Studies on Structures, Dynamics and Interactions with Small Molecules of CNS Regeneration Inhibitory Components Associated with Nogo-A and EphA4

QIN HAINA (M. ENG.)

A THESIS SUBMITTED FOR DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2011

Acknowledgements

My deepest gratitude goes first and foremost to my supervisor, A/P Song Jianxing, for his constant encouragement and guidance. During these years, he has given me valuable suggestions on my project and always be very helpful no matter I am in any difficult situation. I am grateful to him not only for his scientific guidance but also emotional support. His enthusiasm and integral view on research have made a deep impression on me. Without his consistent support and illuminating instructions, it is impossible for me to finish my PhD study at this time.

Second, I would like to express my heartfelt gratitude to all my labmates, Shi Jiahai, Li Minfen, Liu Jingxian, Ran Xiaoyuan, Zhu Wanlong, Huan Xuelu, Wang Wei, Hong Ni, Shaveta Goyal, Lua Shixiong, Miao Linlin, Wang Xin, Ng Huiqi, Gavita Gupta. They gave me such warm friendship and valuable advices to me. In particular I am grateful to Dr Jingsong Fan for NMR experiment training and collecting NMR spectra on the 800 MHz and 500MHz spectrometer.

I also thank my beloved family for their loving considerations and great confidence in me all through these years.

Lastly, I am grateful to National University of Singapore for providing me research scholarship, which enabled me to complete my PhD degree without financial worry in Singapore.

TABLE OF CONTENTS

Abstract

The re-growth of injured neurons in CNS (central nervous system) is largely inhibited by the non-permissive environment around, and indeed several growth inhibitors have been identified so far. My thesis is aimed to study structures, dynamics and protein-protein interactions, as well as protein-small molecule interactions for two CNS regeneration inhibitors: Nogo-A and EphA4 receptor.

Intracellular Nogo-A protein level is believed to correlate with stroke, as well as other neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS) and Alzheimer's disease. Thus, it is of great interest to understand the mechanism of how Nogo-A protein level is regulated *in vivo*. An E3 ubiquitin ligase WWP1 was identified to be a novel interacting partner for Nogo-A both *in vitro* and *in vivo*, and down-regulated Nogo-A protein level by initiating the ubiquitination of Nogo-A. By using CD, ITC, and NMR, we have further conducted extensive studies on all four WWP1 WW domains and their interactions with a Nogo-A peptide carrying the only PPxY motif. Moreover, the solution structure of the best-folded WW4 domain is determined, and the binding-perturbed residues were derived for both WW4 and Nogo-A (650-666) by NMR HSQC titrations. On the basis of the NMR data, the complex model is constructed by HADDOCK 2.0. This study provides rationales as well as a template for further design of molecules to intervene in the WWP1-Nogo-A interaction which may regulate the Nogo-A protein level by controlling its ubiquitination.

EphA4 was proved to play key roles in the inhibition of the regeneration of injured axons, synaptic plasticity, platelet aggregation, and so on. In addition, EphA4 has unique ability to bind all ephrins including 6 A-ephrins and 3 B-ephrins. Therefore, studies of EphA4 structure, dynamics, and its interaction with ephrin ligands and

small molecules will be critical in understanding mechanisms underlying the binding between Eph receptor and ephrin ligands as well as molecule design targeting disease-involved Eph receptors. Both crystal and NMR structures of free EphA4 LBD were resolved, revealing the highly dynamic property of loops that comprising the classical binding pocket. Dynamics study shows that the whole EphA4 ligand binding domain undergoes dramatic conformational exchanges on μs-ms time scales. These results may have crucial implications in understanding why EphA4 owns a unique ability to bind all 9 ephrins. The results with EphA4 dynamics may also help to design and optimize small molecule agonists and antagonists with high affinity and specificity for EphA4. The crystal structure of the EphA4-ephrin-B2 complex was also determined and an additional interaction surface was identified which will enhance the affinity and specificity of the interclass binding. These findings contribute to our understanding of the distinctive binding determinants that characterize selectivity versus promiscuity of Eph receptor-ephrin interactions and suggest that diverse strategies may be needed to design antagonists for effectively disrupting different Eph-ephrin complexes. The first two small molecules which antagonize ephrin-induced effects on EphA4-expressing cells were also presented in our work. Their binding with EphA4 LBD were studied by ITC, NMR and computer docking. Our results demonstrate that the high-affinity ephrin-binding pocket of the Eph receptors is amenable to targeting with small molecule antagonists and suggest avenues for further optimization.

Abbreviations

(RP-)HPLC	(Reversed-Phase) High Performance Liquid
(θ)MRW	Mean Molar Ellipticity per Residue in CD
1J/ 2J / 3J	Scalar Coupling Through One Bond/ Two bonds/
ALS	Amyotrophic Lateral Sclerosis
AU	Asymmetric Unit
CD	Circular Dichroism
CNS	Central Nervous System
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
dαN/ dβN / dNN	NOE Connectivity Between CαH/ CβH/ NH with NH
E.coli	*Escherichia coli*
Eph receptors	Erythropoietin-Producing Human
EphA4 LBD	EphA4 Ligand Binding Domain
ER	Endoplasmic Reticulum
FID	Free Induction Decay
GST	Gluthathione S-transferase
Haddock	High Ambiguity Driven protein-protein Docking

HSQC	Heteronuclear Single Quantum Coherence
IPTG	Isopropyl β-D-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
LB	Luria-Bertani
NMDA receptor	*N*-methyl *D*-aspartate receptor
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauster Effect
NOESY	Nuclear Overhauser Enhancement Spectroscopy
OD	Optical Density
PBS	Phosphate-buffered Saline
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
RMSD	Root Mean-square Deviation
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel
TOCSY	Total Correlation Spectroscopy
UV	Ultraviolet
WWP1	WW domain-containing protein 1

List of Figures

List of Tables

Chapter I INTRODUCTION

1.1 Biological background

1.1.1 CNS injury

CNS (Central Nervous System, including brain and spinal cord) injuries, induced by stroke, traumatic injury, or neurodegenerative diseases, could be permanently disabling because the transected axons in CNS could not regenerate beyond the lesion site. Potential consequences of these CNS injuries include memory loss, inability to concentrate, speech problems, motor and sensory deficits, and behavioural problems. Each year in the United States, more than 2 million people suffer from traumatic brain injuries, over 500,000 people suffer from stroke, and at least 10,000 people suffer from spinal cord injuries. So far, the primary treatments are based on physiotherapy and it is estimated that only 0.9 % of the patients will have completed neurological recovery. Therefore, there is a huge unmet medical need for treating CNS injury.

1.1.2 Mechanisms that inhibit axonal regeneration

It was observed long ago that unlike axons in peripheral nervous system, severed axons in the CNS do not have the ability to grow significant distance (Ramon, 1928). Moreover, the study in 1980 showed that the peripheral nerve implanted in CNS can not grow out of long distance (Weinberg EL, 1980). Thus, the failure of axons in CNS to regenerate is due to the non-permissive environment they inhabit. If the inhibitory factors could be removed or blocked, or if suitable growth-promoting agents added, the axons might re-grow through the lesion site.

Several growth inhibitors have been identified which are either components of the extracellular matrix present in the glial scar or molecules associated with CNS myeline.

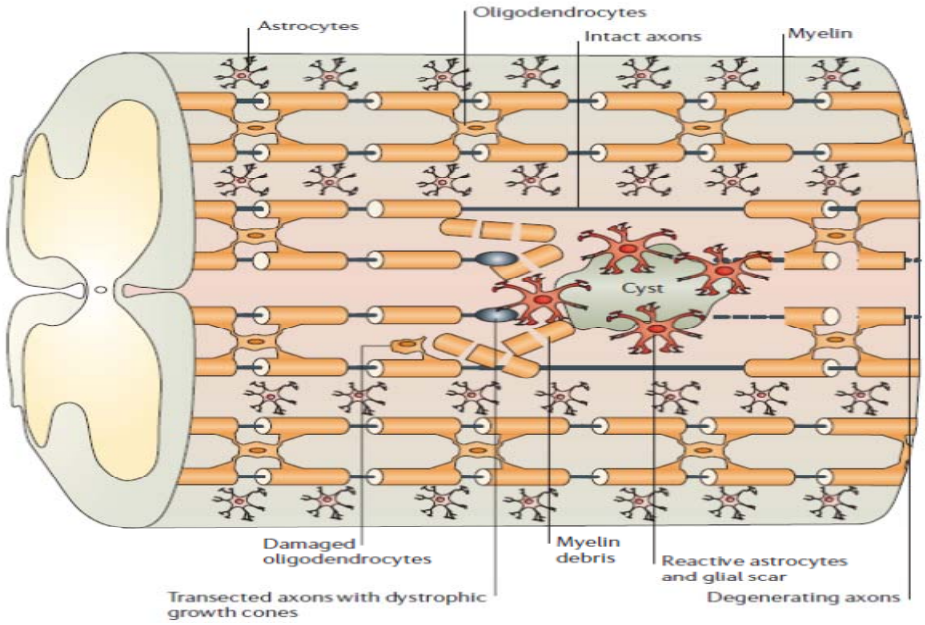


Figure1.1 Schematic representation of the CNS injury site (Glenn Yiu, *et al*, 2006).

1.1.3 Inhibitors from glial scar and associated with CNS myeline

Injury to the adult CNS often results in the transection of nerve fibres and damage to surrounding tissues. The distal ends of the severed axons form characteristic dystrophic growth cones that are exposed to the damaged glial environment. During the early phase of injury, myelin associated inhibitors from intact oligodendrocytes and myelin debris can restrict axon regrowth. Recruitment of inflammatory cells and reactive astrocytes over time leads to the formation of a glial scar, often accompanied by a fluid-filled cyst. This scarring process is associated with the increased release of chondroitin sulphate proteoglycans, which can further limit regeneration. Together,

these molecular inhibitors of the CNS glial environment present a hostile environment for axon (Glenn Yiu *et al*, 2006).

1.1.3.1 Inhibitors by components of the glial scar

Key classes of inhibitory molecules are 1) Chondroitin sulphate proteoglycans; 2) ephrins; 3) semaphorins. All of them are up-regulated after injury. Chondroitin sulphate proteoglycans are components of the CNS extracellular matrix; the precise mechanism by which they cause growth inhibition is not known, but the fact that they do so is well established (McKeon RJ *et al*, 1995; Snow DM *et al*, 1990). Ephrins and their Eph receptors are a family of membrane proteins that are involved in axon guidance during development, but are also present in the CNS in adulthood. Binding of ephrins on one cell to their receptors on another activates bidirectional signalling pathways that in neurons lead to the collapse of the growth cone (Holland SJ *et al*, 1996). Semaphorins are a large family of membrane-bound and secreted proteins that are also involved in axon guidance during development. Upregulated production of Sema-3A by the meningeal cells that migrate in to form the lesion core is again inhibitory to axonal growth (De Winter F *et al*, 2002).

1.1.3.2 Inhibitors associated with CNS myeline

Cultured neurons grow readily on substrates of myelin extracted from peripheral nerve, but not on beds of mature oligodendrocytes or isolated CNS myelin (Schwab ME *et al*, 1988). The proteins Nogo-A (Chen MS *et al*, 2000) and myelin-associated

glycoprotein have been identified as mediators of the growth-suppressive effects of CNS myelin. In addition, oligodendrocytes secrete the inhibitory protein tenascin R.

My thesis puts emphasis on Nogo-A, a widely studied inhibitor, and EphA4, a newly discovered agent involved in neuron regeneration. Thus, the following review will focus on the introduction to these two protein families.

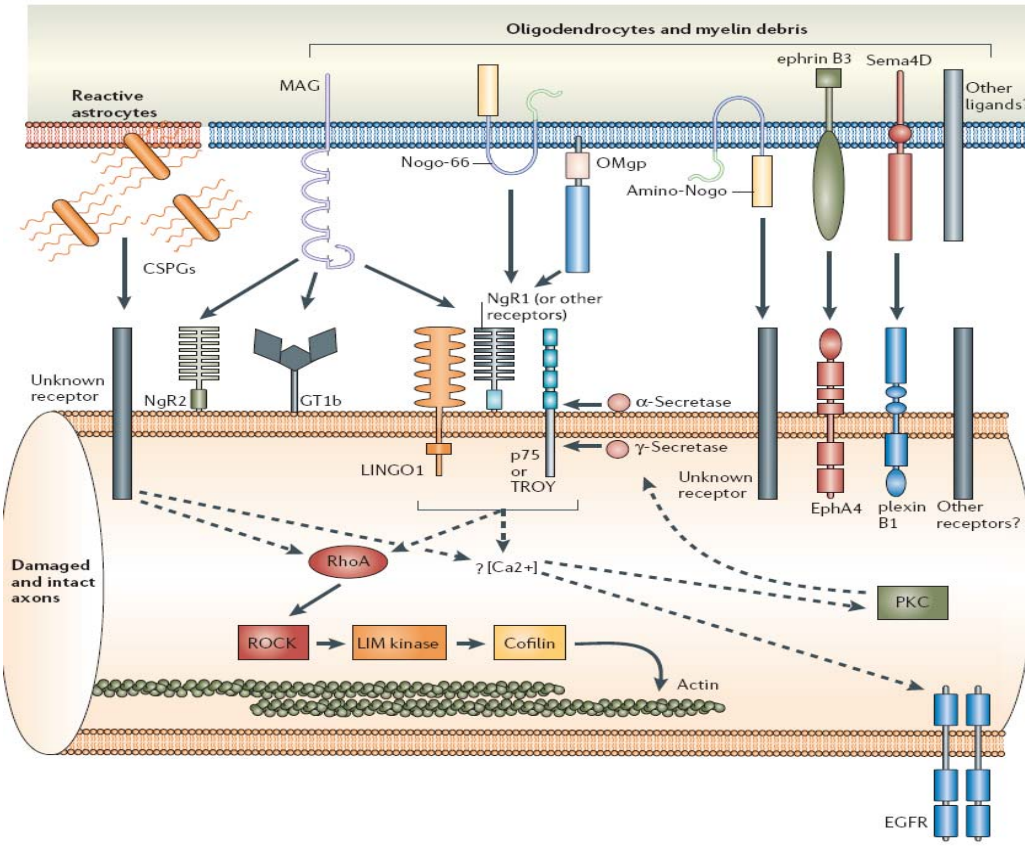


Figure1.2 Glial inhibitors and intracellular signalling mechanism (Glenn Yiu, *et al*, 2006).

1.1.4 NogoA as an Inhibitors of Axon Regeneration in CNS

Nogo is a member of the reticulon family of membrane proteins, and at least three isoforms (Nogo-A, -B and -C) are generated by alternative splicing and promoter usage.

Among these, Nogo-A is best characterized, owing to its high expression in CNS oligodendrocytes (Huber, A. B. *et al*, 2002). Structure–function analyses support the presence of two inhibitory domains: a unique amino-terminal region (amino-Nogo) that is not shared by Nogo-B and Nogo-C (Oertle, T. *et al*, 2003), and a 66 amino acid loop (Nogo-66) that is common to all three isoforms (Prinjha, R. *et al*, 2000). Nogo-66 contains three helices and with the long-range packing between the second and third helix, whereas the amino terminal region was demonstrated as intrinsically unstructured (Li M *et al*, 2006; Li M *et al*, 2004; Li M *et al*, 2006).

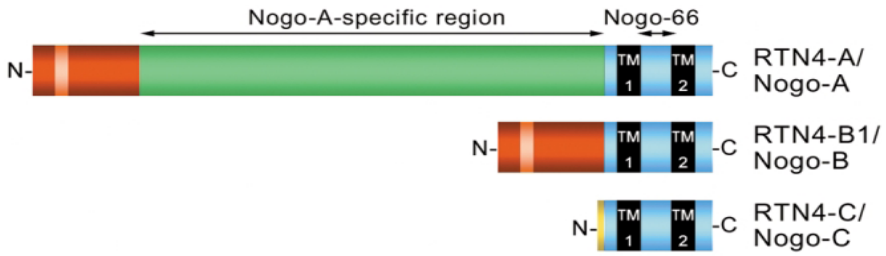


Figure1.3 Schematic representation of Nogo family (Oertle, T. *et al*, 2003)

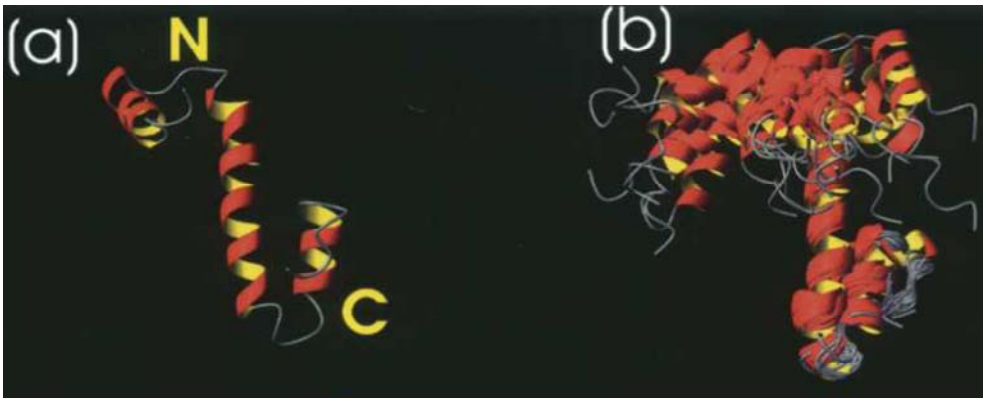


Figure1.4 NMR solution structure of Nogo-66 (Li M *et al*, 2004)

Nogo-A has been demonstrated to be a potent neurite growth inhibitor and plays a key role both in the restriction of axonal regeneration after injury and in structural plasticity of the CNS of higher vertebrates. *In vivo* neutralizing Nogo-A by its antibody has been shown to enhance sprouting and functional recovery after cervical lesion in rat and adult primates. In addition, Nogo-A was also identified to be essential for the tubular network formation of ER (Voeltz G.K. *et al*, 2006). Most recently, a role of Nogo-A in synapse integrity has also been suggested and overexpression of Nogo-A led to destabilization and retraction of nerve terminal (Jokic N. *et al*, 2006).

Intracellular Nogo-A protein level is believed to correlate with stroke (Li S. *et al*, 2006), as well as other neurodegenerative diseases such as ALS (Jokic N. *et al*, 2005) and Alzheimer's disease (Gil V. *et al*, 2006). These observations indicate that the intracellular Nogo-A protein level is essential to the functions of Nogo-A in cell. As a result, it is of significant meaning to know how the Nogo-A protein level is regulated *in vivo*.

Recently, our collaborators found that WWP1, an E3 ubiquitin ligase, can interact with Nogo-A, and initiate the ubiquitination of Nogo-A, subsequently down-regulate the intracellular Nogo-A protein level. As this phenomenon was also observed at the axonal sprouting region of the mice stroke model, WWP1 mediated ubiquitination is likely to play an important role in Nogo-A axon regeneration inhibition. The investigation of how WWP1 interact with Nogo-A would have significant meaning in Nogo-A involved neuron diseases by providing the interaction mechanism information. In my thesis, the binding between WWP1 and Nogo-A was studied by NMR, ITC, and their binding was modeled by molecular docking.

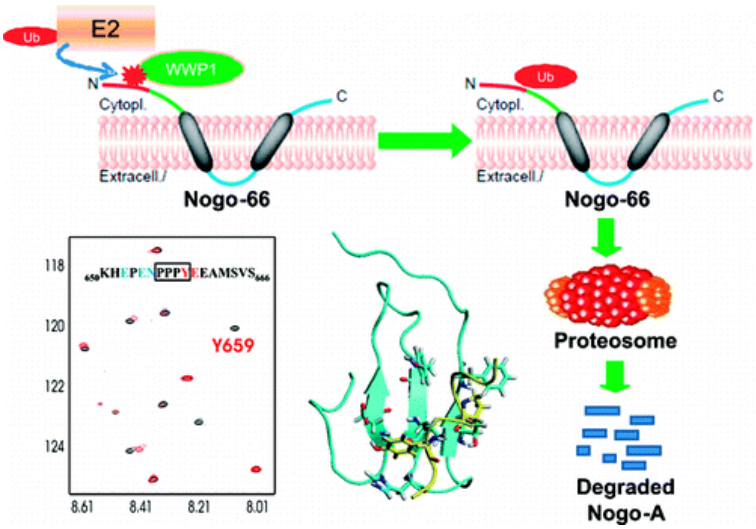


Figure 2.5 Schematic representation of Nogo-A degradation (Qin H. *et al*, 2008)

1.1.3 Eph and ephrin and their function in axon regeneration in CNS

Eph proteins constitute a large family of receptor tyrosine kinases that bind to ligands called ephrins. The Eph and ephrin protein families are each divided into A and B subclasses based on sequence homology, membrane anchorage, and binding preference for each family member. There are 10 EphA and 6 EphB receptors known at present, while 6 ephrin-A and 3 ephrin-B proteins have been identified. Ephrin-A proteins are attached to the cell membrane by a glycophospatidylinositol (GPI) anchor, while ephrin-B proteins have a transmembrane region and a short, highly conserved cytoplasmic tail with a PDZ (postsynaptic density-95/Discs large/zona occludens-1)-binding domain (Martinez A. *et al*, 2005; Song J. *et al*, 2002; Torres R. *et al*, 2008). Eph receptors consist of a highly conserved N-terminal extracellular ligand binding domain, followed by a cysteine-rich domain, two fibronectin III repeats, a juxtamembrane region, and an intracellular kinase domain with a PDZ binding motif (Flanagan JG *et al*, 2008).

In the nervous system, Ephs and ephrins have been most extensively studied for their developmental roles in axon guidance, topographic mapping, hindbrain segmentation, and neural crest cell migration (Wilkinson DG *et al*, 2005; Henkemeyer M. *et al*, 2003). Ephs and ephrins are not only developmental molecules but also play important role in adult nervous system. Evidence showed that, Ephs and ephrins can modulate synaptic function by regulating dendritic spine formation (Henkemeyer M. *et al*, 2003; Ethell IM *et al*, 2001; Murai KK *et al*, 2003), NMDA receptor clustering, and potentiation of calcium influx (Takasu MA *et al*, 2002; Dalva MB *et al*, 2000). Moreover, Ephs and ephrins also are involved in the proliferation and progenitor cells in neurogenic regions (Depaepe V *et al*, 2005; Katakowski M *et al*, 2005; Ricard J *et al*, 2006; Holmberg J *et al*, 2005; Conover JC *et al*, 2000; Aoki M *et al*, 2004).

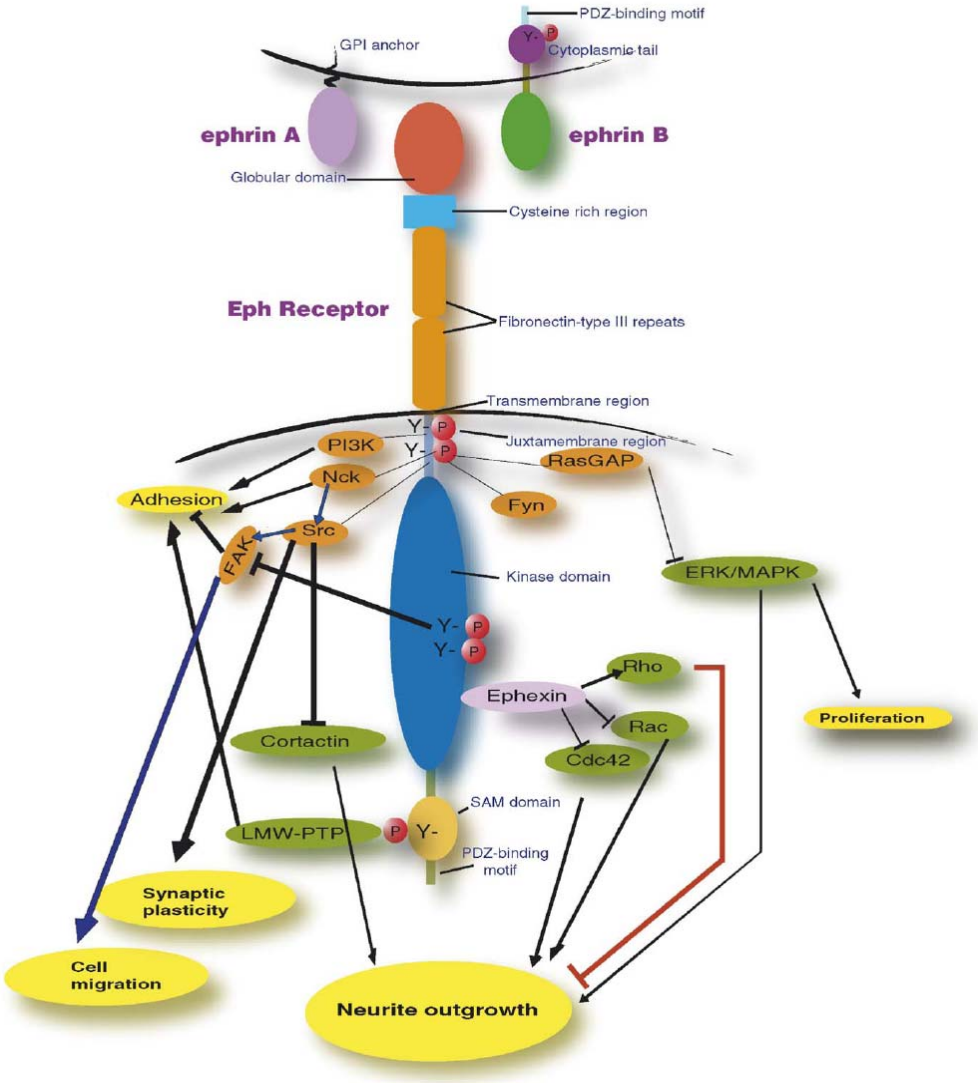


Figure 1.6 Eph receptor structure and signalling (Yona Goldshmit *et al*, 2006)

Outsides the nervous system, ephrinB2 on arterial and EphB4 on venous endothelial cells mediate embryonic vascular assembly and atteriovenous differentiation (Gerety SS *et al*, 1999; Adams RH *et al*, 1999). EphrinB2 and EphA2 also are linked to neovascularization and tumor angiogenesis (Gale NW *et al*, 2001; Shin D *et al*, 2001; Brantley DM *et al*, 2002; Noren NK *et al*, 2004; Brantley-Sieders DM *et al*, 2005). Many Eph and ephrin family members are found in tumors including those of the

breast, lung, colon, prostate, and in glioblastoma as well as melanoma (Surawska H *et al*, 2004), where their expression level is correlated with malignancy to some degree. It is thought that Eph/ephrin promotes tumor metastasis by negatively regulating cell adhesion and enhancing neovascularisation (Wimmer-Kleikamp SH *et al*, 2005; Dodelet VC *et al*, 2000).

1.1.6 Eph/ephrin functions in axon regeneration

In the last several years, much evidence has documented that the Ephs and ephrins play important roles after CNS damage in the brain (Biervert C *et al*, 2001; Moreno-Flores MT *et al*, 1999), optic nerves (Liu X. *et al*, 2006), and spinal cord (Bundesen LQ *et al*, 2003; Fabes J *et al*, 2006; Miranda JD *et al*, 1999; Willson CA *et al*, 2002; Willson CA *et al*, 2003), like EphB2, EphB3, EphA4 and ephrinB2. EphB3 plays a role in retinal ganglion cell axonal plasticity and initial axon attempts at re-growth after injury, ephrinA2 and ephrinA3 can promote neuronal survival and neurite outgrowth after optic nerves injury. In contrast with the promotion effect, some Eph/ephrins directly inhibit the re-growth of injured neuron or indirectly inhibit the re-growth by form the glial scar.

Evidence show that after spinal cord injury, damaged corticospinal tract axons express EphA4, and are surrounded by astrocytes expressing ephrinB2 (Jez Fabes *et al*, 2006). EphA4 is also up-regulated in astrocytes in injured wide-type spinal cord, while regenerating axons express ephrinB3 (Jez Fabes *et al*, 2006). Whereas EphA4 homozygous null mice study showed that EphA4-/- axons were able to cross the lesion site in greater number than in wide-type mice (Yona Goldshmit *et al*, 2004).These evidence indicate that EphA4 may sense the repellent growth signals

from environment by interacting with ephrinB2 and ephrinB3. Interestingly, in the later work, Douglas Benson *et al* found that ephrinB3 showed equally inhibitory activity as other three myelin associated inhibitors for EphA4-positive neurons, which further confirms that the EphA4/ephrinB3 pathway has inhibitory effect on the injured neuron (Benson, M. D. *et al*, 2005).

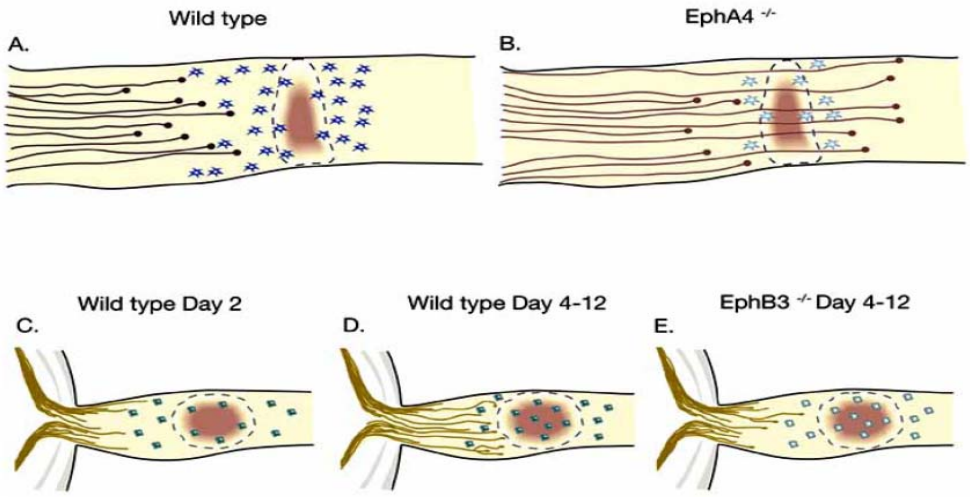


Figure 1.7 Eph and ephrin function after spinal cord and optic nerve injury in mice (Yona Goldshmit *et al*.)

1.1.7 Structure of Eph receptor and its complex with ephrins ligands

Due to the wide distribution *in vitr*o and huge signalling network comprised, the structure studies to Eph receptor family and its binding partners have attracted intensive attention from investigators. These studies have substantially improved our understanding on Eph receptor structure as well as their binding mechanism with ephrin ligands (Himenan *et al*, 1998; Himenan *et al*, 2001; Himenan *et al*, 2004; Chrencik JE *et al*, 2006). From these released structures, Eph receptor ectodomain adopts a Greek key topology constituted by an 11 β-stranded barrel. The concave sheet is comprised of strands C, F, L, H, and I, and the convex sheet of strands D, E,

A, M, G, K, and J, which are connected by loops of variable length. The formation of the complex between Eph receptors and ephrins is centered around the insertion of the solvent exposed ephrin G-H loop into the hydrophobic channel formed by the convex sheet of four β-strands, together with the D-E, J-K and G-H loops of the Eph receptor. These interactions are mostly hydrophobic and, together with an adjacent mostly polar surface region, form the high affinity interface of Eph receptor-ephrin complexes, which is involved in receptor-ephrin dimerization (Himanen JP *et al*, 2001; Himanen JP *et al*, 2004; Chrencik JE *et al*, 2006; Himanen JP *et al*, 2009). Other interfaces contribute to Eph-ephrin binding, including: (1) additional residues on both the receptor and ephrin surfaces, (2) a low affinity interface also located in the binding domains of Eph receptors and ephrins, which was identified in the EphB2-ephrin-B2 complex and appears to mediate the association of two receptor-ephrin dimers (tetramerization) (Himanen JP *et al*, 2004), and (3) an interface involving the cysteinerich region adjacent to the Eph receptor ligand binding domain, which was identified by mutagenesis in EphA3-ephrinA5 complexes but has not been structurally characterized and which might be implicated in higher order clustering.

While Eph receptors interact promiscuously with ephrins of the same class, they rarely interact with ephrins of the other class. A variety of factors appear to contribute to class specificity. B class Eph/ephrin interactions are characterized by a compact conformation, which necessitates considerable structural rearrangements of both the receptor and the ephrin, while EphA receptors and ephrin-A ligands appear to undergo smaller rearrangements when forming a complex (Himanen JP *et al*, 2009). Differences in critical residues located in the interacting regions and sequence differences in the class specificity H-I loop of the Eph receptors seem to also play a role in class specificity (Himanen JP *et al*, 2001; Himanen JP *et al*, 2004; Chrencik JE

et al, 2006; Himanen JP *et al*, 2009). However, examples of interclass binding also exist: EphB2 can bind ephrin-A5 and EphA4 can bind all three ephrin-B ligands (Pasquale EB, 2004).

EphA4 binding to ephrin-B ligands is also weaker than to ephrin-A ligands. However, EphA4-ephrin-B interclass interactions have been shown to be physiologically relevant in many biological systems. For example, EphA4 interaction with ephrin-B1 stabilizes blood clot formation through an integrin-dependent mechanism (Prévost N *et al*, 2005) while EphA4 interaction with ephrin-B2 and/or ephrin-B3 regulates cell sorting in the rhombomeres and branchial arches of the developing hindbrain (Smith A *et al*, 1997; Xu Q *et al*, 1999), somite morphogenesis (Barrios A *et al*, 2003), axon guidance and circuit formation in the developing spinal cord (Kullander K *et al*, 2001; Kullander K *et al*, 2001; Yokoyama N *et al*, 2001; Kullander K *et al*, 2003), and inhibition of axon outgrowth by myelin (Benson MD *et al*, 2005). The distinctive ability of EphA4 to bind both ephrin-A and ephrin-B ligands makes it an attractive model to understand the structural principles underlying the selectivity versus promiscuity of Eph receptor-ephrin interactions, but no structural information has been available for free EphA4 and EphA4-ephrin complexes. In this thesis, high resolution 3D structures of EphA4 and its complexes with ephrin ligands will be determined. This study will reveal structure characterization of EphA4 ligand binding domain and its binding mechanism between EphA4 and its ephrin ligands and find out how the receptors recognize different ephrin ligands with high specificity.

1.1.8 Organic compounds as small antagonist of EphA4

As an important target for drug design, a variety of molecules that inhibit interaction between Eph receptors and ephrins ligands are investigated. A number of peptides indentified by phage display show their selectivity, high binding affinity with some of Eph receptors (Murai KK *et al*, 2003). Other molecules that modulate Eph-ephrin interactions have also been identified, including antibodies and soluble forms of Eph receptors and ephrins extracellular domains (Ireton, R. C. *et al*, 2005; Noren, N. K. *et al*, 2007; Wimmer-Kleikamp, S. H. *et al*, 2005). Several small molecule inhibitors of Eph receptor kinase domain have also been reported (Caligiuri, M. *et al*, 2006; Karaman, M. W. *et al*, 2008; Miyazaki, Y. *et al*, 2008; Kolb, P. *et al*, 2008). These inhibitors occupy the ATP binding pocket of the receptors and are usually broad specificity inhibitors that target different families of tyrosine kinases (Caligiuri, M. *et al*, 2006; Karaman, M. W. *et al*, 2008). Epigallocatechin gallate, a green tea derivative known to inhibit several tyrosine kinases, has also been shown to inhibit EphA receptor-mediated a human umbilical vein endothelial cell migration and capillary-like tube formation, but the mechanism of action of this molecule has not been elucidated. Therefore, the high-affinity ephrin binding pocket of the Eph receptors appears to be an attractive target for design of small molecules capable of inhibiting the Eph receptor signaling by blocking ephrin binding. By high throughput screening approach, Noberini R. et al identified two isomeric 2,5-dimethylpyrrolyl benzoic acid derivatives that selectively inhibit ephrin binding to EphA4 and EphA2 as well as the functions of these receptors in live cells (Noberini R *et al*, 2008). This is a very important start point for drug design of EphA4 receptor involved pathways. It is of significant interest to gain structural insight into the binding interactions between

EphA4 and two molecules, hoping this study will develop small organic antagonist with high binding affinity and specificity.

1.1.9 Dynamics study of proteins

Although by X-ray and NMR, investigators have produced many pictures of protein structures, these static 3D structures alone can not completely explain results from functional biological assays, nor do they necessarily illuminate the path for protein engineering or rational drug design. This is because a three dimensional static structure provides a description of the ground state of the molecule. Macromolecular function is, however, in many cases, highly dependent on excursions to excited molecular states and hence intimately coupled to flexibility. Recently evidence has accumulated to suggest that protein dynamics may play a critical role in the biological functions including signal transmission (Baldwin AJ *et al*, 2009; Henzler-Wildman, K *et al*, 2007; Smock, R.G. *et al*, 2009). Therefore, a complete and much more useful description of the structure of a molecule will require an understanding of how the structure changes with time and bridge the gap between static and dynamic pictures of molecular structure and to demonstrate how motion relates to function. In this thesis, dynamics of free EphA4 ligand binding domain and its complexes with small antagonists will be studied, and more mechanism behind interaction between EphA4 and small antagonists will be revealed.

1.2 Protein NMR

1.2.1 Physical basis of NMR spectroscopy

Atoms and molecules have a variety of quantised energy levels. Many spectroscopic techniques take advantage of transitions between these energy levels with different ΔE values being related to particular frequency-ranges of the electronmagnetic spectrum by equation

$$E = h\nu \qquad \text{(E1.1)}$$

Where h is Planck's constant and ν is frequency.

Nuclear magnetic resonance spectroscopy is a spectroscopic technique which takes advantage of magnetic spin. Magnetic spin is a property of many different types of atoms. Take ^{1}H as example, this nuclei can be regarded as a spinning positively charge. This generates a magnetic field which will have a magnetic spin moment, μ. If an external magnetic field (B_0) is applied to such a nucleus, it can orientate itself either with (parallel) or against (antiparallel) this field like a bar-magnet does in the earth's magnetic field on the macroscopic scale. These two orientations are referred to as spin states and are distinguishable by their different spin quantum numbers, m_I, which are respectively, -1/2, and +1/2. The magnetic spin moment 'wobbles' or processes around the axis of the external magnetic field by an angle, θ, and rotates around this axis with a particular frequency, ω, which is called the Larmore frequency.

The potential energies of the two spins states are given by

(Low energy spin state: m_I=-1/2)

$$E = -\mu B0\sin\theta \qquad \text{(E1.2)}$$

(High energy spin state: m_I=1/2)

$$E = \mu B0\sin\theta \qquad \text{(E1.3)}$$

The energy difference, ΔE, between them is therefore given by

$$\Delta E = 2\mu B0\sin\theta \qquad \text{(E1.4)}$$

ΔE is proportional to the applied magnetic field. Early work on biomolecules used magnetic field strengths of only approximately 40MHz. Modern NMR spectrometers use much larger field strengths (500-800MHz) which give rise to larger ΔE values and yield NMR spectra of much higher resolution.

NMR spectroscopy depends on absorption of electromagnetic radiation from the radiowave part of the spectrum causing the nucleus to undergo a transition from a low to a high energy spin state. The precise value of v required for the transition depends on both the identity of the nucleus and on its precise chemical environment. Because of this, NMR spectra can yield precise information on the structure/composition of biomolecules and on processes in which they are involved.

A wide range of different elements have nuclei which are amenable to study by NMR spectroscopy. Those which are most relevant to the study of biological macromolecules are listed in Table 1.1, the nucleus which is most sensitive to the detection by NMR is hydrogen, and this is by far the most important nucleus for the study of biological macromolecules. Other nuclei such as 15N, 13C are nowadays often detected through their attached protons to take the advantage of sensitivity.

Table1.1 Properties of nuclei of interest on NMR studies of proteins				
Isotope	Spin	Frequency(MHz) at 11.74T	Natural abundance(%)	Relative sensitivity
^{1}H	1/2	500.0	99.98	1.00
^{2}H	1	76.7	1.5×10^{-2}	9.65×10^{-3}
^{3}H	1/2	533.3	0	1.21
^{12}C	0	--	98.89	--
^{13}C	1/2	125.7	1.108	1.59×10^{-2}
^{14}N	1	36.1	99.63	1.01×10^{-3}
^{15}N	1/2	50.7	0.37	1.04×10^{-3}
^{16}O	0	--	100	--
^{17}O	5/2	67.8	3.7×10^{-2}	2.91×10^{-2}
^{19}F	1/2	470.4	100	0.83
^{31}P	1/2	202.4	100	6.63×10^{-2}

The first published NMR spectrum of a biological macromolecule was the 40MHz 1H spectrum of pancreatic ribonuclease reported in 1957. The most that could be deduced from this spectrum was that it was consistent with the amino acid. The subsequent years, perhaps particularly the last twenty years, have seen astonishing developments in instrumentation and methodology which have enormously increased the power of NMR, notably in its application to study the conformations and interactions of biological macromolecules. The most important of these developments include the following:

1. The construction of higher field spectrometers, with a consequent increase in sensitivity and spectral dispersion.

2. The development if pulse Fourier transform methods, in which the radiofrequency

radiation is applied to the sample in the form of a more or less complex sequence of pulses, and the spectrum obtained by Fourier transformation of the response if the nuclear spins to these pulse trains.

3. The development of multi-dimensional NMR, in which resonance intensity is recorded as a function of two, three, or four frequency variables.

1.2.2 Chemical shift

Because instrumental limitations, it is difficult to measure *v* values accurately. To standardize measurements between different NMR spectrometers and different experimental conditions, it is usually include a reference (normally TMS or deuterium signal) with the sample to be analyzed. The frequency corresponding to the resonance condition for each transition in the sample is then expressed as the chemical shift, δ, in parts per million (ppm) as follows:

$$\delta = (\nu_s - \nu_{ref})/\nu_{ref} \times 10^6 \qquad \text{(E1.5)}$$

What makes NMR especially informative is the fact that precise radiation frequency, *v*, corresponding to the resonance condition for each type of nucleus at a given applied magnetic field strength can be affected by its immediate chemical environment. This is due to the magnetic effect of nearby nuclei on that of nuclei undergoing transition. Generally speaking, the factors affect chemical shift are electron density, electronegativity of neighbouring groups and anisotropic induced magnetic field effects. In protein NMR spectrum, each atom with spin will have specific chemical shift, which makes protein studied by NMR possible. Moreover, chemical shift provides useful information in identifying protein secondary structure.

Characteristic chemical shift deviations of Hα, C=O, Cα, and Cβ from random coil values are good indicators for the existence of α-helix or β-sheet (Figure 1.8) (Wishart DS *et al*, 1994) .

