Université de Montréal

Toward the determination of the three-dimensional structure of Tcell PTP using multidimensional NMR spectroscopy

par

Wendy M. Wong

Département de chimie

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)

en chimie

Août, 2000

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé:

Toward the determination of the three-dimensional structure of Tcell PTP using multidimensional NMR spectroscopy

présenté par:

Wendy M. Wong

A été évalué par un jury composé des personnes suivantes:

Joelle Pelletier	président-rapporteur
Yves Aubin	directeur de recherche
Jeffrey Wayne Keillor	codirecteur
William Lubell	membre du jury

Mémoire accepté le:

Abstract

The reversible mechanism of tyrosine phosphorylation is controlled by the dynamic concerted actions of protein tyrosine kinases (PTKs) and phosphatases (PTPs). It is the level of protein tyrosine phosphorylation and dephosphorylation by PTKs and PTPs respectively which regulate various eukaryotic intracellular signaling pathways and mitogenic signaling pathways (Zhang 1998). More specifically, human protein tyrosine phosphatase-1B (PTP-1B) has been recently implicated in the control of type-II diabetes and obesity (Elchebly *et al.* 1999). Type-II diabetes is a metabolic disease characterized by an inability of the body to respond to insulin. Insulin is an important hormone secreted by liver β-pancreatic islet cells in response to increased glucose levels in the blood. Insulin maintains normal blood glucose levels by stimulating glucose influx and metabolism from blood into muscle and adipose tissues. Recently, PTP-1B has been implicated to act as an OFF switch in the insulin signaling pathway by dephosphorylating the insulin receptor and shunting glucose transport from blood to muscles after stimulation of the insulin receptor by insulin (Elchebly *et al.* 1999). Accordingly, this enzyme has been identified as an important therapeutic drug target for the treatment of type-II diabetes. A variety of PTPs exist, therefore, to avoid any unwanted side-effects, it is important that therapeutic agents developed to target PTP-1B are selective towards this enzyme.

To date, approximately one hundred PTPs have been identified but most notably, T-cell PTP (TCPTP). TCPTP, a 45-48 kDa protein, shares ~75% sequence homology within the catalytic domain of PTP-1B. Studies using "substrate-trapping" mutants to

identify PTP substrates have identified PTP-1B and TCPTP as some of the few PTPs capable of binding the autophosphorylated insulin receptor kinase (Wälchli *et al.* 2000). The wild-type versions of these same PTPs were also capable of dephosphorylating the insulin receptor kinase (Wälchli *et al.* 2000). The specific functional role of TCPTP is unknown therefore due to the high sequence homology shared between TCPTP and PTP-1B, their common binding affinity for the insulin receptor kinase and their common ability to dephosphorylate the insulin receptor kinase, TCPTP may represent a significant challenge for the discovery of selective PTP-1B inhibitors for the treatment of type-II diabetes. Using nuclear magnetic resonance (NMR) spectroscopy to study TCPTP may reveal distinct structural differences between TCPTP and PTP-1B that may aid in the design of specific inhibitors towards PTP-1B.

The catalytic domain of TCPTP was cloned into a pFLAG2 expression vector. The construct consisted of residues 1 to 281 of TCPTP whereby the catalytic cysteine residue 216 was mutated to a serine residue. Preliminary NMR studies were performed using this construct to determine feasibility of NMR studies of TCPTP. In order to optimize high yields of protein expression from bacterial growth using isotope-enriched minimal media this gene was then subcloned into various bacterial pET expression systems. Using a series of triple resonance multidimensional NMR experiments with ^{13}C, ^{2}H, ^{15}N-labeled TCPTP C216S (1-281), approximately 50% of the polypeptide backbone resonances were assigned. In anticipation that interproton distance constraints sufficient for three dimensional (3D) structure calculations would be required, a methyl protonation scheme was successfully used to re-introduce additional protons at the methyl groups of

valine, leucine and isoleucine (δ) side chains of ^{13}C, ^{2}H, ^{15}N-labeled TCPTP C216S (1-281).

