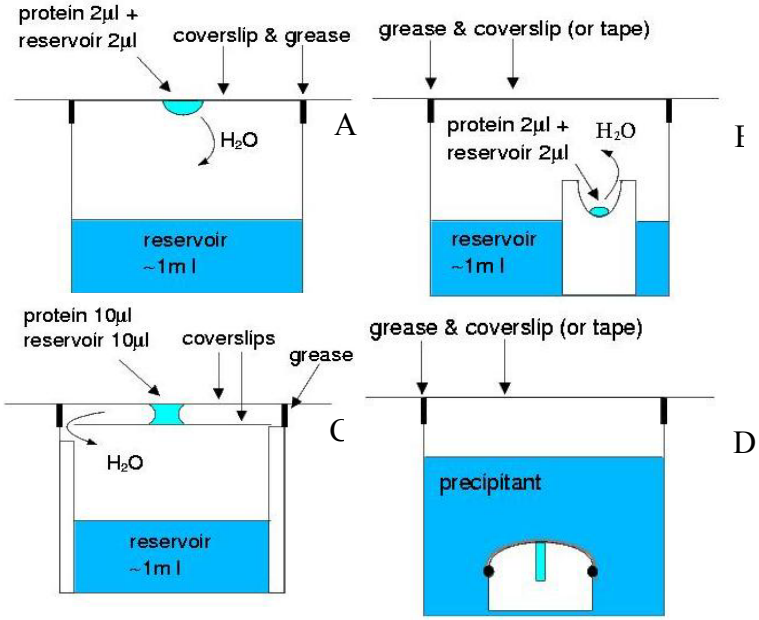

Figure 1.4.1.2 Crystallization techniques. A. The hanging-drop vapor-diffusion. B. The sitting-drop vapor-diffusion. C. Free interface diffusion D. Microdialysis.

1.4.2 X-ray diffraction and data evaluation

Once a single crystal has grown to a suitable size, then it may be used for data collection, which is necessary for further 3D structure determination of a protein molecule. The data collection strategy at synchrotrons or other X-ray sources aims at collecting a complete data set, i.e. one covering all theoretically accessible data points. Electromagnetic waves, such as X-rays, are very suitable for that aim. The X-rays emitted from copper targets bombarded with high energy electrons can emit at several characteristic wavelengths, e.g. CuKα, which has a wavelength of 1.5418 Å. This is very similar to the distance between bonded carbon atoms, so it

well suited to the study of a molecular structure. Since X-rays interact with matter through their interaction with electrons, the final result of a crystallographic experiment is not really an image of the atoms, but a map of the distribution of electrons in a molecule, i.e. an electron density map.

In X-ray crystallography, it is useful to use vectors that will describe properties of the waves. Thus, the length of the vector represents the amplitude of the wave, and the angle it makes with the horizontal axis represents its phase. From the mathematical point of view it is also an advantage to consider a vector in the complex plane, with real and imaginary components, instead of *x* and *y* components as in a real 2D vector. So, the addition of waves can be then represented as the addition of complex numbers.

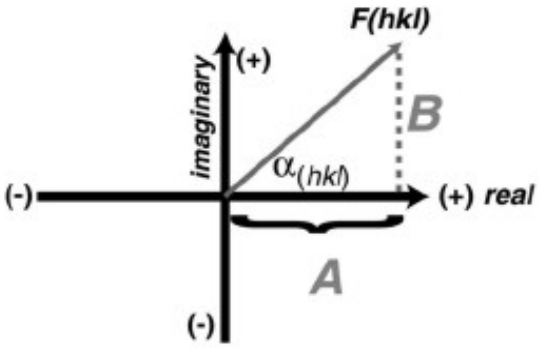

Figure 1.4.2.1 The complex numbers allow description of both amplitude and phase of an *hkl* (see below about Miller indices). The figure is borrowed from: http://dbb.urmc.rochester.edu/labs/wedekind/Wedekind-Lab/Course_Info/BCH412/BCH412_Lect01.pdf

A crystal arranges a huge numbers of molecules in the same orientation, so that scattered waves can add up in phase and raise the signal to a measurable level. A full data set is measured by collecting the diffracted intensities within a special angular range [Arndt et al., 1977]. The relationship between scattering angle and the interplanar spacing is given by Bragg's law (Figure 1.4.2.2). In 1913 the English physicists Sir W.H. Bragg and his son, Sir W.L. Bragg, explained X-ray diffraction as the reflection of X-rays by crystal lattice planes.