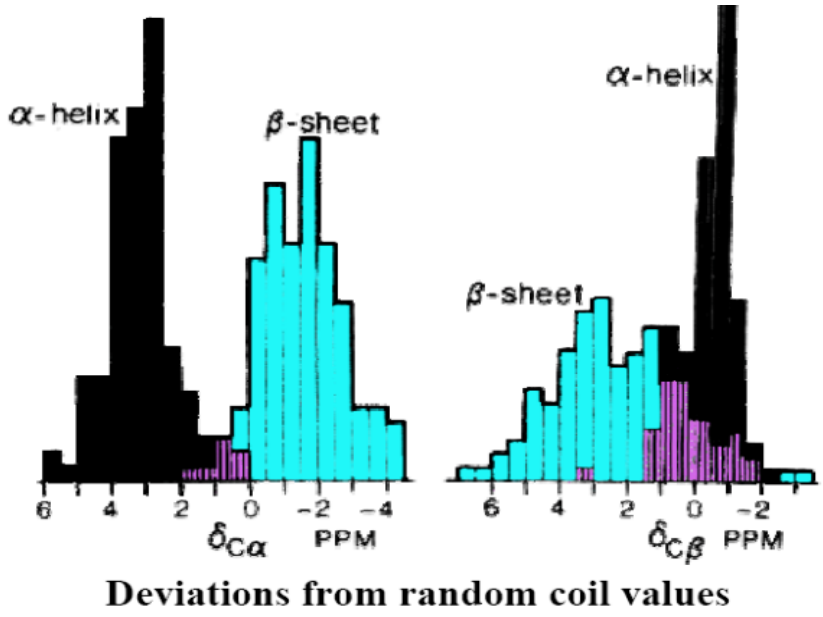


Figure 1.8 Correlation between chemical shift deviation and 2nd structure (Wishart DS *et al*, 1994)

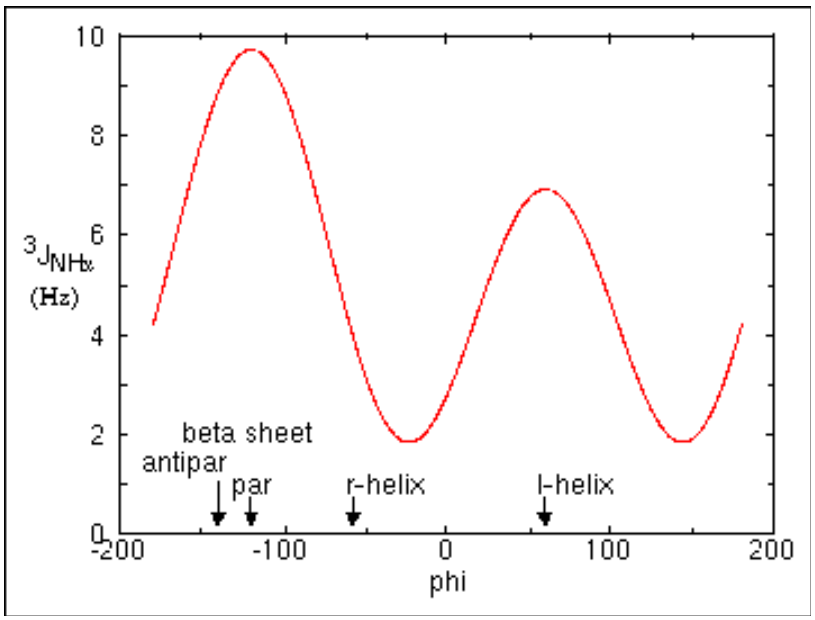


Figure 1.9 Correlation between J-coupling and 2nd structure (A. Pardi *et al*, 1984)

1.2.3 J coupling

Nuclei experiencing the same chemical environment or chemical shift are called equivalent. Those nuclei experiencing different environment or having different chemical shifts are nonequivalent. Nuclei which are close to one another exert an influence on each other's effective magnetic field. This effect shows up in the NMR spectrum when the nuclei are nonequivalent. If the distance between non-equivalent nuclei is less than or equal to three bond lengths, this effect is observable. This effect is called spin-spin coupling or J-coupling. J-coupling contains information about dihedral angles, which can be estimated using the Karplus equation:

$$J(\emptyset) = A\cos^2\emptyset + B\cos\emptyset + C \qquad \text{(E1.6)}$$

Where J is the ^{3}J coupling constant, ϕ is the dihedral angle, and A, B, and C are empirically-derived parameters whose values depend on the atoms and substituents involved.

Similar to chemical shift deviation of residues in structured region from those in random coil, deviation of $^3J_{NHH\alpha}$ values from random coil values provides valuable secondary structural information. In folded proteins, β-Structures are characterized by large coupling constant values in the range 8~10 Hz, while α-helical structures are characterized by values in the range 3~5 Hz. In unfolded proteins, however, the coupling constants are about 6~7.5Hz due to the fact that conformational fluctuation averages the coupling constants (A. Pardi *et al*, 1984).

1.2.4 NOE (Nuclear Overhauster Effect)

Overhauster effect was first discovered by Albert Overhauster in 1953. It is the phenomenon that the transfer of spin polarization from one spin population to another via cross-relaxation in nuclear magnetic resonance spectroscopy. The original Overhauser effect was described in terms of polarization transfer between electron and nuclear spins, but is now mostly used for transfer between nuclear spins, the Nuclear Overhauser Effect (NOE or nOe). A very common application is NOESY (Nuclear Overhauser Effect Spectroscopy), an NMR technique for structure determination of macromolecular motifs.

NOE differs from spin coupling in the respect that NOE is observed through space, not through bonds. Thus, all atoms that are in proximity to each other give a NOE, whereas spin coupling is observed only when the atoms are bonded to same or neighboring atoms. Furthermore, the distance can be derived from the observed NOEs, so that the precise, three-dimensional structure of the molecule can be reconstructed.

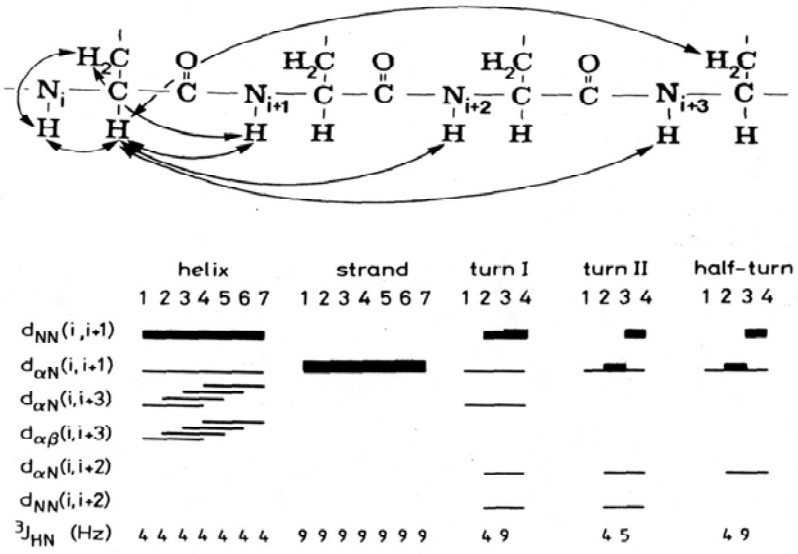


Figure 1.10 NOE patterns associated with protein 2nd structure (Wuthrich K, 1986)

1.2.5 NMR relaxation and protein dynamics

In NMR spectroscopy, the term relaxation describes several processes by which nuclear magnetization prepared in a non-equilibrium state return to the equilibrium distribution. When an excited magnetic moment relaxes back to equilibrium, the z axis, there are two components of this relaxation for isotropic systems in the absence of chemical exchange: longitudinal or spin lattice (T1) and transverse or spin-spin (T2). T1 is always at least slightly slower than T2.

Biomolecules are intrinsically flexible and dynamic systems. These characteristics critically assist them in their quest to perform biological functions. Nuclear magnetic resonance (NMR) spectroscopy can be used to monitor the dynamic behaviour of a protein at a multitude of specific sites. Moreover, protein movements on a broad range of timescales can be monitored using various types of NMR experiments — nuclear spin relaxation rate measurements report the internal motions on fast (sub-nanoseconds) and slow (microseconds to milliseconds) timescales as well as the overall rotational diffusion of the molecule (5–50 nanoseconds), whereas rates of magnetization transfer among protons with different chemical shifts and proton exchange report movements of protein domains on the very slow timescales (milliseconds to days). These features make NMR a unique and powerful tool in studying protein dynamics related to protein functions, and there has been a tremendous growth in these applications since the review by Lewis Kay in *Nature Structural Biology* in 1998. There is an impressive body of evidence indicating that the target binding sites of many proteins are flexible. NMR relaxation measurements are very useful in identifying which residues in a binding site are flexible. Significantly, these measurements are useful even when a high resolution X-ray structure is available, because crystal contacts may quench local motions. In some

instances flexibility at these sites may stabilize the unbound state, while in others it may be necessary for functions.

In NMR relaxation experiments, nuclear spin magnetization is excited by the application of electromagnetic radiofrequency fields and the return of the spin magnetization to thermal equilibrium is monitored using multidimensional NMR experiments. Commonly measured relaxation parameters are longitudinal relaxation times (T1), transverse relaxation times (T2), heteronuclear Overhauser effects (NOEs).

1.2.6 Structure determination by NMR

The chemical shift associated with a nucleus possessing magnetic spin may be altered by its magnetic environment. Bond and nearby nuclei will interact with each other and multiple-dimension NMR can be used to detect these interactions. Therefore, NMR could be a major method for structure determination of biomolecules in solution. The overall strategy for determining structure from 2-D NMR for proteins up to approximately 100 residues was developed by Kurt Wuthrich in the early 1980s. More sophisticated NMR experiments suitable for larger proteins based on this overall approach were developed later. In general, the strategy using NMR to solve structure is mainly composed by 4 steps.

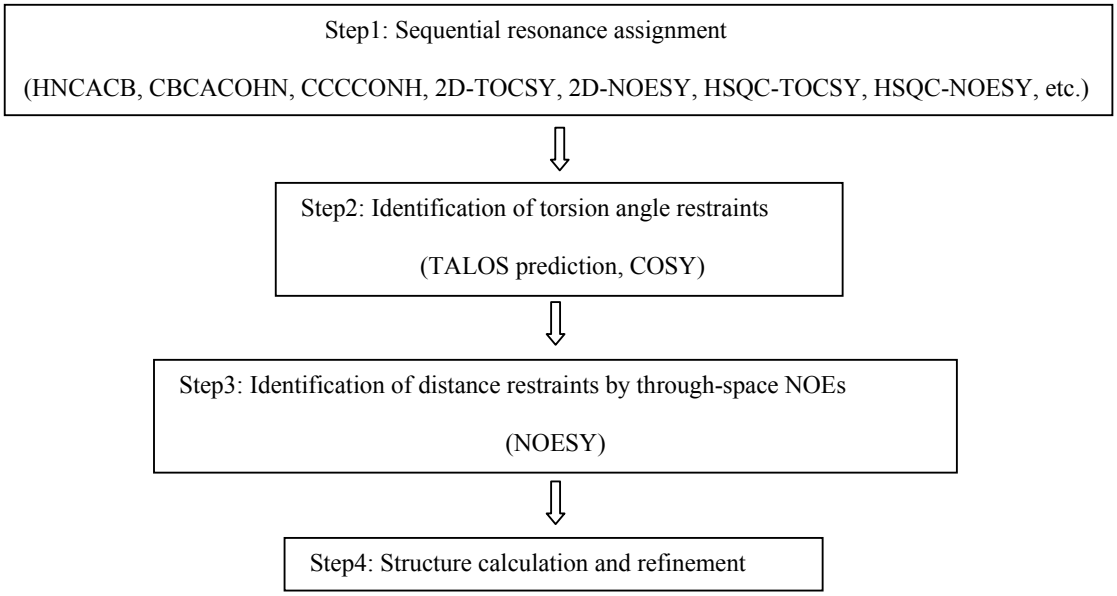


Figure1.11 Flow chart of structure determination

1.2.6.1 Assignment (backbone and side chain) and restraints (distance, dihedral angle)

The most simple and straightforward method of backbone resonance assignment involves the use of 15N, 13C labelled protein and the measurement of CBCACONH and HNCACB spectra. Large proteins give worse NMR spectra, because they tumble more slowly. For this reason the above two experiments of larger proteins (> 150 residues) are often not of sufficient quality to be able to carry out a full assignment. In this case a good option is the use of HNCA, HN(CO)CA, HNCO and HN(CA)CO spectra.

Standard triple resonance backbone assignment of proteins is based on the CBCANNH and CBCA(CO)NNHspectra. The idea is that the CBCANNH correlates each NH group with the Cα and Cβ chemical shifts of its own residue

(strongly) and of the residue preceding (weakly). The CBCA(CO)NNH only correlates the NH group to the preceding Cα and Cβ chemical shifts. The Figure below shows how this can be used to link one NH group to the next into a long chain.

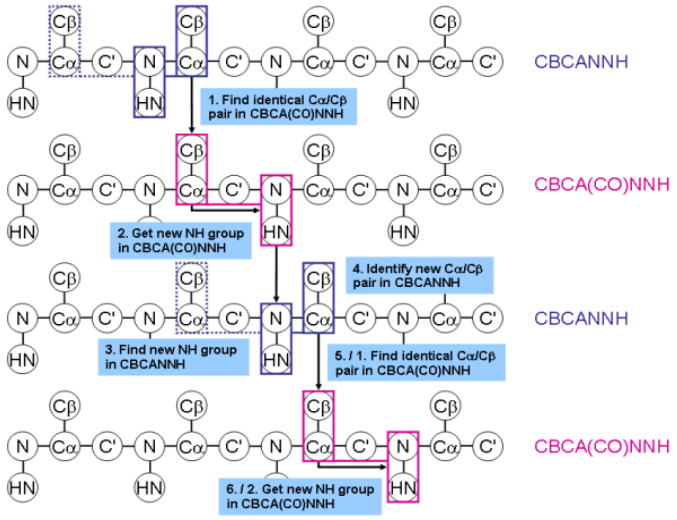


Figure 1.12 Sequential assignment by CBCACONH, HNCACB

In practise, using the CBCANH and CBCA(CO)NH spectra this looks like this (Cαs are shown in dark blue, Cβs in light blue):

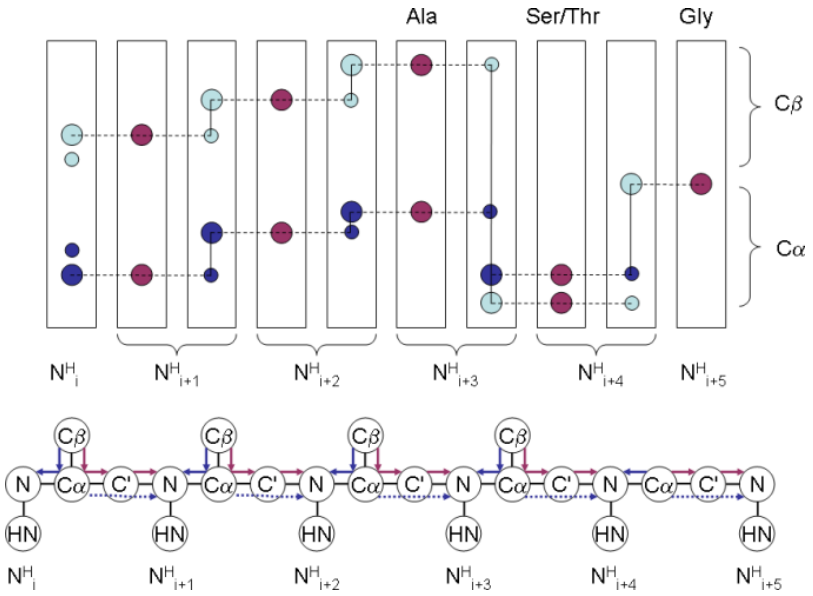


Figure 1.13 Sequential assignment by CBCACONH, HNCACB

After the sequential assignment, the chemical shift of backbone amine, hydrogen, and Cα, Cβ can be collected. Next, we need to do the assignment of side chain. A straight forward method is to begin with a set of HBHA(CO)ONH, HCC(CO)NNH and CC(CO)NNH spectra. These will provide the hydrogen and carbon side-chain chemical shifts for the residue preceding each NH group. For longer side chains not all peaks may neccessarily be visible, so that this may not be sufficient. Therefore, another spectra HCCH-TOCSY is quite useful to identify more chemical shifts because the excellent sensitivity. The HCCH-TOCSY will at any one carbon position show in one dimension the chemical shift of the hydrogen which is attached to the carbon and in another the other hydrogens belonging to that side chain.

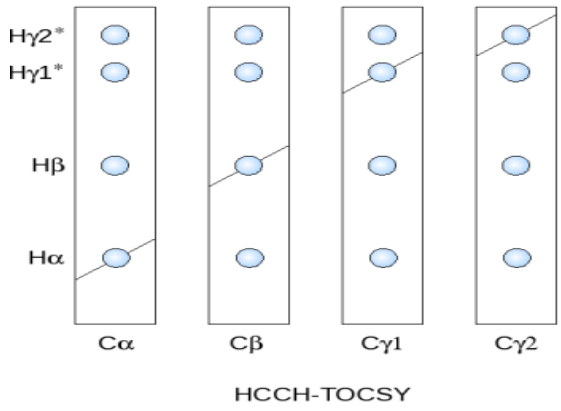


Figure 1.14 Side chain assignment by HCCH-TOCSY

For smaller proteins, it is possible to do the backbone assignment using just ^{15}N-labelled protein. The spectra used for this are the ^{15}N-NOESY-HSQC and the ^{15}N-TOCSY-HSQC. The ^{15}N-NOESY-HSQC will show for each NH group all ^{1}H resonances which are within about 5-7Å of the NH hydrogen. Assignment is done on the assumption that the two neighbouring NH groups are always visible. Thus two NH

groups can be linked because they each have an NOE to the other NH group. Meanwhile, the side chain chemical shifts can also be read from these spectra.

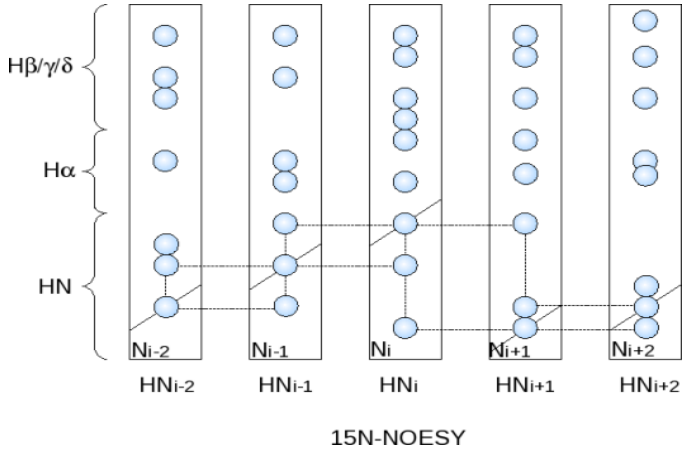


Figure 1.15 NOE assignment by 15N-NOESY

After the assignment of all the isotope atoms and protons, the more important and troublesome part is the assignment of NOE generated by neighbour atoms. Current assignment could be used to generate chemical shift index. The chemical shift index will give clue of the secondary structure of the protein, as shown in figure 1.8, which will be very helpful in the assignment of NOE. NOE pattern was shown in figure 1.10.

After finishing all the assignments, the distance range between protons could be determined from the van der Waal's radii—lower bonds, and the NOE intensity—upper bonds. According to NOE intensity, these NOEs can be categorized into "weak", "middle", "strong" and the corresponding distance are 1.8Å, 2.5 Å, and 5.0 Å.

Dihedral angle restraints are usually calculated by TALOS, which is a free software calculates protein backbone torsion angle by inference from measured Cα and Cβ chemical shifts.

1.2.6.2 Structure calculation and evaluation

There are many procedures for protein tertiary structure determination from NMR data. All these approaches aim to sample conformational space whilst at the same time satisfying a set of constraints. One available approach is model building, which uses the existing structure to determine the structure in question. A very different approach is probabilistic in nature and represents each atom by mean atomic position and variance about this position.

An alternative method is adopted by distance geometry, which is used in my research work and be illustrated in detail here. This method builds structure from internal distances. If a starting structure exists, restrained molecular dynamics can be used to refine it. A random folded starting structure is generated by empirical data, such as bond lengths and bond angles, together with amino acid sequence. Each known parameter will be given potential energy, and it will give a minimal energy if the calculated distance or angle coincides with input values. After each simulation step, the energy potential is recalculated for the new atomic positions and a further step follows. The step is iterated; search the energy hyperspace for a global minimum. After simulation steps at high temperature, the temperature (atomic velocities) is slowly reduced in many steps. The force constants in experimental constraints are raised simultaneously in each step. 20 or more different starting structures with random folds are used to generate a family of structures.

Once an ensemble of structures has been generated, it is necessary to assess them. The first question need to be addressed is how well the resulting structures satisfy the initial set of constraints. Usually a violation greater than 1Å is an indication of a serious problem and causes very high target function. The constraints causing huge

violation should be re-checked in NMR spectra or removed, until the target function decreased to an acceptable value. Verification of whether the backbone torsion angles of the structure lie in the allowed region of is also necessary. The RMS deviation of the ensemble from the esemble average is also commonly used to determine the "goodness" of the structures.

The flowchart of the calculation could be summarized as below:

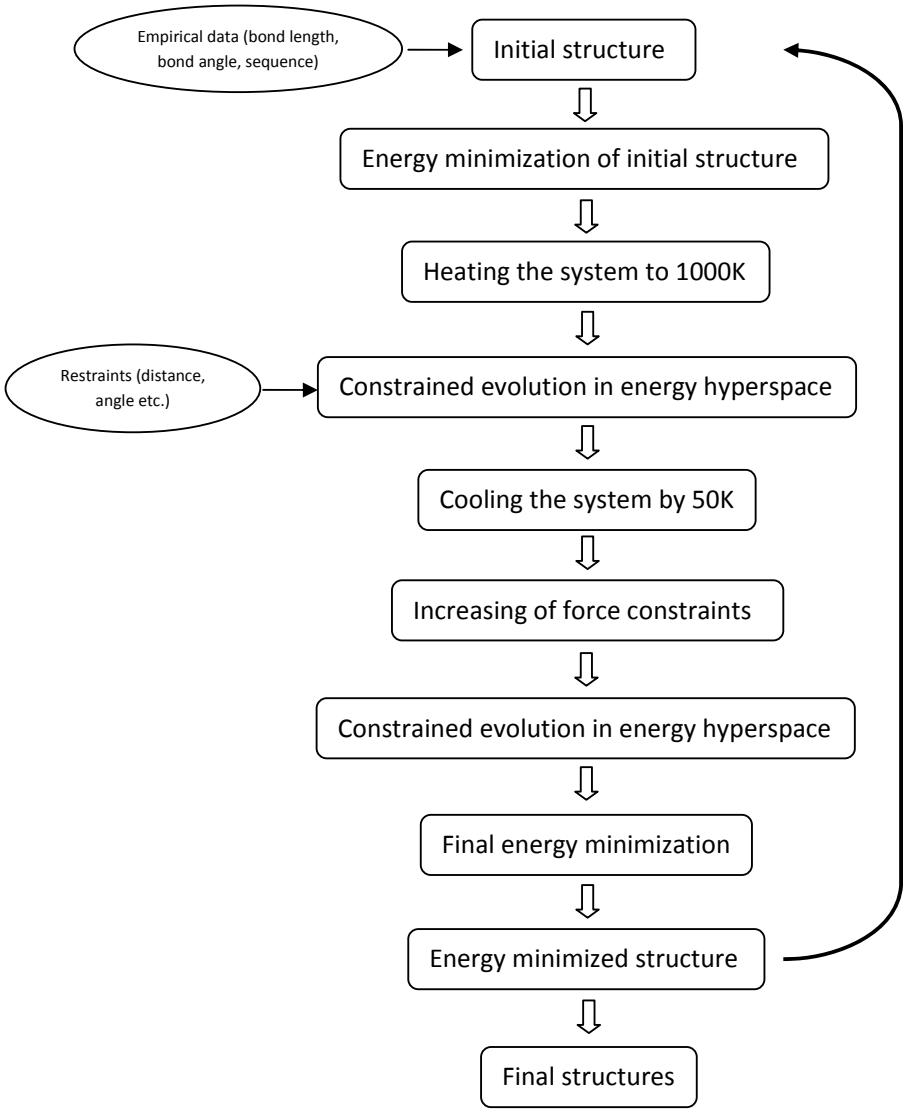


Figure 1.16 Flow chart of structure determination by software

1.3 Objectives and Contributions

Nogo-A is the most intensively studied neuron regeneration inhibitor, and EphA4 is newly discovered an active role in the pathway inhibiting neuron regeneration. No doubly, their structure and binding partner study will help us to understand the molecular mechanism of the interaction with their binding ligand, and provide information in the design of drug to block the inhibition pathway, promoting CNS neuron regeneration. Our research will aim at the structure study of Nogo-A/WWP1, EphA4 structures both in free and complex states, as well as dynamics study for EphA4 and its complexes. Our objectives will focus on the following aspects:

1.3.1 Structural illumination to Nogo-A and its newly discovered binding partner WWP1

Nogo-A is intensively studies due to its inhibitory ability to injured neuron regeneration. Moreover, its protein level is closely related with diseases such as stroke, as well as other neurodegenerative diseases such as ALS and Alzheimer's disease. Recently, our collaborators found that WWP1 interacted with Nogo-A and regulated its protein level, and their binding mechanism will be elucidated in this thesis by structural study. This work definitely is highly important and could provide clue or direction in drug design to Nogo-A protein level regulation.

1.3.2 Structure determination of EphA4 LBD and EphA4/ephrinB2 complex by X-ray

The huge signalling network composed by Eph receptors and ephrin ligands has attracted intensive investigation in recent years. EphA4 is a receptor capable of interacting with ephrins of both classes to generate a diverse spectrum of biological

activities. However, there is no structural study on free EphA4 molecule and its complexes with ephrin ligands. This thesis will focus on the structure determination of EphA4 and its complex with ephrinB2 ligands by X-ray. This work will be very important in understanding the structure and binding mechanism of Eph receptor and ephrin ligands.

1.3.3 Structural study of the complexes between EphA4 LBD and its small organic antagonists

Although some peptides, proteins, and small organic compounds have been identified to interrupt interactions between Eph receptors and ephrin ligands, they are all not suitable for further drug development due to biodegradability, poor specificity etc.. Therefore, the design of small antagonists targeting high affinity binding channel of Eph receptors has attracted intensive attention. Two isomeric 2,5-dimethylpyrrolyl benzoic acid derivatives were identified that can selectively inhibit ephrin binding to EphA4 and EphA2 as well as the functions of these receptors in live cells. This is the first time that people identified organic compounds selectively bind to high affinity binding channel of Eph receptor. In order to improve their specificity and binding affinity, structural insight into the binding interactions between EphA4 and two molecules will be studied. Their binding details will be elucidated by various biophysical methods. Our work will be a very important starting point for drug development targeting high affinity binding channel of Eph receptors.

1.3.4 NMR structure of EphA4 LBD and its dynamics study

With the accumulation of protein structures, people realized that protein is flexible in solution. Moreover, recently, evidence is accumulated to suggest that protein dynamics may play a critical role in the biological functions including signal

transmission. However, most investigators have not included protein dynamics study in their work. Unlike X-ray structure, most of which are rigid, NMR solution structure shows more information of protein dynamics. Therefore, NMR structure and the dynamics of EphA4 LBD will be studied in this thesis. The dynamics study of proteins will definitely gives more information about the protein. In addition, it may help to understand how protein dynamics affect EphA4 signalling regulation.

In order to achieve the objectives described above, protein crystallography and macromolecular NMR will be applied in this study. Some biophysics methods, including ITC, CD, and Biocore, will also be needed. Other techniques, including recombinant protein construction, expression and purification are also required. In the following chapter, these methods will be described in detail.

Chapter II MATERIALS AND METHODS

2.1 Vector construction

The construct of Flag-WWP1 was kindly provided by Pu *et al* (IMCB, Singapore). The DNA fragments corresponding to the respective WW domains were amplified from WWP1 construct. The DNA fragment of Nogo-A peptide was amplified from our vector containing Nogo-A. All the Eph/ephrin DNA fragment were amplified by PCR from human embryonic kidney cell cDNA library (293T). All the fragments and vectors were double digested by BamH1 and Xho1 (New England Biolabs) under 37^{o}C for 3 hours. The digested fragments were purified with QIAquick Gel Extraction Kit (Qiagen). Ligation was carried out by mixing digested vector and fragment with ratio 1:5, keeping at 16^{o}C overnight. Eph/ephrins and fWW fragments were ligated with pET32a, and all the other peptide were ligated with pGEX4T1. Ligation product was transformed into *E. coli.* DH5α. The colony was amplified at 37 ^{o}C overnight, and subsequently extracted plasmid with QIAprep Spin Miniprep Kit 250 (Qiagen). After confirmed by double digestion or PCR, the DNA sequence of positive plasmid was further confirmed by automatic DNA sequencing (ABI). EphA4 mutants were prepared according to protocol of Stratagene Quick-change site mutagenesis kit, and confirmed by DNA sequencing.

2.2 Protein expression and purification

Plasmids of WW domains, Nogo-A peptide were transformed into *E. coli.* BL21 cells. Constructs of Eph receptor and ephrins ligands were transformed into Rosetta Gammi strain to prevent proteins were expressed as inclusion bodies. The cells were cultivated under 37℃ until the OD_{600} value reached 0.6, and IPTG was subsequently added into the broth to reach a final concentration of 0.3 mM to induce protein

expression for 12 hours at 20ºC. Cells were harvested by centrifugation at 5000rpm and lysed by sonication in PBS buffer. The recombinant GST fusion proteins were purified by affinity chromatography with Glutathione Sepharose 4B beads (Pharmacia Biotech) under native condition. The proteins fused with his-tag were purified with Ni-NTA matrices (QIAGEN) under native (Eph and ephrins) or denature (fWW domain) condition. The peptides and proteins were released from beads by thrombin cleavage overnight under room temperature. WW domains and Nogo-A peptide were further purified by HPLC and Eph receptors and ephrins ligands were further purified by FPLC.

To isotope label proteins required for heteronuclear NMR experiments, recombinant proteins were prepared with a similar protocol except for growing the cells in M9 medium with addition of $(^{15}NH_4)_2SO_4$ for ^{15}N-labeling; and $(^{15}NH_4)_2SO_4/^{13}C$-glucose for ^{15}N-/^{13}C- double labeling.

2.3 NMR sample preparation, NMR structure determination, relaxation experiments and data analysis

Sample preparation

NMR samples proteins were prepared by dissolving proteins and peptides in 5 mM phosphate buffer at pH 6.2 with 0.02% (W/V) sodium azine. Eph and ephrin samples were kept in 10mM phosphate buffer at pH 6.3 with 0.02% (W/V) sodium azide.

NMR experiments to structure determination

All NMR experiments were acquired on an 800 MHz Bruker Avance spectrometer equipped with pulse field gradient units at 25 ºC. The NMR spectra acquired for both

backbone and side-chain assignments included ^{1}H-^{15}N HSQC, ^{15}N-edited HSQC-TOCSY, HSQC-NOESY, as well as tripleresonance experiments HNCACB, CBCA(CO)NH, HCCH-TOCSY, and (CCC)CONH. NOE restraints for structure calculation were derived from ^{15}N- and ^{13}C-NOESY spectra. The NOE constraints involved in the aromatic side chains were collected from two-dimensional ^{1}H NOESY spectra in D_2O. For more restraints, two dimensional ^{1}H TOCSY, NOESY in D_2O as well as DQF-COSY in H_2O could be acquired to extract NOE distance and dihedral angle constraints. NMR data was processed by NMRpipe (Delaglio, F. *et al*, 1995) and analyzed by NMRview (Johnson, B. A. *et al*, 1994).

Determination of the NMR structure

In the WW4 structure determination, a set of unambiguous NOE restraints extracted from three-dimensional ^{15}N HSQC-NOESY and two-dimensional ^{1}H NOESY spectra were input to calculate initial structures by CYANA program together with dihedral angle restraints derived from $^{3}J\alpha_{N\text{-}NH}$ coupling constants from a two-dimensional DQF-COSY spectrum. With the availability of the initial structure, more NOE cross-peaks in the two NOESY spectra were assigned with assistance of CYANA followed by a manual confirmation. The 10 structures with lowest target function values accepted by CYANA were checked by PROCHECK and subsequently analyzed by using MolMol and Pymol (http://www.pymol.org).

In the EphA4 structure determination, NOE restraints were extracted from three-dimensional ^{15}N HSQC-NOESY, ^{13}C edited NOESY. Dihedral angel restraints were predicted by TALOS. Hydrogen bonds information was extracted by adding D_2O in freeze dried EphA4 and a set of HSQC spectra were collected after 7min, 2hours, 4hours, and 24hours. The initial structure was calculated by CYANA program, and

more NOEs were confirmed and extracted by taking the initial structure as reference. The 10 structures with lowest target function values accepted by CYANA were checked by PROCHECK and subsequently analyzed by using MolMol and Pymol.

Relaxation data collection and analysis

^{15}N T1, T1ρ relaxation times, {^{1}H}-^{15}N steady state NOE intensities were collected on an 800 MHz Bruker Avance spectrometer equipped with both an actively shielded cryoprobe and pulse field gradient units (Fan *et al*., 2003; Ran *et al*., 2008). Relaxation times T1 were determined by collecting 8 points with delays of 10, 280, 700, 1000, 1100, 1250, and 1400 ms using a recycle delay of 1 s. Relaxation times T1ρ were measured by collecting seven points with delays of 1, 10, 20, 30, 40, 45, and 52 ms using a spin-lock power of 1.6 kHz, a 2.5-s recycle delay. {^{1}H}-^{15}N steady-state NOEs were obtained by recording spectra with and without ^{1}H presaturation of a duration of 3 s plus a relaxation delay of 6 s at 800 MHz.

Relaxation times were fitted as single exponential decays to peak height data. Spin-spin relaxation time T2 was calculated from T1ρ and T1 according to equation:

$$\frac{1}{T1\rho} = \frac{\sin\theta \times \sin\theta}{T1} + \frac{\cos\theta \times \cos\theta}{T2}$$

Where $\theta=\text{atan}(\Delta\omega/\omega_1)$ and $\Delta\omega$, ω_1 are the resonance offset and spin-lock field strength, respectively (Fan *et al*., 2003) . NMR relaxation data were analyzed by "Model-Free" formulism with protein dynamics software suites (Fushman *et al*., 1997). According to Abragam, 1961, relaxation of protonated heteronuclei is dominated by the dipolar interaction with the directly attached ^{1}H spin and by the chemical shift anisotropy mechanism. Relaxation parameters are given by:

$$R_1 = {d^2}/{4}\,[J(\omega_H - \omega_X) + 3J(\omega_X) + 6J(\omega_H + \omega_X)] + c^2 J(\omega_X)$$

$$R_2 = \left({d^2}/{8}\right)[4J(0) + J(\omega_H - \omega_X) + 3J(\omega_X) + 6(J\omega_H) + 6J(\omega_H + \omega_X)]$$

$$+({c^2}/{6})[4J(0) + 3J(\omega_X)] + R_{ex}$$

$$NOE = 1 + (d^2/4R_1)(\gamma_X/\gamma_H)[6J(\omega_H + \omega_X) - J(\omega_H - \omega_X)]$$

In which $d = \mu_0 h\gamma_X\gamma_H < r_{XH}^{-3} >/8\pi^2$, $c = \omega_X\Delta\sigma/\sqrt{3}$, μ_0 is the permeability of free space; h is Planck's constant; γ_X γ_H are the gyromagnetic ratios of ^{1}H and the X spin (X=^{13}C or ^{15}N) respectively; γ_{XH} is the X-H bond length; ω_H and ω_X are the Larmor frequencies of ^{1}H and X spins, respectively; and Δσ is the chemical shift anisotropy of the X spin.

The Model-Free formalism, as described by Lipari and Szabo, and extended by Clore and co-workers, determines the amplitudes and time scales of the intramolecular motions by modeling the spectral density function, $J(\omega)$, as

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1 + (\omega\tau_m)^2} + \frac{({S_f}^2 - S^2)\tau}{1 + (\omega\tau)^2}\right]$$

$$= \frac{2}{5}{S_f}^2\left[\frac{{S_s}^2\tau_m}{1 + (\omega\tau_m)^2} + \frac{(1 - {S_s}^2)\tau}{1 + (\omega\tau)^2}\right]$$

In which, $\tau = \tau_s\tau_m/(\tau_s + \tau_m)$, τ_m is the isotropic rotational correlation time of the molecule, τ_s is the effective correlation tile for internal motions, $S^2 = {S_f}^2{S_s}^2$ is the square of the generalized order parameter characterizing the amplitude of the internal motions, and ${S_f}^2$ and ${S_s}^2$ are the squares of the order parameters for the internal

motions on the fast and slow time scales, respectively. Generalized order parameters represent motions that are described by dynamics on the ns-ps time scale, with values ranging from zero for isotropic internal motions to unity for completely restricted motion in a molecular reference frame.

In order to allow for diverse protein dynamics, several forms of the spectral density function, based on various models of the local motion, were utilized. Including the original Lipari&Szabo approach, assuming fast local motion characterized by the parameters S^2 and τ_{loc}; extended model-free treatment, including both fast ($S_{fast}{}^2, \tau_{fast}$) and slow ($S_{slow}{}^2, \tau_{slow}$)reorientations for the NH bond ($\tau_{fast} \ll \tau_{slow} < \tau_c$); and could also allow for slow, milli- to microsecond dynamics resulting in a conformational exchange contribution, R_{ex}, to the linewidth. The first software used to do modelfree analysis was developed by Palmer A. G., but in this paper, analysis of relaxation data was performed by using software DYNAMICS (Fushman *et al*., 1997; Hall and Fushman, 2003).

The overall rotational diffusion tensors as well as total correlation time (τc) of the EphA4 LBD was determined by ROTDIF (Walker *et al*., 2004; Fushman *et al*., 1997). Isotropic, axially-symmetric and fully anisotropic models for the overall motion were used and then compared. According to the illustration of ROTDIF, fully anisotropic model was finally selected because of smallest Ch^2/df value.

2.4 Crystallization, data collection, and structure determination

The EphA4 ligand-binding domain was prepared at a concentration of 12 mg/ml and crystallized by setting up 2 μl hanging drops at room temperature in well

containing the reservoir solution (18% PEG 4000, 11% isopropanol and 0.1 M Hepes at pH 7.5). Rock-like crystals formed after 4 days and dehydration of the crystals was subsequently performed by moving the coverslips to a new well containing dehydration buffer (18% PEG 4000, 11% isopropanol, 10% glycol and 0.1 M Hepes at pH 7.5).

The X-ray diffraction images for a single crystal were collected by using an in-house Rigaku/MSC FR-E X-ray generator with an R-AXIS IV++ imaging plate detector at the Biopolis shared-equipment facility. The crystal was protected by the cryoprotectant. The data were indexed and scaled using the program d*Trek. After an all-space-group search, the crystals were identified as belonging to the space group P1 with 8 EphA4 molecules per asymmetric unit. The Matthews coefficient was calculated as 2.98 with 58.78% solvent constant and 2.53 with 51.32% solvent constant respectively by CCP4 software package (Collaborative Computational Project, No. 4, 1994)

The structure was determined through the molecular replacement with the search model of the EphA4 ligand-binding domain generated by using our previously-determined free EphA4 structure (3CKH) (Qin *et al*, 2008). This model was completed by manual fitting using the program COOT. The refinement was carried out by program Refmac (Collaborative computational project, No. 4. 1994. The final structure was analyzed by PROCHECK (Laskowski, R. A. *et al*, 1993). All the figures were prepared using the Pymol molecular graphics system (W. L. DeLano, DeLano Scientific LLC, San Carlos, CA).

The EphA4-ephrin-B2 complex was prepared at a concentration of 10 mg/ml and crystallized by setting up 2 μL hanging drops at room temperature in a well containing

the reservoir solution (23.5% PEG 4000, and 0.1 M Tris, 0.2 M MgCl2, pH 7.5). Rock-like crystals formed after 7 days. X-ray diffraction images for a single crystal were collected by using an in-house Bruker X8 Proteum X-ray generator with a CCD detector. The crystal was protected by cryoprotectant (23.5% PEG 4000, and 0.1 M Tris, 0.2 M MgCl2, pH 7.5). The data were indexed and scaled using the program HKL2000 (Otwinowski & Minor, 1997). After an all-space-group search, the crystal was identified as belonging to the space group P2(1) with a=54.651, b=48.711, and c=64.469 with one complex molecule per asymmetric unit. The Matthews coefficient was 2.22 with 44.51% solvent constant.

The model of the EphA4 LBD was generated by the program Phaser (McCoy *et al*, 2005) from the CCP4 suite (Collaborative Computational Project, 23, 1994) using our previously-determined free EphA4 structure (3CKH) as a search model through the molecular replacement method (Qin *et al*, 2008). After determining the EphA4 structure, EphA4 was fixed and the model of ephrin-B2 was generated by the CCP4 suite using a previously determined ephrin-B2 structure (1KYG) as search model. The two models were subsequently combined as search model to find an initial model of the EphA4-ephrin-B2 complex using the program molrep. This model was completed by manual fitting using the program COOT (Emsley and Cowtan, 2004) and refined using the program CNS (Brunger *et al*, 1998) for many iterations. The final structure was analyzed by PROCHECK (Laskowski *et al*, 1993) and the details of the data collection and refinement statistics are shown in Supplementary Table S1. All the figures were prepared using the PyMOL molecular graphics system (W. L. DeLano, DeLano Scientific LLC, San Carlos, CA).

2.5 CD experiments and sample preparation

CD samples were prepared by dissolving proteins and peptides in 5 mM phosphate buffer at pH 6.2 with 0.02% (W/V) sodium azine. Eph and ephrin samples were kept in 10mM phosphate buffer at pH 6.3 with 0.02% (W/V) sodium azine. CD experiments were performed on a Jasco J-810 spectropolarimeter equipped with a thermal controller. The far-UV CD spectra of the WW domains and Eph/ephrin were collected at a protein concentration of ~20 μM at 25ºC, using 1 mm path length cuvette with a 0.1 nm spectral resolution. Data from three independent scans were added and averaged. To assess the stability of WW domains, thermal unfolding was carried out. CD signal at 230nm was monitored as the temperature increased from 5 ºC to 95 ºC. The secondary structure fraction was analyzed by use of CONTINLL program (lamar.colostate.edu/~sreeram/CDPro/ReadMe.htm).

2.6 Isothermal Titration Calorimetry and NMR titration

ITC titration to study protein interaction

All isothermal titration calorimetric (ITC) experiments were performed using a Microcal VP ITC machine. Protein was placed in a 1.8 ml sample cell, while ligands were loaded into a 300 μL syringe. A control experiment with the same parameter setting was also performed to subtract the contribution of the ligand dilution. The titration data after subtracting the results of the control experiment were fitted using the built-in software ORIGIN to obtain thermodynamic binding parameters.

NMR characterization of binding

To characterize binding interactions between proteins and ligands, two dimensional ^{1}H-^{15}N HSQC spectra of the ^{15}N-labeled proteins were acquired at a protein concentration of ~100 μM in the absence or presence of the ligands at different molar ratios. By superimposing the HSQC spectra of the ^{15}N-labeled free protein and mixture of protein and ligand, the shifted HSQC peaks could be identified. The degree of binding-induced perturbation was represented by an integrated chemical shift index calculated by the formula $[(\Delta^{1}H)^{2} + (\Delta^{15}N/4)^{2}]^{1/2}$ ppm (Liu J. *et al*, 2006).

2.7 Docking and modelling

Molecular docking of WW4 and Nogo-A (650-666) was performed by using software Haddock2.0 (High Ambiguity Driven protein-protein Docking, http://www.nmr.chem.uu.nl/haddock/), which makes use of chemical-shift perturbation data to derive the docking while allowing various degrees of flexibility. The docking procedure was performed by three steps. First, randomization and rigid body energy minimization; Second, semi-flexible simulated annealing; Third, flexible explicit solvent refinement.

According to Haddock definition, the solution accessible residues which has larger chemical shift perturbation value (larger than 2 times average value) were set as active residues, and all the residues close to the active residues, and solvent accessible were set as passive residues.

The PDB file of Nogo-A (650-666) was generated by CNS from its sequence. In order to achieve reasonable docking result, all the residues that did not interact with WW4 were removed.

1000 structures were generated during the rigid body docking, and the best 100 structures were selected for semi-flexible simulated annealing. The best 20 structures were selected for further refinement in explicit water. The overall HADDOCK score used at the various stages to rank and select solutions was calculated as a weighted sum of different terms (rigid body stage: 0.01×Evdw+1.0×Eelec+0.01×EAIR-0.01×BSA+1.0×Edesolv; semiflexible refinement: 1.0×Evdw+1.0×Eelec+0.1×EAIR-0.01×BSA+1.0×Edesolv; water-refinement: 1.0×Evdw + 0.2×Eelec +0.1×EAIR+1.0×Edesolv). Here the subscripts are: vdw the van der Waals energy, Elec the electrostatic energy, BSA the buried surface area, AIR the AIR energy

Table1.2: Active residues and flexible regions used in docking of WW domain and nogo-A polyproline peptide

Segid	Active residues	Fully flexible regions
A(WW4)	W9-I11, T28-T30	N1-L5, P34-S39
B(Nogo-A(650-666))	E652-E661	E652-E661

Docking solutions were ranked based on the average HADDOCK score calculated from top solutions. The best 10 solutions were provided as HADDOCK complex structure model.