Sommaire

Diverses fonctions cellulaires, telles que la prolifération, la différentiation cellulaire, le métabolisme, l'organisation cytoskelétale, les interactions intracellulaires, la transcription génétique et la réponse du système immunitaire, sont contrôlées par une série de voies de signalisations impliquant une très grande variété d'interactions protéines-protéines. Ces interactions sont elles-même contrôlées par l'action d'enzymes telles que les kinases et les phosphatases. Plus particulièrement, certaines voies de signalisations sont activées par l'action de kinases des protéines à tyrosine (PTK) ou désactivées par l'action opposée de phosphatases de protéines à tyrosine (PTP).

Dans la voie de signalisation de l'insuline, l'élément initiateur est la liaison de l'insuline à son récepteur. Ceci induit un changement conformationel du récepteur qui se traduit par une autophosphorylation de certains résidus tyrosyles permettant ainsi l'association de diverses molécules, telles que les substrats du récepteur de l'insuline (IRS). Ces protéines sont à leur tour phosphorylées pour ensuite s'associer à d'autres protéines et ainsi propager le signal. Ce dernier vise à activer le transport du glucose sanguin à travers la membrane des cellules du muscle et d'adipocytes. Récemment, la protéase de protéines à tyrosine 1B (PTP-1B) a été reliée au mécanisme d'interruption de la voie de signalisation de l'insuline. Les études d'ablation génique (Elchebly *et al.* 1999) ont démontré que la PTP-1B est responsable de la désactivation du récepteur de l'insuline via la déphosphorylation de celui-ci. En effet, les souris arborant un phénotype PTP-1B $^{-/-}$ ont été soumises à une diète élevée en gras et elles ont démontré une très grande résistance au gain de poids. De leur côté, les souris arborant un phénotype PTP-

1B $^{+/+}$ ont gagné du poids et elles ont développées une résistance à l'insuline. Les conclusions de ces études ont permis d'identifier la PTP-1B comme cible thérapeutique pour le traitement du diabète de type II.

À ce jour, une centaine d'enzymes PTP ont été identifiées rendant le développement d'un inhibiteur sélectif pour la PTP-1B un défi de taille. Parmi les PTP identifiées, la phosphatase des cellules-T (TCPTP) est la plus similaire à la PTP-1B. Sa séquence primaire globale est identique à 65% et les résidus constituant le domaine catalytique sont identiques à 75% (Tiganis *et al.* 1997). De plus, l'étude d'une enzyme mutante (D181A) capable de piéger le substrat a démontré que la PTP-1B et la TCPTP sont les membres d'un très petit groupe d'enzymes capables de se lier au résidus phosphotyrosyles du domaine kinase du récepteur de l'insuline et de le déphosphoryler (Wälchi *et al.* 2000).

La TCPTP humaine est une protéine intracellulaire qui a initalement été isolée à partir d'une librairie d'ADN recombiné issues de cellules-T périphériques humaines (Cool *et al.* 1989). La TCPTP est exprimée à toutes les étapes du développement des mammifères et plus particulièrement dans les lignées de cellules lymphoïdes suggérant ainsi un rôle possible dans la formation des cellules sanguines et les fonctions immunitaires (You-Ten *et al.*1997). Deux épissages différents produisent des polypeptides de poids moléculaire de 48 kDa (TC48) et de 45 kDa (TC45) chez l'humain. Ces deux protéines ne varient qu'à la partie carboxy-terminale, loin du domaine catalytique. À l'instar de la PTP-1B, TC-48 est localisée dans le réticulum endoplastique (Cool *et al.* 1990) tandis que TC45 cible le noyau, et ce, du à la présence d'une séquence bipartite de localisation nucléaire (NLS) (White *et al.*1994). Les travaux de You-Ten et

collaborateurs (You-Ten *et al.*1997) ont démontré que la TCPTP est importante pour le développement des cellules T et B immunocompétantes et elle est critique pour le maintien de l'intégrité du microenvironement de la moelle osseuse nécessaire à la formation des cellules sanguines.