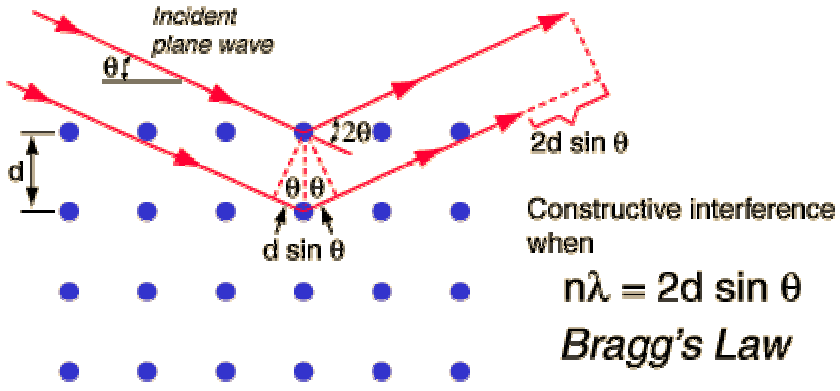

Figure 1.4.2.2 The Bragg's law. The figure is from: [Thornton & Rex, 1993].

Bragg's law is described by a very simple equation (1.4.2.1)[6], where θ (theta) is the scattering angle; d is a distance between lattice planes in a crystal; λ is the wavelength of the incident X-ray beam (Figure 1.4.2.2):

$$\lambda = 2d\sin(\theta) \qquad (1.4.2.1)$$

In Bragg's law, as the angle increases, d must become smaller for the pathlength to remain equal to one wavelength. Rearrangements of the previous equation to:

$$\sin(\theta)/\lambda = 1/(2d) \qquad (1.4.2.2)$$

$$d = \lambda/(2\sin(\theta)) \qquad (1.4.2.3)$$

can help to understand the concept of reciprocal space: the bigger the angle of diffraction[7] the smaller the spacing to which the diffraction pattern is sensitive.

The observed intensity of the diffraction spots can be thought of as corresponding to the size of the reciprocal lattice point. A most useful means to understand the occurrence of diffraction spots is the Ewald construction (Figures 1.4.2.3.A, 1.4.2.3.B, and 1.4.2.3.C). The planes of points in the reciprocal lattice intersect the Ewald sphere to give a circle of points when the diffraction condition is fulfilled. When the planes are aligned perpendicular to the X-ray beam, these circles on the Ewald sphere will project onto circles of spots surrounding the direct beam position, but as the crystal will be rotated (and the reciprocal lattice) the circles on the Ewald sphere will be distorted and will project into what are called lunes of spots (Figure 1.4.2.3.D).

[6] Bragg's law is usually expressed as $n\lambda = 2d\sin(\theta)$. However, from the point of view of the information in the diffraction pattern, it makes sense to choose d so that n is equal to one.

[7] There are two limits of the scattering angle: $\theta = 0$ and 90 degrees.