The models of the EphA4 ligand-binding domains in complex with two antagonistic molecules were constructed by use of the HADDOCK 1.3 in combination with CNS, which is the previous version of HADDOCK2.0. The docking procedure was also performed by three steps: first, randomization and rigid body energy minimization; second, semi-flexible simulated annealing; third, flexible explicit solvent refinement.

To conduct the docking, several invisible residues over the loop regions were added to the EphA4 crystal structures by COOT (Emsley, P. *et al*, 2004) and then the obtained structures were subjected to several rounds of energy minimization by PHENIX (Adams, P.D. *et al*, 2002). Subsequently, hydrogen atoms were added to the structures by use of the CNS protocol. On the other hand, the geometric coordinates and parameters for the two small molecules were generated and energy-minimized by the on-line PRODRG server (http://davapc1.bioch.dundee.ac.uk/prodrg/).

All EphA4 residues with a chemical shift perturbation greater than the threshold value of 0.08 (2.5 times of the standard deviation) were set to be "active" residues while neighbors of active residues were defined as ''passive residues'' according to HADDOCK definition. These active residues included Gln43 on the E β-strand, Ile31-Met32 and Ile39 on the D-E loop, and Asp123 and Ile131-Gly132 on the J-K loop. Furthermore, all residues with heteronuclear NOE intensities of less than 0.7 were found to be located on the N- and C- termini, or on the loops, and thus set to be "fully-flexible" during the molecular docking. One thousand structures were generated during the rigid body docking, and the best 50 structures were selected for semi-flexible simulated annealing, followed by water refinement. Three structures with the lowest energies were selected for detailed analysis and display.

Chapter III RESULTS AND DISCUSSION

3.1 WWP1 and Nogo-A Interaction

3.1.1 Identification of WWP1 as a novel binding partner for Nogo-A

GST fused Nogo-A specific fragment (567-774) was used as bait in a phage display screening. The screening was performed on a human brain cDNA library displayed by a T7 phage vector for five rounds of affinity selection. The enrichment after each round of selection was analyzed by performing PCR using T7 primers. A single PCR product was clearly emerging after 3rd round of selection. The phages display of this specific protein become dominant in the 4th and 5th selection. Sequencing of this PCR product revealed that this cDNA encoded a WWP1 fragment containing residues 252–388. Examination of WWP1 fragment from phage display shows that this fragment contains WW domain, which has been extensively documented to bind to proline riched consensus sequence with the motif PPxY. Nogo-A fragment contains PPPY motif (656-659), thus, we believed that Nogo-A fragment interacts with WWP1 through WW domain via PPPY motif. The mutation of Y→A totally abolished the binding between WWP1 and Nogo-A, which confirmed our hypothesis. WWP1 contains four WW domains, as well as a long loop from residue 346 to 531, as shown in figure 3.1. To investigate the binding mechanism between WWP1 and Nogo-A, we prepared the whole WW domain fragment (346-531), called fWW protein, and four isolated WW domain (WW1, WW2, WW3 and WW4 corresponding to WWP1 residues 346-385, 380-417, 453-491, and 494-531 respectively), and proline rich motif on Nogo-A (650-666).

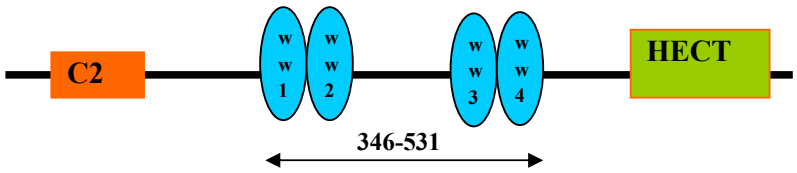


Figure 3.1 Schematic representation of the modular structure of WWP1 protein

3.1.2 Preliminary CD and NMR characterization

The secondary structure of fWW protein was first assessed by far-UV CD spectroscopy, as seen in Figure3.2a, the fragment had a positive CD signal at ~230nm, but negative signal at ~ 190nm, indicating that the fWW protein consists of the β-turn secondary structure, together with some loops. Secondary structure prediction in Table 3.1 also confirms this observation. We have prepared 15N isotope-labeled fWW protein and subsequently acquired its 1H-15N HSQC spectrum. The fWW protein had a relatively dispersed HSQC spectrum, with some peaks even located at ~9.5 ppm (Figure 3.2b). However, only ~80 resonance peaks could be detected in the spectrum, much fewer than its number of residues (186). In particular, of these peaks, ~40 had very strong intensities while the rest had much weaker intensities. This observation implies that the strong HSQC peaks may result from the ~40-residue linker region between the second and third WW domains which might be highly unstructured. Nevertheless, as seen in Figure 3.2b, upon addition of the Nogo-A(650-666) peptide, many HSQC peaks of the fWW protein shifted significantly, indicating that it was able to interact with Nogo-A. Unfortunately, due to the very limited manifestation of the HSQC peaks as well as a strong tendency to aggregate at a high concentration, detailed NMR study is not feasible for the fWW protein. However, as presented in panels c and d of Figure 3.2 and Table 3.2, the thermodynamic parameters for the binding of fWW to the Nogo-A peptide were successfully obtained by using the ITC titration. Interestingly, the fWW protein containing all four WW domains is capable of binding to Nogo-A with a dissociation constant Kd of 1.68 μM.

The structural properties of the four isolated WW domains were also assessed by far-UV CD spectroscopy. As shown in Figure 3.2a, all four WW domains had a positive CD signal at ~230 and a negative one at ~208 nm, indicating that all of them

possessed β-turn/sheet secondary structures, in complete agreement with the classic three-stranded β-fold conserved in all WW domains.

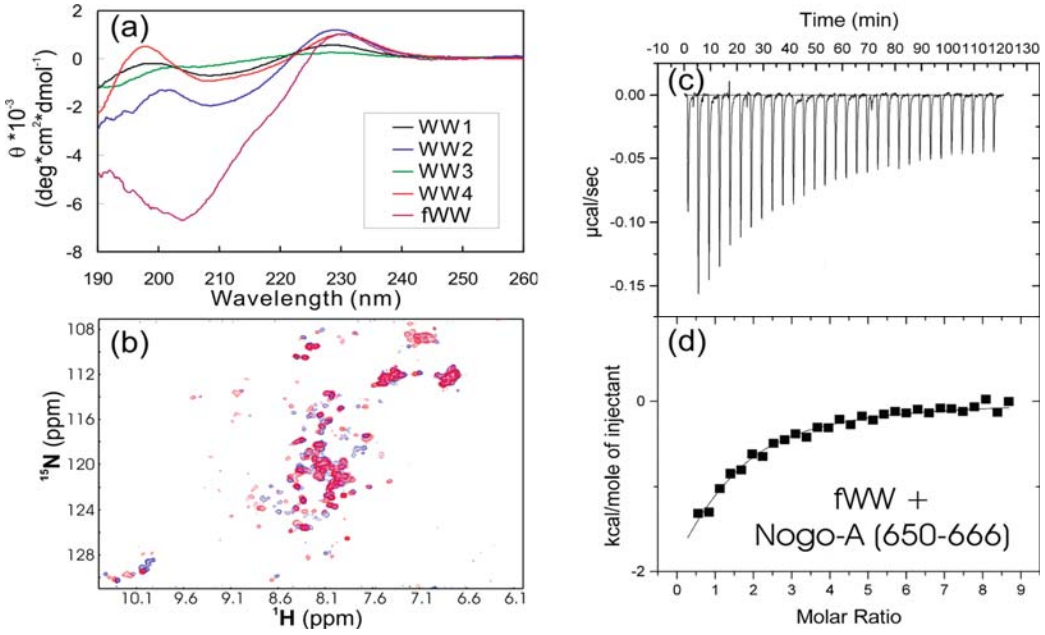


Figure 3.2 Structural and binding characterization of the fWW protein. (a) Far-UV CD spectra of fWW and four isolated WW domains at protein concentrations of 20 μM in 5 mM phosphate buffer (pH 6.2) at 25 °C. (b) Superimposition of the HSQC spectra of the 15Nlabeled fWW protein in the absence (blue) and presence (red) of Nogo-A(650-666) at a molar ratio of 1:4 (fWW:peptide). Spectra were acquired in 5 mM phosphate buffer (pH 6.2) at 25^{o}C on an 800 MHz Bruker NMR spectrometer. (c) ITC titration profiles of the binding reaction of the fWW protein with Nogo-A (650-666). (d) Integrated values for reaction heats with subtraction of the corresponding blank results normalized by the amount of ligand injected vs the Nogo-A:fWW molar ratio. The detailed conditions and setting of the ITC experiments are presented in Materials and Methods as well as Table 3.2.

To visualize structural properties of four isolated WW domains by NMR spectroscopy, all proteins were ^{15}N-isotope-labelled and purified by HPLC. Subsequently, two-dimensional ^{1}H-^{15}N HSQC spectra were collected on an 800 MHz NMR spectrometer. Interestingly, based on their HSQC spectra it appeared that four WW domains owned very different structural features. WW1 and WW2 were only partially-folded which had narrowly-dispersed HSQC spectra with only about half of the resonance peaks detectable (Figure 3.3a-3.3b). On the other hand, as seen in

Figure 3.3d, WW4 was the best-folded with uniform HSQC peak intensities in a well-dispersed spectrum (~3 ppm for ^{1}H and 26 ppm for ^{15}N dimensions). With regard to WW3, although it was folded as evident from its well-dispersed HSQC spectrum (Figure 3.3c), its peak intensities were not uniform, indicating that some regions might undergo conformational exchanges on μs-ms time scale.

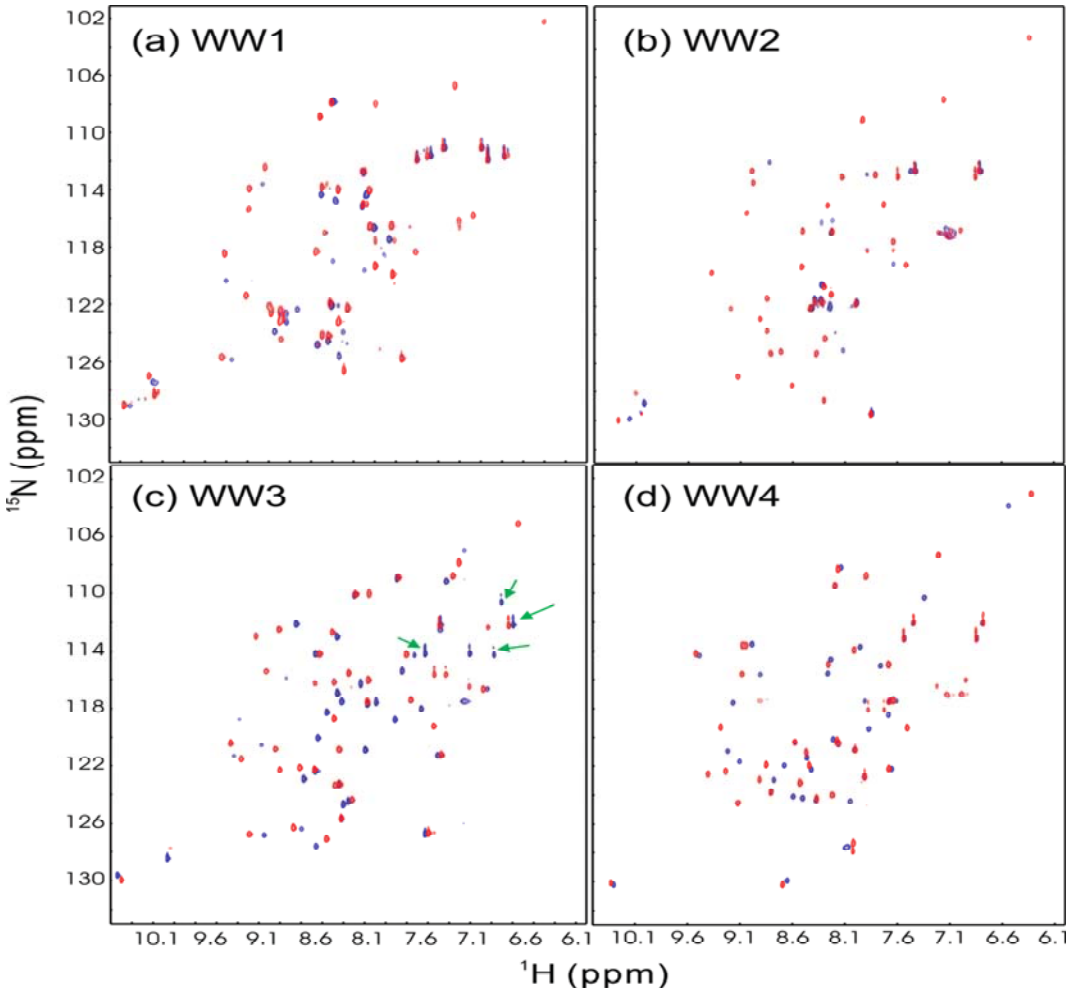


Figure 3.3 Binding of four ^{15}N-labeled WW domains with Nogo-A (650-666). Superimposition of the HSQC spectra of ^{15}N-labeled WW1(a), WW2(b), WW3(c), and WW4(d) in the absence (blue) and presence (red) of Nogo-A(650-666) at a molar ratio of 1:4 (WW:peptide). All spectra were recorded in a 5 mMphosphate buffer (pH 6.2) at 25 °C on an 800 MHz Bruker NMR spectrometer. The doublets of some HSQC peaks of the complexed WW3 domain are indicated.

The binding interactions between WW domains and the Nogo-A (650-666) peptide were assessed by NMR HSQC titrations. Very interestingly, although as shown above, the fWW protein and four WW domains had very-different HSQC spectra, addition of Nogo-A (650-666) induced dramatic HSQC-peak shifts for all of them (Figure 3.3), indicating that all of them were able to bind the Nogo-A (650-666) peptide. In particular, for WW1 and WW2 domains, the introduction of Nogo-A (650-666) resulted in manifestation of HSQC peaks for almost all non-proline residues, clearly indicating that upon binding, the WW1 and WW2 domains underwent significant conformational changes from partially-folded to folded states. The addition of Nogo-A to fWW protein did not induce dramatic chemical shift changes to the residues with strong intensity. This might because the residues with strong intensity may locate at random coil regions, which had weak interaction with the peptide.

3.1.3 ITC measurements of binding parameters

To quantitatively characterize binding interactions between WW domains and Nogo-A (650-666), isothermal calorimetric titrations were conducted to measure their thermodynamic binding constants. The raw titration data were shown in Figure 3.4 while binding parameters obtained by data-fitting were presented in Table 3.2. Very strikingly, although fWW protein and four WW domains possessed very differential structural characteristics, they held very similar affinities to the Nogo-A (650-666) peptide, with dissociation constants ranging from ~1.0 to 4 μM, which could be ranked within a high affinity category among previously-documented WW-ligand interactions[70]. From the ITC profile of fWW protein (figure 3.2C), the four WW domains seemed bind to Nogo-A peptide as a whole protein without binding

preference to different WW domains. This phenomenon may due to the similar dissociation constants of four WW domains. They cooperate with each other than compete to each other in signalling transduction.

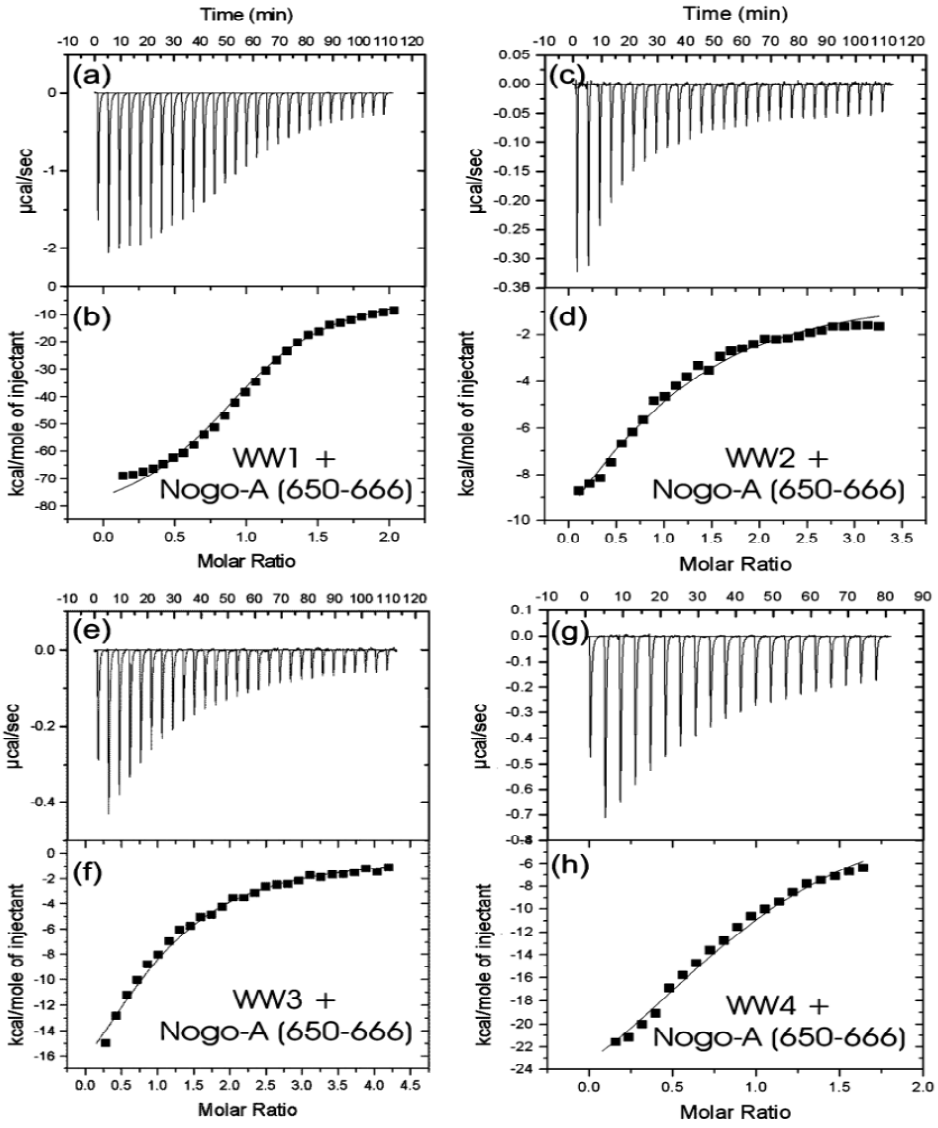


Figure 3.4 ITC characterization. ITC titration profiles of the binding reactions of the WW1 (a), WW2 (c), WW3 (e), and WW4 (g) domains with Nogo-A(650-666). Integrated values for reaction heats with subtraction of the corresponding blank results normalized by the amount of ligand injected vs the Nogo-A:WW1 (b), Nogo-A:WW2 (d), Nogo-A:WW3 (f), and Nogo-A:WW4 (h) molar ratio. The detailed conditions and settings of the ITC experiments are presented in Materials and Methods and Table 3.2.

3.1.4 NMR characterization of binding interactions

The detailed binding of Nogo-A (650-666) to WW domains was characterized by monitoring the HSQC peak shifts of the peptide induced by adding different WW domains. Interestingly, four WW domains induced very similar shift patterns for HSQC peaks of Nogo-A (650-666), indicating that they interacted with Nogo-A in a similar manner. As exemplified in Figure 3.5, upon adding WW4 domain, the HSQC peaks of the characteristic residue Tyr659 with its neighbouring residue Glu660 totally disappeared, indicating the binding would induce dramatic conformational exchanges over this region on the μs-ms time scale. It is worthwhile to note that residues Glu652, Glu654 and Asn655 N-terminal to the PPPY motif also underwent significant shifts, indicating that these residues might also play an important role in mediating the binding affinity and specificity.

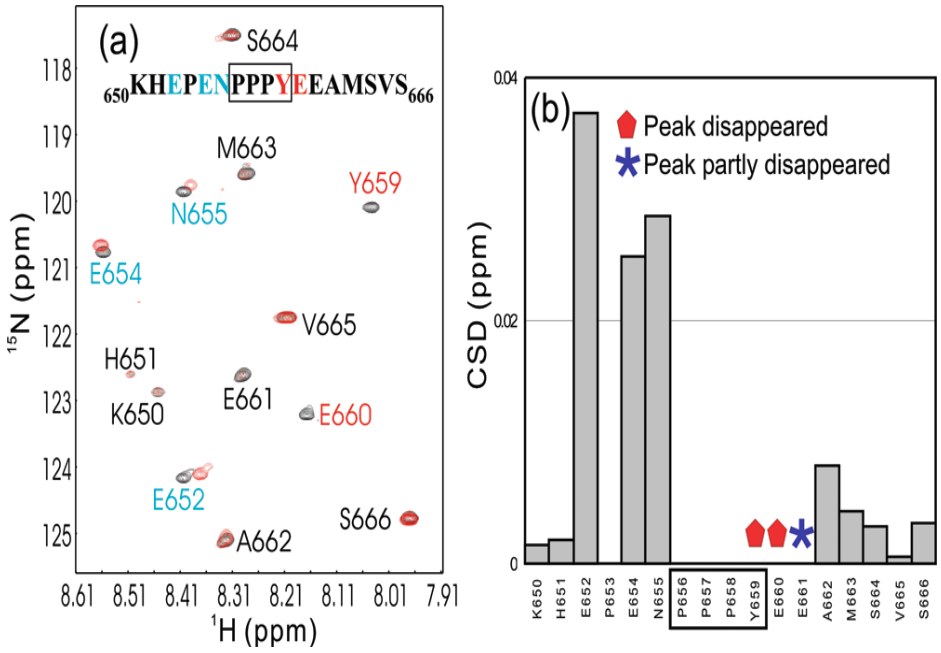


Figure 3.5 Binding of ^{15}N-labeled Nogo-A(650-666) with the WW4 domain. (a) Superimposition of the HSQC spectra of the 15N-labeled Nogo-A(650-666) peptide in the absence (black) and presence (red) of the WW4 domain at a 1:4 peptide:WW4 molar ratio. All spectra were recorded in a 5 mM phosphate buffer (pH 6.2) at 25^{o}C on an 800 MHz NMR spectrometer. The sequential assignments of Nogo-A(650-666) were made by analyzing heteronuclear HSQC-TOCSY and HSQC-NOESY spectra. The sequence of Nogo-A(650-666) was included, and the two residues (Tyr659 and Glu660) with HSQC peaks that completely disappeared upon binding are labeled in red. Residues Glu652, Glu654, and Asn655 with significant peak shifts upon binding are labeled in cyan. (b) Residue-specific chemical shift difference (CSD) of Nogo-A (650-666) induced by binding to the WW4 domain. The bar values were calculated from the shifts observed in Figure 5a using the formula described in Materials and Methods. The characteristic WW domain binding PPPY motif in Nogo-A (650-666) is boxed.

3.1.5 Three dimensional structure and binding interface of the WW4 domain

It will be more straightforward if we determine the 3D structure of fWW protein and Nogo-A peptide. However, from the above NMR and CD results, fWW protein was identified to possess poor structure, making the structure determination infeasible. Moreover, out of four WW domains, WW1, WW2 and WW3 had extensive conformational exchanges on μs-ms time scale, thus not suitable for determination of three-dimensional structure. On the other hand, we also attempted to determine the WW structure complexed with Nogo-A (650-666) but failed because the binding induced severe conformational exchanges on μs-ms time scale as exemplified in Figure 3.3. Therefore, in this study we decided to determine the solution structure of WW4 and subsequently derived its binding interface with Nogo-A (650-666) by NMR HSQC titration. Figure 3.6a showed the ^{1}H-^{15}N HSQC spectra of the WW4 domain in the absence and in the presence of Nogo-A (650-666) at a ratio of 1:4 (WW4/peptide) while Figure 3.6b presented the residue-specific shifts of the WW4 HSQC peaks upon binding to the peptide. Interestingly, the significantly-perturbed residues (with shift index >0.2) were located over two regions, one over Trp9-Glu10-Ile11 and another over Thr28-Thr29-Thr30. This result is overall in agreement with previous observations that two discrete WW regions were mainly responsible for binding to the ligand, one centralized around the first Trp residue and another around the second Trp residue. However, it is very interesting to note that in WW4, the most-perturbed residues were centralized around Thr28-Thr29-Thr30 but not Phe31 as usually observed.

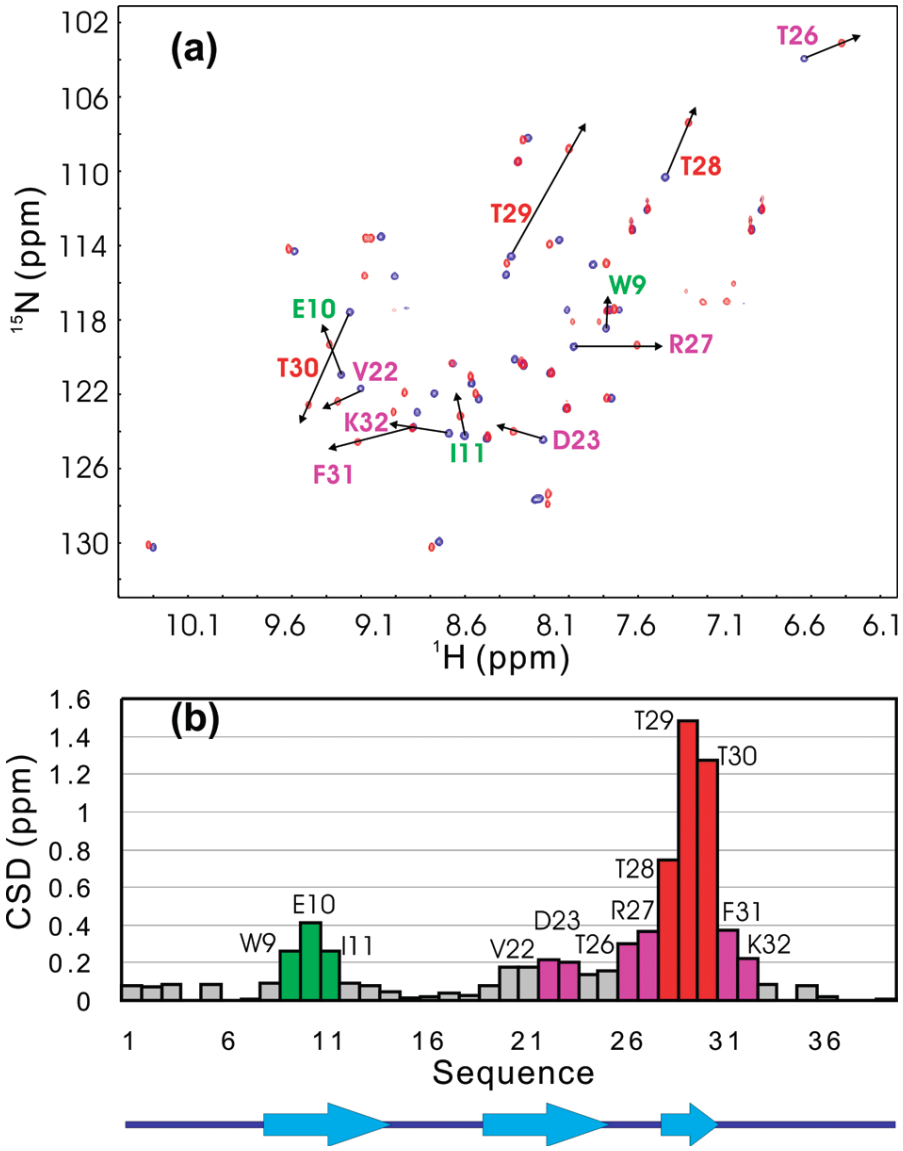


Figure3.6 Binding of the 15N-labeled WW4 domain with Nogo-A (650-666). (a) Superimposition of the HSQC spectra of the 15Nlabeled WW4 domain in the absence (blue) and presence (red) of Nogo-A (650-666) at a 1:4 WW4:peptide molar ratio. All spectra were acquired in 5 mM phosphate buffer (pH 6.2) at 25^{o}C on an 800 MHz NMR spectrometer. The sequential assignments of WW4 were made by analyzing heteronuclear HSQC-TOCSY and HSQC-NOESY spectra. The residues with significant peak shifts upon binding are labelled in the spectra. (b) Residue-specific chemical shift difference (CSD) of WW4 induced by binding to Nogo-A (650-666). The bar values were calculated from the shifts observed in Figure 3.6A using the formula described in Materials and Methods. Red was used for coloring residues with changes larger than 0.6, pink for residues with changes larger than 0.2 but smaller than 0.6 if on the third β-strand, and green for residues on the first β-strand.

By using experimental NMR distance and dihedral angle constraints, the three-dimensional structure of the WW4 domain was determined. Figure 3.7a shows the 10 solution structures with the lowest target functions while Table 3.3 presents their structural statistics. As seen in Figure 3.7a, the WW4 domain adopts an antiparallel three-stranded β-sheet fold common to all WW domains[94-96], with the first β-strand over residues Gly8-Tyr13, second over Arg19-His24 and third from Thr28-Phe31 respectively. Interestingly, in the WWP1 WW4 domain there is a short C-terminal 3-10 helix formed over residues Pro34-Asn36 which is not conserved in the classic WW fold. In the 10 WW4 solution structures, the secondary structure regions are well-defined, with rms deviations of 1.37 Å for all atoms, 1.12 Å for the heavy atoms and 0.25 Å for the backbone atoms respectively.

Based on the HSQC titration data for both WW4 domain and Nogo-A peptide, we have constructed the model of the WW4 domain in complex with the Nogo-A peptide by the well-established HADDOCK 2.0 program with the best structural solution shown in Figures 3.7b and 3.7c. As seen in Figures 3.7a and 3.7b, the Nogo-A peptide binds to the WWP1 WW4 domain in a manner very similar to those previously reported for other WW-PPxY complexes. Briefly, the major WW4 residues in contact with the Nogo-A peptides include two aromatic residues Tyr20 and Phe31, Val22, His24 and Thr29 which are highly conserved and thus thought to be common residues critical for the ligand binding. On the other hand, the residue Glu10 located in the first β-strand also has a close contact with Tyr659 sidechain. Interestingly, in WW domains such as from Yap and Rsp proteins, the position corresponding to Glu10 is occupied by the negatively-charged Glu residue. By contrast, in some other WW domains from Ykb2, Db10 and Yfx1 proteins, this position resides in a positively-charged Lys residue. As seen in Figure 3.6, upon binding to Nogo-A, Glu10 in the

WW4 domain is also significantly perturbed, implying that it may play an important role in the binding of the WWP1 WW4 domain to Nogo-A. The result that the mutation of Tyr659 completely abolished the binding ability of the Nogo-A peptide strongly highlights the critical role of Tyr659 in binding with the WWP1 WW4 domain. Indeed, as seen in Figure 3.7c, the Tyr659 sidechain deeply inserts into a pocket formed by the residues Glu10 and His24, thus explaining why the WW4 residue Glu10 was significantly perturbed by binding to the Nogo-A peptide (Figure 6). Furthermore, in the complex model, the Nogo-A residue Pro656 has a close contact with the WW4 residue Thr29, thus rationalizing the observation that upon binding to the Nogo-A peptide, the residues centralizing Thr28-Thr29-Thr30 were significantly perturbed (Figure 3.6).

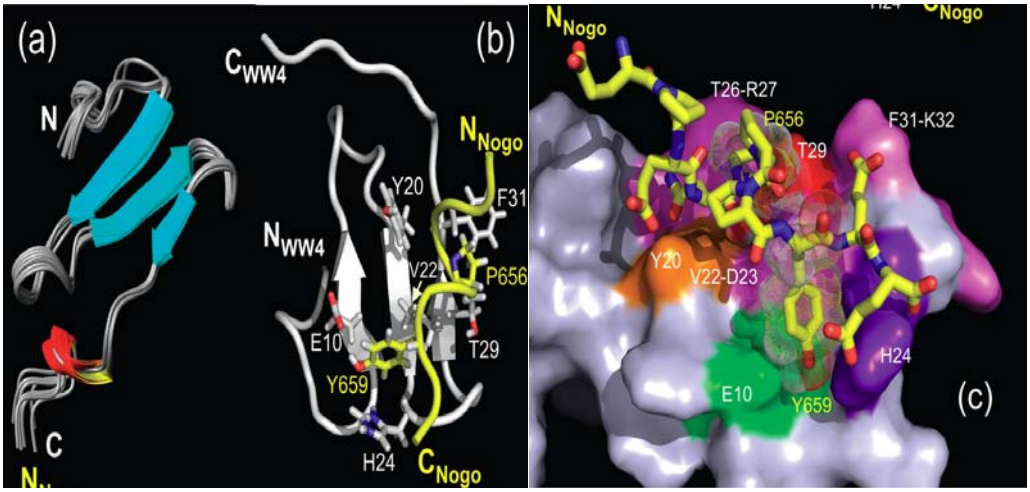


Figure 3.7 Structures of the free and complexed WW4 domains. (a) Superimposition of the 10 selected NMR structures of the WW4 domain having the lowest target functions in ribbon mode. (b) Best docking solution of the WW4 domain (gray) in complex with the Nogo-A peptide (yellow). The side chains of the binding-important residues are displayed as sticks and labeled for both the WW4 domain and Nogo-A peptide. (c) Surface representation of the WW4 domain in complex with the Nogo-A peptide in stick mode. The binding-perturbed WW4 residues are colored as in Figure 6b. Additional residue His24 is colored purple and Tyr20 brown.

3.1.6 Discussion

Nogo-A is a multifunctional protein which has been implicated in a variety of diverse and important biological processes, including inhibition of neural regeneration (Schwab ME *et al*, 1988), participation in ER shaping (Voeltz G.K. *et al*, 2006), and stabilization of neuromuscular junction (Jokic N. *et al*, 2006). In particular, changes in the intracellular Nogo-A protein level have been directly associated with stroke (Li S. *et al*, 2006) as well as other neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS) (Jokic N. *et al*, 2005) and Alzheimer's disease (Gil V. *et al*, 2006). These observations clearly indicate that the Nogo-A protein level represents a critical factor in regulation of its functions. However, so far how the Nogo-A protein level is regulated in cells remains unknown. On the other hand, Nogo-A is extensively believed to interact with many other proteins, and consequently, identification of such novel binding partners holds key implications in understanding the functional roles of Nogo-A as well as designing molecules of therapeutic interest.

In this study, we first successfully identified WWP1 to be a novel binding partner of Nogo-A. In particular, our results demonstrate that WWP1 forms a complex with the endogenous Nogo-A both *in vitro* and *in vivo*. WWP1 belongs to the E3 ubiquitin ligase family which is the critical component that provides specificity to the Ub system. More specifically WWP1 utilizes its WW modular domains to specifically bind to the PPxY motif in the substrate and consequently promotes ligation of ubiquitin to the substrate. Currently, we have obtained the *in vivo* results revealing that WWP1 indeed binds Nogo-A to promote Nogo-A ubiquitination, thus regulating the Nogo-A protein level. Given the critical role of the Nogo-A protein level in various human diseases, the Nogo-A-WWP1 interaction interface may represent a promising target for developing molecules of medical interest. So far, no structural

study has been reported for the WWP1 WW domains. Given the crucial role of the interaction between the WWP1 WW domains and PPxY motif in its substrates, we further conducted extensive investigations on all four WWP1 WW domains as well as their interactions with the Nogo-A(650-666) peptide carrying the only Nogo-A PPxY motif by using ITC, CD, and NMR spectroscopy. The results show that despite containing a large unstructured region in the 186-residue fWW protein consisting of all four WW domains, it is able to bind to the Nogo-A (650-666) peptide with a high affinity, with a *K*d of 1.68μM. On the other hand, surprisingly, the CD and NMR characterization reveals that four isolated WWP1 WW domains have differential structural properties. While the fourth WW domain is well-folded, the first and second WW domains undergo extensive conformational exchanges on the microsecond to millisecond time scale, likely due to a slight unfolding or/and a slight dynamic aggregation, which causes many HSQC peaks to be undetectable. Nevertheless, as revealed by ITC measurements, all four WWP1 WW domains are capable of binding to Nogo-A(650-666) with similar affinities, with *K*d values ranging from 1.03 to 3.85 μM, which can be ranked within a high-affinity category among previously documented WW-ligand interactions. Very interestingly, as shown by NMR characterization, the partially folded WW1 and WW2 domains suddenly undergo significant conformational transitions to become well-structured upon binding to Nogo-A (650-666). This phenomenon is rarely observed for the native WW domains, and previously, only one engineered Trp17-to-Phe mutant of the YAP WW domain was found to be partially unfolded in the free state but became structured upon binding to the PPxY ligand. Further structure determination of the best-folded WW4 domain by NMR spectroscopy shows that it adopts a common WW fold with an antiparallel three-stranded β-sheet, with an additional C-terminal 3_{10}-helix. Detailed

HSQC titrations indicate that the Nogo-A binding interface of the WW4 domain is mostly constituted by residues centralized over the Thr28-Thr29-Thr30 sequence on the third β-strand and over the Trp9-Glu10-Ile11 sequence on the first strand. Furthermore, the complex model generated from molecular docking suggests that the Nogo-A peptide binds to the WW4 domain in a mode very similar to those previously reported. More importantly, the complex model rationalizes the HSQC titration results as well as the mutation result that Tyr665 of Nogo-A (650-666) is absolutely indispensable for the binding to the WW4 domain. Recently, we have demonstrated that the whole 1016-residue N-terminus of human Nogo-A was intrinsically unstructured. Therefore, *in vivo* this WW-binding motif of Nogo-A might be already accessible to WWP1 without needing to be further unfolded.

In conclusion, we have identified WWP1, an E3 ubiquitin ligase, as a novel binding partner for Nogo-A which may regulate the Nogo-A protein level *in vivo*. With consideration that the protein level of Nogo-A is a key factor regulating its functions, and in particular correlated with human diseases, the interaction interface between the WWP1 WW domains and the Nogo-A PPPY motif might represent a promising target for design of molecules of significant medical interest. In this regard, this study also provides rationales as well as a template Nogo-A (650-666) for further drug design.

Table3.1 Secondary Structure Fraction Prediction Based on far-UV CD Spectrum

	fWW	WW1	WW2	WW3	WW4
α-Helix	0.059	0.004	0.004	0.034	0.047
β-strand	0.106	0.427	0.430	0.413	0.433
β-turn	0.466	0.225	0.212	0.321	0.167
Random coil	0.369	0.345	0.355	0.361	0.357

Table 3.2: Thermodynamic Parameters of the Binding Interactions between four WWP1 WW Domains and Nogo-A (650-666) as Measured by Isothermal Titration Calorimetry (ITC)

Syringe	Cell	Buffer	Volumn of each injection (μL)	K_a (M^{-1})	K_d (μM)	Stoichiometry (n)	ΔS (cal/mol)	ΔH (cal/mol)
Nogo-A (650-666) (0.2mM)	WW1 (0.01mM)	Phosphate (5mM, pH 6.2)	5	$9.70\times10^5 \pm 4.5\times10^4$	1.03	1.07 ± 0	-222	$-7.434\times10^4 \pm 555$
Nogo-A (650-666) (0.16mM)	WW2 (0.005mM)	Phosphate (5mM, pH 6.2)	5	$2.60\times10^5 \pm 1.6\times10^4$	3.85	0.9 ± 0	-29.3	$-1.613\times10^4 \pm 383$
Nogo-A (650-666) (0.16mM)	WW3 (0.005mM)	Phosphate (5mM, pH 6.2)	5	$5.41\times10^5 \pm 6.5\times10^4$	1.85	0.95 ± 0	-54.2	$-2.399\times10^4 \pm 809$
Nogo-A (650-666) (0.16mM)	WW4 (0.007mM)	Phosphate (5mM, pH 6.2)	5	$6.68\times10^5 \pm 2.7\times10^4$	1.50	0.942 ± 0.0076	-57.7	$-2.514\times10^4 \pm 331$
Nogo-A (650-666) (0.5mM)	fWW (0.005mM)	Phosphate (5mM, pH 6.2)	5	$1.30\times10^5 \pm 0.9\times10^4$	7.7	1.000 ± 0	8.763	$-0.436\times10^4 \pm 123.5$
Nogo-A (650-666) (0.5mM)	fWW (0.005mM)	Phosphate (5mM, pH 6.2)	5	$5.938\times10^5 \pm 6.823\times10^4$	1.684	4.000 ± 0	23.08	-994 ± 17.78

Table 3.3 Structural statistics for 10 selected NMR structures of the WW4 domain

Experimental constraints for structure calculation	
NOE restraints	
Total	263
Intra residue	119
Sequential	49
Medium	17
Long-range	78
Dihedral angle constraints	54
CYANA target function	0.63 ± 0.035
Distance violations (>0.20 Å)	0
Dihedral angle violations (>5º)	0
Ramachandran statistics (%)	
Most favored:	71.0
Additionally allowed:	22.6
Generously allowed:	6.5
Disallowed:	0.0
Root mean square deviation (Å)	**Secondary structure regions (7-14,18-24,27-32)**
All atoms	1.28
Heavy atoms	1.04
Backbone atoms	0.20

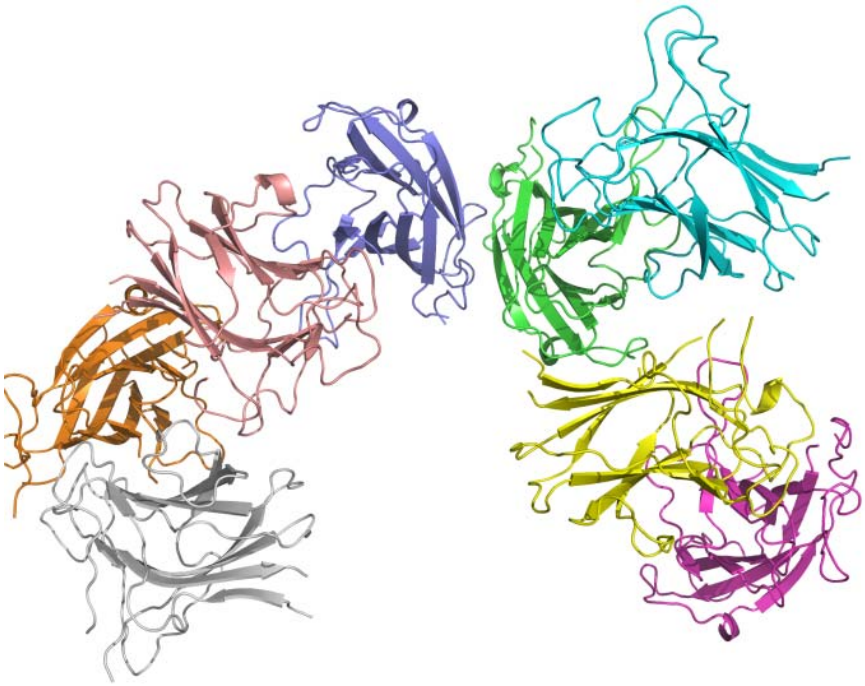
3.2 Sixteen Structures in Two Crystals Reflect the Highly Dynamic Property of the Loops of EphA4 Ligand Binding Domain

Two *apo* EphA4 receptor crystals were obtained in our lab and their structures were determined at 2.4Å and 2.9 Å, respectively. What make these crystals unique is that there are 8 EphA4 molecules in one asymmetric unit (AU). Previous works of Eph receptors show that complex structure of EphB2 and ephrinB2 (Himenan *et al*, 2001), EphA2 and ephrinA1 (Himenan *et al*, 2009) will form complex clusters in one single asymmetric unit, but this kind of protein cluster has never been observed in free Eph receptors. Comparison of these structures showed that even in a single AU, with the presence of packing force, A-C, D-E, G-H, and J-K loops still could adopt various conformations, indicating they are highly flexible. More importantly, both open and closed conformations of J-K loop coexist in these structures, indicating that, many different conformations coexist over the loop regions, which may be ready for binding to different ligands with slight rearrangements. It also should be pointed out that J-K loop adopted open conformation only when G-H loop of neighbour EphA4 molecule was inserted into the classical binding channel, which may help us to understand more on induce and fit mechanism.