La localisation commune de la PTP-1B et TC48, la grande similitude de leur domaine catalytique et le rôle incertain de cette dernière en font une candidate intéressante pour l'étude de la spécificité des inhibiteurs développés contre la PTP-1B. Plusieurs structures tridimensionnelles de la PTP-1B en présence d'inhibiteur ont été élucidées par diffraction des rayon-X (RX) ou par résonance magnétique nucléaire (RMN). Celles-ci permettent d'expliquer à l'échelle atomique les déterminants structuraux influençant l'inhibition de la PTP-1B. À ce jour, de telles études ne sont pas disponibles dans le cas de la TCPTP, principalement à cause de la difficulté à produire des cristaux en vue d'utiliser la diffractions RX. C'est pourquoi, l'étude des interactions inhibiteur-enzyme impliquant la TCPTP a été entrepris en utilisant les toutes dernières techniques de RMN triple résonance.

La spectroscopie RMN appliquée aux protéines permet d'extraire des paramètres structuraux de la TCPTP en solution. Ces paramètres fourniront une information utile à l'identification et à la compréhension des différences structurales impliquées dans les interactions enzyme-inhibiteur de la PTP-1B et de la TCPTP. Ces connaissances pourront ensuite être mise à profit par la chimie médicinale dans le développement d'inhibiteurs sélectifs de la PTP-1B.

La stratégie employée pour l'étude des protéines par spectroscopie RMN se résume en trois points. Premièrement, obtenir un échantillon de la protéine d'intérêt qui

doit être soluble à des concentrations d'environ 1mM et stable. Deuxièment, l'attribution des tous les signaux RMN des noyaux détectés et troisièment, mesurer les paramètres structuraux tels que les déplacements chimiques, les distances interprotons dérivées de la NOE, et les couplages dipolaires résiduels. Afin de déterminer la faisabilités de telles études RMN sur le domaine catalytique de la TCPTP, un échantillon marqué à l'azote-15 a été préparé et une corrélation hétéronucléaire 2D-^{1}H-^{15}N-HSQC a été enregistrée. Le spectre RMN a mis en évidence toutes les caractéristiques d'un polypeptide bien replié et stable dans les conditions de concentration, pH et température choisies pour cette expérience préliminaire. Ces résultats ont donc permis de confirmer la faisabilité d'utiliser la spectroscopie RMN pour élucider la structure tridimensionnelle de la TCPTP.

Les techniques de RMN triple résonance appliquées à une protéine de la taille de la TCPTP (35 kDa) nécessitent la production d'échantillons triplement marqués au deutérium, carbone-13 et azote-15. Ces techniques permettent de tirer avantage des larges fenêtres spectrales de ces noyaux permettant de disperser la très grande quantité de signaux sur plusieurs dimensions augmentant ainsi la résolution. De plus, le marquage au deutérium est impératif pour permettre l'étude de protéines de la taille de la TCPTP. En effet, le culbutage plus lent du polypeptide en solution est caractérisé par un signal RMN des carbones de trop courte durée ne pouvant être détecté par les techniques à triple résonance. Par conséquent, le protocole utilisé pour la production de l'échantillon utilisé pour l'étude préliminaire a été optimisé pour maximiser la production d'échantillons qui rencontrent les exigences de la RMN. De façon générale, une protéine est produite en utilisant un système bactérien. *Escherichia coli* est l'hôte de préférence car celle-ci peut être cultivées croître dans un milieu minimum n'utilisant que le glucose et le chlorure

d'ammonium comme seules sources de carbone et d'azote, respectivement. L'incorporation de deutérium s'effectue simplement en substituant l'eau pour l'oxide de deutérium (D_2O). L'utilisation du D_2O comme solvant dans le milieu de culture ralenti significativement la croissance des bactéries. Il est donc nécessaire d'utiliser un système d'expression hautement efficace. Pour ce faire, le gène du domaine catalytique de la TCPTP a été introduit dans plusieurs systèmes d'expression de type pET. Ceux-ci mettent à profit la grande efficacité de la machinerie d'expression du virus T7 qui produit de grande quantité de la protéine cible, et ce, dans des milieux de culture minimums. Ensuite, le protocole de purification a été optimisé pour obtenir les échantillons le plus efficacement possible tout en maximisant les rendements.