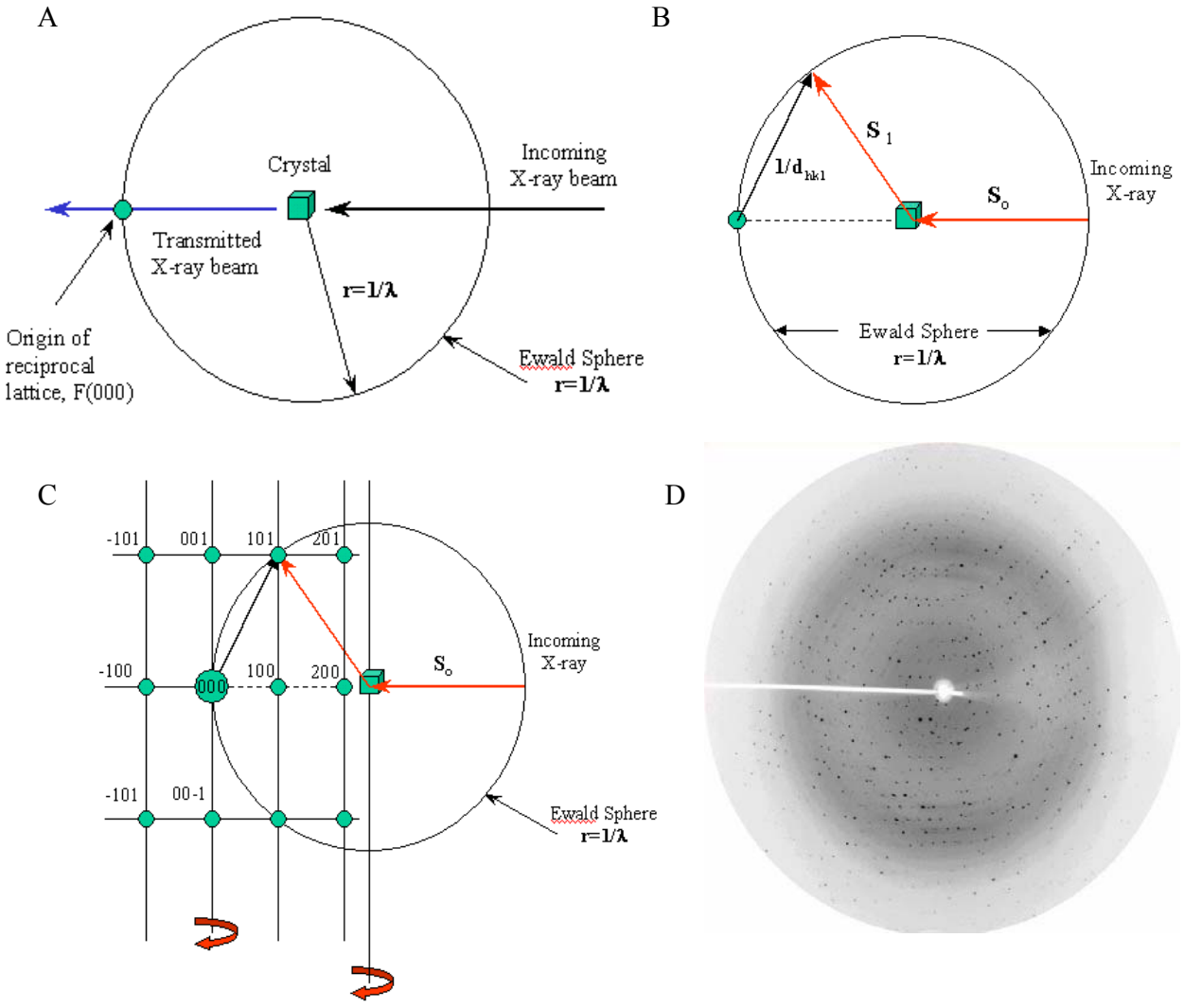

Figure 1.4.2.3 The Ewald construction. A. The Ewald sphere has a radius of 1/λ. The origin of the reciprocal lattice lies in the transmitted beam, at the edge of the Ewald sphere. B. The diffraction maxima (reflections, diffraction spots) occur when Bragg’s equation in vector form (see below Figure 1.4.2.4) is satisfied. This condition occurs whenever a reciprocal lattice point lies exactly on the Ewald sphere. As assumed already, the chance for this to occur is modest, and one needs to rotate the crystal in order to move more reciprocal lattice points (*hkl* or Miller indices) through the Ewald sphere. The reciprocal lattice is rotated along the vertical axis of the drawing. It is accomplished by rotating the crystal along the same axis. C. Turning the reciprocal lattice through the Ewald sphere in the beginning, only planes (101) and (10-1) give rise to a reflection. After the reciprocal lattice will be turned a bit (which actually means turning the crystal around the same axis), the reciprocal lattice point 201 will enter the sphere and create a diffraction spot. D. Typical diffraction pattern. From: http://www-structure.llnl.gov/Xray/tutorial/ewald.htm