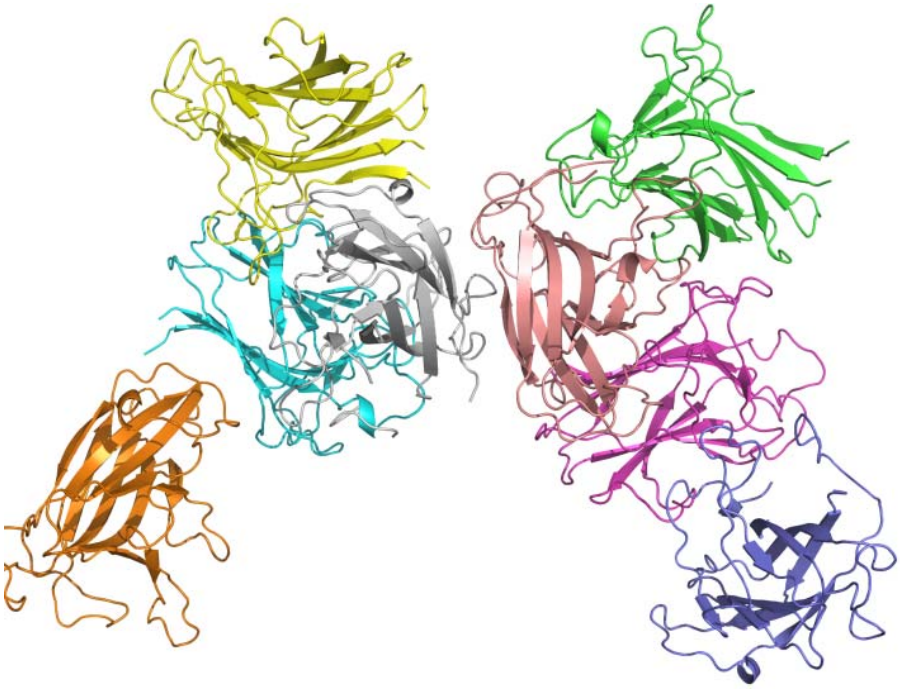
3.2.1 16 structures determined from two EphA4 LBD crystals

We refer these two crystals as crystal1 and crystal2. Crystal1 is determined with resolution 2.9 Å, and crystal2 is determined with resolution 2.4 Å. There are 8 molecules in one asymmetric unit (AU) in both crystals. Therefore, 16 EphA4 ligand binding domain structures were captured at the same time. The details of the data collection and refinement statistics are shown in Table3.4. The overall topology of these EphA4 structures is identical to previous Eph receptor structures. It is composed of 11 anti-parallel β-strands organized in jellyroll β-sandwich architecture, with several loops with various lengths. Most of the electron density map is complete with

side chain, but the map quality of D-E and J-K loops is relatively poor. In some regions, only the map of backbone can be traced, which are substituted as poly-A in the model built, while in some regions, the map is completely invisible. As shown in Figure 3.8, the arrangement of these EphA4 molecules is very complicated. In general, they are interacting with the molecules of the same AU or the molecules from neighbour AU through polar contacts or hydrophobic interactions. In crystal1, the polar contacts are between chain A and G, C and D, E and H, F and H. The hydrophobic contacts are between G-H loop and classical binding channel. G-H loop of chain D and F inserted into binding channel of chain C and H respectively. There are also contacts between the molecules of different AU. The classical binding channel of chain A and G were occupied by G-H loop from another AU. In crystal2, polar contacts mainly located at G-H, H-I, and part of J-K loops, as well as G and K strands. Chain A interact with chain C and E; chain B interact with chain D, H and F; chain C interact with G and E; chain D interact with chain B and F. All the molecules in crystal2 adopt closed conformation because their binding channels are left unoccupied. There are contacts between different AU, but still no interaction between G-H loop and classical binding channel.



Crystal 1



Crystal 2

Figure3.8 Pattern of EphA4 lignad binding domain clusters

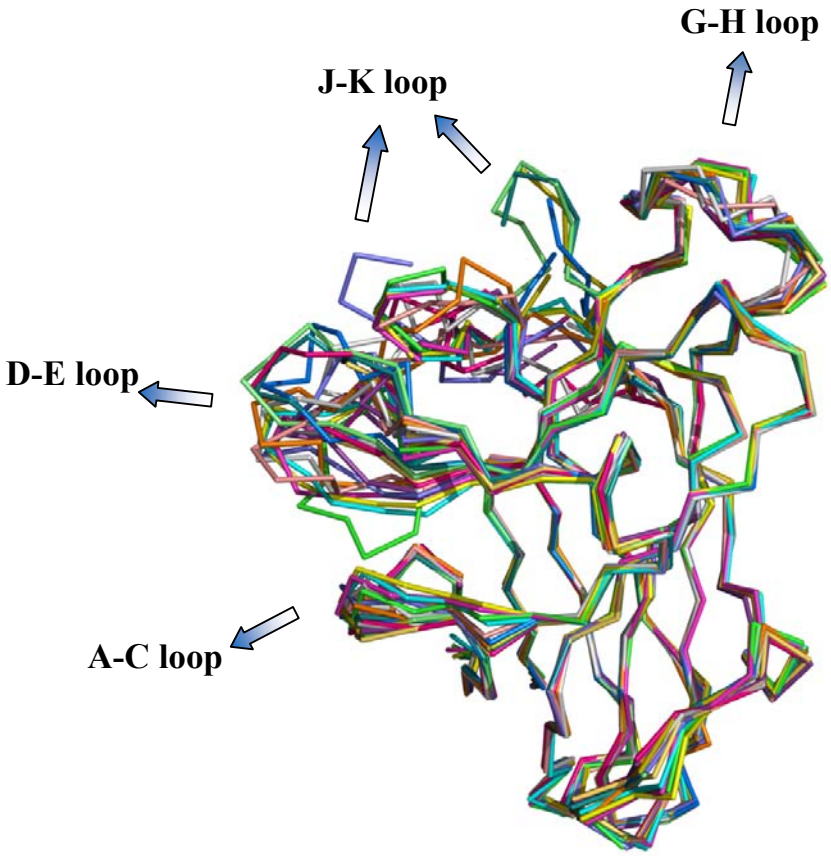


Figure3.9 Comparison between 16 EphA4 LBD structures. The loops are labelled in arrow. J-K loop took dramatic different conformations in different structures.

3.2.2 Comparison between 16 structures and previous EphA4 structures

Comparison of these structures shows that except for the loop regions, the regions with secondary structure are highly superimposed. The mean global backbone RMSD of these structures is 1.48±0.54Å, while mean global heavy atom RMSD is 1.94 ± 0.54 Å. When only secondary structure regions were considered, the mean global backbone RMSD is 0.23±0.07 Å, while mean global heavy atom RMSD is 0.42±0.11 Å. The loop regions adopted various conformations, showing their high flexibility.

The mean global backbone RMSD of these regions is 2.69±1.04 Å, and the mean global heavy atom RMSD is 3.61±1.13 Å (Figure3.9).

The previous released EphA4 ligands binding domain structures and the structures in this paper were compared in Figure3.10. Both free and complex states of EphA4 ligand binding domain were included. It can be observed that D-E and J-K loops have various conformations. Based on the analysis to the released EphA4 LBD structures, it could be concluded that J-K loop, the fragment from D123 toN139 (according to the sequence of our construct), adopts dramatic different conformations in free and bound states (Figure 3.10). Compared with these EphA4 LBD structures, the EphA4 structures we resolved here could be categorized into two categories. One is free and fully closed conformation, which is identical with the conformation of free EphA4 LBD dimer (3CKH, Qin *et al*, 2008; 2WO1, Bowden A.T. *et al*, 2009) (shown in red in Figure3.10). The other category has opened J-K loops, which is identical with the conformation in EphA4/ephrin complexes (3GXU, Qin *et al*, 2010; 2WO2, Bowden A.T. *et al*, 2009; 2WO3, Bowden A.T. *et al*, 2009) (shown in blue in Figure3.10). On the other hand, all D-E loops adopt closed conformation, which is identical with the conformations of free EphA4 LBD (shown in red in Figure3.10). When we examined the interaction between these EphA4 molecules, we found that some G-H loops inserted into the classical binding channel of neighbour molecules, thus leading to the closed to open states transition of J-K loops. All D-E loops taking closed conformation probably because the lacking of polar contacts on D, E strands of neighbour molecules. These polar contacts are present between A-C loop, D, E strands of EphA4 and C, G, and F strands of ephrinB2 ligands (Figure 3.11), as well as other Eph receptors and their ephrin ligands (Himanen *et al*, 2001; Chrencik J.E. *et al*, 2008; Bowden T.A. *et al*, 2009; Qin *et al*, 2009). It should be noticed that the

structure with PDB ID 3CKH has similar conformation. 3CKH has two EphA4 LBD in one AU, one is fully closed, the other one has opened J-K loop because the G-H loop of neighbour molecule is inserted. D-E loop also has closed conformation because of lacking polar contacts on D, E strands. Under each category, these loops show various conformations, indicating that these loops are highly dynamic. The electron density maps of open and closed conformations are shown in Figure3.12. These maps are clear enough to define the conformation of these loops.

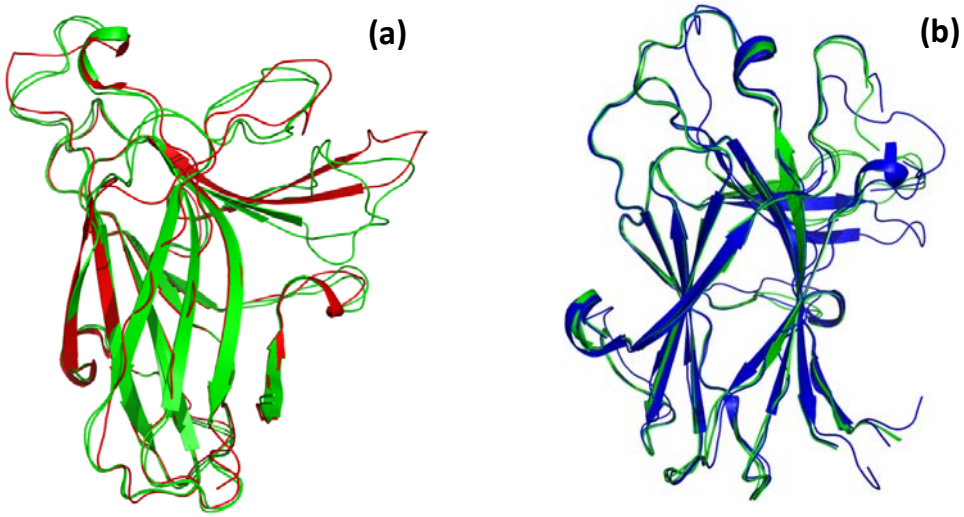


Figure3.10 16 EphA4 LBD structures presented here versus open and closed conformations in previous publications. In order to have a clear review, two structures were selected to compare in each conformation. (a), closed conformation, green: two EphA4 LBD structures in this thesis with open conformation; red: closed conformation in previous study (Chain B of 2WO1); (b), open conformation, green: two EphA4 LBD structures in this thesis with open conformation; blue: Open conformation in previous studies (2WO2 and 2WO3).

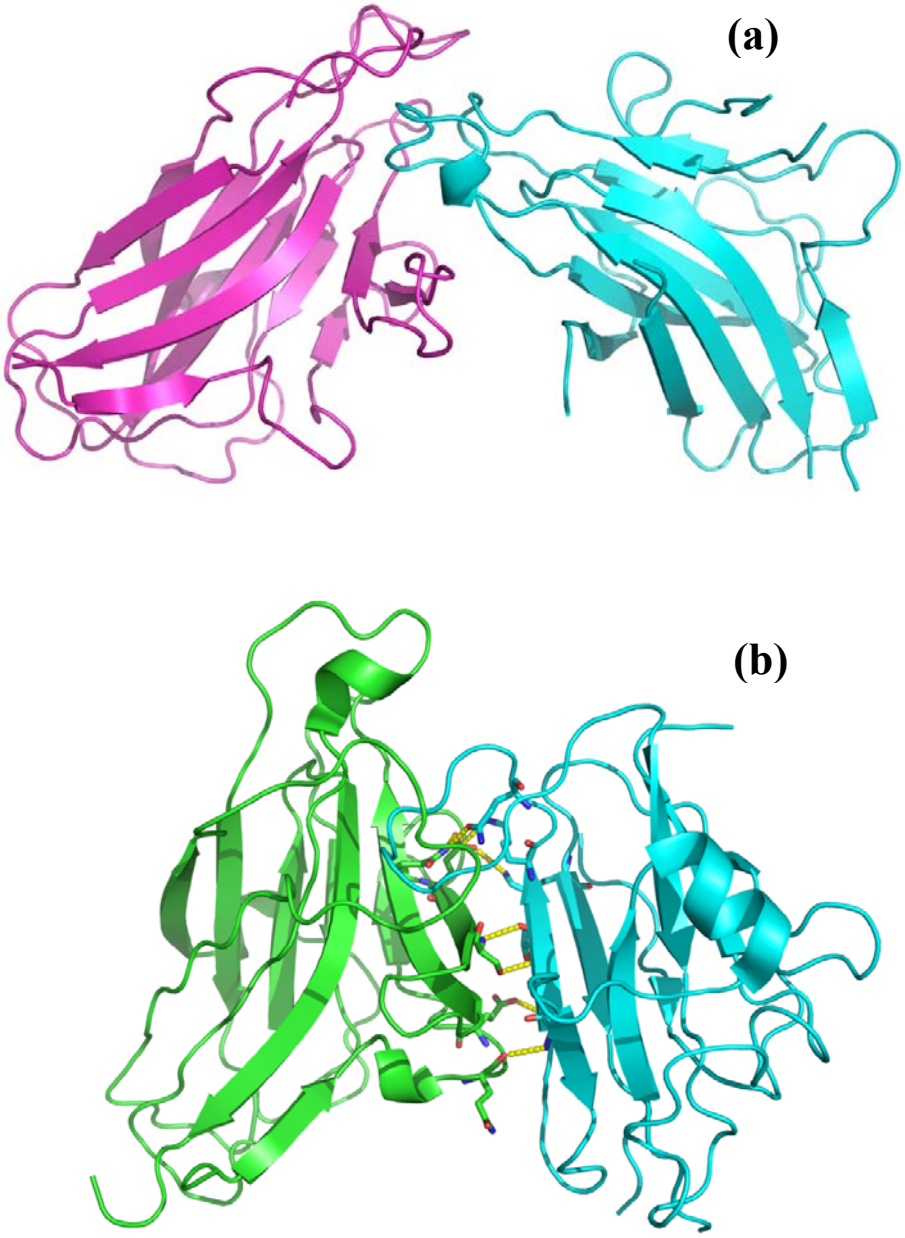


Figure 3.11 Comparison of the interactions between EphA4/EphA4, and EphA4/ephrinB2. (a): Interaction between EphA4 LBD; (b): Interaction between EphA4 LBD (green) and ephrinB2 ectodomain (Cyan). Yellow dashes represent hydrogen bonds (Qin *et al*, 2009)

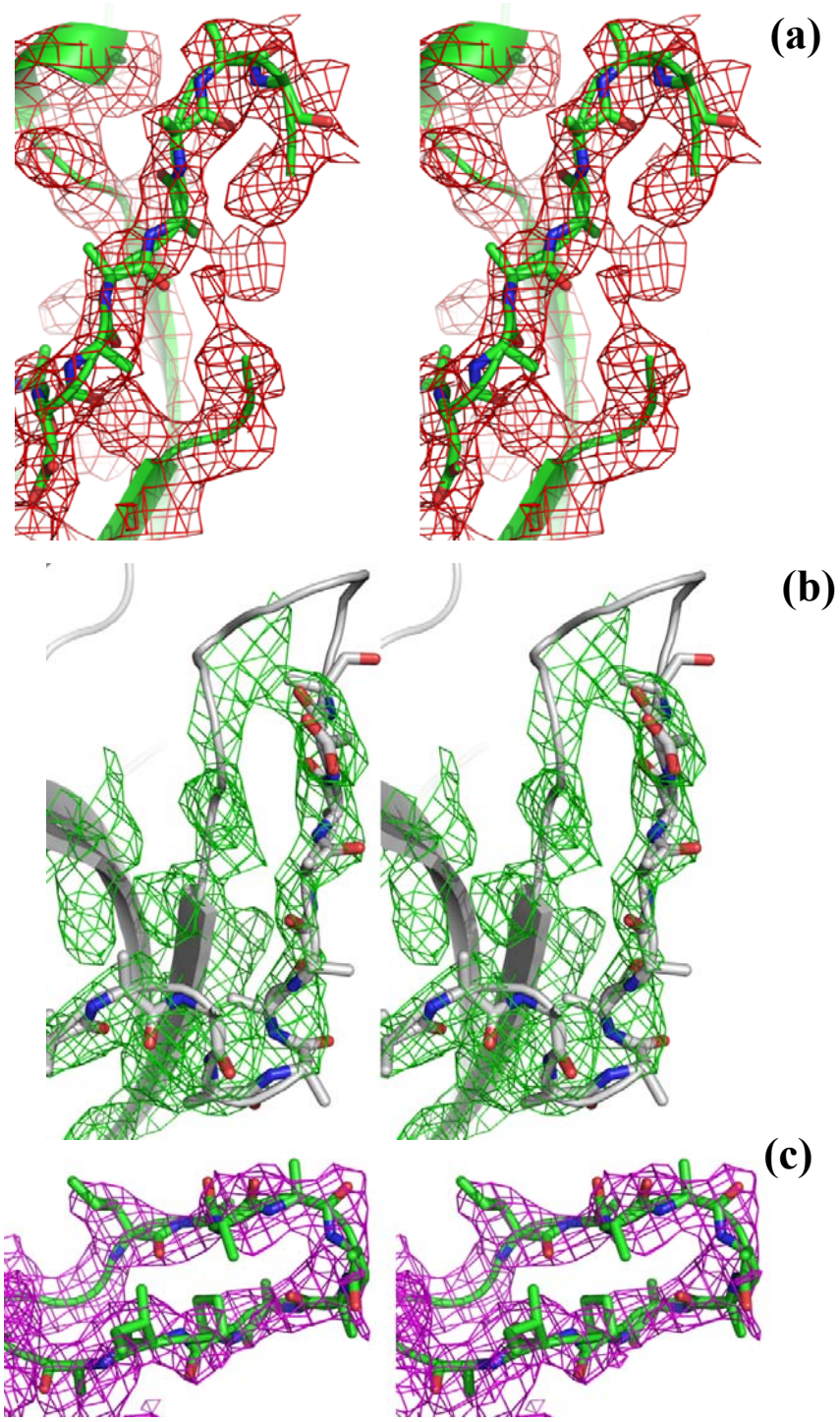


Figure 3.12 Stereo view of J-K and D-E loops built into the simulated annealing 2Fo-Fc electron density map contoured at 1.0σ. (a). J-K loop in open conformation and its electron density map, blue ribbon is from another EphA4 molecule; (b). J-K loop in closed conformation and its electron density map; (c). D-E loop and its electron density map. These density maps are clear enough to define the conformations of these loops.

3.2.3 Discussion

Although there are plentiful works on structure study of Eph receptors and Eph/ephrin complexes, this is the first time to observe that so many free Eph receptors contained in one AU. These two crystals enable us to capture 16 different EphA4 ligand binding domain structures at the same time. The comparison of these structures shows all the loop regions of free EphA4 LBD are flexible. D-E and J-K loops, which comprise of the classical binding channel, take various conformations in these structures, indicating that even in free EphA4 LBD, various conformations coexist, which may be ready to bind different kinds of ephrin ligands with slight rearrangement. It is interesting to note that EphB2, the only other Eph receptor which can bind both A- and B-class of ephrin ligands, also undergo dramatic conformational change over D-E and J-K loops (Y. Goldgur *et al*, 2009). These observations suggest that the ligand promiscuity may directly correlates with the structural flexibility of the ligand binding domain of Eph receptors.

Another interesting result derived from these structures is that, although both open and closed conformations coexist over J-K loop, J-K loop adopts open conformation only when it is interacted by a G-H loop from other EphA4 LBD.

Based on these results, the binding between Eph and ephrin does not follow lock and key mechanism (Fischer E. 1894), in which two rigid binding partners are of complementary shape, whereas, it may be explained by induce and fit mechanism (Koshland, D.E. 1958).

Comparison of these structures show that both D-E and J-K loops are highly dynamic and various conformations are coexist, however, the transition from closed to open conformation relies on the interaction from other molecules. For example, D-E loop keeps closed conformation in all the structures, probably because there is no

enough interaction to switch D-E loop from closed to open. Some of J-K loops also take closed conformation and there is no molecule interacts with them. However, when G-H loop is present, J-K loop will transit from closed to open conformation, just like what we observed in these structures.

Another important point revealed in this paper is that protein dynamics is a universal phenomenon and it may be critical in mediating protein-protein interactions. Our previous work on ephrinB2 dynamics shows that protein dynamics could be a very important complementary aspect to protein structures because most loop regions in X-ray structures are rigid, while in fact they are highly dynamic and critical in protein-protein interactions (Ran *et al*, 2008). Compared with X-ray structures, NMR structures can reflect protein dynamics because loop regions will have more randomized conformations. NMR is also a unique tool to study protein dynamics. It is of great interests to solve the NMR structure of Eph receptors and study their dynamics in both free and complex states. Those results could be critical to understand binding and signalling mechanisms of Eph/ephrin family.

Table 3.4 Crystallographic data and refinement statistics for the EphA4 structures

Data collection	**Crytal1**	**Crystal2**
Wavelength (Å)	1.5418	1.5418
Resolution limit (Å)	50-2.9	50-2.4
Space group	P1	P1
Cell parameters		
a, b, c (Å)	53.212, 70.621, 126.985	46.881, 70.030, 123.103
α, β, γ (°)	90.011, 90.036, 89.999	89.982, 89.972, 89.990
Unique reflections	75555	98655
Completeness	97.7%	94.6
Redundancy	1.1	1.8
Linear R-factor	0.082	0.094
Refinement		
Resolution range (Å)	25.0-3.0	25.0-2.6
R_{work}	0.236	0.237
Number of Reflections/test	34290/1816	48082/4071
R_{free}	0.312	0.262
Rmsd bond lengths (Å)	0.013	0.011
Rmsd bond angles (deg)	1.613	2.063
Ramachandran plot		
Favored, %	75.9	72.9
Allowed, %	18.2	21.4
Generously allowed, %	3.5	3.2
Disallowed, %	2.4	2.4

3.3 Structure Characterization of EphA4-ephrinB2 Complex Reveals New Features Enabling Eph-ephrin Binding Promiscuity

3.3.1 Crystal structure of the EphA4-ephrin-B2 complex

In the present study, we have determined the crystal structure of the EphA4 ligand-binding domain in complex with the ephrinB2 ectodomain at 2.5 Å resolution with a final R-factor of 0.23 (R_{free} = 0.285). Details of data collection and refinement statistics are summarized in Table 3.5.

Table 3.5 Crystallographic data and refinement statistics for the EphA4-ephrinB2 complex structure

Data collection	
Wavelength (Å)	1.5418
Resolution limit (Å)	50-2.5
Space group	P2(1)
Cell parameters	
a (Å)	54.651
b (Å)	48.711
c (Å)	64.469
α (°)	90.000
β (°)	110.434
γ (°)	90.000
Unique reflections	11071
Completeness	98.8%
Redundancy	4.0
Linear R-factor	0.144
Refinement	
Resolution range (Å)	25.0-2.5
R_{work} **	0.23
Number of Reflections	11189
R_{free} ***	0.285
Rmsd bond lengths (Å)	0.006
Rmsd bond angles (deg)	1.457
Ramachandran plot	
Favored, %	71.1
Allowed, %	25.6
Generously allowed, %	2.3
Disallowed, %	1.1

In the final model, the EphA4-ephrinB2 complex exists as a heterodimer (Figure 3.13a), consistent with size exclusion FPLC chromatography analysis indicating that the two proteins form a 1:1 heterodimer at concentrations of up to 500 μM. Some residues were not detectable in the complex due to poor electron density probably resulting from local flexibility, including EphA4 G-H loop residues Gly85-Val86-Met87 and H-I loop residue Arg110 (the numbering is according to the residues in the EphA4 construct, as shown in Figure 3.16) as well as ephrinB2 C-D loop residues Asp69-Ser70-Lys71-Thr72-Val73-Gly74 and Lys96 (the numbering is according to the mouse ephrinB2 sequence for consistency with previously published structures). Importantly, however, all EphA4 D-E and J-K loop residues are visible in the complex even though some of them were previously undetectable in the free EphA4 structure (Qin *et al*, 2008). The overall structure of the EphA4-ephrinB2 complex is architecturally similar to those of other Eph-ephrin complexes previously determined, with root mean square (RMS) deviations between equivalent Cα positions of 1.56 Å for the EphB2-ephrinB2 complex, 1.98 Å for the EphB4-ephrinB2 complex, 1.28 Å for the EphB2-ephrinA5 complex, and 2.36 Å for the deposited EphA2-ephrinA1 complex. However, large structural variations were observed in the EphA4-ephrinB2 complex as compared with other Eph-ephrin complexes over the loop regions directly involved in binding, which include the EphA4 A-C, D-E, G-H and J-K loops (Figure 3.13b) and the ephrin-B2 G-H loop (Figure 3.13c).

We previously determined the crystal structure of the EphA4 ligand-binding domain in the free state and demonstrated that there are two conformers (designated as molecule A and molecule B) in one asymmetric unit (Qin *et al*, 2008). The two EphA4 molecules are almost superimposable over the whole sequence, except for a substantial structural difference in the J-K loop. The J-K loop of molecule A does not

have any regular secondary structure while the J-K loop of molecule B contains a short β-sheet (Figure3.13b). This difference was attributed to the different packing of the two molecules in the neighbouring asymmetric units. Structural comparison of EphA4 bound to ephrinB2 with the two free EphA4 molecules reveals that the J-K loop of ligand-bound EphA4 is much more similar to that of molecule A than molecule B because it does not contain the short β-sheet formed by residues Phe126-Val129 and Met136-Asn139 (Figure 3.13b). The structure of EphA4 in complex with ephrinB2 has RMS deviations of 1.08 Å from the equivalent Cα positions of the free EphA4 molecule A, while RMS deviations are 1.13Å from the equivalent Cα positions of the unbound EphA4 molecule B. Four EphA4 loops (A-C, D-E, G-H and J-K) undergo substantial movements towards ephrinB2 upon binding (Figure 3.13b). For example, the Cα atom of Glu14 in the A-C loop shifts by 2.08 Å, the C α atom of Pro84 in the G-H loop shifts by 3.06 Å and, strikingly, the Cα atom of Glu34 in the J-K loop shifts by 10.34Å. EphrinB2 undergoes less dramatic structural changes upon binding to EphA4. The EphA4-bound ephrinB2 structure has RMS deviations of 1.06 Å from the equivalent Cα positions of ephrinB2 in the free state (1IKO), 0.91 Å from the Cα positions of ephrinB2 bound to EphB2 (1KGY), 0.80 Å from the Cα positions of ephrinB2 bound to EphB4 (2HLE), and 0.89Å and 0.80 Å from the Cα positions of ephrinB2 bound to the G attachment glycoproteins of the Nipah (2VSK) and Hendra (2VSM) viruses respectively (Figure3.13c). Relatively large conformational changes among different ephrin-B2 structures are observed only in the F-G and G-H loops. Interestingly, it appears that the G-H loop of ephrinB2 has somewhat different conformations when bound to an Eph receptor versus the G attachment glycoproteins of Nipah and Hendra viruses (Bowden T.A. *et al*, 2008).

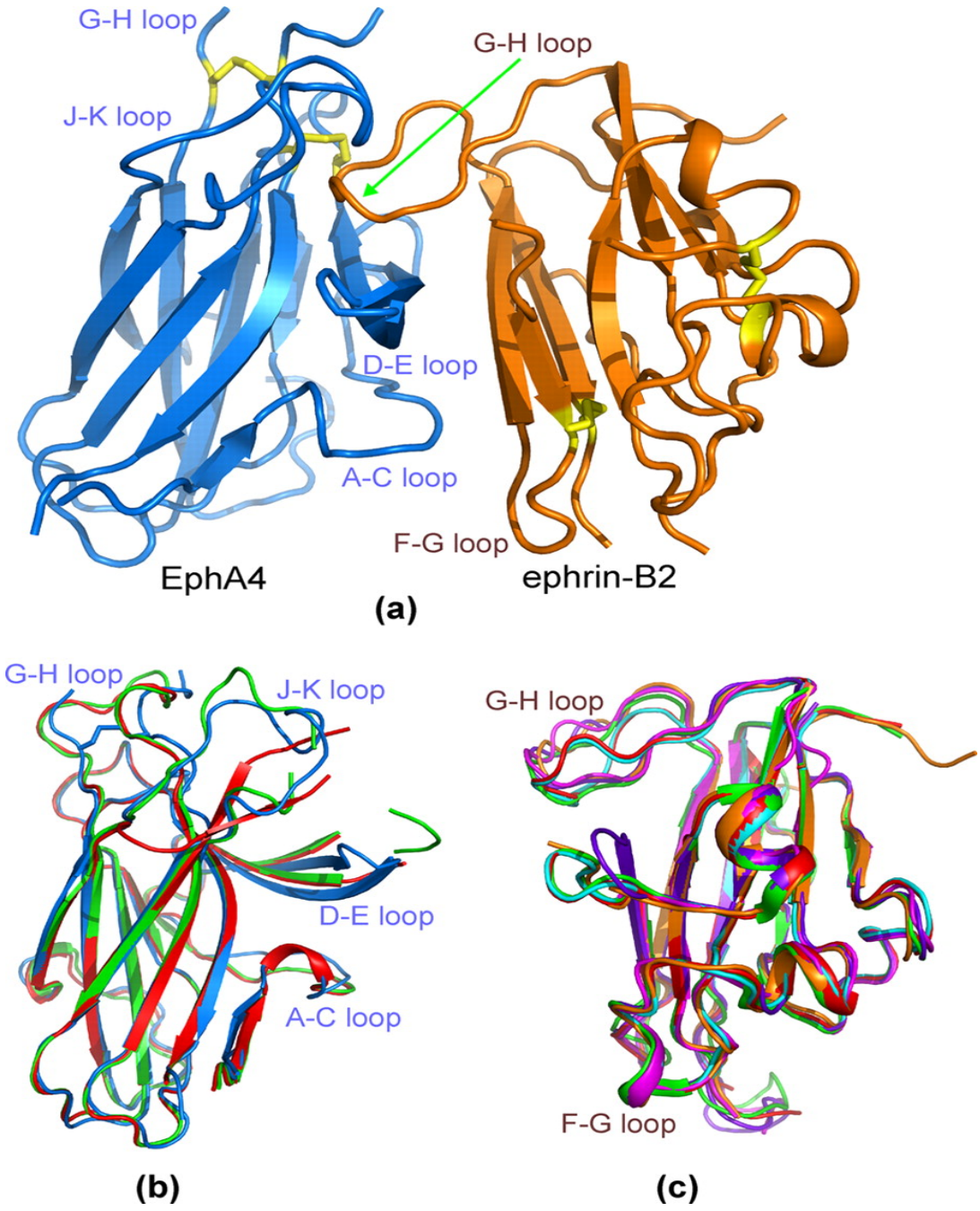


Figure 3.13. Crystal structure of the EphA4-ephrin-B2 complex. (a). Overall structure of the EphA4-ephrin-B2 complex. (b). Superimposition of the EphA4 structures in complex with ephrin-B2 (blue); free EphA4 structures A (green) and B (red) we previously determined. (c) Superimposition of ephrin-B2 structures in complex with EphA4 (brown); with EphB2 (brown, 1KGY); with EphB4 (purplish blue 2HLE); with Hendra viral attachment protein (cyan, 2VSK); with Nipha viral attachment protein (red, 2VSM) and free ephrin-B2 (green, 1IKO).

3.3.2 Binding interface of the EphA4-ephrin-B2 complex

The dissociation constant (K_d) for EphA4-ephrinB2 binding measured by isothermal calorimetry is 203 nM (Figure 3.14 and Table 3.6). This binding affinity is much weaker than those between EphB2 and ephrinB2 (22 nM) and between EphB4 and ephrinB2 (40 nM) also measured by isothermal calorimetry.

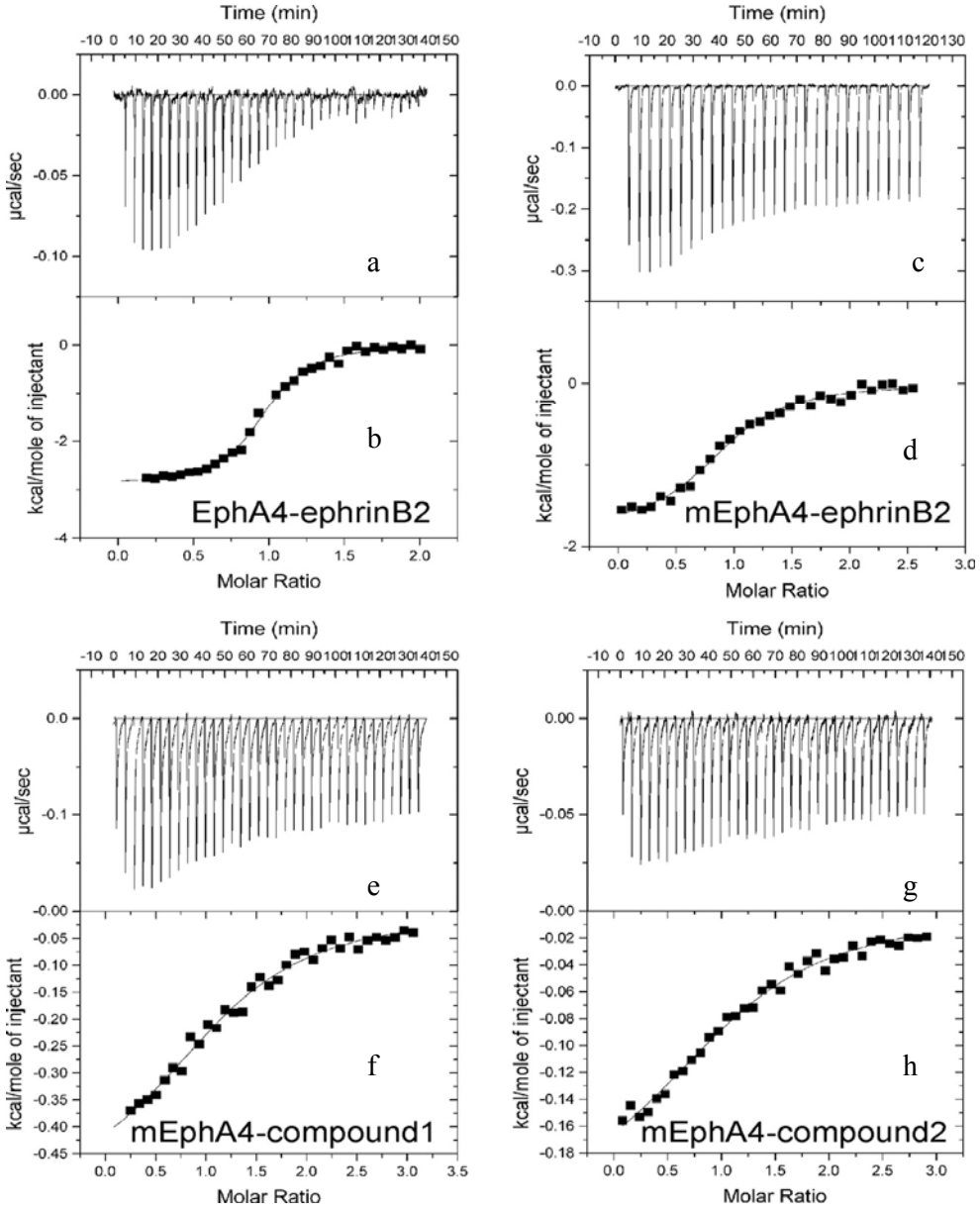


Figure3.14. ITC characterization of WT-EphA4 and mutated EphA4 binding with ephrinB2 and compound1 and 2. (a). The ITC titration profiles of the binding reaction of the EphA4 ligand-binding domain with the ephrinB2 ectodomain and integrated values for reaction heats with subtraction of the corresponding blank results normalized by the amount of ligand injected versus molar ratio of EphA4/ephrinB2 (b). (c). The ITC titration profiles of the binding reaction of the mutated EphA4 ligand-binding domain with the ephrinB2 ectodomain and integrated values for reaction heats with subtraction of the corresponding blank results normalized by the amount of ligand injected versus molar ratio of EphA4 mutant/ephrinB2 (d). (e). The ITC titration profiles of the binding reaction of the mutated EphA4 ligand-binding domain with the compound 1 and integrated values for reaction heats with subtraction of the corresponding blank results normalized by the amount of ligand injected versus molar ratio of C1/EphA4 mutant (f). (g). The ITC titration profiles of the binding reaction of the mutated EphA4 ligand-binding domain with the compound 2 and integrated values for reaction heats with subtraction of the corresponding blank results normalized by the amount of ligand injected versus molar ratio of C2/EphA4 mutant (h). The detailed conditions and setting of the ITC experiments are presented in Materials and Methods as well as Table 3.6.

Table 3.6 : Thermodynamic Parameters of the Binding Interactions between EphA-WT, EphA4 mutant and their ligands as Measured by Isothermal Titration Calorimetry (ITC)

Syringe	Cell	Buffer	Volume of injection (μL)	K_a (M^{-1})	K_d (μM)	Stoichiometry(n)	ΔS (cal/mol*K)	ΔH (kcal/mol)
EphA4 (0.2mM)	ephrinB2 (0.01mM)	Phosphate (10mM, pH 6.2)	5	$4.928\times10^{6}\pm 4.698\times10^{5}$	0.20	1.000 ± 0.000	20.92	-2.892 ± 0.035
EphA4M (0.6mM)	ephrinB2 (0.03mM)	Phosphate (10mM, pH 6.2)	5	$4.255\times10^{5}\pm 5.454\times10^{4}$	2.35	0.918 ± 0.021	19.70	-1.806 ± 0.053
Compound1* (2mM)	EphA4 (0.07mM)	Phosphate (10mM, pH 6.2)	5	$4.893\times10^{4}\pm 5.071\times10^{3}$	20.44	1.000 ± 0.000	18.11	-1.001 ± 0.027
Compound1 (2mM)	EphA4M (0.07mM)	Phosphate (10mM, pH 6.2)	5	$4.950\times10^{4}\pm 3.093\times10^{3}$	20.20	1.200 ± 0.000	19.77	-0.513 ± 0.008
Compound2* (2mM)	EphA4 (0.07mM)	Phosphate (10mM, pH 6.2)	5	$3.785\times10^{4}\pm 7.575\times10^{3}$	26.42	1.000 ± 0.000	20.15	-0.237 ± 0.013
Compound2 (2mM)	EphA4M (0.07mM)	Phosphate (10mM, pH 6.2)	5	$3.935\times10^{4}\pm 2.137\times10^{3}$	25.41	0.978 ± 0.000	20.38	-0.192 ± 0.003

*The data is from 3.3 The binding characterization of EphA4 ectodomain and its small antagonists

To understand the structural basis of the differences in ephrinB2 binding affinity for EphA4 versus EphB receptors, we performed a detailed analysis of the binding interface between EphA4 and ephrinB2.

The EphA4-ephrinB2 interface centers around the ephrinB2 G-H loop, which is inserted into a hydrophobic channel of the EphA4 receptor, as observed previously for other Eph-ephrin complexes (Figure 3.15a). The EphA4 D, E and J β-strands serve as the sides of the channel, and the G and M strands form the back of the channel, which is further capped by the EphA4 G-H loop. Interactions between the ephrinB2 G-H loop and the EphA4 channel appear dominated by Van der Waals contacts. In particular, the side chains of residues Leu124 and Trp125 at the tip of the ephrinB2 G-H loop establish extensive hydrophobic interactions with the EphA4 hydrophobic side chains of Ile31 in the D strand, Val167 in the M strand and Phe126, Val129, Ile135 and Leu138 in J-K loop. Furthermore, the aromatic ring of ephrinB2 Phe120 establishes hydrophobic interactions with the side chains of EphA4 Leu83 and Pro84. In addition, Pro122 of ephrinB2 is in direct contact with Ala165 in the M strand of EphA4 as well as the EphA4 disulfide bridge between Cys45 and Cys163. Besides the hydrophobic contacts, there are two polar interactions between EphA4 and ephrinB2 in the channel. The first is a salt bridge between ephrinB2 Glu128 and EphA4 Arg78, a residue that is conserved in all Eph receptors except EphB4 (in which the equivalent residue is Leu) (6). The second is a side-chain hydrogen bond between ephrinB2 Gln118 and EphA4 Gln43 (Figure 3.15b).

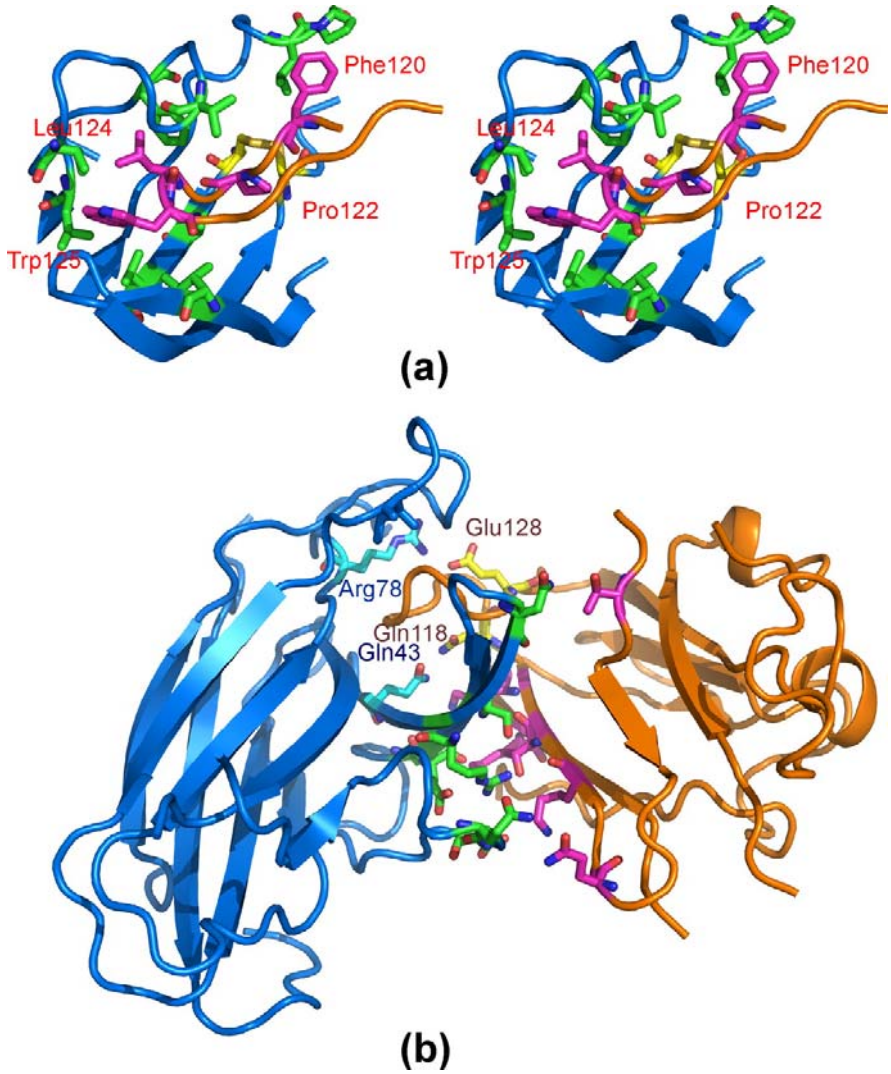


Figure 3.15. Anatomy of the EphA4-ephrin-B2 binding interface. (a). Hydrophobic interactions between EphA4 and ephrin-B2. The ephrin-B2 hydrophobic residues are labeled. (b). Polar interactions between EphA4 and ephrin-B2. The residues involved in polar interactions within the binding channel are labeled: a side chain hydrogen bond between EphA4 Gln43 and ephrin-B2 Gln118; and a side chain salt bridge between EphA4 Arg78 and ephrin-B2 Glu128.