La première étape des études RMN consiste en l'attribution de tous les signaux RMN de la TCPTP qui seraient susceptibles de fournir de l'information sur sa structure tridimensionnelle. Une série de cinq expériences RMN à trois dimensions permettent d'attribuer la résonance des noyaux formant le squelette peptidique (HN, N, C_α and C_β). L'analyse des données a permi l'attribution que de la moitié du squelette en raison de la faible intensité d'un trop grand nombre de signaux. Les données ont été enregistrées de nouveaux à une température plus élevée (30 °C) mais avec un gain marginal en sensibilité. Ce résultat semble suggérer que si des conditions d'échantillons permettant d'enregistrer les données RMN à plus haute température (35-37 °C) existait, il serait possible de compléter l'attribution des résonance de la TCPTP.

La détermination de la structure tridimensionnelle de la TCPTP nécessite la mesure de distance interprotons dérivées de l'effet NOE. Étant donné que tous les protons non-labiles ont été remplacés par des deutériums, seules les distances NH-NH

peuvent être mesurées. Gardner et Kay (Gardner and Kay 1997) ont démontré que les distances NH-NH ne suffisent pas à elle seules à calculer des structures précises d'une protéines. C'est pourquoi ils ont proposé un nouveau protocole de marquage istopique. Celui-ci consiste à réintroduire des protons sur les groupements méthyles des chaînes latérales des résidus valines, leucines et isoleucines (position δ seulement) en plus du triple marquage (^{2}H,^{13}C et ^{15}N). Un échantillon marqué de cette façon permet la mesure de distances NH-NH, NH-CH_3 et CH_3-CH_3 fournissant ainsi un nombre suffisant de données expérimentales pour calculer des structures de bonne qualité. Malgré une attribution incomplète des résonances du squelette peptidique, le protocole de marquage des groupements méthyles a été appliqué à la TCPTP. En plus de permettre la mesure de distances interprotons, l'attribution des chaînes latérales des résidus valines, leucines et isoleucines aidera à compléter l'attribution des résonances du squelette peptidique.

Summary

Protein tyrosine phosphatases (PTPs) constitute a superfamily of structurally and functionally diverse enzymes that work in coordination with protein tyrosine kinases (PTKs) in regulation of various cellular processes such as cell cycle functions, proliferation and differentiation, metabolism, cytoskeletal organization, cell-cell interactions, gene transcription, and the immune response (Tonks and Neel 1996). It is the level of tyrosine phosphorylation by PTKs which controls these intracellular signaling pathways. In contrast, PTPs are hydrolytic enzymes that dephosphorylate tyrosine residues in phosphoprotein substrates and in turn either potentiate or antagonize these PTK-dependent signaling pathways.

Recently, PTP-1B has been identified to have a major role in the negative regulation of the insulin signaling pathway. When insulin binds to its receptor, tyrosine residues of the insulin receptor are autophosphorylated. Activation of the insulin receptor ensues and the intrinsic insulin receptor tyrosine kinase in turn phosphorylates the various insulin receptor substrate (IRS) proteins in order to propagate the insulin signaling event (White and Kahn 1994). Once the insulin signaling pathway is activated glucose is actively transported from the blood into the muscle and adipocytes for storage. PTP-1B acts by dephosphorylating the activated insulin receptor thus preventing the transport of glucose from the blood into the muscle and adipocytes for metabolism. Studies of PTP-1B $^{-/-}$ knockout mice placed on a high fat diet have demonstrated enhanced insulin sensitivity and resistance to weight gain (Elchebly *et al.* 1999). In contrast, their wild-type littermates were observed to gain weight and become insulin insensitive (Elchebly *et*

al. 1999). Consequently, PTP-1B has been identified as a potential therapeutic drug target for the treatment of type-II diabetes and obesity (Elchebly *et al.* 1999).

Approximately one hundred PTP enzymes have been identified thus far, making the development of specific inhibitors to PTP-1B a great challenge. Most notably, T-cell protein tyrosine phosphatase (TCPTP) displays overall 65% sequence identity to PTP-1B and 74% sequence identity within the conserved catalytic domain of PTP-1B (Tiganis *et al.* 1997). Furthermore, studies using PTP "substrate-trapping" mutants (D181A) for autophosphorylated insulin receptor kinase (IRK) identified PTP-1B and TCPTP, amongst a few other enzymes, to be capable of binding and dephosphorylating the insulin receptor kinase (Wälchi *et al.* 2000).