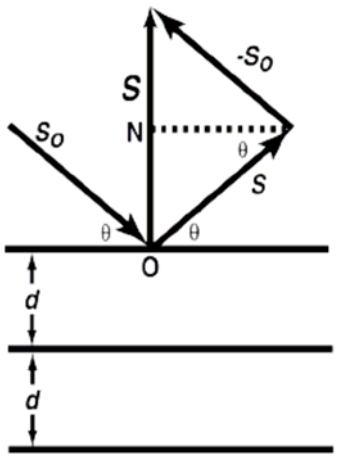

Figure 1.4.2.4 The Bragg's construct for the diffraction vector (*S*), which is perpendicular to the Bragg's planes. $S = s - s_o$ and since the magnitude of s is 1/λ and according to the Bragg's law (Eqn. 1.4.2.1) one writes for the diffraction vector $|S| = 1/d$. In following equations the diffraction vector is denoted as *r** (reciprocal space). The figure is from: http://dbb.urmc.rochester.edu/labs/wedekind/Wedekind-Lab/Course_Info/BCH412/BCH412_Lect01.pdf

Diffraction occurs one photon at a time. The probability that the photon will be scattered in any particular direction is given by the square of the amplitude of the sum of the scattered waves. Therefore, when the intensity of a diffraction spot (which is proportional to the number of photons in that spot) is measured, one takes the square root as part of determining the amplitude for the electron density calculations. The measured intensities are related to the electron density of the macromolecule in the crystal by a Fourier transform as follows:

$$\rho(r) = \int_{S^*} F(r^*)\exp(-2\pi i r^* r)dr^* \qquad (1.4.2.4)$$

$$F(r^*) = |F(r^*)|\exp(i\alpha(r^*)) \qquad (1.4.2.5)$$

$$|F(r^*)| = c\sqrt{I(r^*)} \qquad (1.4.2.6)$$

where $|F|$ is modulus of structure factor; F is the structure factor, α is the phase angle of the structure factor; r^* is a vector in reciprocal space (*hkl*); r is the corresponding vector in real space (*xyz*) and S^* is the scattering vector in reciprocal space. Despite the property that the Fourier transform can be inverted, there is a phase problem in crystallography. To compute the inverse Fourier transform, one needs not only the amplitudes, but also the phase of the diffracted waves. In the diffraction experiment only the intensity is measured, the phase is lost. Unfortunately, there is no practical way to measure the relative phase angles for the different diffracted spots. The phase is very important for structure determination, and solutions for retrieving it will be described in the next section.

There are many programs that allow the semi-automatic evaluation of the diffraction patterns produced by protein crystals: for example, MARXDS [Kabsch 1988] and DENZO/SCALEPACK [Otwinowski & Minor, 1997]. A list of probable lattices, cell parameters and space groups can be obtained by choosing a set of reflections produced by the crystal. The integration of the intensities can be automatically attained by computers after the calculation of the crystal orientation. These intensities have to be corrected for effects like absorption, polarization, and crystal radiation damage. After merging and rescaling the measured frames, a final list of diffraction intensities or amplitudes ($|F_o|^2 = I_o$) is obtained. The completeness of a data set is usually reported in % of the observed data (not using any intensity-based cutoff) compared to possible data in the asymmetric unit of the reciprocal space.

1.4.3 The phase problem

From the diffraction pattern the values of I_{hkl} can be obtained. Since $I_{hkl} \propto |F_{hkl}|^2$, the amplitudes $|F_{hkl}|$ can be obtained. The phase can be not measured directly, so that one needs to deduce it indirectly. There are several ways to solve the phase problem: (i) using heavy atom isomorphous replacement (SIR, MIR); (ii) exploiting the anomalous effect with and without isomorphous replacement (SIRAS, MIRAS, MAD, SAD), (iii) using an already available 3D structure of a similar protein (Molecular Replacement, MR), or (iv) by statistical, "direct" methods [Dodson 2003; Gonzáles 2003; and Murray & Garman, 2003]. Direct methods are very valuable, but they require a complete diffraction data set measured to atomic resolution (< 1 Å). Heavy atom-based phasing methods depend on the successful soaking of the native crystal in solutions of a heavy atom salt, co-crystallization, substituting the protein's methionines by selenomethionines or exposing the crystal to a noble gas atmosphere [Kestenbaum 1998; Schiltz et al., 1994; and Soltis et al., 1997]. There are automated computer programs including SOLVE [Terwilliger & Berendzen, 1999] and RESOLVE [Terwilliger 2000; Terwilliger 2002] that yield crystallographic solutions for MIR, SAD, and MAD with the use of statistical density modification, local pattern matching, automated model building, automated ligand fitting, and prime-and-switch minimum bias phasing.