Interestingly, in the previously described EphB2-ephrinB2 complex (Himanen *et al*, 2001), Tyr57 in the D-strand of mouse EphB2 (corresponding to Tyr50 in human sequence, Figure 3.16) engages in an aromatic-hydrophobic interactions with Leu160 in the J-K loop of EphB2 and Leu127 in the G-H loop of ephrinB2 (Figure 3.18a). These interactions lead to the formation of a hydrophobic patch, which is absent in the

EphA4-ephrinB2 complex (Figure 3.18a) because the EphA4 residue structurally equivalent to EphB2 Tyr57 is a Met (Figure 3.18). The absence of this hydrophobic patch in the EphA4-ephrinB2 complex appears to be responsible for the 10.3 Å movement of the EphA4 J-K loop away from ephrinB2 compared to the equivalent region in the EphB2-ephrinB2 complex. As a consequence, the J-K loop of EphA4 interacts more loosely with the ephrinB2 G-H loop than the J-K loop of EphB2. This may at least in part account for the relatively low binding affinity between EphA4 and ephrinB2. In the EphB4 receptor, the residue corresponding to EphB2 Tyr57 is also a Met, and the hydrophobic patch observed in the EphB2-ephrinB2 complex is also absent. However, the EphB4 J-K loop forms a short two-stranded β-sheet with a Pro at the β-turn position, which establishes additional contacts with the ephrin-B2 G-H loop, such as those between EphB4 Pro151 and ephrinB2 Phe120. These additional contacts, together with the other contacts between ephrinB2 Phe120/Pro122 and EphB4 Leu95 (which corresponds to an Arg in all other Eph receptors), may partly compensate for the absence of the hydrophobic patch and still yield a high binding affinity for the EphB4-ephrinB2 interaction. There is also a second contact region in the EphA4-ephrinB2 interface, structurally separate from the channel and involving extensive surface polar contacts mediated by a network of hydrogen bonds and salt bridges between the upper surface of EphA4 (A-C loop and D and E strands) and ephrinB2 C, G and F strands (Figure 3.18b). These polar interactions include hydrogen bonds between Gln109L and Gln12R (where L indicates the ligand and R the receptor), Thr114L and Ser30R/Arg40R, Lys60L and Glu28R, Gln118L and Gln43R, and Thr99L and Asn36R as well as salt bridges between Lys112L and Glu14R, Lys116L and Glu27R, and Glu128L and Arg78R. Within this surface contact region, Gln12 and Glu14 of EphA4 together with Gln109 and Lys112 of ephrinB2

would be predicted to play a particularly critical role (Figure 3.18b). More specifically, Gln12 in the A-C loop of EphA4 forms a side-chain hydrogen bond with Gln109 in the F-G loop of ephrinB2, and Glu14 in the A-C loop of EphA4 forms a side-chain salt bridge with Lys112 in the G β-strand of ephrinB2. Among the Eph-ephrin complex structures reported so far, similar contacts were observed in the EphB2-ephrinB2 and EphB4-ephrinB2 complexes, but not in the EphA2-ephrinA1 complex or the EphB2-ephrinA5 interclass complex (Figure3.17).

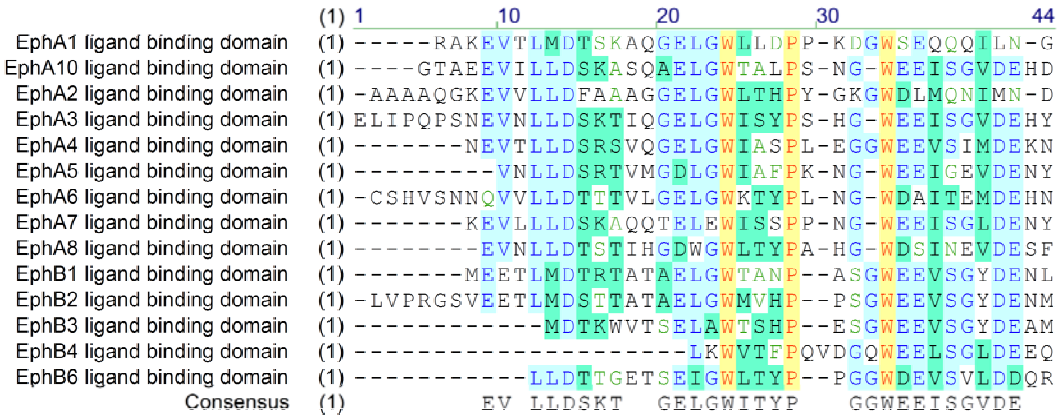


Figure 3.16. Sequence alignment of part of the ligand binding domains of Eph Receptors

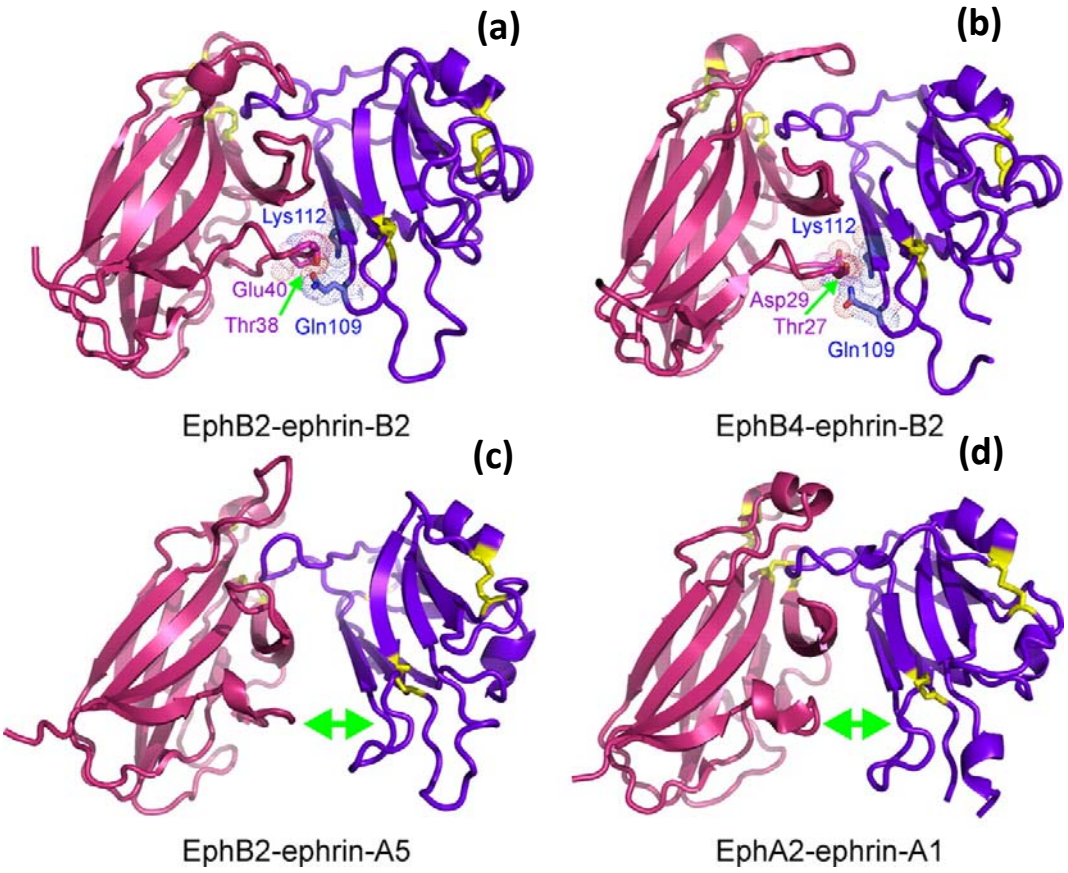


Figure 3.17. Structures of the Eph-ephrin complexes. (a) The EphB2-ephrin-B2 (1KGY) complex structure with the structurally equivalent contact region for that observed in the EphA4-ephrin-B2 as shown in Figure 4b. (b) The EphB4-ephrin-B2 (2HLE) complex structure with the structurally equivalent contact region. (c) The EphB2-ephrin-A5 (1SHW) complex structure without the structurally equivalent contact region. (d) The EphA2-ephrin-A1 (3CZU) complex structure without the structurally equivalent contact region.

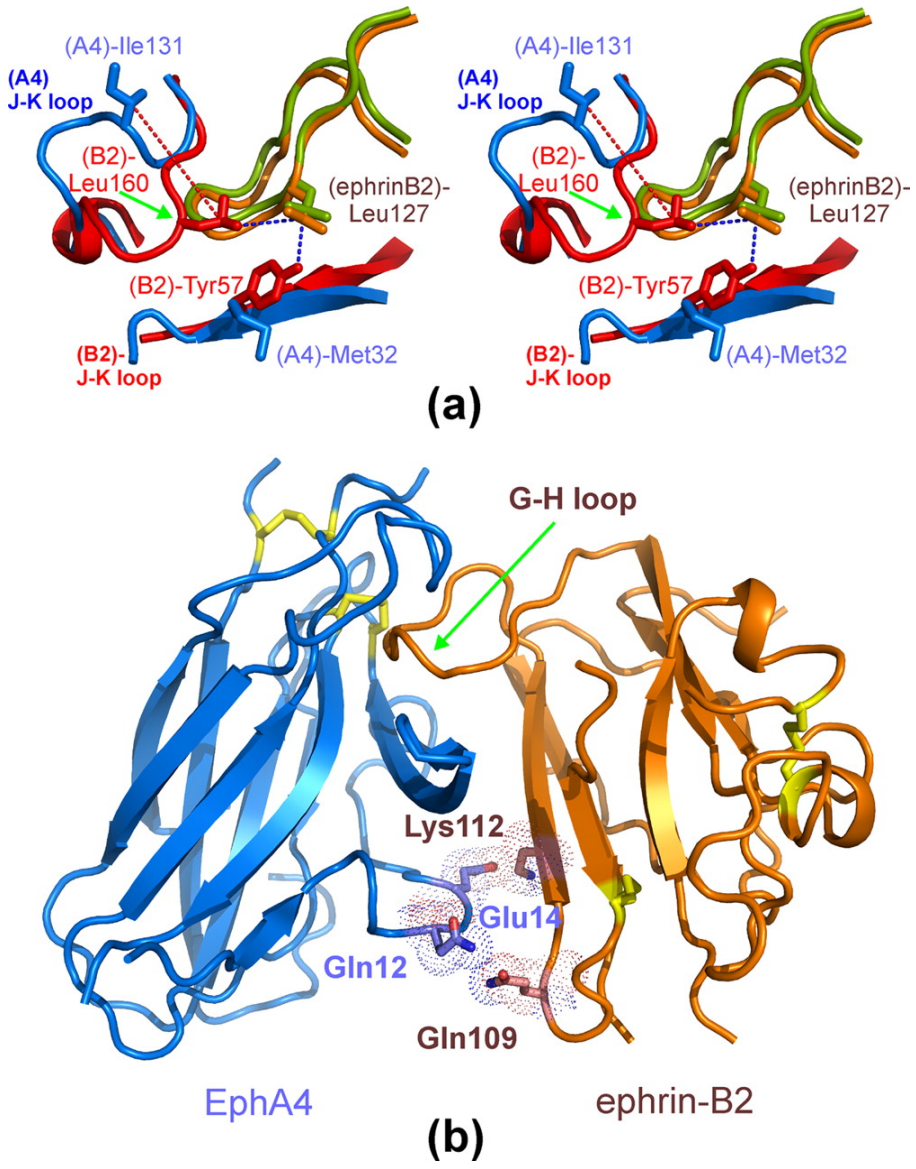


Figure 3.18. Unique features for the EphA4-ephrin-B2 complex. (a). Comparison of the conformations of the Eph receptor J-K loops in the EphA4-ephrin-B2 and EphB2-ephrin-B2 (1KGY) complexes. (b). A contact region identified in the EphA4-ephrin-B2 complex formed between EphA4 residues Gln12, Glu14 and ephrin-B2 residues Gln109, Lys112.

3.3.3 Ligand-binding properties of the EphA4 Gln12/Glu14 mutant

To investigate the contribution of the surface contact region to the binding affinity and promiscuity between EphA4 and ephrinB2, we replaced with alanines the residues of EphA4 and ephrin-B2 involved in the interface and expressed the mutant proteins in *E. coli*. Mutation of Gln12 and Glu14 in the A-C loop of EphA4 did not result in significant overall conformational changes, as judged by circular dichroism and NMR characterization (Figures 3.19b and 3.19c). Nevertheless, the affinity of the EphA4 mutant for ephrinB2 is ~ 10-fold lower than that of wild-type EphA4 (Figure 3.14 and Table 3.6). Interestingly, while the entropy change (ΔS) associated with the binding of ephrinB2 to mutant EphA4 remains mostly unchanged compared to wild-type EphA4, the enthalpy change (ΔH) is significantly lower. This implies that the hydrophobic interactions between the ephrin-B2 G-H loop and the EphA4 channel are very similar for both wild-type and mutant EphA4. However, mutation of EphA4 Gln12 and Glu14 to Ala disrupts the polar surface interactions with ephrinB2, thus leading to a significant difference in ΔH. Two small molecule antagonists designated compound 1 and compound 2 were previously found to target the ephrins binding channel of EphA4, thus antagonizing the binding of several ephrins. We found that the K_d and ΔS values for the binding of the two compounds to mutant and wild-type EphA4 are very similar (Figure 3.14 and Table 3.6). Furthermore, the binding of the two compounds perturbs the same residues of wild-type and mutant EphA4 in NMR HSQC titrations (spectra not shown). These results indicate that mutation of Gln12 and Glu14 to Ala does not detectably affect the EphA4 ligand-binding channel, which is the target of the two compounds.

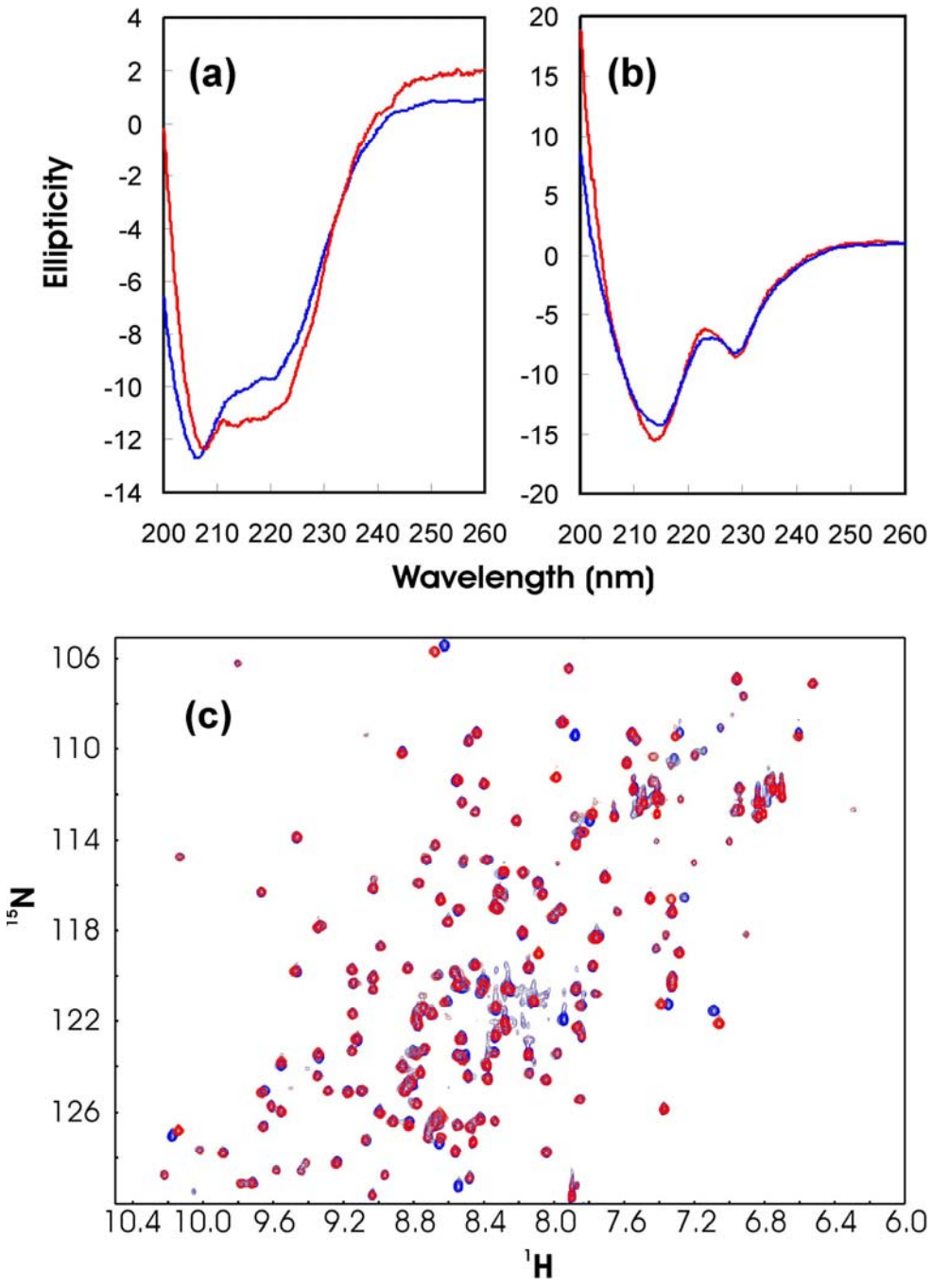


Figure 3.19. CD and NMR characterization of EphA4 and ephrin-B2 mutants. (a). The far-UV CD curves of the wild-type ephrin-B2 (blue) and its mutant (red). (b). The far-UV CD curves of the wild-type EphA4 (blue) and its mutant (red). (c). Two dimensional 1H-15N NMR HSQC spectra of the wild-type EphA4 (blue) and mutated EphA4 (red).

To investigate the contribution of the surface contact region to the binding affinity of EphA4 for different ephrins, we performed ELISA binding assays using the EphA4 ligand-binding domain fused to alkaline phosphatase (AP) and ephrin Fc fusion proteins generated in eukaryotic cells. The K_d values for the binding of mutated EphA4 to ephrinA1, ephrinA4 and ephrinA5 are very similar to those of wild type EphA4, suggesting that Gln12 and Glu14 do not play an important role in ephrinA binding. In contrast, mutant EphA4 shows a ~10-fold lower affinity for ephrinB2 and ephrinB3 compared to wild-type EphA4, demonstrating a critical role of the two residues (Figure 3.20). It should be noted that the apparent K_d values measured in the ELISA assays are much lower than those obtained by isothermal titration calorimetry. This is due to the dimeric nature of alkaline phosphatase fusion proteins, which therefore have increased avidity. Furthermore, the concentration of fusion proteins calculated based on alkaline phosphatase activity may be underestimated. Nevertheless, these results clearly indicate that EphA4 Gln12 and Glu14 do not play an important role in the binding of ephrinA ligands, whereas they are critical for the interclass binding of ephrinB ligands.

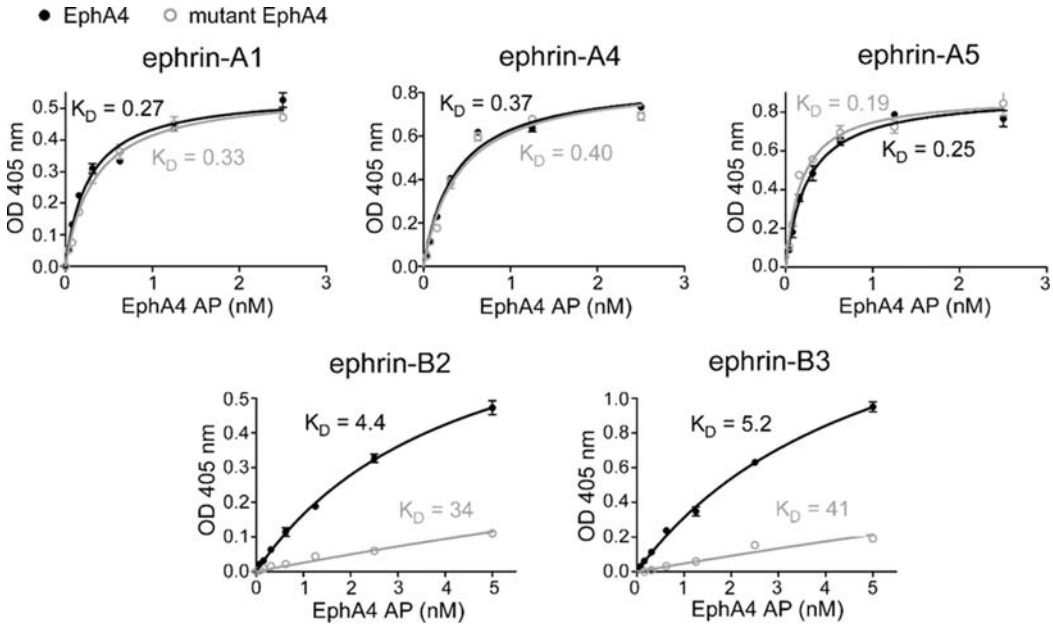


Figure 3.20. Binding properties of EphA4 and its mutant to different ephrin ligands. (a) Binding curves and K_d values of EphA4 (black) and EphA4 mutant (grey) to ephrin-A1. (b) Binding curves and K_d values of EphA4 (black) and EphA4 mutant (grey) to ephrin-A4. (c) Binding curves and K_d values of EphA4 (black) and EphA4 mutant (grey) to ephrin-A5. (d) Binding curves and K_d values of EphA4 (black) and EphA4 mutant (grey) to ephrin-B2. (e) Binding curves and K_d values of EphA4 (black) and EphA4 mutant (grey) to ephrin-B3. (Data from collaborator's group, R. Noberini, *JBC*, 2009)

3.3.4 Receptor-binding properties of the ephrin-B2Gln109/Glu112 mutant

Structural characterization by circular dichroism suggests that ephrinB2 appears to become more helical when Gln109 and Lys112 are mutated to Ala (Figure 3.19a). This is reasonable given that Lys112 is located on the G β-strand and Ala is known to be a helix-inducing residue. Despite these structural changes, the affinity of mutant ephrinB2 AP for EphB receptors in ELISA assays was reduced by only 2-4 fold compared to wild-type ephrinB2 AP (Figure 3.21). This suggests that Gln109 and Lys112 play a relatively minor role in EphB receptor binding and that most of the binding affinity depends on the interaction of the ephrinB2 G-H loop with the EphB

channel. In contrast, the affinity of the mutant ephrinB2 for EphA4 was much more compromised and the binding was barely detectable (Figure 3.21), consistent with the effect of the complementary mutations in EphA4 and a key importance of the surface contact region in EphA4-ephrinB2 interclass binding.

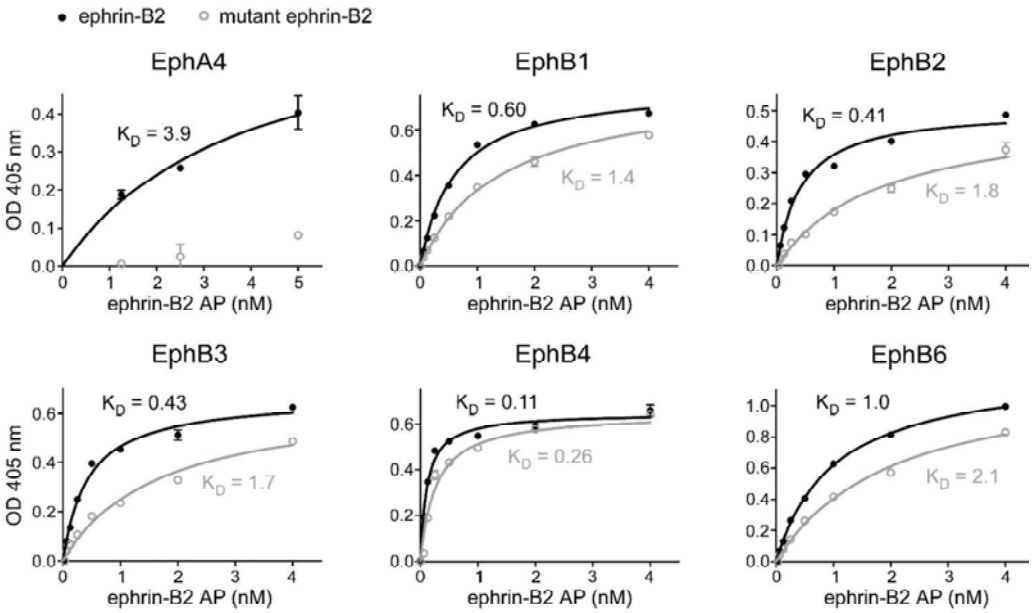


Figure 3.21. Binding properties of wide type and mutated ephrin ligands with Eph receptors. (a) Binding curves and Kd values of wide type (black) and mutated (grey) ephrinB2 with EphA4. (b) Binding curves and Kd values of wide type (black) and mutated (grey) ephrinB2 with EphB1. (c) Binding curves and Kd values of wide type (black) and mutated (grey) ephrinB2 with EphB2. (d) Binding curves and Kd values of wide type (black) and mutated (grey) ephrinB2 with EphB3. (e) Binding curves and Kd values of wide type (black) and mutated (grey) ephrinB2 with EphB4. (f) Binding curves and Kd values of wide type (black) and mutated (grey) ephrinB2 with EphB6. (Data from collaborator's group, R. Noberini, *JBC*, 2009)

3.3.5 NMR visualization of structural perturbations occurring in EphA4 upon ephrinB2 binding

We also used NMR spectroscopy to gain further insights into the interaction of wild-type and mutated EphA4 with ephrinB2 in solution. In NMR HSQC titrations, the binding between wild-type EphA4 and ephrinB2 was saturated at a molar ratio of 1:1.5 (EphA4/ephrin-B2), consistent with the relatively strong binding affinity between the two proteins. Surprisingly, only a small portion of EphA4 HSQC peaks shifted while the majority disappeared upon ephrinB2 binding (Figure 3.22a).

This observation implies that the binding to ephrinB2 would provoke significant conformational exchanges on the μs-ms time scale over the whole EphA4 ligand-binding domain as we previously observed on the Nck2 SH2 domain upon binding to phosphorylated ephrin-B2 cytoplasmic domains. Furthermore, mapping the perturbed EphA4 residues onto the EphA4-ephrinB2 complex structure shows that they are distributed over almost the entire protein (Figure 3.23a).

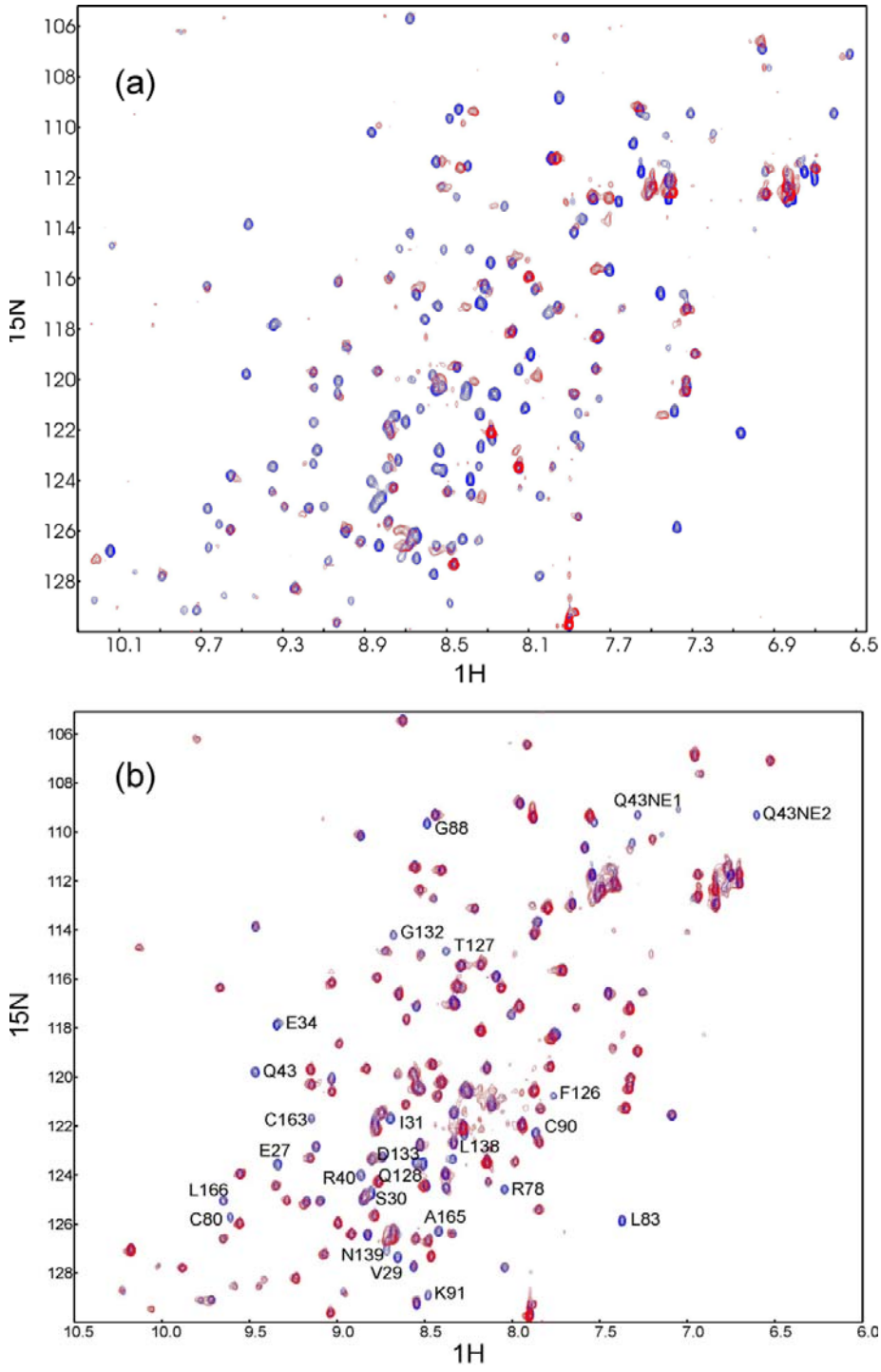


Figure 3.22. NMR HSQC mapping of the binding interfaces. (a). Superimposition of the NMR HSQC spectra of the wild-type EphA4 in the absence (blue) and presence (red) of ephrin-B2 at a molar ratio of 1:1.5 (EphA4/ephrin-B2). (b). Superimposition of the NMR HSQC spectra of the mutated EphA4 in the absence (blue) and presence (red) of ephrin-B2 at a molar ratio of 1:3 (EphA4 mutant/ephrin-B2). The assignments of the disappeared peaks are labelled.

On the other hand, the binding between mutant EphA4 and ephrinB2 is saturated at a molar ratio of 1:3 (mutant EphA4/ephrinB2), consistent with the weaker binding between the two proteins. Strikingly, only a small portion of the mutant EphA4 residues were perturbed by the binding of ephrinB2 (Figure 3.22b). Furthermore, all of the perturbed residues center around the binding interface with the ephrinB2 G-H loop (Figures 3.23b). In contrast to wild-type EphA4, residues in the A-C loop of mutant EphA4 were not perturbed even at a molar ratio of 1:8. Almost all the perturbed residues in the mutant EphA4 have direct contacts with ephrinB2 in the crystal structure (Figures 3.15 and 3.23), including Glu27, Val29, Ser30 and Ile31 in the D β-strand; Arg40 and Gln43 in the E β-strand; Ala165 in the M β-strand; Glu34 in the D-E loop; Phe126, Thr127, Gln128, Gly132, Asp133, Leu138 and Asn 139 in the J-K loop; and Arg78, Cys80, Leu83, Cys90 and Lys91 in the G-H loop. Interestingly, we also observed a significant perturbation of Cys163, which forms a disulfide bond with Cys45. This disulfide bond is conserved in different Eph receptors and has been implicated in the binding of both ephrins and antagonistic peptides.

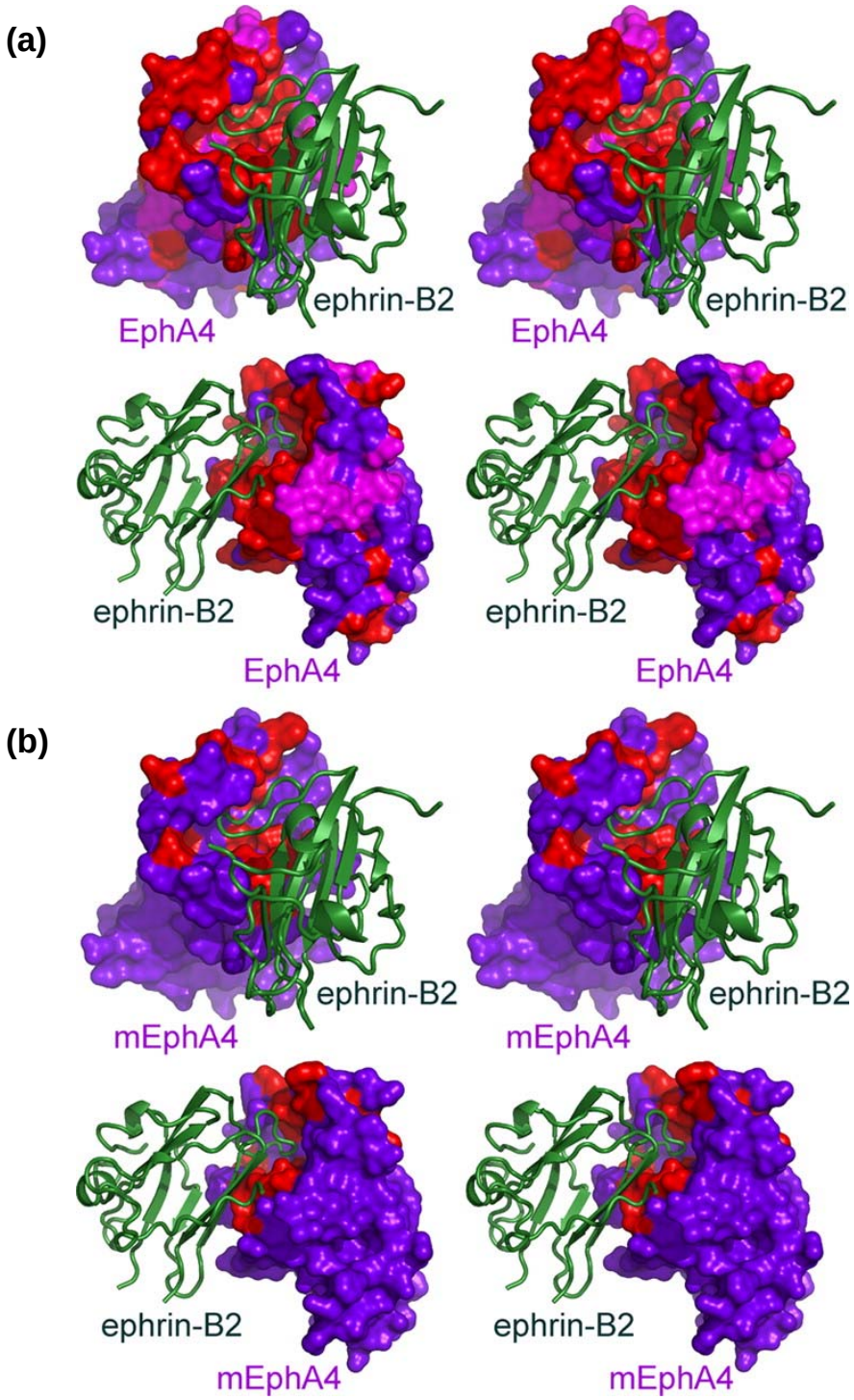


Figure 3.23. Binding interfaces mapped out by NMR. (a). The binding interface of the wild-type EphA4 with ephrin-B2 mapped out by NMR HSQC titrations. (b). The binding interface of the mutated EphA4 with ephrin-B2. The shifted residues are colored in pink and disappeared in red. The EphA4 is in surface mode and ephrin-B2 in ribbon.

3.3.6 Discussion

The binding promiscuity between Eph receptors and ephrin ligands appears to be a key strategy enabling Eph-ephrin signaling networks to control a wide array of biological functions. Understanding the structural principles governing promiscuity versus selectivity of Eph receptor-ephrin binding is therefore an important step to elucidate the mechanisms underlying the biological functions of the Eph system and is also critical for the design of antagonists to target Eph-ephrin interactions. EphA4 is the only Eph receptor that can bind with substantial affinity all ephrin-A and ephrin-B ligands (Pasquale EB, 2004). Here we report the structure of the EphA4-ephrin-B2 complex, which represents the first structure of a complex between an EphA receptor and an ephrin-B ligand.

The overall architecture of the EphA4-ephrin-B2 complex is very similar to those previously determined for other Eph-ephrin complexes. The high affinity interface of the complex can be divided into two relatively-independent regions. One mostly involves the hydrophobic interactions between the EphA4 ligand-binding channel and the ephrin-B2 G-H loop. The other, which was observed in the EphB-ephrin-B complexes, but was greatly reduced in the EphA2-ephrin-A1 complex and absent in the EphB2-ephrin-A5 complex, involves polar interactions between surface residues: Gln12 and Glu14 in EphA4 and Gln109 and Lys112 in ephrin-B2. In the EphA4-ephrin-B2 complex, we did not obtain evidence for a distinct lower affinity binding interface that may mediate tetramerization of two EphA4-ephrin-B2 heterodimers similar to what has been described for the EphB2-ephrin-A5 complex. Such interface is also not evident in the EphB4-ephrin-B2 and EphA2-ephrin-A1 complexes.

Several notable structural variations are observed in critical regions of the high affinity EphA4-ephrin-B2 dimerization interface upon complex formation. Several EphA4 regions – including the D and E β-strands as well as the A-C, D-E, G-H and J-K loops – undergo significant conformational rearrangements to accommodate the ephrin-B2 ligand. In particular, the short β-strand observed in one of the conformers of the free EphA4 is completely rearranged into a loop suitable for interacting with the ephrin-B2 G-H loop. This observation is in agreement with previous studies with EphB receptors, which show that significant structural rearrangements occur in these regions to enable the formation of Eph-ephrin complexes . In contrast, only minor structural changes have been observed upon formation of the EphA2-ephrin-A1 complex.

The EphA4-ephrin-B2 structure also explains the relatively weak affinity of the interclass EphA4-ephrin-B2 binding (K_d of 203 nM) compared with intraclass interactions, such as those of EphB2 with ephrin-B2 (K_d of 22 nM) and EphB4 with ephrin-B2 (K_d of 40 nM). Because of sequence variations, a hydrophobic patch present in the EphB2-ephrin-B2 complex is absent in the EphA4-ephrin-B2 complex. As a consequence, half of the EphA4 J-K loop remains open and does not make any contacts with the ephrin-B2 G-H loop (Figure 3.19a). Given that the structurally-equivalent EphB2 loop region in the complex with ephrin-A5 is more open and the binding affinity of EphB2 for ephrin-A5 is even lower (K_d of 320 nM), this feature likely accounts at least in part for the low binding affinity between EphA4 and ephrin-B2. However, comparing K_d values for wild-type and mutated proteins shows that the surface contacts formed between EphA4 residues Gln12 and Glu14 and ephrin-B2 residues Gln109 and Lys112 increase by about 10 folds the binding affinity and thus play a critical role in interclass EphA4-ephrin-B2 binding.

Mutation of the two EphA4 residues involved in the interface yields an EphA4 mutant that does not undergo any global structural changes or modifications in the ephrin-binding channel. This is evident from the unchanged binding affinity of the EphA4 mutant for two small molecule antagonists that bind within the channel. Interestingly, the surface contact region of EphA4 does not appear to be necessary for the intraclass binding with ephrin-A ligands, which probably involves a more intimate fit between the EphA4 ligand-binding channel and the ephrin-A G-H loop. This is the case for the EphA2-ephrin-A1 complex, where the polar surface contact region is extremely reduced. Similarly, this region is likely not present or minimal in EphA4-ephrin-A complexes. On the other hand, Gln109 and Lys112 of ephrin-B2 have been shown to also interact with residues on the surface of EphB receptors, which could explain the 2-3 fold decrease in the affinity of mutant ephrin-B2 for EphB receptors. However, we cannot exclude that the decrease in binding may be due to the changes in the overall conformation of the mutant ephrin observed by CD spectroscopy. Nevertheless, the much more pronounced impairment in the binding of mutant ephrin-B2 to EphA4 than to EphB receptors suggests that the surface contact region is much more critical for the interclass EphA4-ephrin-B2 binding than the intraclass EphB-ephrin-B2 binding. Interestingly, the residue corresponding to ephrin-B2 Gln109 is a Leu in ephrin-B3, and therefore cannot be involved in the hydrogen bond with EphA4 Gln12 that instead is present in the EphA4-ephrin-B2 complex. However, the binding affinity of EphA4 mutant for ephrin-B2 or ephrin-B3 is equally reduced compared to wild-type EphA4, suggesting that the salt bridge between ephrin-B2 Lys 112 and EphA4 Gln 14 might play a more important role in EphA4-ephrin-B interclass binding.

The availability of both wild-type and mutated EphA4 receptors has given us the unique opportunity to use NMR spectroscopy to gain new insight into the dynamics of the EphA4-ephrin-B2 complex in solution, which could not be achieved from analysis of the more static crystallized complex. Interestingly, the NMR data show that many of the EphA4 residues perturbed upon ephrin-B2 binding are located outside the high affinity EphA4-ephrin-B2 binding interface. For example, many residues in the EphA4 L and H strands are far away from the binding interface but they are significantly perturbed. Interestingly, in the crystal structures these residues do not appear to undergo any significant conformational changes upon ephrin-B2 binding (Figure 3.14b). The observed perturbations by NMR may be explained by changes in the dynamics of various EphA4 regions upon ephrin-B2 binding. In contrast, once the interactions mediated by EphA4 Gln12 and Glu14 are removed, the residues perturbed by the binding of ephrin-B2 are only limited to the EphA4 interface in direct contact with ephrin-B2. Thus, contacts mediated by Gln12 and/or Glu14 have far-reaching effects over the entire EphA4 ligand-binding domain. It is tempting to speculate that these dynamic perturbations may reflect changes in EphA4 that affect its biological function upon ephrin binding.

The new EphA4-ephrin-B2 complex structure that we have characterized, together with those previously determined for other Eph receptor-ephrin pairs, highlights a surprising diversity in the use of the two regions of the high affinity interface to accomplish intraclass or interclass Eph receptor-ephrin binding. For example, it appears that intraclass binding is mediated almost exclusively (A class) or predominantly (B class) by the hydrophobic Eph channel/ephrin G-H loop region of the interface. Accordingly, the fit of the ephrin G-H loop into the Eph channel is more intimate for the A than the B class. EphB2-ephrin-A5 interclass binding relies only on

the channel region of the interface, although the loose interclass fit of the EphB channel and the ephrin-A G-H loop results in the lowest binding affinity among the complexes structurally characterized so far. Interestingly, we found that EphA4-ephrin-B2 interclass binding uses a unique strategy, where very weak binding through the Eph channel/ephrin G-H loop is supplemented by interactions in the polar surface contact region of the interface. In this manner, EphA4 achieves the highest promiscuity among the Eph receptors.

As a consequence of this variability in Eph receptor-ephrin interfaces, the design of antagonists to target Eph-ephrin interactions may be more challenging than previously thought and diverse strategies may be needed depending on the Eph receptor and the ephrin involved. Consistent with this notion, we previously found that two small molecule antagonists that target the ephrin-binding channel of EphA4 inhibit the binding of some ephrins but not others. For example, we did not detect inhibition of ephrin-B2 and ephrin-A4 binding to EphA4, at concentrations that completely inhibited the binding of other ephrins. The structural information we obtained and the effects of the EphA4 and ephrin-B2 mutations also suggest that it is possible to selectively inhibit EphA4 binding to ephrin-B but not ephrin-A ligands by disrupting the polar surface region of the high affinity interface, while the binding of ephrin-A ligands to EphA4 may be selectively inhibited by disrupting appropriate contacts in the channel region. Such strategies may help dissect the biological roles of intraclass versus interclass EphA4-ephrin binding and guide more selective approaches for the design of EphA4 inhibitors.