Human TCPTP is an intracellular protein that was originally isolated from a human peripheral T-cell cDNA library (Cool *et al.* 1989). TCPTP has been found to be expressed at all stages of mammalian development and most notably in lymphoid cell lineages suggesting its possible role in both hematopoiesis and immunity (You-Ten *et al.* 1997). Two splice variants give rise to a 48-kDa (TC48) and a 45-kDa (TC45) form of human TCPTP which differ only in their non-catalytic C-terminal. TC48, like PTP-1B, localizes to the endoplasmic reticulum (ER) (Cool *et al.* 1990) whereas TC45 is targeted to the nucleus due to the presence of a bipartite nuclear localization sequence (NLS) at the C-terminal (White and Kahn 1994). TCPTP has been shown to be important for the development of immunocompetent T- and B-cells and are critical for maintaining integrity of the bone marrow microenvironment that is crucial for hematopoiesis (You-Ten *et al.* 1997).

It is the common subcellular localization, high degree of sequence similarity shared between PTP-1B and TC48 and the uncertain role of the latter which makes TCPTP an interesting candidate in studying the specificity of therapeutic drugs developed to target PTP-1B. The x-ray crystal structure of PTP-1B has been elucidated and studies on the structural properties and functional role of PTP-1B have been studied however great difficulty has been encountered in the attempt of crystallizing TCPTP for structural studies. Consequently, TCPTP represents a possible candidate for structural studies using NMR spectroscopy. Moreover, structural studies of PTP-1B using NMR spectroscopy are being studied concurrently within the same laboratory.

NMR spectroscopy is a sensitive and versatile technique that may be used to elucidate macromolecular structure to atomic resolution. Using NMR spectroscopy to study TCPTP may also reveal distinct structural differences between PTP-1B and TCPTP which may be significant and aid in the design of selective inhibitors towards PTP-1B. The basis of NMR is that when NMR active nuclei are placed in a strong external magnetic field, a unique signal characteristic of the type of nucleus and the local chemical environment surrounding this nucleus arises. As a result any slight perturbation to the system, such as the binding of inhibitors to a target protein, may be detected and studied. Using NMR spectroscopy to study TCPTP-inhibitor interactions may reveal residues important in these interactions and may further assist in the discovery of specific inhibitors for PTP-1B. The strategy for performing structural studies of TCPTP using NMR spectroscopy consist of three major steps; 1) obtaining protein samples that are suitable for NMR studies, 2) resonance assignment of all NMR active nuclei, and 3)

measurement of structure-dependent parameters, such as interproton distances (NOE) and residual dipolar coupling chemical shifts.

Foremost, to determine feasibility of NMR structural studies with TCPTP a preliminary 2D ^{1}H-^{15}N HSQC (heteronuclear single quantum coherence) of a TCPTP construct made available by Dr. E. Asante-Appiah was acquired. The TCPTP construct was expressed from a pFLAG2 expression vector and consisted of the catalytic domain of TCPTP, residues 1 to 281, whereby the catalytic cysteine residue 216 was mutated to a serine residue (C216S) in order to avoid oxidation of the active site cysteine residue. The resulting 2D ^{1}H-^{15}N HSQC of ^{15}N-labeled TCPTP C216S (1-281) demonstrated a well resolved, homogeneous spectrum which was indicative of a stable, well-folded protein amenable to NMR solution studies. To obtain a protein sample of TCPTP C216S (1-281) suitable for NMR studies, efficient production of soluble protein had to be generated by bacterial expression and the purification protocol had to be optimized to obtain high quantities of pure protein. Most importantly, to avoid heterogeneity of the NMR spectra, samples of up to 1-2 mM concentration had to be prepared without aggregation. Furthermore, TCPTP is a large 32 kDa protein whose slow tumbling rate in solution is characterized by a fast decay of the NMR signals due to dipole-dipole interactions of the NMR active nuclei with nearby protons. This phenomenon degrades the spectroscopic quality resulting in decreased sensitivity and resolution. Consequently, to study TCPTP using multidimensional NMR techniques, enrichment with ^{13}C and ^{15}N NMR-active isotopes is required in order to take advantage of heteronuclear coupling and a respectively larger spectral window. To circumvent the loss of NMR signal intensity due to relaxations through proton dipole-dipole interactions, replacement of the non-labile