1.4.3.1 Molecular replacement

Crystallographic phasing can be much easier if the structure is going to be solved by MR. This method can be applied if a model of sufficient structural similarity with the target structure (usually characterized by a root-mean-square deviation (r.m.s.d.) of < 2.0 Å) is available. If the model is a homologous protein, it must usually have at least 40% sequence identity with the target structure. If a search model is very poor, MR will yield poor phases, and refinement of the structure may be a challenge.

To build an atomic model of the new crystal form, it is necessary to work out how the model should be oriented and positioned in the new cell. Thus, in the MR method one tries to calculate a rotation matrix (C) and a translation vector (t) that have to be applied to the model, so that its most similar regions would overlap in real space relating the 3D structures of the known search model (X') and the unknown structure (X) by:

$$X = [C]X' + t \qquad (1.4.3.1.1)$$

Traditional MR methods are based on the properties of the Patterson function, which can be computed without phase information. A Patterson can also be computed from a trial atomic model and compared to the observed Patterson[8]:

$$P_{model}(u) = \int_S \rho(r) * \rho(r+u)dr \qquad \text{where: } u = (u,v,w), \quad r = (x,y,z) \qquad (1.4.3.1.2)$$

$$P_{cryst}(u) = \frac{1}{V}\sum_S |F(s)|^2 \cos[2\pi u * s] \qquad \text{where: } u = (u,v,w), \quad s = (h,k,l) \qquad (1.4.3.1.3)$$

where (u, v, w) are fractional coordinates of a Patterson cell which may be expressed by the vector u. The concept of the integral in Eqn. 1.4.3.1.2 only has significant values when u is a vector relating one atom to another.

In this simplistic approach which is given by Eqn. 1.4.3.1.2 and Eqn. 1.4.3.1.3, the MR problem has six dimensions (three parameters to specify orientation and three to specify position), which makes for a very large problem. Fortunately, the Patterson can be divided into parts that are sensitive to only some of these parameters. The usual strategy is to look at these parts separately, reducing the dimensionality of the problem. Remembering that the Patterson map is a vector map, one may find maps that are dominated by intramolecular vectors (from one atom in the molecule to another atom in the same molecule), which depend only on the

[8] One must know the Patterson function is the convolution of an atomic structure with its inverse.

orientation of the molecule, and not on its position in the cell. These maps are the basis of the rotation function. Intermolecular vectors depend both on the orientation of the molecule and on its position so, once the orientation is known, these can be exploited in the translation function. The intramolecular vectors will be shorter than the intermolecular vectors, so the rotation function can be computed using only the part of the Patterson map near the origin[9]. When the model is oriented correctly and placed in the correct position in the unit cell, the two Pattersons should be similar.

Rotation function. Before starting MR, it is strongly recommended to calculate a self-rotation function (Eqn. 1.4.3.1.4) to check for the presence of non-crystallographic symmetry (NCS).

$$R_{self}(C) = \int_V P_{cryst}(u) * P_{cryst}(Cu)du \qquad (1.4.3.1.4)$$

There are two exceptions where the self-rotation function does not correctly show NCS. The first case is when the NCS operator is parallel to a crystallographic symmetry operator of the same order; e.g. a 2-fold NCS parallel to a 2-, 4-, or 6-fold crystallographic axis, or a 3-fold NCS parallel to a crystallographic 3-, or 6-fold. In this case the orientation of the NCS-related molecules is identical to the orientation of molecules related by crystallographic symmetry and therefore cannot be distinguished. A second side-effect is that translational symmetry may be generated. NCS has been introduced into X-PLOR in two different ways [Brünger 1988].

The second exception is twinning that occurs when a crystalline specimen consists of multiple domains, which are mutually reoriented according to a specimen transformation that does not belong to the symmetry operations of the crystal point group, but is related in some way to the crystal lattice [Koch 1992][10]. These crystals are also oriented in such a way that their diffraction patterns exactly overlap each other. In the case like this every twin will make an independent contribution to the structure factors and, therefore, to the peaks in the Patterson. One can statistically detect twinning to perform de-twinning for data from twinned crystals or one can try to grow a new single crystal. Some ideas, with particular examples, how to solve this problem are reviewed in [Dauter 2003; Terwisscha van Scheltinga et al., 2003].

[9] The integration is typically carried out over a volume that eliminates all Patterson vectors beyond a certain radius, called also Patterson integration radius. The effect of the origin peak can be removed by subtracting its contribution from the Patterson function [Müller et al., 1995].