3.4 Interactions of EphA4 Ligand Binding Domain with Two Small Molecule Antagonists

As mentioned in introduction, the critical roles of EphA4 in various physiological and pathological processes validate this receptor as a promising target for the development of small molecule drugs to treat human diseases. Until now, by targeting Eph-ephrin binding pocket, only two small molecules have been identified by a high throughput screening, which are able to antagonize ephrin-induced effects in EphA4-expressing cells (Noberini R *et al*, 2008). It is of significant interest to gain structural insight into the binding interactions between the two small molecules and the EphA4 ligand-binding domain, with the ultimate goal to develop small molecule antagonists capable of inhibiting Eph-ephrin binding with high affinity and specificity. In our study, we characterized its binding to two antagonistic small molecules namely 4- and 5-(2,5 dimethyl-pyrrol-1-yl)-2-hydroxy-benzoic acid by use of isothermal titration calorimetry (ITC), circular dichroism (CD), nuclear magnetic resonance (NMR) spectroscopy and computational docking.

3.4.1 Binding interactions characterized by ITC and CD

Recently, two isomeric 2,5-dimethylpyrrolyl benzoic acid derivatives have been identified by a high throughput screening to antagonize ephrin-induced effects in EphA4-expressing cells (Noberini R *et al*, 2008). To assess whether the two small molecules directly interact with the EphA4 ligand-binding domain, we utilized ITC to measure their thermodynamic binding parameters. Strikingly, by using a high EphA4 concentration (70 μM), we succeeded in obtaining these parameters, which are presented in Table 3.7, thus clearly confirming that the two small molecules do interact with the ligand-binding domain of EphA4. Interestingly, the two compounds have similar binding affinities for the EphA4 ligand binding domain, with K_d values

of 20.4 μM for compound 1 and 26.4 μM for compound 2, but their binding causes different enthalpy changes (*ΔH* values of -1,001 for compound 1 and -237 cal for compound 2).

Table 3.7: Thermodynamic parameters for the binding interactions between EphA4 and two small molecules by ITC

Syringe	Cell	Buffer	Volume injection (**μL**)	K_a **(M^{-1})**	K_d **(μM)**	Stoichiometry **(n)**	*ΔS* **(cal/mol*K)**	*ΔH* **(cal/mol)**
Compound1 (2 mM)	**EphA4** (70 μM)	Phosphate (10 mM, pH 6.3)	5	4.893×10^4 ± 5071	20.4	1	18.11	-1001 ± 26.62
Compound2 (2 mM)	**EphA4** (70 μM)	Phosphate (10 mM, pH 6.3)	5	3.785×10^4 ± 7575	26.4	1	20.15	-237.0 ± 12.99

Compound1: 4-(2,5-Dimethyl-pyrrol-1-yl)-2-hydroxyl-benzonic acid; Compound2: 5-(2,5-Dimethyl-pyrrol-1-yl)-2-hydroxyl-benzonic acid

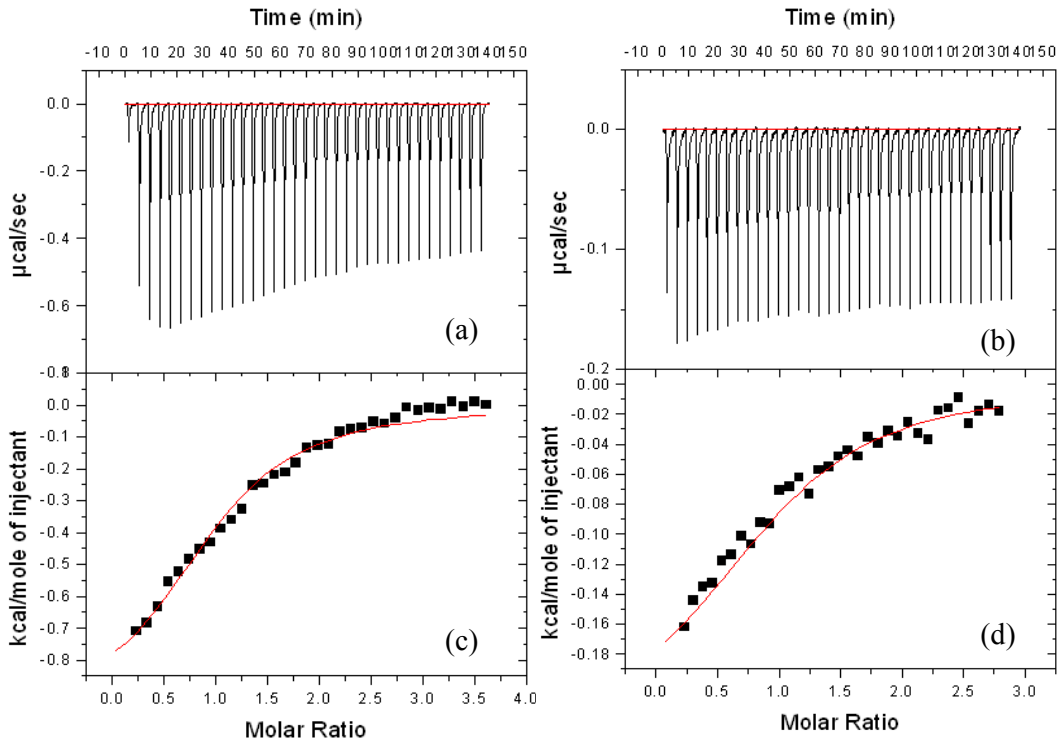


Figure 3.24. The ITC titration profiles of the binding reaction of the EphA4 ligand-binding domain with compound 1: (a); and with compound 2: (c). Integrated values for reaction heats with subtraction of the corresponding blank results normalized by the amount of ligand injected versus molar ratio of compound 1/EphA4 (b) and of compound 2/EphA4(d).

Far-UV CD spectroscopy was also used to monitor the overall structural changes in the EphA4 ligand-binding domain upon binding of the two molecules. As seen in Figure 3.25a, no significant difference is detected between the far-UV CD spectra of EphA4 in the absence and in the presence of the two molecules at a molar ratio of 1:6 (EphA4:compound). This result implies that no significant changes in secondary structure occurred in the EphA4 ligand binding domain upon binding, which is consistent with the relatively weak binding affinity of the two molecules.

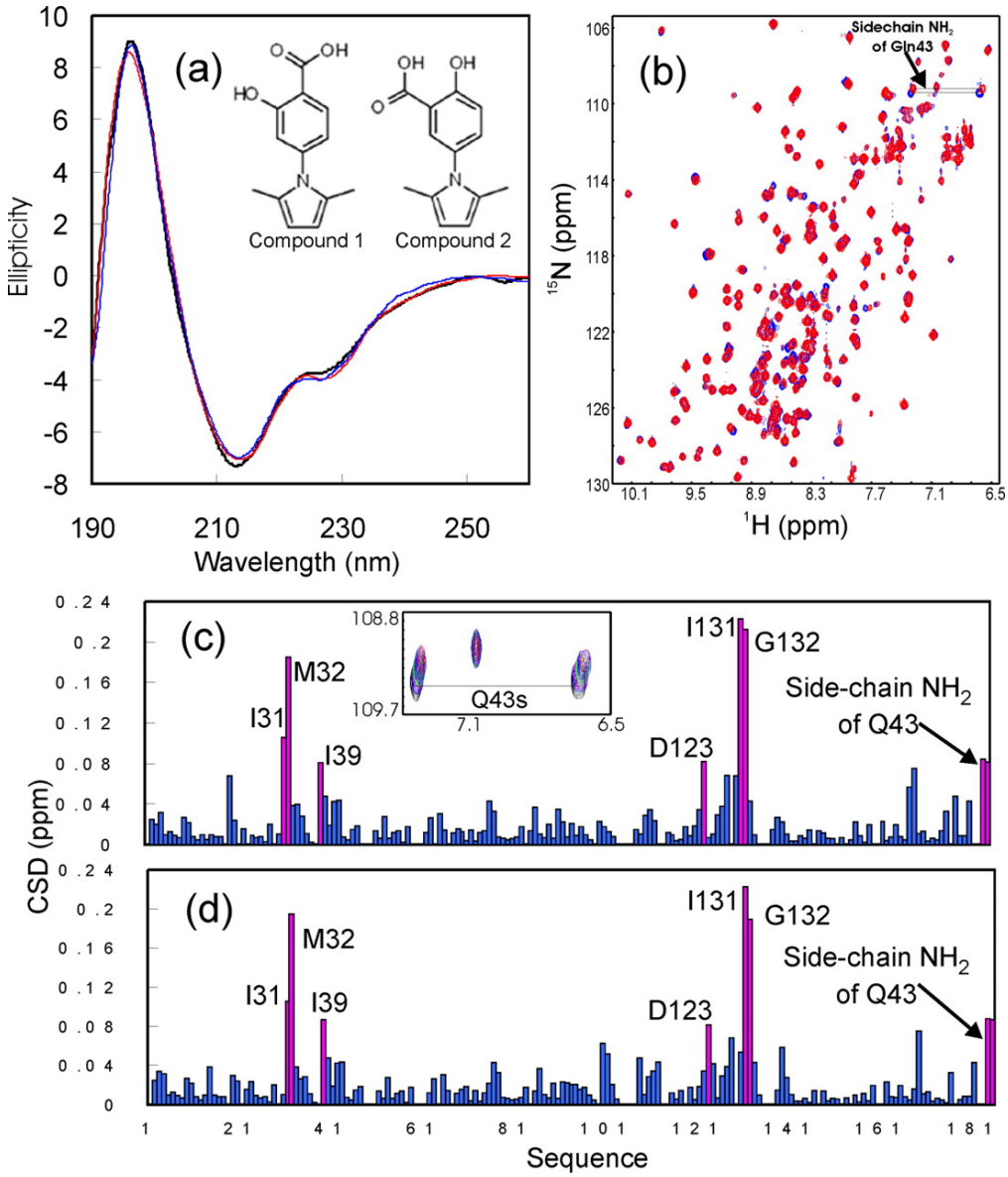


Figure 3.25. Characterization of the interactions with two small molecule antagonists. (a) Far-UV CD spectra of the EphA4 ligand-binding domain in the absence (black) and in the presence of compound 1 (blue) or compound 2 (red) at a molar ratio of 1:6 (EphA4:compound). (b) The ^{1}H-^{15}N NMR HSQC spectra of the EphA4 ligand-binding domain in the absence (blue) and in the presence of compound 1 (red) at a molar ratio of 1:6 (EphA4:compound). (c) Residue-specific chemical shift differences (CSD) of the EphA4 ligand-binding domain in the presence of compound 1 (red) at a molar ratio of 1:6 (EphA4:compound). (d) Residue-specific chemical shift differences (CSD) of the EphA4 ligand-binding domain in the presence of compound 2 (red) at a molar ratio of 1:6 (EphA4:compound2).

3.4.2 Binding interactions characterized by NMR

Because the two small molecules have medium binding affinity for EphA4, it is challenging to obtain stable receptor-compound complexes for co-crystallization. We therefore decided to probe their binding interactions using NMR spectroscopy, which is highly sensitive to weak binding. We prepared $^{15}N/^{13}C$ double-labeled EphA4, collected a series of three-dimensional heteronuclear NMR spectra at a protein concentration of 500 μM, and completed the sequential assignments. As evident from the very large dispersions in both dimensions (3.7 ppm for ^{1}H and 25 ppm for ^{15}N) of the HSQC spectrum (Figure 3.25b), the EphA4 ligand-binding domain is well-folded in solution. Only one set of HSQC peaks was observed for all the EphA4 residues, suggesting that the asymmetric dimer observed in the crystals does not exist in solution on the NMR time scale.

We subsequently used NMR HSQC titrations to detect the EphA4 residues that were perturbed by the binding of two compounds. Since the chemical shift value of a NMR active atom is sensitive to its chemical environment, chemical shift perturbation analysis upon titration of ligands represents a powerful method for identifying residues that directly contact the ligands or that are indirectly affected by the binding event. Two-dimensional ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled EphA4 were recorded to monitor the changes of the HSQC cross-peaks of the amide groups induced by successive additions of the two compounds. We observed a gradual shift of the EphA4 HSQC peaks, correlating with the increased concentrations of the two molecules, which suggests that the free and bound EphA4 molecules undergo a fast exchange on the chemical shift timescale. This allowed assignment of the resonances in the complex by following the shifts in the EphA4 cross-peaks upon gradual addition of increasing amounts of two compounds.

The shift of many HSQC peaks was largely saturated at a molar ratio of 1:6 (EphA4:compounds). Therefore, to identify the interaction surfaces, the chemical shift differences (CSD) between the free state and the bound state in the presence of a 6-fold excess of the two compounds were calculated and plotted versus the EphA4 sequence (Figures 3.25c and 3.25d). The two compounds induced a similar shift patterns for the EphA4 residues and most EphA4 residues did not experience large chemical-shift perturbations, indicating that the two compounds did not alter the overall structure of EphA4, consistent with the far-UV CD results shown in Figure 3.25a. Interestingly, we observed only 8 resonance peaks with significant CSD (deviating more than 2.5 standard deviations from the mean CSD), including residues Ile31-Met32 and Ile39 located in the D-E loop, Gln43 in the E β-strand, and Asp123 and Ile131-Gly132 in the J-K loop. Since the E β-strand and the D-E and J-K loops have been previously demonstrated to be key components of the high-affinity ephrin-binding channel of the Eph receptors, the NMR titration results thus suggest that the two molecules bind to the high-affinity ephrin-binding channel of EphA4.

Further attempts to identify the intermolecular NOE connectivity between EphA4 and the ligand molecules were not successful because the presence of the molecules appeared to induce significant intermediate conformational exchanges on the NMR timescale over the involved side-chains, which even caused the disappearance of their own intermolecular and intraresidue NOEs. On the other hand, with progressive addition of the EphA4 protein, all ^{1}H resonance peaks of the two molecules underwent line broadening and gradual shifting in one-dimensional NMR spectra. This indicates that the free and bound forms of the two molecules were in fast exchange on the chemical shift timescale and also suggests that the entire molecules

were either directly or indirectly affected by binding to EphA4, consistent with their small size.

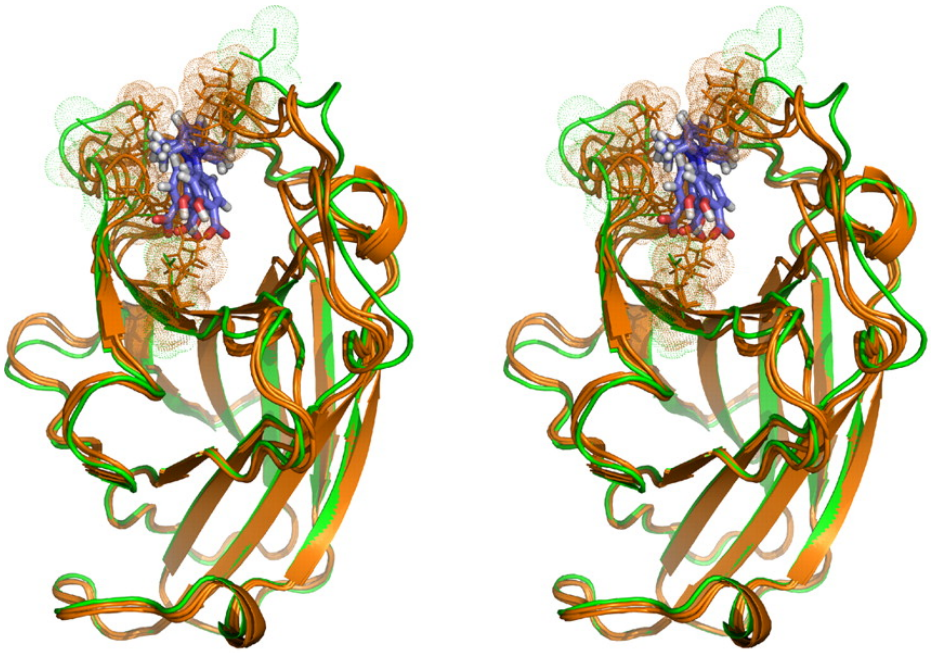
3.4.3 Molecular docking

The absence of intermolecular NOEs between the EphA4 ligand-binding domain and the two molecules made it impossible to calculate the structures of their complexes with NMR distance constraints. As an alternative, we used the HADDOCK docking strategy to construct models of EphA4 in complex with the two molecules. HADDOCK is a recent but well-established docking procedure that makes use of NMR chemical shift perturbation data in conjunction with the CNS program to drive the molecular docking of protein-protein and protein-small molecule complexes. Here we used the x-ray structures solved in our lab to built up the docking model. As described in our previous publication (Qin *et al*, 2008), in crystal, each asymmetric unit contains two EphA4 molecules A and B, which show large structural differences in the J-K loop. We have therefore analyzed the NMR Cα, Cβ and Hα chemical shifts of the EphA4 ligand-binding domain in solution and found that the four residues Phe126-Val129 in the J-K loop show a preference to form a short β-strand, as observed in molecule B. However, it appears that in solution the J-K loop might undergo an exchange between the two conformations observed in the crystal because the chemical shift deviations of the Phe126-Val129 residues are not as large as those for the well-formed β-strands in the same EphA4 ligand-binding domain.

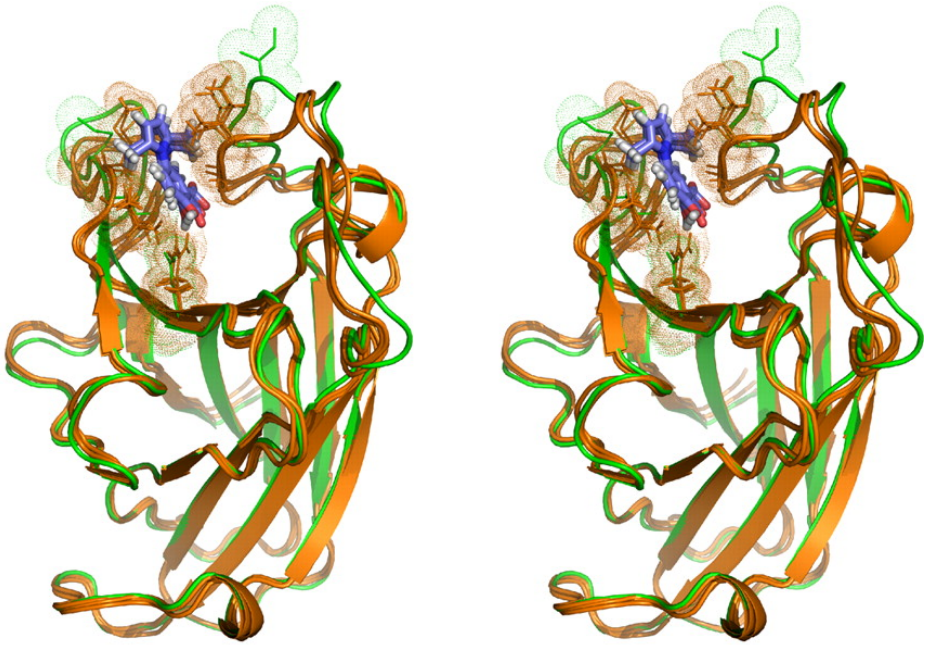
To capture the binding properties of the small molecules with EphA4 structures, we used both conformations (molecule A and molecule B) to construct the complex models by using the HADDOCK docking procedure and obtained models of

EphA4 and two compounds. From the structures obtained from the computational docking, we selected the three with the lowest energies for further display and analysis (Figures 3.26a and 3.26b). As revealed from complex models, the initial EphA4 structures only need some local conformational rearrangements to accommodate the two small molecules. As such, the average rms deviations between the 3 selected structures and the initial structure are relatively small: only ~2.0 (all protein atoms) and 1.1 Å (protein backbone atoms) for EphA4(A)-compound 1; ~2.1 (all protein atoms) and 1.2 Å (protein backbone atoms) for EphA4(A)-compound 2; ~1.9 (all protein atoms) and 1.0 Å (protein backbone atoms) for EphA4 (B)-compound 1; and ~1.8 (all protein atoms) and 1.0 Å (protein backbone atoms) for EphA4 (B)-compound 2. If not considering the D-E and J-K loops, the rms deviation values reduce to ~0.8 (all protein atoms) and 0.3 Å (protein backbone atoms) for EphA4(A)-compound 1; ~0.8 (all protein atoms) and 0.3 Å (protein backbone atoms) for EphA4(A)-compound 2; ~0.9 (all protein atoms) and 0.4 Å (protein backbone atoms) for EphA4 (B)-compound 1; and ~0.8 (all protein atoms) and 0.3 Å (protein backbone atoms) for EphA4 (B)-compound 2.

As seen in Figures 3.26 and Figure 3.27, The two small molecules interact mainly with residues Ile31-Met32 in the D-E loop, Gln43 in the D-E β-strand and Ile131-Gly132 in the J-K loops, all of which have significant chemical shift differences (CSDs) in the NMR HSQC titration (Figures 3.25c and 3.25d). In contrast, despite being set as “active residues” in the docking, residues Ile39 on the D-E loop and Asp123 on the J-K loop do not have direct contact with two small molecules in these models, implying that the chemical shift perturbations observed for these two residues are the secondary effect of binding-induced structural rearrangements of the D-E and J-K loops.

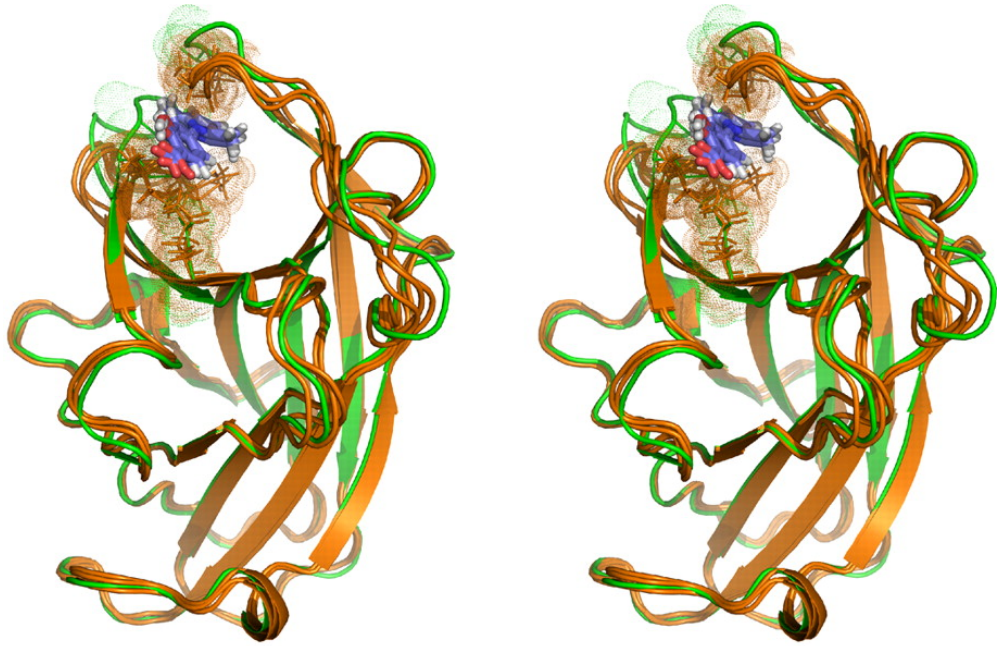


(a) EphA4(A)-Compound 1

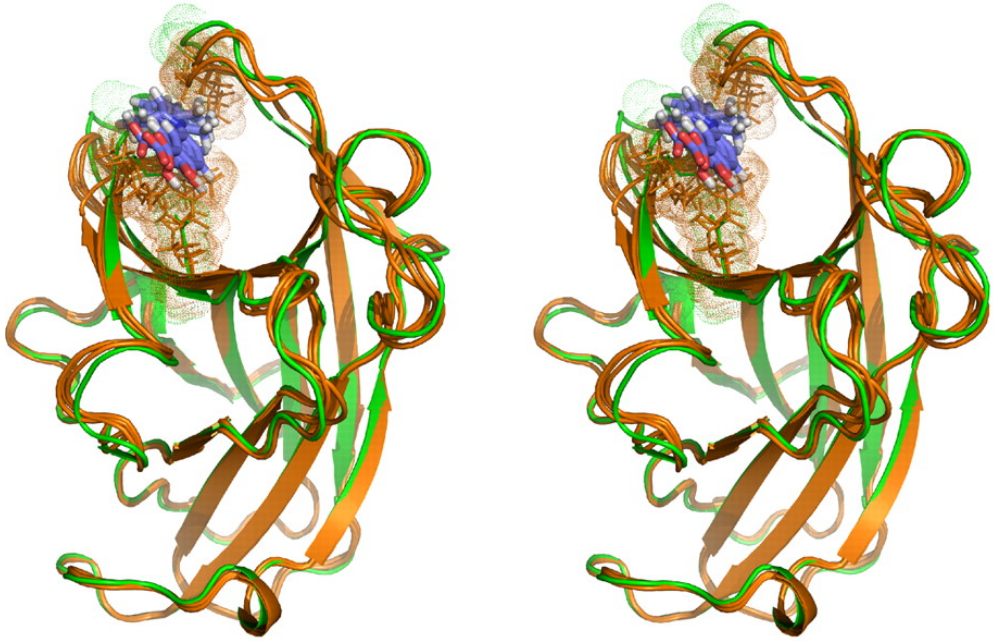


(b) EphA4(A)-Compound 2

Figure 3.26. Models of EphA4 (chainA) in complex with small molecule antagonists. (a) Stereo view of the superinposition of the unbound EphA4 (green) with its 3 selected docking models (brown) in complex with compound 1. (b) Stereo view of the superinposition of the unbound EphA4 (green) with its 3 selected docking models (brown) in complex with compound 2. Both sticks and dots are used to highlight residues Ile31-Met32 in the D-E loop, Gln43 in the E β-strand and Ile131-Gly132 in the J-K loop.



(a) EphA4(B)-Compound 1



(b) EphA4(B)-Compound 2

Figure 3.27. Models of EphA4 (chainB) in complex with small molecule antagonists. (a) Stereo view of the superinposition of the unbound EphA4 (green) with its 3 selected docking models (brown) in complex with compound 1. (b) Stereo view of the superinposition of the unbound EphA4 (green) with its 3 selected docking models (brown) in complex with compound 2. Both sticks and dots are used to highlight residues Ile31-Met32 in the D-E loop, Gln43 in the E β-strand and Ile131-Gly132 in the J-K loop.

As shown in Figure 3.28, a close examination of all the model structures reveals that the pyrrole and benzene rings of the two small molecules stack onto the hydrophobic surface formed by residues Ile31 and Met32 in the D-E loop. Moreover, the pyrrole ring is sandwiched by the hydrophobic side chains of Ile31-Met32 in the D-E loop and those of Ile131 in the J-K loop. On the other hand, one of the methyl groups on the pyrrole ring inserts into the hydrophobic patch between the Ile31 and Met32 side chains and the other methyl group is in close contact with the Ile131 side chain. These interactions emphasize the importance of the two methyl groups on the pyrrole ring, which is completely consistent with the structure-activity relationship analysis of a series of small molecules with the 2,5-dimethylpyrrolyl benzene scaffold (Noberini R *et al*, 2008).

In all 12 selected models, the carboxylic and hydroxyl groups on the benzene ring always orient towards the side chain of the EphA4 residue Gln43. Detailed analysis indicates that in all these models at least one hydrogen bond forms between the oxygen atoms of the carboxylic or hydroxyl groups and the side chain amide protons of Gln43. In some structures, even two hydrogen bonds can be identified between them. This observation may explain why removal of either the carboxylic or the hydroxyl group causes a dramatic loss in the activity of the modified compounds (Noberini R *et al*, 2008). Taken together, the docking results imply that both pyrrole and benzene rings, and the two methyl groups on the pyrrole ring, as well as the carboxylic and hydroxyl groups on the benzene ring are all critical for the binding of small molecules with a 2,5-dimethylpyrrolyl benzene scaffold to the EphA4 ligand-binding domain.

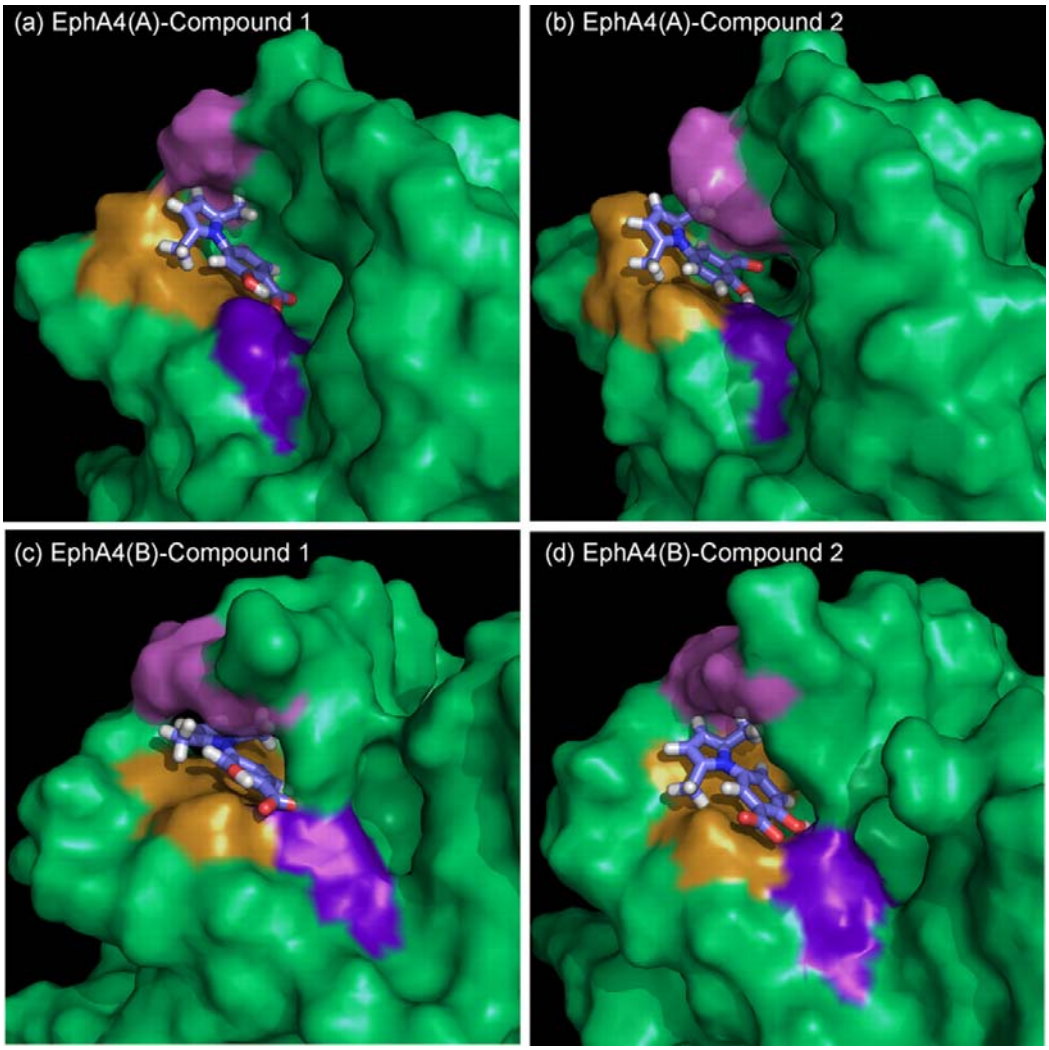


Figure 3.28. EphA4 binding pocket for the small molecule antagonists. Surface representation of the EphA4 binding pocket of the docking model with the lowest energy. (a) EphA4 structure A with compound 1; (c) EphA4 structure B with compound1; (b) EphA4 structure A with compound 2; (d) EphA4 structure B with compound2 The small molecule antagonists are represented by sticks and oxygen atoms are colored in red. Brown is used for residues Ile31-Met32 in the D-E loop, blue purple for Gln43 in the E β-strand and violet for Ile131-Gly132 in the J-K loop.

3.4.4 Discussion

The extensive involvement of the Eph receptor-ephrin interaction in various pathologies suggests that the main interface between the two proteins may serve as a promising new target for drug development. Previous studies reveal that the Eph receptor-ephrin interaction is mediated by two binding sites in the ligand-binding domain of the Eph receptor. One is a high affinity binding site, which includes a hydrophobic channel that is mainly constituted by the convex sheet of four β-strands and the D-E and J-K loops and that accommodates the protruding G-H loop of the ephrin. The other is a separate low affinity binding site (Himanen *et al*, 2007; Himanen *et al*, 2001; Himanen *et al*, 2004; Chrencik JE *et al*, 2007). In particular, the high affinity hydrophobic channel of the receptor appears to be highly amendable for targeting by small molecule antagonists. However, previously-identified small molecules including a natural product from green tea all seem to target the intracellular kinase domain of the Eph receptors. Only now two small molecules with a 2,5-dimethylpyrrolyl benzene scaffold have been successfully identified in a high throughput screen. The fact that the two compounds competitively inhibit ephrin binding to EphA4 result strongly suggests that the two compounds occupy the ephrin-binding channel, thus directly competing with ephrins in binding with the EphA4 receptor. Therefore, it was of significant interest to define the structural mechanism by which the two compounds interact with the EphA4 receptor.

To achieve this, we solved the EphA4 ligand binding domain structure in the free state by crystallography. This represents the first structure of an Eph receptor of the A class. In the crystal, each asymmetric unit contains two EphA4 molecules with a novel intermolecular packing interface. On the other hand, the EphA4 ligand-binding domain was found to be monomeric in solution by NMR spectroscopy. Interestingly,

in the crystal the two EphA4 molecules in the same unit show some large structural differences in the J-K loop due to their differential packing interactions with other EphA4 molecules in the neighboring asymmetric units.

We have used isothermal titration calorimetry, circular dichroism, NMR and computational docking to characterize the possible binding interactions of the EphA4 ligand-binding domain with the two small molecules that inhibit the binding of peptide and ephrin ligands. The isothermal titration calorimetry results show that both small molecules bind to the EphA4 ligand-binding domain with similar K_d values in the micromolar range. On the other hand, consistent with the modest binding affinity of the compounds, the circular dichroism results indicate that binding of the two small molecules does not induce significant structural changes in the EphA4 ligand-binding domain. To identify the EphA4 residues involved in the binding of the two small molecules, we have collected a large set of NMR spectra and succeeded in obtaining sequential assignments. This allowed us to identify the EphA4 residues that are significantly perturbed upon binding of the two small molecules by performing NMR HSQC titrations. Interestingly, only a few EphA4 residues showed significant perturbations upon binding, which include residues Ile31-Met32 in the D-E loop, Gln43 in the E β-strand, and Ile131-Gly132 in the J-K loop, in agreement with the small sizes of the two small molecules.

We further used the well-established HADDOCK docking procedure to construct models of the EphA4 ligand-binding domain in complex with the two small molecules. The docking results indicate that both molecules occupy a cavity of the high-affinity ephrin binding channel of EphA4 in a similar manner, by interacting mainly with EphA4 residues in the E strand and the D-E and J-K loops. The results also reveal that all three building blocks of the 2,5-dimethylpyrrolyl benzene scaffold, namely the

dimethylpyrrole ring, the benzene ring, and the carboxylic/hydroxyl groups on the benzene ring, are crucial for binding to the EphA4 ligand-binding domain. The pyrrole and benzene rings appear to play a key role in establishing stacked aromatic-hydrophobic interactions with Ile31-Met32 on the D-E loop and Ile131 on the J-K loop. The two methyl groups on the pyrrole ring further anchor the small molecules in between the D-E and J-K loops by using one methyl group to interact with the hydrophobic side chains of Ile31-Met32 and the other to interact with the hydrophobic side chain of Ile131. Furthermore, the carboxylic and hydroxyl groups on the benzene ring are involved in hydrogen bonding to the side-chain amide protons of Gln43 in EphA4, thus providing additional contacts with EphA4 as well as dictating the orientation of the small molecules in the complexes. Consequently, the docking models provide the structural rationale for the results of an extensive study on the structure-activity relationship of small molecules with a pyrrolyl benzene scaffold as EphA4 ligand-binding antagonists.

Our results shed light on how such small molecules are capable of selectively targeting only EphA4 and the closely related EphA2 receptor. Sequence alignment reveals that some of the EphA4 residues that are perturbed by the binding are not conserved in other Eph receptors. In particular, residues Ile31-Met32 are only presented in EphA4 and EphA2 but not other Eph receptors, which may be at least partly responsible for the high binding-selectivity of the two molecules for the EphA4 and EphA2 receptors. Therefore, in the future, targeting the less-conserved residues within the ephrin binding channel may represent a feasible and efficient strategy to design small molecules with the ability to specifically target an individual Eph receptor of therapeutic interest.

Our results may also explain why the two small molecules bind to EphA4 with a relatively weak affinity. First, EphA4 residues Ile31-Met32 and Ile131, which are critical for binding, are from the D-E and J-K loops. These loops are relatively flexible, as indicated by previous crystal structures and our NMR ^{15}N heteronuclear NOE data in the following section. Second, as shown in Figure 3.27a, the two small molecules only occupy a portion of the EphA4 ligand-binding channel, which in EphB2 and EphB4 is occupied by the tip of the G-H loop of the ephrin ligands (Himanen *et al*, 2007; Himanen *et al*, 2001; Himanen *et al*, 2004; Chrencik JE *et al*, 2007). In contrast, interactions occurring outside of the high-affinity binding pocket of the Eph receptor are totally absent in the case of the small molecules. These interactions include those between the ephrin G β-strand and the Eph receptor D and E β-strands and A-C loop. Even within the high-affinity binding channel, a large portion of the key Eph receptor-ephrin interactions is absent in the EphA4-small molecule complexes due to the small size of the dimethylpirrole derivatives. For example, NMR titrations did not detect strong interactions between the two small molecules and the EphA4 G and M β-strands. Furthermore, as shown in Figure 3.27b, the interaction interface between EphA4 and the two compounds is also smaller than the interaction interfaces between the EphB2 and EphB4 receptors and their respective peptide ligands (Chrencik JE *et al*, 2006; Chrencik JE *et al*, 2007). For example, the two small molecules do not interact with the EphA4 disulfide bridge linking Cys45 and Cys53, whereas this interaction was found to be conserved in all the EphB structures in complex with either ephrins or antagonistic peptides (Chrencik JE *et al*, 2007).

In conclusion, our studies confirm the binding interaction between the EphA4 ligand-binding domain and two novel small molecule antagonists with a 2,5-

dimethylpyrrolyl benzene scaffold. Furthermore, we utilized NMR titrations to map out the residues involved in the interaction and used this information to construct models of the EphA4 ligand-binding domain in complex with the two small molecules. These models provide a structural rational for the results of an extensive structure-activity study on a large set of small molecules with a pyrrolyl benzene scaffold and for the high binding selectivity but relatively weak affinity of the compounds. Based on our model, we propose that modifications to enhance interactions with the EphA4 G and M β-strands may represent a promising direction to improve the binding activity and specificity of the EphA4 antagonists with a 2,5-dimethylpyrrolyl benzene scaffold.

3.5 NMR Structure and Dynamics of EphA4 Ligand Binding Domain

NMR and crystallography are two major techniques to resolve protein 3D structure with high resolution. Crystallography is preferred by researchers because of low cost, high accuracy, and relatively less workload on data analysis. However, protein dynamic information could be lost due to crystal packing force, but dynamic information could be essential in the understanding of protein function. NMR spectroscopy can resolve this issue because NMR data is collected from solution sample. Therefore, the NMR solution structure of EphA4 ligand binding domain was also solved in my thesis. In section 3.2, we have reported 16 EphA4 LBD structures from two protein crystals, and structure comparison implies that the loops of EphA4 LBD are highly flexible. As a complementary aspect for static 3D structure, the dynamics of EphA4 ligand binding domain is also studied in this section.

3.5.1 Generation and structural properties of the EphA4 LBD

Constructs containing EphA4 ectodomain DNA sequence were transformed in *E. Coli.* System (Rosetta gammi, Invitrogen) and EphA4 LBD with correct disulfide bonds formation could be produced. Preparation of the protein was illustrated in Chapter2.

Preliminary structural study of EphA4 LBD was characterized by far-UV and near-UV CD. As seen from far-UV CD spectrum (Figure3.29a), EphA4 LBD has a negative peak at 212nm, but a positive peak at 200nm, indicating EphA4 LBD is very well behaved and mainly consists of β-strands. Near-UV CD spectrum of EphA4 LBD changed dramatically upon adding 6M GuHCl to denaturalizing the protein, indicating that EphA4 has tight tertiary packing. As seen from Figure3.29b, there is a positive peak around 290nm and a negative peak around 270nm under native

condition but these two peaks disappeared when protein was denatured by 6M GuHCl. To further visualize the structure characters of EphA4, ^{15}N isotope labelled EphA4 LBD was prepared and ^{1}H-^{15}N HSQC spectrum was collected (Figure3.29c). From the spectrum, it could be observed that EphA4 LBD has a well-dispersed HSQC spectrum with ~3.75 ppm for ^{1}H dimension and ~25ppm for ^{15}N dimension, indicating the protein is suitable for further study by NMR. However, the peak intensities of HSQC spectrun are not uniform, indicating that some regions might undergo conformational exchanges on μs-ms time scale. HSQC spectrum also revealed that EphA4 ligand binding domain only has one set of HSQC peaks, proving that EphA4 LBD is monomeric in solution. The dimerization formed in crystal is probably due to the crystallization packing force.

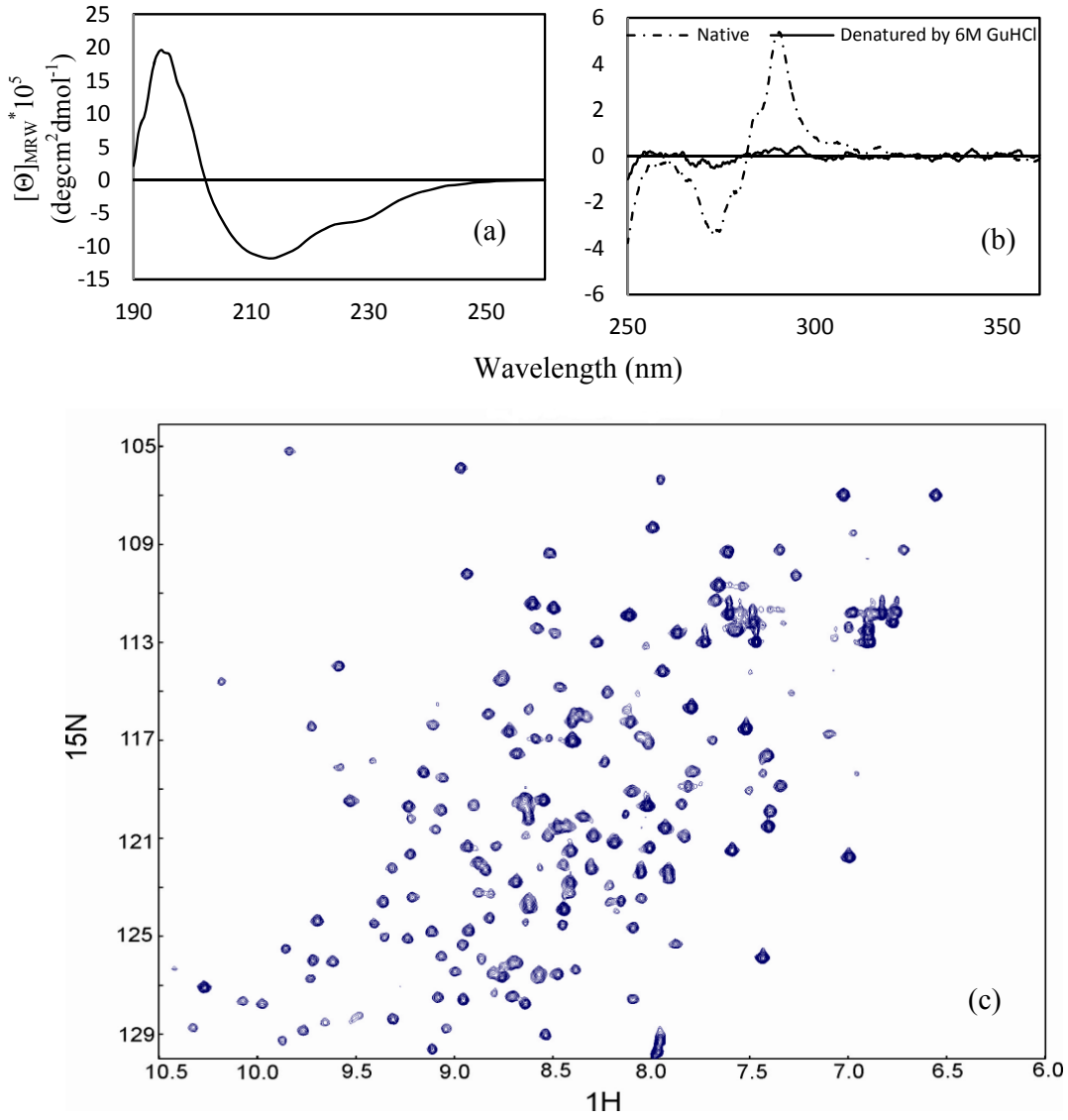


Figure 3.29. Structure characterization of EphA4 LBD by CD and NMR spectrum (a) Far-UV CD, the spectrum was collected in 10mM phosphate buffer, pH6.3; (b) Near-UV CD, the spectrum of native EphA4 LBD was collected in 10mM phosphate buffer, pH6.3. The spectrum of denatured EphA4 LBD was collected again after 6M GuHCl was added in the sample; (c) ^{1}H-^{15}N HSQC spectrum of EphA4 LBD. The spectrum was collected under 25^{o}C in 10mM phosphate buffer, pH6.3.