protons with deuterium is also required. Replacement of carbon-bound protons with deuterium extends the ^{13}C relaxation rates which results in increased survival of the NMR signals, more efficient magnetization transfer and hence improved spectral quality. As a result, bacterial expression of the protein also had to be optimal for uniform labeling of TCPTP with ^{13}C, ^{2}H and ^{15}N isotopes.

Once a triply labeled sample of TCPTP was obtained, triple resonance multidimensional NMR experiments were used for assignment of the signals arising from the polypeptide backbone. By assigning the chemical shift resonance of each NMR active nuclei to a specific amino acid in the primary sequence, a "fingerprint" was obtained to facilitate identification of those residues important for TCPTP-inhibitor interactions. A series of four triple-resonance NMR experiments with ^{2}H-decoupling were used to assign the polypeptide backbone atoms $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and ^{1}HN and ^{15}N chemical shift resonances of ^{13}C, ^{2}H, ^{15}N-labeled TCPTP C216S (1-281).

Ultimately, determination of the three dimensional structure of TCPTP relies on the measurements of proton-proton distances through the nuclear Overhauser effect (NOE). Since all but the exchangeable amide protons were replaced by deuterons, only NH-NH distances can be measured. Gardner and Kay (1997) showed that these do not provide a sufficient amount of distance constraints to calculate an accurate 3D structure. To overcome this difficulty, they developed a labeling scheme that introduced protons on methyl groups of valine (Val), leucine (Leu) and isoleucine (Ile-δ) side chains. Samples labeled in this fashion allow for the measurement of NH-NH, NH-CH_3 and CH_3-CH_3 interproton distance constraints for 3D structure calculations. Moreover, assignment of

the protonated methyl groups of Val, Leu and Ile(δ) side chains may also supplement the resonance assignment of the polypeptide backbone.

Using multidimensional NMR experiments of triply labeled (^{13}C, ^{2}H, ^{15}N) TCPTP C216S (1-281) almost 50% of the chemical shift resonances of the polypeptide backbone atoms have been assigned. Success of the methyl-protonation labeling scheme of Val, Leu and Ile-δ side chains was confirmed with a constant time 2D ^{13}C-^{1}H HSQC. Ambiguities of the polypeptide backbone assignment may be further alleviated with data obtained from experiments that assign methyl group-containing side chains (Val, Leu and Ile-δ). To achieve complete sequential assignment of the polypeptide backbone of TCPTP, selective labeling with specific ^{15}N-labeled amino acid types may be required. Alternatively, protein engineering to further stabilize TCPTP may be required since protein engineering of PTP-1B was essential before NMR structural studies could even proceed. For PTP-1B protein engineering whereby all the cysteine residues were mutated to alanine residues and two disulfide bridges were introduced to anchor the protein together resulted in significant improvement in the stability of PTP-1B for NMR structural studies. Considering TCPTP C216S (1-281) is quite stable protein engineering of TCPTP in a similar fashion would only help improve the stability of TCPTP.

Table of Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

A	adenine
Arg	arginine
Asp	aspartate
ATP	adenosine triphosphate
B_1	applied magnetic field (perpendicular to B_0)
B_0	external static magnetic field
*Bam*HI	*Bam*HI restriction endonuclease
C216S	cysteine residue 216 mutated to serine residue
C	cytosine
cDNA	complementary deoxyribonucleic acid
CSA	chemical shift anisotropy
CT	constant time
Cys	cysteine
D181A	aspartate residue 181 mutated to alanine (PTP substrate trapping mutant)
DD	dipole-dipole
dNTPs	dideoxynucleoside triphosphates
DSP	dual-specificity phosphatase
DTT	DL-dithiothreitol
DTT-d_{10}	deuterated DL-dithiothreitol
E	energy
EDTA	ethylenediaminetetraacetic acid