[10] More about twinning can be read in [Parsons 2003]. Merohedral twinning has been reported for several protein crystals [Yeates 1997; Chandra et al., 1999; Dumas et al., 1999; and Yeates & Fam, 1999].

Rossmann and Blow [Rossmann & Blow, 1962] proposed the function in Eqn. 1.4.3.1.5 to quantify the agreement between search and target Pattersons. The cross-rotation function will be calculated in a polar or Eulerian angle system to maximize the correlation between two Patterson functions:

$$R_{cross}(C) = \int_V P_{cryst}(u) * P_{\mathrm{mod}\,el}(Cu)du \qquad (1.4.3.1.5)$$

Several parameters may affect the sensitivity of the rotation function: (i) resolution; (ii) B/E values[11]; (iii) radius of integration; (iv) step size; (v) search model box size.

Translation function. The success of the translation function depends on how good the search model is and how accurate the rotation function solution is. After the orientation of the search model has been found, one has to place it at the correct position in the unit cell. The next step is to locate the translation vector with components (t_x, t_y, t_z) that relates the correctly oriented model to the unknown crystal structure. A modified Patterson function, which is part of a translation function $T(t)$ that only considers the cross-vectors, i.e. those relating two different crystal symmetry mates, can accomplish this:

$$P_{12}(u) = \int_S \rho 1(r) * \rho 2(r+u)dr, \qquad \text{where: } u = (u,v,w), \qquad r = (x,y,z) \quad (1.4.3.1.6)$$

$$T(t) = \int_V P_{cryst}(u) * P_{12}(u,t)du, \qquad \text{where: } t = (t_x, t_y, t_z) \qquad (1.4.3.1.7)$$

Crowther and *Blow* [Crowther & Blow, 1967] have originally developed this translation function, which can show maxima for the correct translation solution. The fast rotation and translation functions implemented in the program AMoRe reduce the time needed for an MR search from several hours to minutes [Navaza 1987; Navaza 1990; Navaza 1993; and Navaza 1994].

The rotation and translation solutions together with packing and NCS considerations should result in a correctly placed model [Harada et al., 1981; Tong 1996]. The correct position of the search model is defined relative to the symmetry axes in the crystal. For tricky problems one could consider to first do some rigid body refinement of the oriented search model.

[11] Instead of applying resolution cutoff, one can also apply a B value (B-factor) to weaken (positive B) or strengthen (negative B) the high resolution terms. Alternatively, one can use E value, also called normalized structure factor that do not decrease with resolution. The sharpening of a data places more emphasis on the high resolution terms, similar to applying a negative B value.

1.4.4 Refinement and validation

The computational procedures of refinement and structure validation can be formulated as the stereochemically-constrained or restrained non-linear optimization of a target function, which measures the agreement between observed data and data computed from an atomic model. In the 1940s, it was demonstrated that crystal structures could be refined by minimizing the sum of the squares of the deviations between the observed and calculated amplitudes or intensities. This type of crystallographic structure refinement only succeeds within a limited radius of convergence, i.e. when the current model is not too different from the true structure. Thus, this method is not capable of sampling molecular conformations thoroughly enough to find the most optimal model, the global minimum of a target function (some more information of well known multiple minima problem could be read in [Brünger et al., 1997]). In the last few years, maximum likelihood (ML) methods have been introduced and have quickly caught on, because often they are more successful [Read 2003]. The concept of ML takes into account model error, model incompleteness, and errors in the experimental data (errors of the phase estimation) [Pannu & Read, 1996; Adams et al., 1997].

ML and cross-validation. A more appropriate target for macromolecular refinement can be obtained through a ML formulation, where the best model is most consistent with the observations. Consistency is measured statistically by the probability that the observations should have been made. If the model is changed to make the observations more probable, the likelihood increases, indicating that the model has improved. The probabilities have to include the effects of all sources of error, including not just measurement errors, but also errors in the model itself. The effects of model errors (misplaced atoms and missing atoms) on the calculated structure factors are first quantified with σ_A values, which correspond roughly to the fraction of each structure factor that is expected to be correct [Pannu & Read,1996; Read 1986]. To achieve an improvement over the least-squares residual, cross-validation was used for the computation of σ_A, necessitating its calculation with a randomly selected test set of diffraction data that was never included in the refinement process. The cross-validated σ_A values are then used to compute the expected value of $|F_o|$ and the corresponding variance (σ_{ML}^{cv2}). As the model improves, its errors clearly decrease, which means the probabilities become sharper. The sharpening of probabilities also increases the likelihood, as long as they are no sharper than appropriate.