3.5.2 Chemical shift assignment of EphA4 LBD

High-quality heteronuclear NMR data of the 183-residue EphA4 LBD was successfully acquired on an 800 MHz spectrometer with both ^{15}N- and ^{15}N-/^{13}C-labeled protein samples. Detailed analysis of these data, in particular triple-resonance experiments led to almost complete backbone assignments for all 183 residues except 6Pro, Asn3, Glu4, Val5, Leu17, Ile20, Met50, Glu51, Asp61, Trp62, Ile63, Arg65, Gln69, Asp106, Lys107, Glu108, Arg109, Gln115, Phe116, Val131, R136, Ile137, Met138, Tyr175 whose HSQC peaks could not be observed under the experimental conditions used for structure determination. Chemical shifts for ^{13}C and non-labile hydrogens were completed for all sidechain atoms except Ile20, Met50, Glu51, Asp61, Trp62, Lys107, Glu108, Arg109, Gln115, Phe116, R136, Ile137 whose ^{15}N-^{13}C correlations are unavailable. Chemical shift assignments for all hydrophobic residues (8 Ala, 15Val, 13 Ile, 13 Leu, 8 Phe, 4 Trp, 4Met) were almost completed except for atoms of Ile20, Met50, Trp62, Ile137; HD, HG atoms of Leu7, Leu8; CD, CG, of Leu58, Ile119; HG atoms of Ile63.

3.5.3 Secondary structure characterization by chemical shift

It is well known that the chemical shift deviation of Cα, Cβ, Hα, and C=O from random coil value reflects the secondary structure of protein (Wishart DS *et al*, 1994). After finishing all the chemical shift assignment, the deviation was calculated to derive the secondary structure of EphA4 LBD to guide NOE assignment.

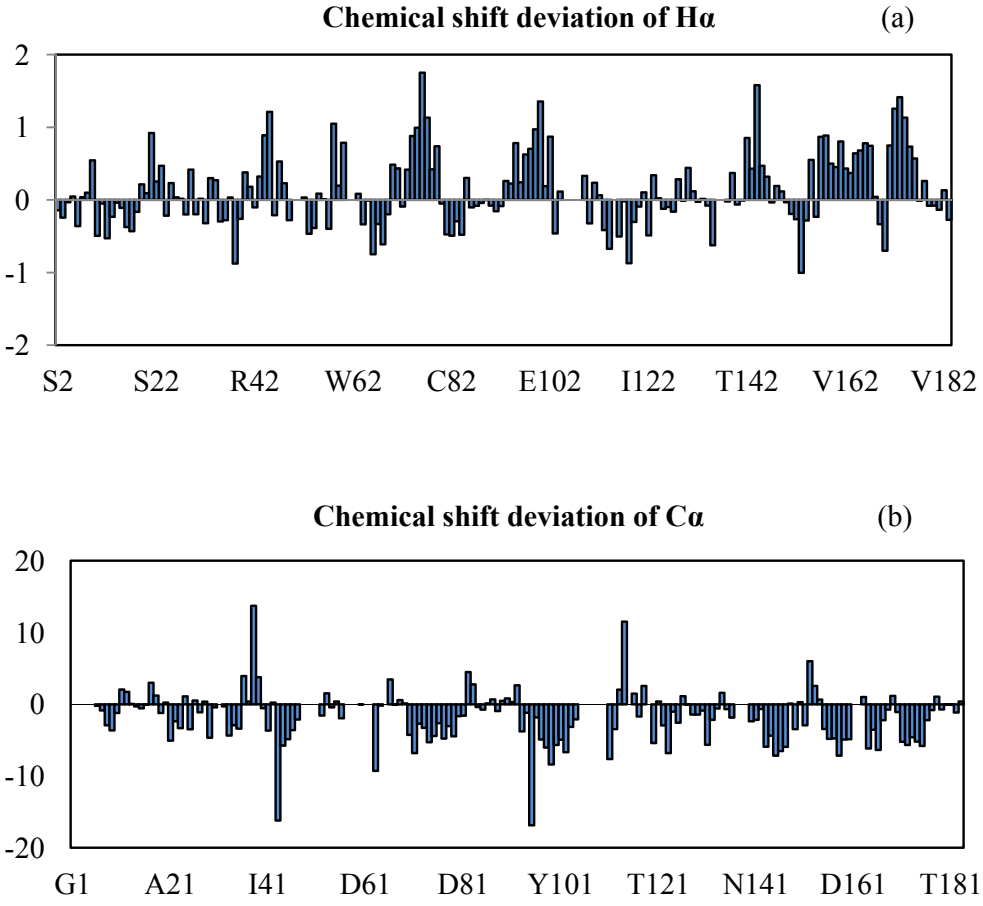


Figure3.30. EphA4 LBD chemical shift deviation from random coil value provides insights in its secondary structure. (a) Hα Chemical shift deviation of EphA4 LBD from random coil value; (b) Cα chemical shift deviation of EphA4 LBD from random coil value.

As seen from Figure3.30a, the deviation of Hα is positive in most regions and the deviation of Cα is negative in most regions, indicating that these regions are β-strands. Interestingly, the positive regions of Figure3.30a are consistent with the negative regions of Figure3.30b, implying that the secondary structure tendency derived from chemical shift is correct. From Figure3.30, it can be concluded that like other Eph receptors, EphA4 is also composed of β-strands together with some loops.

3.5.4 NMR solution structure of EphA4 LBD

After assigning the chemical shift to all possible atoms, NOE are also assigned by HSQC-NOESY and ^{13}C-edited NOESY experiments. The NMR structure of the EphA4 LBD was calculated by CYANA program taking restraints from NOE distances, TALOS-predicted dihedral angles, and hydrogen bond constraints (Table 3.8). An ensemble of 10 structures with the lowest target functions were selected from the final CYANA calculation and analyzed by procheck. As seen in Figure 3.31, the NMR structure of the EphA4 LBD adopts the same Greek key topology constituted by an 11 β-stranded barrel as previously revealed for the Eph receptor LBD by crystallographic studies. The concave sheet is comprised of strands C, F, L, H, and I, and the convex sheet of strands D, E, A, M, G, K, and J, which are connected by loops of variable length. As shown in table 3.8 and Figure 3.31, the residues over the secondary structure regions are well-defined in all 10 NMR structures, with a backbone RMS deviation of 1.38 Å.

Figure 3.31. Structure ensemble of EphA4 ligand binding domain solved by NMR spectroscopy

Table 3.8: Structural statistics for 10 selected NMR structures of the EphA4

Experimental constraints for structure calculation	
NOE restraints	
Total	1517
Intra residue	493
Sequential	481
Medium	103
Long-range	440
Dihedral angle constraints	238
	phi 123
	psi 115
CYANA target function	1.71 ± 0.32
Distance violations (>0.20 Å)	5
Dihedral angle violations (>5º)	0
Ramachandran statistics (%)	
Most favored:	75.6
Additionally allowed:	19.5
Generously allowed:	3.7
Disallowed:	1.2
Root mean square deviation (Å)	Secondary structure regions (41-46,56-60, 71-79,
	97-100, 141-147, 157-161, 166-173)
Backbone atoms	0.86± 0.22 Å
Heavy atoms	1.87 ± 0.39 Å

Figure3.32 shows the NOE distribution of EphA4 LBD. As it is mainly composed of β-strands, more than half of NOEs are sequential and long range NOEs. The residues that have a lot of long range NOEs are also the residues on β-strands. During NOE assignment, a lot of β-sheet characterized NOEs, such as $d_{\alpha N}(i,i+1)$, $d_{NN}(i,j)$, $d_{\alpha N}(i,j)$, and $d_{\alpha\alpha}(i,j)$, were identified and assigned.

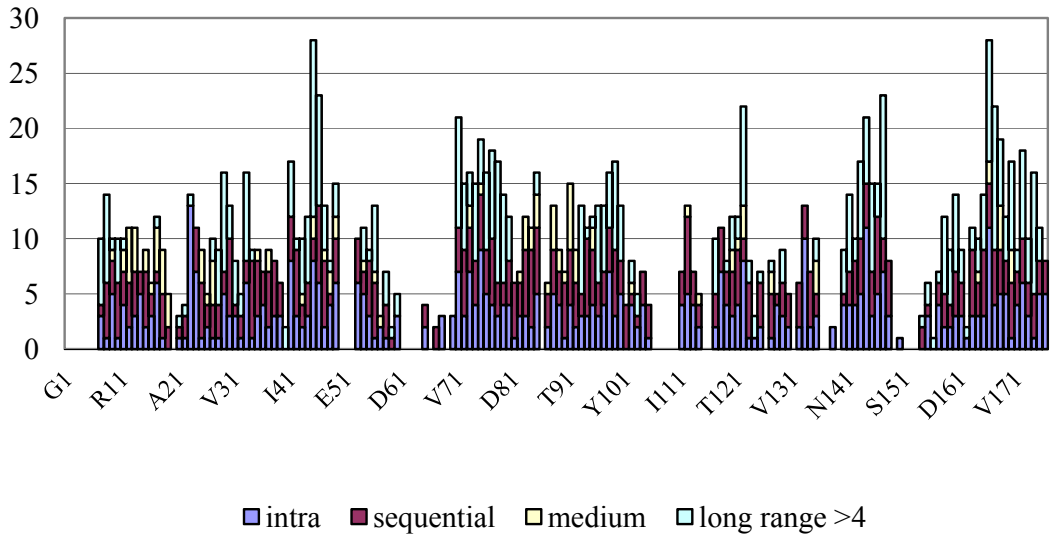


Figure 3.32. NOE distribution of EphA4 LBD

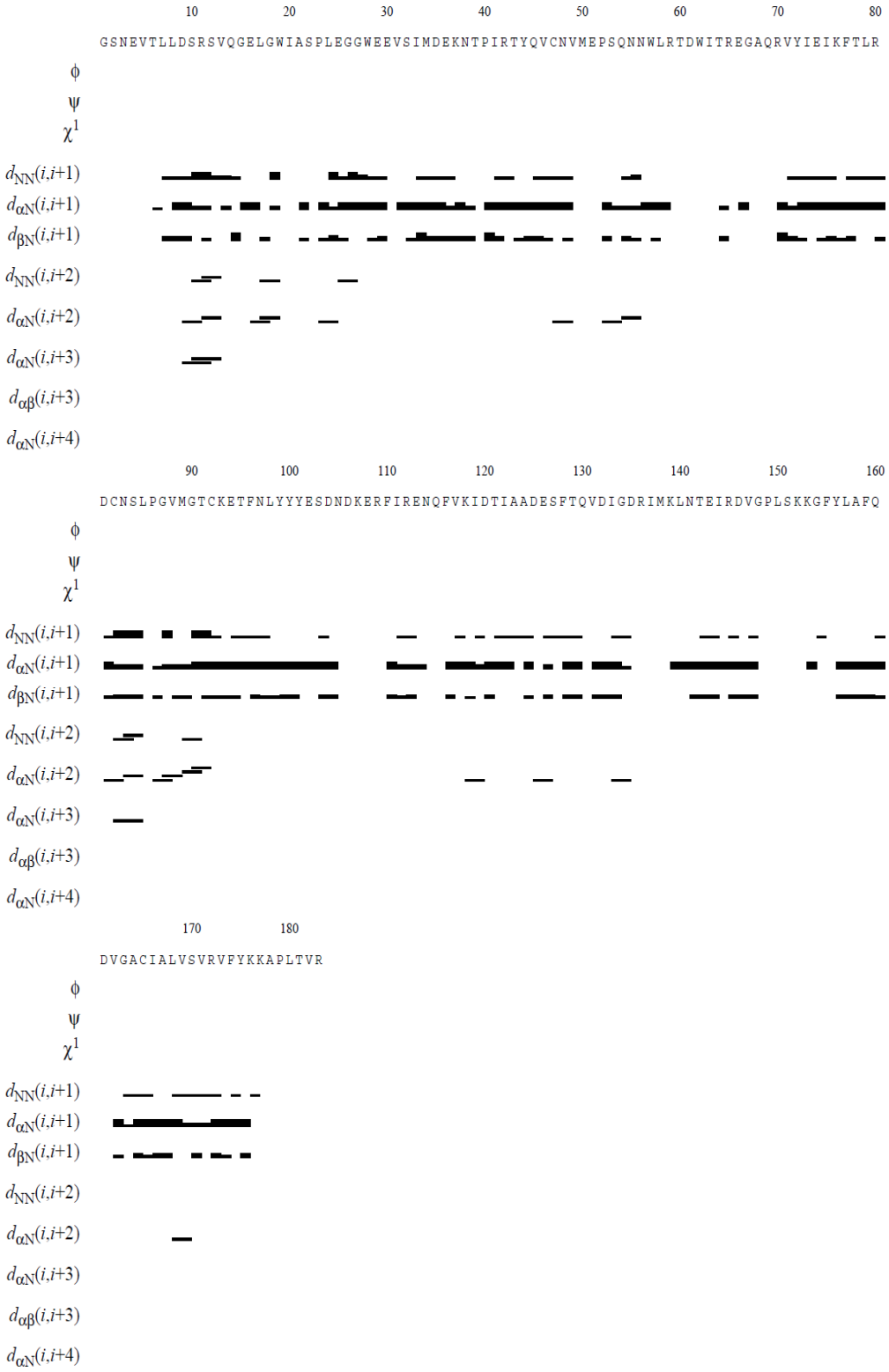


Figure3.33. Sequential NOEs plotted against amino acid sequence

Sequential NOEs plot against amino acid sequence is also provided in Figure3.33. Strong $d_{\alpha N}(i,i+1)$ and weak $d_{NN}(i,i+1)$ indicate that most regions of EphA4 LBD are β-strands.

3.5.5 Comparison of NMR solution structure and crystal structure

The crystal structures of free EphA4 LBD were also resolved by our group and Bowden. *et al*. independently (Qin *et al*, 2008; Bowden *et al*, 2009). From the structure resolved in our lab, one asymmetric unit contains two EphA4 LBD molecules packed against each other to form asymmetric dimer. In Bowden's structure, there are two EphA4 monomers in one asymmetric unit. The comparison of crystal structures shows that except the 12 highly rigid β-strands, the loops are highly flexible, especially D-E, J-K, G-H loops, they seem more flexible than other loops (Figure 3.34). Like we discussed in section 3.2, Chain A of 3CKH adopts open conformation because its binding channel is occupied by G-H loop of Chain B from neighbour AU.

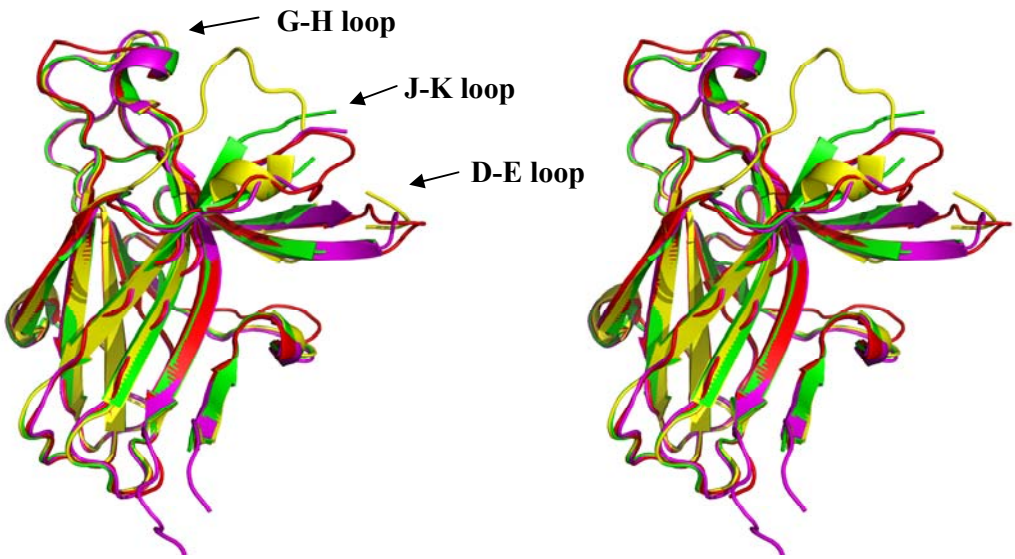


Figure 3.34. Comparison of EphA4 ligand-binding domain crystal structures. Green: Ribbon representation of EphA4 chainB (3CKH); Yellow: Ribbon representation of EphA4 chainA (3CKH); Red: Ribbon representation of EphA4 chainB (2WO1); Magentas: Ribbon representation of EphA4 chainA (2WO1).

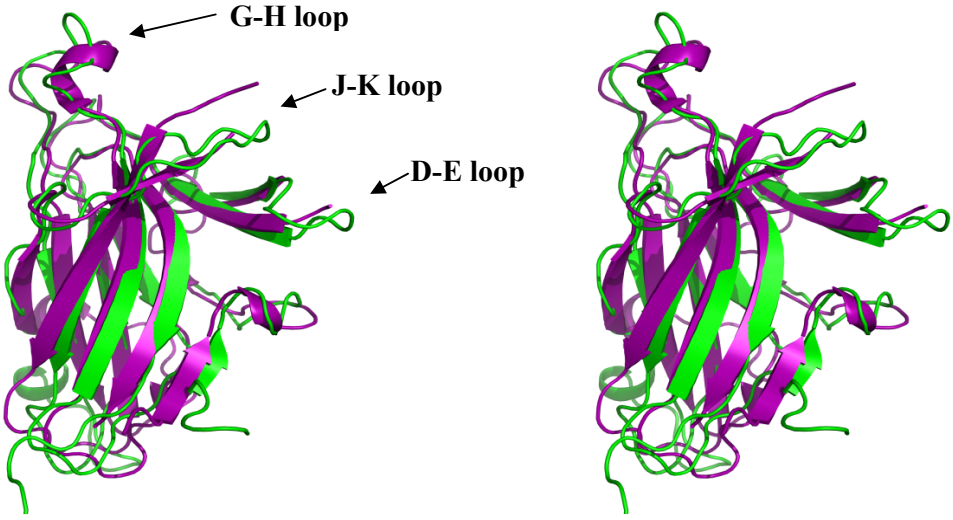


Figure 3.35. Comparison between NMR solution structure and X-ray structure. Green: average NMR structure ensemble of EphA4 LBD; Purple: Chain B (open conformation) of crystal structure 3CKH. Functional loops are indicated by arrow.

In Figure3.35, NMR structure and crystal structure of EphA4 LBD are compared. If only superimposed over the β-barrel regions, these structures superimposed very well, with a backbone RMS deviation of 1.16 Å, and all atoms RMSD is 1.651 Å. If superimposed the whole structure, the backbone RMS deviation is 2.103 Å, and all atoms RMSD is 2.946 Å. The dramatic differences come from loop regions, in particular those function related loops, such as D-E, J-K, and G-H loops. As we discussed in previous section 3.2, these loops are highly flexible and probably this character makes EphA4 ligand binding domain can bind to both subclasses of ephrin ligands, and makes dramatic conformational change to accommodate G-H loop from ephrin logands. The other significant difference from crystal structures is the orientation of the second short 3_{10} helix over residue R109 to N114, probably due to the lack of long range NOEs between the helical and β-strands nearby. In addition, only one set of HSQC peaks were observed for all the EphA4 residues, suggesting that the asymmetric dimer observed in the crystals does not exist in solution on the NMR time scale. This is consistent with the results of section3.2. Only when enough force is provided, J-K loop will transit from closed to open conformation. Otherwise, it will adopt closed conformation

3.5.6 Dynamics study of free EphA4 and analysis of relaxation data

^{15}N NMR relaxation data is shown in Figure 3.36 and Table 3.9, which are very enlightening to pinpoint the dynamics of the protein local enviroment on the pico- to nano-second timescale as well as dynamic aggregation. In particular, $\{^{1}H\}$-^{15}N steady-state NOE (hNOE) offers a reliable measure to the backbone flexibility. As seen in Figure 3.36a, most residues forming secondary structures have hNOE values

larger than 0.75, indicating that they have significantly-limited backbone motion. However, in addition to the N- and C-terminal residues with negative hNOE values which clearly indicate both termini are extremely-flexible, three additional regions, the J-K, G-H, and K-L loops, also have low hNOE values. The first region over the J-K loop residues D125 to L140 has an average hNOE value of 0.6 while the second region over the G-H loop residues L85 to Y91 holds an average hNOE value of 0.58. The third region over the K-L loop residues L150-K153 holds an average hNOE value of 0.03, indicating this region is also extremely flexible.

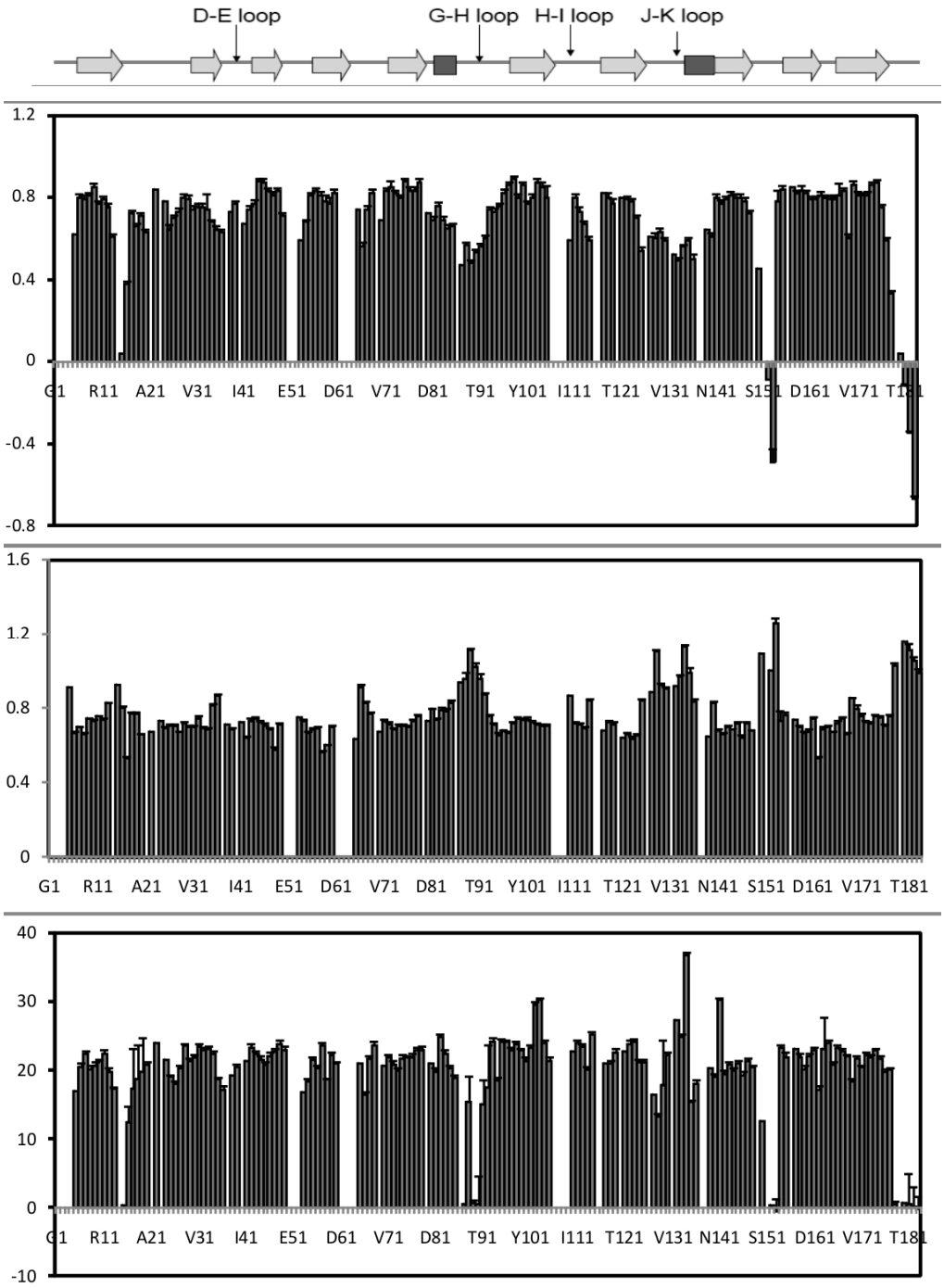


Figure 3.36. The 15N NMR backbone relaxation data of the EphA4 ligand binding domain in the 10 mM phosphate buffer (pH 6.3) at 25^{o}C collected on a 800-MHz Bruker Avance NMR spectrometer. (a). {1H}–15N steady-state NOE intensity bars ; (b) 15N T1 (longitudinal) relaxation times, (c). 15N T2 (transverse) relaxation times. The secondary structure segments identified from the NMR structure are indicated, and functionally-critical C-D, G-H, and J-K loops are labeled.

3.5.7 Modelfree analysis of relaxation data

NMR relaxation data was analyzed by "Model-Free" formulism with protein dynamics software suites (Fushman et al., 1997). Due to the overlap or/and weak intensity of many resonance peaks resulting from the relatively-large size and presence of many exposed loop residues, 127 out of 176 non-Proline peaks are appreciable for the quantitative analysis. The analysis generated squared generalized order parameters, S^2, which reflect the backbone rigidity (Table 3.10, Figure3.38). Similar with relaxation data, the regions with secondary structure have higher S^2, showing its rigidity at ps-ns time scale; while the loop regions have smaller S^2, showing its flexibility at ps-ns time scale.

Model-free analysis also outputs Rex values, which reflects conformational exchanges on μs-ms time scale. Strikingly, from figure 3.38b, the whole EphA4 LBD seems undergo conformational exchange on μs-ms time scale. Those residues over D-E, G-H, and J-K loops have larger Rex values, indicating these regions have more significant conformational exchange on μs-ms time scale. This property probably is essential for the open and closed conformations transition of EphA4 LBD.

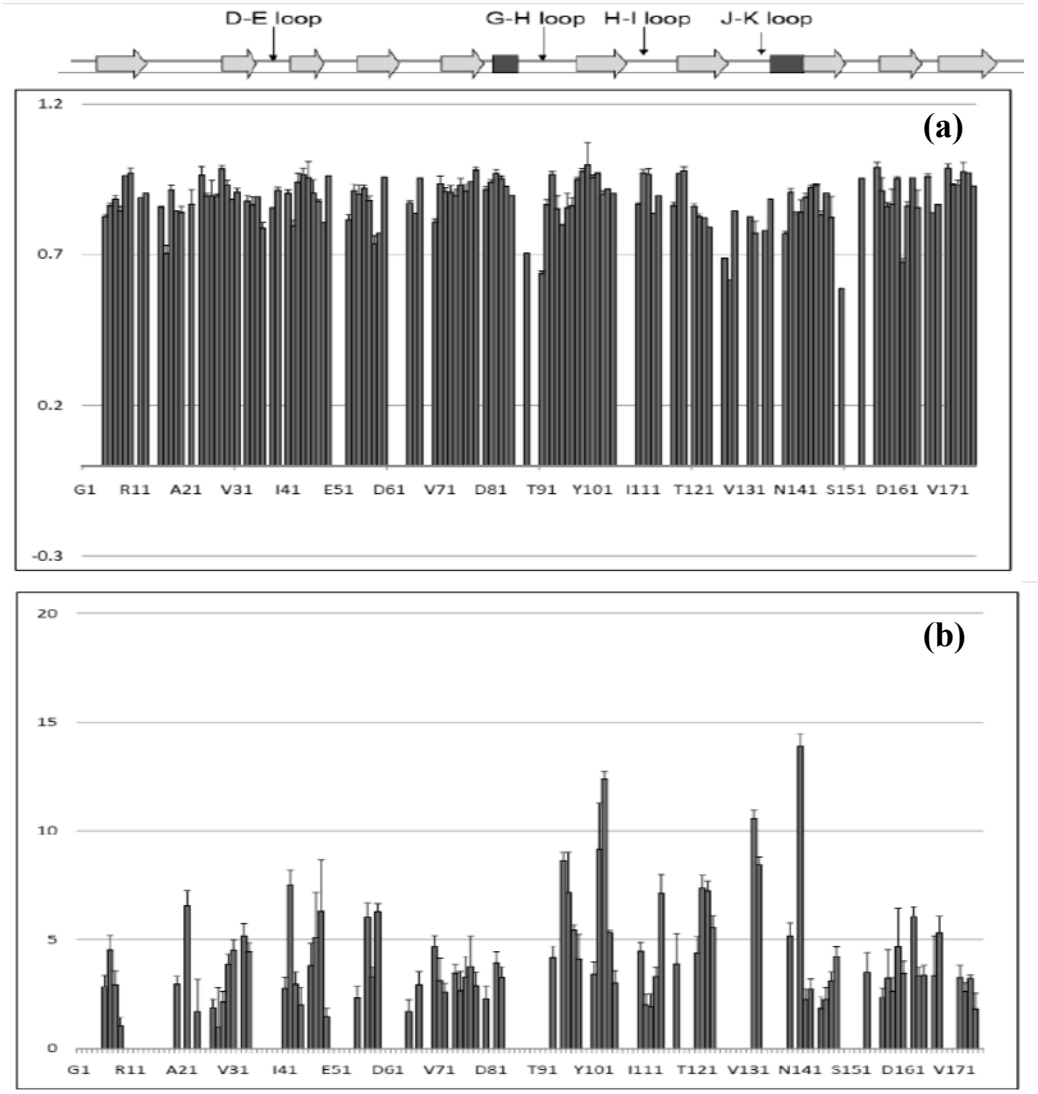


Figure3.38. Backbone order parameter (S^2) and Rex determined from ^{15}N relaxation data using fully-anisotropic model. (a) S2; (b) Rex.

3.5.8 Discussion

Due to its ability to bind with both classes of ephrin ligands, EphA4 receptor is an attractive target in understanding the binding promiscuity and specificity between Eph receptor and ephrin ligands. Several structures of free and bounded EphA4 LBD have been released, but there is no dramatic structural difference between EphA4 and other Eph receptors. In the past decades, protein dynamics has attracted attention of researchers because it is thought to play important role in protein function and is a very important supplementary aspect to static structure. In our previous study, we solved NMR structure of ephrinB2 and proved that the function involving loops of ephrinB2 ectodomain are highly dynamic in solution, which is different from X-ray structure (Ran *et al*, 2008). Although several X-ray structures of EphA4 have been resolved, local dynamic motion could be quenched by packing force during crystallization. Therefore, NMR solution structure of EphA4 LBD and its dynamic study were carried out in this section, which were helpful in understanding why EphA4 can bind both classes of ephrin ligands.

The NMR structure of EphA4 LBD shows that EphA4 LBD has rigid β-sheet barrel core and flexible loops, which is consistent with X-ray structures. The loops forming classical binding channel, D-E, J-K, and G-H loops, are more flexible than other loops. This is probably correlated with its ability to make huge rearrangement to accommodate G-H loop from ephrin ligands.

^{15}N relaxation data of EphA4 ligand binding domain was collected and the analysis of the relaxation data was carried out by Model-free formulism. Like what we observed from structure, the dynamic data shows that D-E, G-H, and J-K loops are highly dynamic in solution. Model-free analysis to relaxation data generated squared

order parameter S^2 and Rex, which reflect the dynamic motion of protein in ps-ns, and μs-ms time scale, respectively. The regions with secondary structure has higher S^2, indicating their rigidity, while loop regions has lower S^2, indicating their flexibility.

Model-free analysis also outputs Rex values, which reflects conformational exchanges on μs-ms time scale. Our results showed that the residues at D-E, G-H, and J-K loops have higher Rex, indicating these regions have conformational exchange at μs-ms time scale. This result is consistent with the conclusion in section 3.2. There are various conformations existed in 16 EphA4 LBD structures. J-K loop would adopt open or closed state depending on whether there is G-H loop present. This conformational exchange may be important for EphA4 LBD in the transition from closed to open states. In addition, the residues over other regions also have smaller Rex, indicating these regions also undergo minor conformational exchange on μs-ms time scale. Based on our results, the whole molecule of EphA4 LBD seems to undergo conformational exchange on μs-ms time scale. Probably this is the reason why EphA4 LBD can interact with different types of ephrin ligands. It is interesting that EphB2, the only another receptor that bind both subclasses of ephrin ligands, its J-K and D-E loops are also undergo dramatic conformational change. It could be suggested that the ligand promiscuity of Eph receptors could be directly correlated with structural flexibility.

Table3.9: Relaxation data of EphA4 LBD

Resid	R1	R1 error	R2	R2 error	NOE	NOE error
G1						
S2						
N3.						
E4.						
V5	0.000914	1.06E-05	0.017014	0.000538	0.62	0.0179
T6	0.000666	0.000007	0.020526	0.000485	0.8	0.0231
L7	0.000697	6.2E-06	0.022421	0.000661	0.79	0.0148
L8	0.00066	6.7E-06	0.020154	0.000615	0.81	0.0177
D9	0.000741	1.33E-05	0.020685	0.000365	0.85	0.0132
S10	0.00073	6.5E-06	0.021294	0.000531	0.77	0.0146
R11						
S12	0.00074	8.8E-06	0.019741	0.000348	0.75	0.0143
V13	0.000824	8.3E-06	0.01732	0.000443	0.61	0.0194
Q14						
G15						
E16	0.0008	8.9E-06	0.012466	0.005752	0.38	0.0163
L17	0.000532	8.6E-06	0.017359	0.005021	0.72	0.0154
G18	0.00077	6.9E-06	0.018766	0.00492	0.66	0.0134
W19	0.00077	4.3E-06	0.019842	0.00056	0.71	0.0152
I20	0.000658	1.12E-05	0.020783	0.000235	0.63	0.0159
A21						
S22	0.000675	1.06E-05	0.024048	0.000383	0.84	0.0167
P23						
L24	0.000731	9.5E-06	0.021536	0.000268	0.78	0.029
E25	0.000695	9.8E-06	0.019138	0.000323	0.64	0.0124
G26	0.000705	7.8E-06	0.018062	0.000377	0.7	0.0173
G27	0.000704	7.8E-06	0.020277	0.000345	0.73	0.0161
W28	0.00067	1.01E-05	0.023553	0.00039	0.8	0.0246
E29	0.000718	6.1E-06	0.021402	0.000466	0.79	0.0194
E30	0.000701	7.5E-06	0.021876	0.000403	0.74	0.0209
V31	0.000701	1.23E-05	0.023402	0.000358	0.75	0.0192
S32						
I33	0.000693	1.02E-05	0.023108	0.000525	0.74	0.0136
M34	0.000683	0.00001	0.022362	0.00029	0.68	0.0236
D35	0.000814	1.05E-05	0.018692	0.000471	0.64	0.0136
E36	0.000867	4.8E-06	0.017212	0.000517	0.63	0.0254
K37						
N38	0.000711	9.9E-06	0.019251	0.000409	0.73	0.0153
T39	0.000684	9.7E-06	0.020521	0.000404	0.77	0.0266
P40						
I41	0.000721	7.8E-06	0.021399	0.000521	0.67	0.0175

R42	0.000642	1.31E-05	0.023322	0.000562	0.74	0.0165
T43	0.000731	0.000009	0.02236	0.000499	0.77	0.0147
Y44	0.000746	9.7E-06	0.021585	0.000406	0.88	0.0206
Q45	0.000723	1.34E-05	0.020789	0.000626	0.87	0.0154
V46	0.000707	8.5E-06	0.022019	0.000499	0.83	0.0212
C47	0.000684	1.16E-05	0.022763	0.00049	0.81	0.019
N48	0.000578	1.03E-05	0.023887	0.00071	0.83	0.0135
V49	0.000714	0.000009	0.022884	0.000349	0.71	0.0168
M50						
E51						
P52						
S53	0.000751	7.4E-06	0.0168	0.000404	0.59	0.0131
Q54	0.00073	6.6E-06	0.01843	0.000339	0.68	0.0143
N55	0.000669	1.19E-05	0.021544	0.000414	0.81	0.0159
N56	0.000686	9.6E-06	0.020313	0.000556	0.83	0.0216
W57	0.000692	1.06E-05	0.023588	0.00016	0.81	0.0318
L58	0.000565	6.7E-06	0.018656	0.000408	0.78	0.0292
R59	0.0006	9.9E-06	0.022204	0.000273	0.77	0.018
T60	0.000698	1.18E-05	0.021011	0.00053	0.82	0.0178
D61						
W62						
I63						
T64						
R65	0.000634	1.46E-05	0.020987	0.000367	0.74	0.0235
E66	0.000912	8.3E-06	0.016488	0.000491	0.56	0.0202
G67	0.000826	0.00001	0.021627	0.000603	0.74	0.0195
A68	0.00077	1.36E-05	0.023642	0.000574	0.82	0.0151
Q69						
R70	0.00067	8.3E-06	0.020617	0.000457	0.69	0.019
V71	0.000729	1.34E-05	0.021869	0.000663	0.83	0.0281
Y72	0.000712	9.8E-06	0.020811	0.000444	0.85	0.0125
I73	0.000685	8.7E-06	0.019873	0.000611	0.82	0.0141
E74	0.000704	9.6E-06	0.021636	0.000313	0.8	0.0133
I75	0.000707	1.06E-05	0.021842	0.000501	0.88	0.0227
K76	0.000701	9.9E-06	0.021908	0.000526	0.83	0.0245
F77	0.00073	9.7E-06	0.022868	0.000711	0.83	0.0211
T78	0.000754	1.64E-05	0.022914	0.000374	0.87	0.0191
L79						
R80	0.000729	7.1E-06	0.021037	0.000561	0.72	0.0188
D81	0.000793	5.4E-06	0.019843	0.000356	0.69	0.0178
C82	0.000739	1.53E-05	0.024879	0.000501	0.76	0.0159
N83	0.000787	1.49E-05	0.02247	0.000486	0.69	0.0184
S84	0.000781	0.000016	0.020244	0.000355	0.65	0.0145
L85	0.000827	9.2E-06	0.018941	0.000476	0.66	0.0177

P86						
G87						
V88	0.000954	1.13E-05	0.015387	0.000423	0.57	0.0141
M89						
G90						
T91	0.000956	5.8E-06	0.015012	0.006258	0.56	0.017
C92	0.000874	8.7E-06	0.01751	0.000522	0.6	0.0164
K93	0.000757	1.09E-05	0.024213	0.000515	0.74	0.0177
E94	0.000709	1.47E-05	0.018543	0.000361	0.73	0.0139
T95	0.000656	8.5E-06	0.02421	0.000354	0.76	0.0202
F96	0.000676	1.09E-05	0.02408	0.000535	0.82	0.0176
N97	0.000668	6.1E-06	0.022868	0.00056	0.86	0.012
L98	0.00072	1.46E-05	0.023751	0.000216	0.89	0.0195
Y99	0.000741	1.02E-05	0.022918	0.00059	0.8	0.0174
Y100	0.000735	1.92E-05	0.021304	0.000416	0.86	0.0145
Y101	0.000735	6.7E-06	0.023294	0.000519	0.77	0.0158
E102	0.000725	1.34E-05	0.02955	0.000339	0.8	0.0202
S103	0.00071	8.2E-06	0.030225	0.000374	0.87	0.0244
D104	0.000704	0.000005	0.024055	0.000541	0.85	0.0551
N105	0.000708	9.3E-06	0.021351	0.000555	0.8	0.0142
D106						
K107						
E108						
R109						
F110	0.000865	1.28E-05	0.02268	0.000393	0.59	0.0158
I111	0.000717	1.11E-05	0.024061	0.000462	0.8	0.0215
R112	0.000712	9.2E-06	0.0234	0.000551	0.73	0.0143
E113	0.00069	8.9E-06	0.020039	0.000363	0.67	0.023
N114	0.00084	3.18E-05	0.025278	0.000798	0.59	0.0193
Q115						
F116						
V117	0.000679	0.00001	0.021042	0.000513	0.82	0.0224
K118	0.000721	1.28E-05	0.020916	0.000743	0.8	0.0227
I119	0.000713	1.08E-05	0.022499	0.000628	0.77	0.02
D120						
T121	0.00064	9.2E-06	0.022733	0.000691	0.8	0.0164
I122	0.000662	1.34E-05	0.023767	0.000491	0.79	0.0151
A123	0.000637	7.8E-06	0.02418	0.000392	0.78	0.0124
A124	0.000652	8.2E-06	0.021178	0.000518	0.7	0.0193
D125						
E126						
S127	0.000886	9.2E-06	0.016379	0.000425	0.61	0.0293
F128	0.001108	1.06E-05	0.013284	0.006522	0.6	0.0182
T129	0.000922	1.19E-05	0.017837	0.000426	0.63	0.0162

Q130						
V131						
D132	0.000918	8.3E-06	0.027361	0.000376	0.52	0.0141
I133	0.000969	1.07E-05	0.024917	0.00031	0.49	0.0114
G134	0.001131	2.63E-05	0.036838	0.000223	0.56	0.016
D135	0.000991	1.68E-05	0.015397	0.00055	0.59	0.0231
R136	0.000834	8.8E-06	0.018099	0.000373	0.5	0.0137
I137						
M138						
K139	0.000646	7.9E-06	0.020376	0.000511	0.64	0.0206
L140	0.00083	0.000009	0.01907	0.000378	0.61	0.0184
N141	0.000678	6.1E-06	0.030244	0.000583	0.8	0.0165
T142	0.000661	9.9E-06	0.019453	0.000357	0.77	0.0184
E143	0.000697	0.000005	0.020829	0.000402	0.79	0.0166
I144	0.000687	1.37E-05	0.019951	0.000484	0.81	0.0154
R145	0.000716	1.06E-05	0.021028	0.000464	0.8	0.017
D146	0.000643	5.7E-06	0.019333	0.000501	0.8	0.0192
V147	0.000719	4.6E-06	0.021316	0.000379	0.78	0.0178
G148	0.000679	8.8E-06	0.020392	0.000362	0.72	0.0477
P149						
L150	0.001095	1.16E-05	0.012572	0.006245	0.45	0.0317
S151						
K152						
K153						
G154	0.000731	2.01E-05	0.023299	0.000776	0.78	0.0161
F155	0.000762	0.000011	0.021894	0.00039	0.84	0.0204
Y156						
L157	0.00074	1.08E-05	0.023025	0.000539	0.85	0.0161
A158	0.000698	1.06E-05	0.021919	0.000565	0.82	0.0259
F159	0.000665	1.39E-05	0.020162	0.00046	0.83	0.0174
Q160	0.000675	1.05E-05	0.022111	0.000479	0.82	0.0169
D161	0.000746	1.18E-05	0.022873	0.000508	0.79	0.0132
V162	0.00053	1.66E-05	0.017221	0.004649	0.79	0.0198
G163	0.000686	4.3E-06	0.02317	0.000427	0.81	0.0123
A164	0.000701	8.1E-06	0.023987	0.000336	0.8	0.0147
C165	0.00067	1.26E-05	0.020842	0.000358	0.79	0.0191
I166						
A167	0.000744	9.7E-06	0.022756	0.000297	0.81	0.0147
L168	0.00066	7.9E-06	0.022074	0.000441	0.83	0.0246
V169	0.000851	2.13E-05	0.018397	0.000476	0.6	0.0207
S170	0.000794	1.22E-05	0.021749	0.000293	0.86	0.0166
V171	0.000758	9.5E-06	0.02049	0.000462	0.81	0.0112
R172	0.000725	1.28E-05	0.022262	0.000493	0.81	0.0164
V173	0.000716	0.000012	0.021905	0.000251	0.81	0.0137