ER	endoplasmic reticulum
F_1, F_2, F_3	spectral windows
FID	free induction decay
FTIR	Fourier transform infrared
G	guanine
GLUT4	glucose transporter
Gly	glycine
h	Planck's constant
His	histidine
HNCA	three dimensional experiment to correlate ^{1}H, ^{15}N and $^{13}C_{\alpha}$ of the intraresidue (*i*) and the $^{13}C_{\alpha}$ of the previous residue (*i-1*)
HNCACB	three dimensional experiment to correlate ^{1}H, ^{15}N, and $^{13}C_{\beta}$ of the intraresidue (*i*) and the $^{13}C_{\beta}$ of the previous residue (*i-1*)
HNCO	three dimensional experiment to correlate ^{1}H, ^{15}N, and ^{13}C-carbonyl of the intraresidue (*i*)
HN(CO)CA	three dimensional experiment to correlate ^{1}H and ^{15}N of the intraresidue (*i*) and $^{13}C_{\alpha}$ of the previous residue (*i-1*)
HN(COCA)CB	three dimensional experiment to correlate ^{1}H and ^{15}N of the intraresidue (*i*) and $^{13}C_{\beta}$ of the previous residue (*i-1*)
HSQC	heteronuclear single quantum coherence
I	nuclear spin quantum number
IgG	immunoglobulin G
Ile	isoleucine
INEPT	insensitive nuclei enhanced by polarization transfer
IPTG	isopropyl-β-D-thiogalactopyranoside

IR	insulin receptor
IRK	insulin receptor kinase
IRS	insulin receptor substrate
J	coupling constant
J	angular momentum
J(ω)	spectral density function
k	Boltzmann constant
LB	Luria broth
Leu	leucine
LMW-PTP	low molecular weight protein tyrosine phosphatase
Lys	lysine
M_0	bulk magnetization vector
M_z	bulk magnetization vector along *z*-axis
mRNA	messenger ribonucleic acid
MWCO	molecular weight cut-off
*Nde*I	*Nde*I restriction endonuclease
NIDDM	non-insulin dependent diabetes mellitus
NLS	nuclear localization sequence
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
PCR	polymerase chain reaction
Phe	phenylalanine
PI	phosphoinositol

PLCC-SH2	SH2 domain of phospholipase-Cγ
pNPP	*p*-nitrophenyl phosphate
P-O	phosphate-oxygen
ppm	parts per million
PTK	protein tyrosine kinase
PTP	protein tyrosine phosphatase
PTP-1B	protein tyrosine phosphatase-1B
PTP-1B $^{-/-}$	protein tyrosine phosphatase-1B gene removed
rf	radio-frequency
*Sal*I	*Sal*I restriction endonuclease
SAR	structure-activity relationships
SDS	sodium dodecyl sulfate
Ser	serine
SH2	Src-homology 2
SPI	selective polarization inversion
t_1, t_2	time periods
T_1	time constant describing longitudinal relaxation
T_2	time constant describing transverse relaxation
τ_c	rotational correlation time
T	thymine
TBE	Tris-borate buffer
TCPTP	Tcell protein tyrosine phosphatase
TC45	Tcell protein tyrosine phosphatase - 45 kDa splice variant

TC48	Tcell protein tyrosine phosphatase - 48 kDa splice variant
TEMED	*N,N,N',N'*-tetramethylethylenediamine
Thr	threonine
Tris-HCl	Tris[hydroxymethyl]aminomethane hydrochloride
TROSY	transverse relaxation optimized spectroscopy
T-TBS	Tween-Tris buffered saline
Tyr	tyrosine
Val	valine
μ	magnetic moment
ν	frequency
ω_0	Larmor frequency
σ	chemical shift tensor
σ_0	isotropic chemical shift tensor
γ	magnetogyric ratio
2D	two-dimensional
3D	three-dimensional

Acknowledgments

I would like to express great gratitude to my supervisor, Dr. Yves Aubin, for sharing his knowledge and providing continuous help, support and guidance throughout my studies. Dr. Ernest Asante-Appiah is thankfully acknowledged for his help in the recombinant DNA section of this project. Dr. Lewis E. Kay is also gratefully acknowledged for providing the NMR pulse sequences and Dr. Bruce Johnson for providing the program NMRView. To my co-supervisor, Dr. Jeffrey Keillor, thank you for your helpful discussions and words of encouragement. To the remaining NMR group which consists of Dr. Laird Trimble and the few yet unforgettable students thank you for your friendship and for making my research studies at Merck Frosst & Co. even more enjoyable. Finally, I am most thankful to my family and friends for their continued love, support and understanding.