The ML refinement, implemented in REFMAC [Murshudov et al., 1997] and in the last versions of XPLOR/CNS [Brünger 1992; Adams et al., 1997], assumes that the best model is most consistent with the observations.

Measure of the quality of a crystal structure. Fitting a model to electron density, refining it, and validating the final result are processes which have to cope with the poor observation-to-parameter ratio that influences macromolecular structure determination. Sometimes it is possible to overfit or misfit the diffraction data because of the poorness of this ratio, and then an incorrect model can be refined to quite good *R* values but with gross errors [Bränden & Jones, 1990]. To indirectly increase the data-to-parameter ratio, one can incorporate the chemical information, i.e. bond length and bond angle restraints, which could be obtained from libraries of ideal values obtained from high-resolution structures of small model compounds. As a further complication, model bias in the electron density maps may complicate the process of manual refitting between cycles of automated refinement. Moreover, this problem is exacerbated by potentially overfitting the diffraction data in the refinement process. The crystallographic residual *R′* (the least-squared residual for the diffraction data), the most commonly used value to assess the correctness of the refinement, is written as:

$$R' = \sum_{h,k,l} \left(\left| F_{obs}(h,k,l) \right| - k \left| F_{calc}(h,k,l) \right| \right)^2 \qquad (1.4.4.1)$$

It is a linear function of the negative logarithm of the likelihood of the atomic model assuming that all observations are independent and normally distributed. *h ,k, l* are the reciprocal lattice points of the crystal, $\left|F_{obs}(h,k,l)\right|$ and $\left|F_{calc}(h,k,l)\right|$ are the observed and calculated structure factor amplitudes, respectively. *k* is a scale factor to bring the F_{obs} on the same scale as the F_{calc}. The least-squares residual does not account for the effects of phase errors in the calculated structure factors, so it is poorly justified when the model is far away from the correct one or incomplete.

Traditional measure for the quality of a structure is, however, the *R* value, which is also closely related to the *R′*, and described by the following formula:

$$R = \frac{\sum_{h,k,l} \left\| F_{obs}(h,k,l) - |k| F_{calc}(h,k,l) \right\|}{\sum_{h,k,l} \left| F_{obs}(h,k,l) \right|} \qquad (1.4.4.2)$$

R can be made arbitrarily small by increasing the number of model parameters and subsequent refinement against *R′*, that is the diffraction data can be overfitted without improvement or even reducing the information content of the atomic model.

Validation. Axel Brünger [Brünger 1992] has introduced the free-*R* value as a tool to avoid overfitting the diffraction data. In crystallography, cross-validation is a standard method, which helps to distinguish between refinement that fits the model to signal, and refinement that fits the model to noise in the data. Cross-validation, in the form of the free-*R* value, can be used to detect overfitting. The free-*R* value is the residual calculated on a set of data that were not included in the refinement, and this value is free from model bias (unless there is NCS present).

$$R_T^{free} = \frac{\sum_{(h,k,l)\in T} \left\| F_{obs}(h,k,l) \right| - k \left| F_{calc}(h,k,l) \right\|}{\sum_{(h,k,l)\in T} \left| F_{obs}(h,k,l) \right|} \qquad (1.4.4.3)$$

where the *R* value is computed for the *T* set of reflections. *T* is omitted in the modeling process (e.g. crystallographic refinement). The refinement is considered productive if the free-*R* value decreases significantly (>2%). Another parameter to watch is the difference between *R* and R^{free}. One would expect that free-*R* value is less prone to overfitting than *R*. Difference between the two *R* factors should not exceed 5% or the structure is tending towards overdetermined. There are several further checks to be taken into consideration: (i) atomic bumps; (ii) forbidden bond lengths/angles; (iii) peptide torsional angles or unusual side chain conformations; (iv) rotamers and surroundings. These are only a few examples that have to be examined in order to avoid introduction of unwanted errors which are difficult to correct in later stages of the refinement [Vriend 1990].