F174	0.000755	1.07E-05	0.023005	0.000446	0.86	0.0138
Y175	0.000747	7.6E-06	0.021668	0.00048	0.87	0.0134
K176	0.000708	1.39E-05	0.019809	0.000293	0.75	0.0163

Table 3.10: Modelfree analysis results of EphA4 LBD

Residue	S^2	S^2 error	τ_e	τ_e error	Rex	Rex error	S_f^2	S_f^2 error	model	χ^2
G1										
S2										
N3.										
E4.										
V5	0.82625	0.028509	0.83962	0.13488	0	0	1	0	1	5.9682
T6	0.86266	0.010435	0.011161	0.003741	2.8438	0.53403	1	0	3	5.61E-09
L7	0.88486	0.008609	0.016403	0.003585	4.5202	0.69208	1	0	3	1.17E-09
L8	0.84618	0.008758	0.0081	0.002416	2.9319	0.63539	1	0	3	3.90E-08
D9	0.96266	0.013088	0	0	1.0599	0.36976	1	0	2	0.25557
S10	0.97236	0.01458	0.1123	0.090479	0	0	1	0	1	1.5709
R11										
S12	0.88928	0.015442	0.82584	0.094815	0	0	0.94918	0.012345	4	5.27E-07
V13	0.9056	0.026513	0.44584	0.14227	0	0	1	0	1	6.1935
Q14										
G15										
E16	0.85851	0.009792	0.3029	0.039194	0	0	1	0	1	0.92318
L17	0.70504	0.007892	0.010021	0.000795	0	0	1	0	1	0.20825
G18	0.91611	0.002771	0.59446	0.018815	0	0	1	0	1	0.13388
W19	0.84552	0.025637	0.95906	0.080242	0	0	0.94713	0.020016	4	5.89E-07
I20	0.84124	0.01452	0.03921	0.00492	2.9551	0.38373	1	0	3	6.01E-08
A21										
S22	0.86755	0.019906	0	0	6.5742	0.68674	1	0	2	0.98689
P23										
L24	0.96411	0.049274	0	0	1.6889	1.4824	1	0	2	7.02E+00
E25	0.89426	0.023811	0.062015	0.019155	0	0	1	0	1	2.66E+00
G26	0.89612	0.028636	0.046374	0.0183	0	0	1	0	1	4.33E+00
G27	0.8946	0.010574	0.035089	0.005354	1.8644	0.39327	1	0	3	4.84E-09
W28	0.98548	0.052053	0	0	0.97858	1.8217	1	0	2	5.4756
E29	0.93125	0.008797	0.028657	0.007785	2.1418	0.48923	1	0	3	3.37E-08
E30	0.88392	0.010315	0.028976	0.004923	3.8852	0.45679	1	0	3	5.28E-09
V31	0.90838	0.016501	0.033823	0.010266	4.5083	0.48555	1	0	3	1.25E-08
S32										
I33	0.87911	0.01309	0.02746	0.004493	5.1792	0.57426	1	0	3	1.35E-08
M34	0.86546	0.014388	0.03731	0.00615	4.4535	0.41499	1	0	3	1.66E-08
D35	0.89232	0.018912	0.64932	0.1056	0	0	1	0	1	3.34E+00
E36	0.78846	0.026162	0.85936	0.070987	0	0	0.94305	0.019941	4	7.34E-06

K37										
N38	0.85634	0.019535	0.77051	0.099185	0	0	0.91999	0.015255	4	4.20E-07
T39	0.9144	0.017239	0.56409	0.27693	0	0	0.94601	0.013877	4	1.09E-07
P40										
I41	0.90388	0.010458	0.060749	0.009259	2.7461	0.53569	1	0	3	2.99E-08
R42	0.7945	0.016333	0.014795	0.002385	7.5169	0.67689	1	0	3	1.32E-08
T43	0.94197	0.012036	0.046051	0.014714	2.957	0.55739	1	0	3	3.98E-09
Y44	0.96674	0.021099	0	0	2.0128	0.80338	1	0	2	1.2823
Q45	0.95547	0.029795	0	0	0	0	1	0	0	7.855
V46	0.90622	0.020396	0	0	3.8169	1.0229	1	0	2	1.57E+00
C47	0.87842	0.055019	0	0	5.1034	2.0727	1	0	2	6.0101
N48	0.80755	0.043112	0	0	6.3004	2.3687	1	0	2	4.07E+00
V49	0.96194	0.005443	0.27493	0.090494	1.4695	0.36343	1	0	3	3.41E-09
M50										
E51										
P52										
S53	0.81646	0.021474	0.46554	0.060948	0	0	0.90803	0.015153	4	6.25E-07
Q54	0.91281	0.01783	0.066005	0.018455	0	0	1	0	1	2.39E+00
N55	0.90151	0.016543	0.012955	0.004524	2.3213	0.5587	1	0	3	6.59E-09
N56	0.92102	0.021973	0	0	0	0	1	0	0	7.438
W57	0.8811	0.029742	0	0	6.0227	0.68009	1	0	2	2.1405
L58	0.73744	0.00995	0.006689	0.001835	3.2887	0.43411	1	0	3	3.91E-08
R59	0.77225	0.013584	0.009256	0.001984	6.2916	0.37487	1	0	3	4.09E-08
T60	0.95797	0.025547	0	0	0	0	1	0	0	7.28E+00
D61										
W62										
I63										
T64										
R65	0.87241	0.019258	0.023851	0.006669	1.6936	0.54061	1	0	3	8.32E-09
E66	0.83881	0.014183	0.63695	0.052238	0	0	1	0	1	1.56E+00
G67	0.95522	0.007822	0.47901	0.1044	2.9377	0.60645	1	0	3	1.70E-09
A68										
Q69										
R70	0.80893	0.010834	0.023864	0.002816	4.6835	0.51785	1	0	3	2.51E-08
V71	0.9364	0.024207	0	0	3.1194	1.0312	1	0	2	8.98E-01
Y72	0.91083	0.009952	0	0	2.595	0.39818	1	0	2	0.27627
I73	0.90797	0.025091	0	0	0	0	1	0	0	11.074
E74	0.89671	0.012665	0.015681	0.004125	3.4656	0.40844	1	0	3	8.03E-09
I75	0.93046	0.021306	0	0	2.65	0.88734	1	0	2	1.0489
K76	0.91194	0.020948	0	0	3.2835	0.95597	1	0	2	1.1845
F77	0.94321	0.023751	0	0	3.7667	1.4086	1	0	2	1.5911
T78	0.98115	0.022462	0	0	2.8945	0.6011	1	0	2	0.48748
L79										
R80	0.91744	0.009739	0.051355	0.008231	2.2778	0.59966	1	0	3	5.16E-08

D81	0.94156	0.009146	0.46179	0.079331	0	0	1	0	1	1.3736
C82	0.97064	0.008418	0.38528	0.19173	3.9369	0.51657	1	0	3	7.48E-09
N83	0.95282	0.008317	0.2668	0.1149	3.2566	0.49025	1	0	3	1.48E-08
S84	0.92708	0.012869	0.46906	0.10666	0	0	1	0	1	8.52E-01
L85	0.89739	0.00864	0.67816	0.054505	0	0	1	0	1	0.73315
P86										
G87										
V88	0.7059	0.021881	0.89293	0.035437	0	0	0.9299	0.01726	4	7.38E-06
M89										
G90										
T91	0.63842	0.08765	0.96952	0.14306	0	0	0.89499	0.14219	4	6.57E-08
C92	0.86615	0.010809	0.64276	0.047222	0	0	1	0	1	1.091
K93	0.96682	0.008411	0.16228	0.11134	4.1635	0.5401	1	0	3	9.66E-09
E94	0.85187	0.01735	0.6843	0.10913	0	0	0.9082	0.013357	4	4.46E-07
T95	0.79931	0.010939	0.012779	0.002526	8.6239	0.41107	1	0	3	3.03E-07
F96	0.85534	0.043095	0	0	7.1763	1.8604	1	0	2	4.2988
N97	0.86235	0.003299	0	0	5.4494	0.24231	1	0	2	0.078673
L98	0.95026	0.049303	0	0	4.1065	1.1519	1	0	2	2.9034
Y99	0.97732	0.02712	0.66478	0.82157	0	0	1	0	1	7.29E+00
Y100	0.99859	0.006959	0	0	0	0	1	0	0	0.39886
Y101	0.95679	0.009153	0.064658	0.020738	3.423	0.55119	1	0	3	3.42E-08
E102	0.9725	0.075627	0	0	9.151	2.1369	1	0	2	7.9378
S103	0.90087	0.008639	0	0	12.392	0.34884	1	0	2	0.30473
D104	0.91951	0.001208	0	0	5.3305	0.10206	1	0	2	1.50E-02
N105	0.90463	0.011962	0.017111	0.004631	3.0111	0.57583	1	0	3	2.30E-08
D106										
K107										
E108										
R109										
F110	0.86773	0.00888	0.62316	0.037652	4.4702	0.41888	1	0	3	9.57E-09
I111	0.97271	0.007751	0.72522	0.28209	2.0106	0.48978	1	0	3	1.56E-08
R112	0.96752	0.005271	0.25771	0.11013	1.9117	0.60599	1	0	3	1.50E-08
E113	0.8393	0.011411	0.033123	0.003982	3.306	0.40882	1	0	3	1.71E-08
N114	0.89653	0.01884	0.47287	0.12479	7.1334	0.87516	1	0	3	8.10E-09
Q115										
F116										
V117	0.8625	0.031223	0	0	3.8939	1.3943	1	0	2	2.66E+00
K118	0.97128	0.010346	0.060763	0.027332	0	0	1	0	1	1.94E-01
I119	0.97923	0.01256	0.21291	0.53714	0	0	1	0	1	2.53E+00
D120										
T121	0.86012	0.012558	0.010576	0.003014	4.3983	0.75747	1	0	3	4.45E-09
I122	0.82666	0.017254	0.010298	0.002357	7.362	0.61042	1	0	3	6.18E-09
A123	0.82307	0.010193	0.011204	0.001813	7.2519	0.44216	1	0	3	2.67E-08
A124	0.79136	0.010436	0.01994	0.002542	5.5506	0.56475	1	0	3	1.77E-09

D125										
E126										
S127	0.68812	0.018883	1.004	0.075588	0	0	0.89483	0.016243	4	5.16E-07
F128	0.61639	0.08323	1.1083	0.14888	0	0	0.91465	0.13342	4	4.38E-08
T129	0.84613	0.000918	0.81071	0.004449	0	0	1	0	1	5.76E-03
Q130										
V131										
D132	0.82665	0.006998	0.60905	0.020703	10.566	0.39314	1	0	3	9.48E-09
I133	0.77191	0.008169	0.70515	0.018693	8.4515	0.34145	1	0	3	1.06E-07
G134										
D135	0.78095	0.040312	0.85263	0.14591	0	0	1	0	1	5.31E+00
R136	0.88488	0.011647	0.35204	0.057898	0	0	1	0	1	1.386
I137										
M138										
K139	0.77025	0.010179	0.024934	0.00235	5.1873	0.58971	1	0	3	1.11E-08
L140	0.90766	0.016974	0.47751	0.10187	0	0	1	0	1	2.57E+00
N141	0.83646	0.008132	0.009469	0.002409	13.888	0.57675	1	0	3	8.65E-09
T142	0.84205	0.013293	0.014729	0.003047	2.2453	0.47573	1	0	3	3.27E-08
E143	0.89023	0.007274	0.017228	0.003903	2.7149	0.47188	1	0	3	1.64E-08
I144	0.92414	0.041258	0	0	0	0	1	0	0	1.68E+01
R145	0.92987	0.014353	0.023707	0.008923	1.8283	0.55395	1	0	3	1.00E-08
D146	0.83152	0.00833	0.008771	0.002363	2.2503	0.54698	1	0	3	6.97E-09
V147	0.90575	0.007054	0.023891	0.005004	3.1097	0.40973	1	0	3	9.40E-09
G148	0.8238	0.014296	0.02147	0.006532	4.2205	0.45274	1	0	3	1.02E-08
P149										
L150	0.5881	0.06951	0.84562	0.10992	0	0	0.93047	0.12252	4	2.12E-08
S151										
K152					0	0		0	1	43.242
K153										
G154	0.95398	0.021698	0.052013	0.14572	3.4949	0.92045	1	0	3	3.38E-09
F155										
Y156										
L157	0.9894	0.009176	0	0	2.3618	0.39307	1	0	2	0.18361
A158	0.91298	0.029601	0	0	3.2142	1.3289	1	0	2	2.01E+00
F159	0.85978	0.019071	0.005117	0.002026	2.6322	0.62154	1	0	3	1.10E-08
Q160	0.86723	0.043452	0	0	4.6738	1.7784	1	0	2	4.69E+00
D161	0.95254	0.014636	0.043986	0.041151	3.4554	0.55721	1	0	3	3.44E-09
V162	0.67617	0.051164	0	0	0	0	1	0	0	1.19E+01
G163	0.86121	0.006121	0.009314	0.002158	6.067	0.43884	1	0	3	7.31E-08
A164	0.95602	0.01132	0.037871	0.016668	3.3313	0.40789	1	0	3	3.52E-08
C165	0.85621	0.016105	0.012595	0.003561	3.3761	0.46122	1	0	3	1.23E-08
I166										
A167	0.95948	0.059148	0	0	3.3542	1.8357	1	0	2	10.053
L168	0.84091	0.016142	0	0	5.3143	0.78005	1	0	2	1.16E+00

V169	0.86622	0.010558	0.65748	0.048176	0	0	1	0	1	3.21E-01
S170										
V171	0.98741	0.000109	0.22031	0.012016	0	0	1	0	1	0.00058
R172	0.9319	0.017103	0.020249	0.009528	3.2486	0.58878	1	0	3	1.17E-08
V173	0.9329	0.015045	0.020349	0.008336	2.6245	0.39565	1	0	3	2.81E-08
F174	0.97678	0.005069	0	0	3.2018	0.192	1	0	2	0.059311
Y175	0.97377	0.014547	0	0	1.8074	0.74914	1	0	2	0.98451
K176	0.92838	0.031314	0.045302	0.029105	0	0	1	0	1	3.6082

Chapter IV CONCLUSION AND FUTURE WORK

4.1 Summary

My thesis is composed of two sections: Structural and interaction study of Nogo-A and structural and interaction study of EphA4 receptor and its ephrin ligands, small molecule antagonists, as well as dynamics study to free EphA4.

Nogo-A has been proposed to play key roles in a variety of important biological processes. More recently, a change in Nogo-A level has been observed in several neural diseases such as stroke, Alzheimers' and Amyotrophic lateral sclerosis (ALS). However, the mechanism regulating Nogo-A levels in cells remains unknown. By using T7 phage display, our collaborators identified an E3 ubiquitin ligase WWP1, which can interact with Nogo-A. Further study implied that the interaction may trigger the ubiquitination and degradation of Nogo-A. In the present work, we gave a detailed picture on how WWP1 interacts with Nogo-A proline rich domain.

Eph/ephrin comprises the largest protein kinase family. The bidirectional signaling pathways initiated by their interaction are involved in various bioprocesses. EphA4 receptor attracts intensive study not only for its important role in function mediation, but also for its ability to bind both ephrin ligands types. However, there is still limited structure information on this receptor and its complexes. This thesis therefore gave substantial research on EphA4 ligand binding domain structure, the structures of EphA4 complexes with ephrin ligands, small organic antagonists, as well as its dynamics study by NMR.

4.2 Key contributions

WWP1 interact with Nogo-A through interaction between WW domains and PPxY motif

WWP1 and its isolated 4 WW domains bind with Nogo-A proline rich motif with a K_d of micro molar range. NMR structure of WW4 domain was resolved, and its complex model with Nogo-A peptide was derived by molecular docking. For the first time, we present the evidence that WWP1 is a novel interacting partner of Nogo-A. Following ubiquitination and protein degradation maybe one of the important mechanisms in regulating Nogo-A levels and functions in disease states. Our study provides rationales as well as template for further design of molecule to intervene in the WWP1-Nogo-A interaction and further regulate Nogo-A protein level by controlling its ubiqutination.

16 EphA4 LBD crystal structures show highly dynamic loops and dramatic conformational exchange of J-K loop

16 structures of EphA4 ligand binding domain from two protein crystals were resolved with resolution of 2.4Å and 2.9Å. These crystals were special because one asymmetric unit contained 8 molecules, which has never been observed in previous works. This observation enabled us to capture the images of 16 EphA4 molecules at the same time. Strikingly, 16 EphA4 structures are almost identical over the secondary structure regions but show high variations over the D-E, G-H and J-K loops. Based on this result, it is likely that, even for the free EphA4 ligand domain, many different conformations coexist over the loop regions, which may be ready for binding to different ligands with slight rearrangements. In addition, through the careful analysis to loop conformation and contact between molecules, we found that the transition from closed state to open state needs interaction from other molecule. Otherwise, the molecule will prefer closed state with flexibility at certain degree.

Based on our results, the interaction between Eph receptors and ephrin ligands seems to follow induce-fit mechanism. How does the external interaction help the transition of J-K loop from closed to open conformation is still under investigation. Due to the poor resolution of protein crystals, electron density map of some residues of several critical loops were missed or only backbone could be traced, but those missed map will not affect our conclusion because those visible residues have given enough electron density map to define the loop conformations. In addition, the residues on J-K loop defining open or closed states are from A121 to S126, and this fragment is absent in only two structures among these 16 structures, and the rest 14 structures show clearly which state the structure adopts. These results also suggest how important protein dynamics is in understanding protein-protein interaction and it could be interesting to investigate further in future.

A newly discovered interaction surface by crystal structure of EphA4/ephrinB2

A very interesting phenomenon of interaction between Eph receptors and ephrins ligands is that Eph receptors interact promiscuously with ephrins of the same class. They rarely interact with ephrins of the other class. However, EphA4 can bind all ephrin-A and B ligands. The distinctive ability of EphA4 makes it an attractive model to understand the structural principles underlying the selectivity versus promiscuity of Eph receptor-ephrin interactions. In our study, we presented the first EphA4-ephrin complex structure and identified a polar contact region structurally separated from the ephrin-binding channel, which is critical for EphA4-ephrin-B2 binding. The results showed that EphA4 uses different strategies for binding ephrin-A versus ephrin-B ligands, thus achieving remarkable promiscuity. We also characterized the EphA4-ephrin-B2 complex in solution by NMR spectroscopy, which represents the first NMR visualization of an Eph-ephrin complex. These findings contribute to our

understanding of the distinctive binding determinants that characterize selectivity versus promiscuity of Eph receptor-ephrin interactions and suggest that diverse strategies may be needed to design antagonists for effectively disrupting different Eph-ephrin complexes.

First picture on the structure study of EphA4 ligand binding domain complex with small molecule antagonist

Due to the critical role in wide function spectrum, Eph receptors and ephrins ligands, have become a very important target for drug design. The high affinity binding pocket of Eph receptors is a very attractive target for designing of small molecule antagonist to block ephrins ligand binding. In my thesis, we presented the first small molecule antagonists that can inhibit EphA4 and EphA2, and gained structure insight into the binding interaction between small molecules and EphA4. Our study gave solid evidence to confirm that these compounds indeed bind to the classical binding pocket of EphA4 ligand binding domain. The molecular docking models provided a structural rational for the results of an extensive structure-activity study on a large set of small molecules with a pyrrolyl benzene scaffold and for the high binding selectivity but relatively weak affinity of the compounds. Based on our results, to optimize the small antagonists to achieve higher affinity and specificity will be an interesting direction to go.

First NMR solution structure of EphA4 LBD and dynamics study to free EphA4 LBD

X-ray crystallography has become the most important tool in determining the protein structure, but the packing force during crystallization will quench the local motion of protein. NMR spectroscopy is capable of providing solution structure and dynamics information of protein, further reveals the structure function relationship. In this study, we resolved the first NMR solution structure of Eph receptors. This

structure shows that the loops of EphA4 are highly dynamic, especially functional related D-E and J-K loops. This results also explained why D-E and J-K loops adopted different orientation in different X-ray structure. Further dynamics study of free EphA4 LBD shows that the loops are highly flexible and have dramatic conformational exchange on μs-ms time scale. This property could be important for EphA4 LBD in the transition from closed to open state. While the regions with secondary structure are rigid in both ps-ns but show minor conformational exchange on μs-ms time scales. In another word, the whole molecule of EphA4 LBD undergoes minor conformational exchange on μs-ms time scales, which is probably important for EphA4 to bind different classes of ephrin ligands.

REFERENCE

Abragam, A. (1961). Principles of Nuclear Magnetism. Oxford, Clarendon Press.

Adams RH, Wilkinson GA, Weiss C, Diella F, Gale NW, Deutsch U, et al. *Genes Dev*, 13, 295-306 (1999).

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.-W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K. and Terwilliger, T.C. (2002) *Acta Cryst*. D58, 1948-1954.

Altieri, A. S., Hilton, D. P. and Byrd, R. A. (1995) Association of Biomolecular systems via pulsed field gradient NMR self-diffusion measurements. J. Am. Chem. Soc. 117, 7566-7567.

Aoki M, Yamashita T, Tohyama M. *J Biol Chem*, 279, 32643-32650 (2004).

Baldwin AJ, Kay LE. NMR spectroscopy brings invisible protein states into focus. (2009) ***Nat Chem Biol.*** 5, 808-14.

Barrios A, Poole RJ, Durbin L, Brennan C, Holder N, Wilson SW. (2003) *Curr Biol.* 13: 1571-82.

Benson MD, Romero MI, Lush ME, Lu QR, Henkemeyer M, Parada LF. (2005) Proc Natl Acad Sci U S A. 102, 10694-9.

Biervert C, Horvath E, Fahrig T. *Neurosci Lett*, 315, 25-8 (2001).

Bourgin C, Murai KK, Richter M, Pasquale EB. (2007) J Cell Biol. 178, 1295-307.

Bowden TA, Aricescu AR, Gilbert RJ, Grimes JM, Jones EY, Stuart DI. (2008) *Nat Struct Mol Biol.* 15: 567-72.

Bowden TA, Aricescu AR, Nettleship JE, Siebold C, Rahman-Huq N, Owens RJ, Stuart DI, Jones EY. Structure. 2009 Oct 14;17(10):1386-97.

Brantley DM, Cheng N, Thompson EJ, Lin Q, Brekken RA, Thorpe PE, *et al*. *Oncogene,* 21, 7011-7026 (2002).

Brantley-Sieders DM, Fang WB, Hicks DJ, Zhuang G, Shyr Y, Chen J. *Faseb J*, 19, 1884-1906 (2005).

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L. , Gros, P. , Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998). D54: 905–921.

Bundesen LQ, Scheel TA, Bregman BS, Kromer LF. *J Neurosci,* 23, 7789-800 (2003).

Caligiuri, M., Molz, L., Liu, Q., Kaplan, F., Xu, J. P., Majeti, J. Z., Ramos-Kelsey, R., Murthi, K., Lievens, S., Tavernier, J., and Kley, N. (2006) *Chem.Biol.* 13, 711–722

Chen MS, Huber AB, van der Haar ME, et al.. *Nature* 2000;403:434-9.

Chong, P. A., Lin, H., Wrana, J. L., and Forman-Kay, J. D. (2006) *J. Biol. Chem. 281*, 17069–17075.

Chrencik JE, Brooun A, Kraus ML, Recht MI, Kolatkar AR, Han GW, Seifert JM, Widmer H, Auer M, Kuhn P. (2006) *J Biol Chem.* 281, 28185-92.

Clore, G. M.,Szabo, A.,Bax, A.,Kay, L. E.,Driscoll, P. C. & Gronenborn, A. M. (1990). *J. Am. Chem. Soc*. 112, 4989-4991.

Collaborative computational project, No. 4. 1994.. *Acta Cryst*. D50, 760-763.

Conover JC, Doetsch F, Garcia-Verdugo JM, Gale NW, Yancopoulos GD, Alvarez-Buylla A. *Nat Neurosci*, 3, 1091-1097 (2000).

Dalva MB, Takasu MA, Lin MZ, Shamah SM, Hu L, Gale NW, *et al*. *Cell*, 103, 945-956 (2000).

De Vries SJ, van Dijk AD, Krzeminski M, van Dijk M, Thureau A, Hsu V, Wassenaar T, Bonvin AM. (2007). *Proteins.* 69, 726-733.

De Winter F, Oudega M, Lankhorst AJ, et al. *Exp Neurol* 2002;175:61-75.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995), *J. Biomol. NMR* 6, 277-293.

Depaepe V, Suarez-Gonzalez N, Dufour A, Passante L, Gorski JA, Jones KR, *et al*. *Nature*, 435, 1244-1250 (2005).

Dodelet VC, Pasquale EB. *Oncogene*, 19, 5614-9 (2000).

Dominguez C, Boelens R, Bonvin AM, (2003). *J. Am. Chem. Soc*., 125, 1731-1737.

Du, J., Fu, C., and Sretavan, D. W. (2007) Current pharmaceutical design 13, 2507-2518

Einbond, A., and Sudol, M. (1996). *FEBS Lett. 384*, 1–8.

Emsley, P., and K. Cowtan. (2004). *Acta Crystallogr*. D 60:2126–2132.

Ethell IM, Irie F, Kalo MS, Couchman JR, Pasquale EB, Yamaguchi Y. *Neuron*, 31, 1001-1013 (2001).

Fabes J, Anderson P, Yanez-Munoz RJ, Thrasher A, Brennan C, Bolsover S. *Eur J Neurosci*, 23, 1721-30 (2006).

Fan D, Zheng Y, Yang D, Wang J. *J Biol Chem*. 2003 Jun 6;278(23):21212-20.

Fischer E. (1894). "Einfluss der Configuration auf die Wirkung der Enzyme". Ber. Dt. Chem. Ges. 27: 2985–93.

Flanagan, J. G., and Vanderhaeghen, P. (1998) Annu. ReV. Neurosci. 21, 309-345.

Fushman D, Cahill S, Cowburn D. (1997). *J. Mol. Biol.* 266 173-194;

Gale NW, Baluk P, Pan L, Kwan M, Holash J, DeChiara TM, *et al*. *Dev Biol*, 230, 151-160 (2001).

Gallarda BW, Bonanomi D, Müller D, Brown A, Alaynick WA, Andrews SE, Lemke G, Pfaff SL, Marquardt T. (2008) Science. 320, 233-6.

Gerety SS, Wang HU, Chen ZF, Anderson DJ.. *Mol Cell,* 4, 403-414 (1999).

Gil V., Nicolas O., Mingorance A. *et al*. *J Neuropathol Exp Neurol.* 65, 433-444 (2006).

Glenn Yiu and Zhigang He. *Neuronscience*. 7, 617-627 (2006).

Goldgur Y, Paavilainen S, Nikolov D, Himanen JP. (2009) *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 65: 71-4.

Goldshmit, Y., Galea, M. P., Wise, G., Bartlett, P. F., and Turnley, A. M. (2004) *J Neurosci* 24, 10064-10073

Hall JB, Fushman D, (2003). *J. Biomol. NMR* 27, 261-275;

Henkemeyer M, Itkis OS, Ngo M, Hickmott PW, Ethell IM. *J Cell Biol*, 163, 1313-1326 (2003).

Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. (2007) *Nature* 450, 964–972.

Himanen JP, Chumley MJ, Lackmann M, Li C, Barton WA, Jeffrey PD, Vearing C, Geleick D, Feldheim DA, Boyd AW, Henkemeyer M, Nikolov DB. (2004) *Nat Neurosci.* 7, 501-9.

Himanen JP, Goldgur Y, Miao H, Myshkin E, Guo H, Buck M, Nguyen M, Rajashankar KR, Wang B, Nikolov DB. (2009) *EMBO Rep*. 10: 722-8.

Himanen JP, Henkemeyer M, Nikolov DB. (1998) *Nature*. 396: 486-91.

Himanen JP, Rajashankar KR, Lackmann M, Cowan CA, Henkemeyer M, Nikolov DB.(2001) *Nature* 414, 933–938.

Himanen JP, Yermekbayeva L, Janes PW, Walker JR, Xu K, Atapattu L, Rajashankar KR, Mensinga A, Lackmann M, Nikolov DB, Dhe-Paganon S. *PNAS.* 2010 Jun 15;107(24):10860-5.

Holland SJ, Gale NW, Mbamalu G, et al. *Nature* 1996;383:722-5.

Holmberg J, Armulik A, Senti KA, Edoff K, Spalding K, Momma S, *et al*. *Genes Dev*, 19, 462-471 (2005).

Huber, A. B., Weinmann, O., Brosamle, C., Oertle, T. & Schwab, M. E. *J. Neurosci.* 22, 3553–3567 (2002).

Ilsley, J. L., Sudol, M., and Winder, S. J. (2002) *Cell Signalling 14*, 183–189.

Ireton, R. C., and Chen, J. (2005) *Curr. Cancer Drug Targets* 5, 149–157.

Ja¨ger, M., Zhang, Y., Bieschke, J., Nguyen, H., Dendle, M., Bowman, M. E., Noel, J. P., Gruebele, M., and Kelly, J. W. (2006) *Proc. Natl. Acad. Sci. U.S.A. 103*, 10648–10653.

Jansma, A. L., Kirkpatrick, J. P. Hsu, A. R., Handel, T. M. and Nietlispach, D. *J. Bio. Chem*. 2010, 285, 19, 14424-14437

Jez Fabes, Patrick Anderson, Rafael J. Ya´n˜ez-Mun˜oz, Adrian Thrasher, Caroline Brennan and Stephen Bolsover, *European Journal of Neuroscience*, 23, 1721–1730 (2006).

Johnson, B. A., and Blevins, R. A. (1994) NMRView: A computer program for the visualization and analysis of NMR data, *J. Biomol. NMR* 4, 603-614.

Jokic N., Gonzalez de Aguilar J.L., Dimou L., Lin S., Fergani A., Ruegg M.A., Schwab M.E., Dupuis L., Loeffler J.P. *EMBO Rep.* 7, 1162-1167 (2006).

Jokic N., Gonzalez de Aguilar J.L., Pradat P.F., *et al*. *Ann Neurol.* 57, 553-556 (2005).

Kanelis, V., Bruce, M. C., Skrynnikov, N. R., Rotin, D., and Forman-Kay, J. D. (2006). *Structure 14*, 543–553.

Karaman, M. W., Herrgard, S., Treiber, D. K., Gallant, P., Atteridge, C. E., Campbell, B. T., Chan, K. W., Ciceri, P., Davis, M. I., Edeen, P. T., Faraoni, R., Floyd, M., Hunt, J. P., Lockhart, D. J., Milanov, Z. V., Morrison, M. J., Pallares, G., Patel, H. K., Pritchard, S., Wodicka, L. M., and Zarrinkar, P. P. (2008) *Nat. Biotechnol.* 26, 127–132

Karplus, M. & Kuriyan, J. (2005) .*Proc. Natl. Acad.Sci.* USA 102, 6679–6685.

Katakowski M, Zhang Z, Decarvalho AC, Chopp M. *Neurosci Lett*, 385, 204-209 (2005).

Koepf, E. K., Petrassi, H. M., Ratnaswamy, G., Huff, M. E., Sudol, M., and Kelly, J. W. (1999) *Biochemistry 38*, 14338–14351.

Kolb, P., Kipouros, C. B., Huang, D., and Caflisch, A. (2008) *Proteins* 73, 11–18.

Koshland, D.E. 1958. Proc. Natl. Acad. Sci. 44: 98–104.

Kullander K, Butt SJ, Lebret JM, Lundfald L, Restrepo CE, Rydström A, Klein R, Kiehn O. (2003) *Science*. 299: 1889-92.

Kullander K, Croll SD, Zimmer M, Pan L, McClain J, Hughes V, Zabski S, DeChiara TM, Klein R, Yancopoulos GD, Gale NW. (2001) *Genes Dev.* 15: 877-88.

Kullander K, Mather NK, Diella F, Dottori M, Boyd AW, Klein R. (2001) *Neuron*. 29: 73-84.

Lafont V, Armstrong AA, Ohtaka H, Kiso Y, Mario Amzel L, Freire E. (2007) *Chem Biol Drug Des.* 69: 413-22.

Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton. (1993). *J. Appl. Cryst.*26: 283–291.

Li M, Song J. *Proteins: structure, function and bioinformatics*. (2006).

Li M., Liu J., and Song J. *Protein Sci.* 15, 1835–1841, (2006).

Li M., Shi J., Wei Z., Teng F.Y., Tang B.L., and Song J.. *Eur. J. Biochem.* 271, 3512–3522 (2004).

Li S. and Carmichael S.T. *Neurobiology of Disease,* 23, 362-373 (2006).

Lipari, G. & Szabo, A. (1982). *J. Am. Chem. Soc*. 104, 4559-4570.

Lipari, G. & Szabo, A. (1982*). J. Am. Chem. Soc*. 104, 4546-4559

Liu J., Li M., Ran X., Fan J., and Song J. (2006) *Biochemistry.* 45, 7171–7184.

Liu X, Hawkes E, Ishimaru T, Tran T, Sretavan DW. *J Neurosci*, 26, 3087-101 (2006).

Macias, M. J., Hyvo¨nen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., and Oschkinat, H. (1996). *Nature 382*, 646–649.

Macias, M. J., Wiesner, S., and Sudol, M. (2002). *FEBS Lett. 513*, 30–37.

Martinez A, Soriano E. *Brain Res Brain Res Rev.* 49, 211-226 (2005).

McKeon RJ, Höke A, Silver J. *Exp Neurol* 1995;136:32-43.

Miranda JD, White LA, Marcillo AE, Willson CA, Jagid J, Whittemore SR. *Exp Neurol.,* 156, 218-22 (1999).

Miyazaki, Y., Maeda, Y., Sato, H., Nakano, M., and Mellor, G. W. (2008) *Bioorg. Med. Chem. Lett.* 18, 1967–1971

Monod, J., Wyman, J., and Changeux, J.P. 1965. J. Mol. Biol. 12: 88–118.

Moreno-Flores MT, Wandosell F. *Neuroscience,* 91, 193-201 (1999).

Murai KK, Nguyen LN, Koolpe M, McLennan R, Krull CE, Pasquale EB. *Mol Cell Neurosci*. 2003 Dec;24(4):1000-11.

Noberini R, Koolpe M, Peddibhotla S, Dahl R, Su Y, Cosford ND, Roth GP, Pasquale EB. *J Biol Chem*. 2008 Oct 24; 283(43): 29461-72.

Noren NK, Lu M, Freeman AL, Koolpe M, Pasquale EB. *Proc Natl Acad Sci USA*, 101, 5583-5588 (2004).

Noren, N. K., and Pasquale, E. B. (2007) *Cancer Res.* 67, 3994–3997.

Oertle, T. et al. *J. Neurosci.*23, 5393–5406 (2003).

Otwinowski, Z., and W. Minor. (1997). In C. W. Carter, Jr., and R. M. Sweet (ed.), Methods in enzymology, 276, 307–326. Academic Press, New York, NY.

Palmer, A. G.,Rance, M. & Wright, P. E. (1991). *J. Am. Chem. Soc*. 113, 4371-4380.

Pardi, L. Billeter, K. Wuthrich, *J. Mol. Biol*., (1984) 180: 741-751.

Pasquale EB. (2004) *Nat Neurosci.* 7: 417-8.

Pasquale, E. B. (2005) *Nat. Rev. Mol. Cell Biol*. 6, 462-475.

Pires, J. R., Taha-Nejad, F., Toepert, F., Ast, T., Hoffmu¨ller, U., Schneider-Mergener, J., Ku¨hne, R., Macias, M. J., and Oschkinat, H. (2001). *J. Mol. Biol. 314*, 1147–1156.

Popovych N, Sun S, Ebright RH, Kalodimos CG. (2006) *Nat Struct Mol Biol*. 13, 831-8.

Prévost N, Woulfe DS, Jiang H, Stalker TJ, Marchese P, Ruggeri ZM, Brass LF. (2005) *Proc Natl Acad Sci U S A*. 102: 9820-5.

Prinjha, R. et al.. *Nature* 403, 383–284 (2000).

Qin H, Noberini R, Huan X, Shi J, Pasquale EB, Song J. (2010). *J Biol Chem.* 285, 644-654.

Qin H, Pu HX, Li M, Ahmed S, Song J. *Biochemistry*. 2008 Dec 23;47(51):13647-58.

Qin H, Shi J, Noberini R, Pasquale EB, Song J. (2008). *J Biol Chem.* 283, 29473-84.

Ramon y Cajal S. *Degeneration & regeneration of the nervous system*. London: Oxford University Press, 1928.

Ran X, Qin H, Liu J, Fan JS, Shi J and Song J (2008). *Proteins*. 72, 1019–1029.

Ran X. and Song J. (2005) *J Biol Chem.* 280: 19205-12.

Ricard J, Salinas J, Garcia L, Liebl DJ. *Mol Cell Neurosci*, 31, 713-722 (2006).

Schwab ME, Caroni P. *J Neurosci* 1988;8:2381-93.

Shi J, Han N, Lim L, Lua S, Sivaraman J. Wang L, Mu Y and Song J. (2011). *PLoS Computational Biology*. In press.

Shin D, Garcia-Cardena G, Hayashi S, Gerety S, Asahara T, Stavrakis G, *et al*. *Dev Biol*, 230, 139-150 (2001).

Singla N, Goldgur Y, Xu K, Paavilainen S, Nikolov DB, Himanen JP. *Biochem Biophys Res Commun*. 2010 Sep 3;399(4):555-9. Epub 2010 Aug 1.

Smith A, Robinson V, Patel K, Wilkinson DG. (1997) *Curr Biol.* 7: 561-70.

Smock, R.G. & Gierasch, L.M. (2009) Sending signals dynamically. *Science* 324, 198–203.

Snow DM, Lemmon V, Carrino DA, Caplan AI, Silver J. *Exp Neurol* 1990;109:111-30.

Song J, Vranken W, Xu P, Gingras R, Noyce RS, Yu Z, et al. *Biochemistry* 41, 10942-10949 (2002).

Song, J. (2003) *J. Biol. Chem*. 278, 24714-24720.

Surawska H, Ma PC, Salgia R. *Cytokine Growth Factor Rev,* 15, 419-433 (2004).

Takasu MA, Dalva MB, Zigmond RE, Greenberg ME. *Science*, 295, 491-495 (2002).

Torres R, Firestein BL, Dong H, Staudinger J, Olson EN, Huganir RL, et al. *Neuron.* 21, 1453-1463 (1998).

Velazquez-Campoy A, Leavitt SA, Freire E. (2004) *Methods Mol Biol.* 261: 35-54.

Voeltz G.K., Prinz W.A., Shibata Y., Rist J.M. and Rapoport T.A. *Cell* 124, 573–586 (2006).

Walker O, Varadan R, Fushman D. (2004), *J. Magn. Reson*. 168, 336-345.

Weinberg EL, Raine CS. *Brain Res.* 1980 Sep 29;198(1):1-11.

Wilkinson DG. *Nat Rev Neurosci,* 2, 155-164 (2001).

Willson CA, Irizarry-Ramirez M, Gaskins HE, Cruz-Orengo L, Figueroa JD, Whittemore SR, et al. *Cell Transplant*, 11, 229-39 (2002).

Willson CA, Miranda JD, Foster RD, Onifer SM, Whittemore SR. *Cell Transplant,* 12, 279-90 (2003).

Wimmer-Kleikamp SH, Lackmann M. *IUBMB Life*, 57, 421-431 (2005).

Wishart DS, Sykes BD. (1994). *Methods Enzymol.* 239:363-92.

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

Xu Q, Mellitzer G, Robinson V, Wilkinson DG. (1999) *Nature*. 399: 267-71.

Yokoyama N, Romero MI, Cowan CA, Galvan P, Helmbacher F, Charnay P, Parada LF, Henkemeyer M. (2001) *Neuron*. 29: 85-97.

Yona Goldshmit, Mary P. Galea, Graham Wise, Perry F. Bartlett, AnnM. Turnley. *The Journal of Neuroscience*, 45, 10064 –10073 (2004).

Yona Goldshmit, Samuel McLenachan, Ann Turnley, *Brain Research Reviews*. 52 (2006), 327-345.

Y. Goldgur, S. Paavilainen, D. Nikolov, J.P. Himanen, *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun*. 65 (2009) 71–74.

PUBLICATION

1. Xiaoyuan Ran, **Haina Qin (Co-first author)**, Jingxian Liu, Jing-song Fan and Jiahai Shi; Jianxing Song, NMR Structure and Dynamic of Human EphrinB2 Ectodomian: The functionally-critical C-D and G-H loops are highly-dynamic in solution. ***PROTEINS: Structure, Function, and Bioinformatics***, 2008, Aug 15; 72(3):1019-29.

2. **Haina Qin**, Helen X. Pu, Minfen Li, Sohail Ahmed and Jianxing Song, Identification and structural mechanism for a novel interaction between a ubiquitin ligase WWP1 and Nogo-A, a key inhibitor for central nervous system regeneration. ***Biochemistry***. 2008 Dec 23; 47(51):13647-58.

3. **Qin Haina**, Noberini Robert, Huan Xuelu, Shi Jiahai, Pasquale EB, Song Jianxing. Structural characterization of the EphA4-ephrin-B2 complex reveals new features enabling Eph-ephrin binding promiscuity.***J Biol Chem***. 2009 Oct 29 [Epub ahead of print]

4. **Qin Haina**, Shi Jiahai, Noberini Robert, Pasquale EB, Song Jianxing, Crystal structure and NMR binding reveal that two small molecule antagonists target the high affinity ephrin-binding channel of the EphA4 receptor. ***J Biol Chem***. 2008 Oct 24; 283(43):29473-84.

5. **Qin Haina**, Chen F, Huan X, Machida S, Song J, Yuan YA. Structure of the Arabidopsis thaliana DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. ***RNA***. 2010 Mar;16(3):474-81.

6. Liu Jingxian, Zhu Wanlong, **Qin Haina**, Song Jianxing, NMR studies reveal a novel mode for hFADD to bind with the unstructured hRTN3 which initiates the ER-stress activated apoptosis. ***Biochem Biophys Res Commun***. 2009 Jun 12; 383(4):433-9.

7. Li Minfen, Li Yan, Liao Xuanhao, Liu Jingxian, **Qin Haina**, Xiao Zhicheng, Song Jianxing, Rational design, solution conformation and identification of functional residues of the soluble and structured Nogo-54, which mimics Nogo-66 in inhibiting the CNS neurite outgrowth. ***Biochem Biophys Res Commun***. 2008 Sep 5; 373(4):498-503.

8. Minfen Li , Jingxian Liu , Xiaoyuan Ran, Miaoqing Fang , Jiahai Shi , **Haina Qin** , June-Mui Goh and Jianxing Song, Resurrecting Abandoned Proteins with Pure Water: CD and NMR Studies of Protein Fragments Solubilized in Salt-Free Water, ***Biophysical Journal*** 2006 (91):4201-4209.

9. Aaron Petty, Eugene Myshkin, **Haina Qin (Co-first Author)**, Hui Miao, Hong Guo, Gregory P. Tochtrop, Jer-Tsong Hsieh, Phillip Page, Lili Liu, Daniel J. Lindner, Chayan Acharya, Alexander D. MacKerell, Jr., Eckhard Ficker, Jianxing Song, and Bingcheng Wang, Doxazosin, a Small Molecule Agonist for EphA2 Kinase, Inhibits Tumor Cell Migration *in vitro* and Suppresses Distal Metastasis *in vivo*, ***Cancer Cell***. (Under Review)