The Natural Sciences and Engineering Research Council of Canada and Merck Frosst Canada & Co. are gratefully acknowledged for providing me with the industrial postgraduate scholarship.

Chapter 1

Protein Tyrosine Phosphatases

I. Introduction

Protein tyrosine phosphatases (PTPs) constitute a superfamily of structurally diverse set of signal transducing enzymes that work in coordination with protein tyrosine kinases (PTKs) in regulating various cellular processes. Various eukaryotic signal transduction pathways are triggered both by phosphorylating and dephosphorylating events (Barford *et al.* 1995). These pathways are important for several cell cycle functions such as proliferation and differentiation, metabolism, cytoskeletal organization, cell-cell interactions, gene transcription, and the immune response (Tonks and Neel 1996). The reversible mechanism of protein phosphorylation is modulated by the opposing dynamic actions of protein tyrosine kinases and phosphatases. PTKs are enzymes that promote signaling responses by catalyzing the transfer of the γ-phosphoryl group of ATP to the 4-hydroxyl of tyrosyl residues within specific protein/peptide substrates. In contrast, PTPs are hydrolytic enzymes that catalyze the dephosphorylation of tyrosine residues in phosphoprotein substrates which can in turn either potentiate or antagonize PTK-dependent signaling pathways. The protein phosphatases may be categorized into three structurally distinct families: the cytoplasmic and receptor-like tyrosine specific PTPs, the dual-specificity phosphatases (DSPs) and the low-molecular weight phosphatases (LMW-PTPs) (Zhang 1998). PTPs show exclusive substrate specificity for phosphotyrosyl proteins whereas DSPs can utilize protein substrates containing phosphotyrosine, as well as phosphoserine/threonine. Similarly, LMW-PTPs show substrate specificity primarily toward phosphotyrosine however they also show weak activities toward phosphoserine and phosphothreonine. There is little amino acid

sequence identity among the three subfamilies of phosphatases although they are all characterized by the ~240 amino acid conserved catalytic domain which includes the 11 residue PTP signature sequence motif (H/V)**C**$(X)_5$**R**(S/T) containing the essential catalytic cysteine and arginine residues (Barford *et al.* 1995).

Overall more than 100 PTPs have been identified so far (Zhang 1998). They exist as soluble cytosolic, non-receptor forms or receptor-like transmembrane forms, where the functional diversity within the PTP family is generated from the structural variation of the noncatalytic regulating and targeting sequences attached to either the amino or carboxyl terminal of the PTP domain (Barford *et al.* 1995; Zhang 1998). The receptor-like PTPs have an extracellular domain, a single transmembrane region and one to two cytoplasmic PTPase domains. The membrane-proximal PTPase domain usually exhibits phosphatase activity (Zhang 1998). The cytosolic PTPs have a single catalytic domain flanked by non-catalytic sequences which may regulate PTP activity directly and/or target the enzyme to particular subcellular localizations and substrates (Barford *et al.* 1995).

II. Protein Tyrosine Phosphatase-1B

Structure and function

Protein tyrosine phosphatase-1B (PTP-1B) is the first PTP to be purified to homogeneity from human placenta. It is a 37 kDa protein (321 amino acids) consisting predominantly of the catalytic domain (Tonks *et al.* 1988*a*, 1988*b*). Isolation of its cDNA illustrated that this purified protein was derived from a full-length molecule of 435 amino acid residues, containing a C-terminal regulatory segment fused to the catalytic domain (Guan *et al.* 1990; Chernoff *et al.* 1990; Brown-Shimer *et al.* 1990). As expected, PTP-1B contained the conserved PTP domain (residues 30-278) and a noncatalytic carboxyl-terminal. The extreme C-terminal 35 residues were composed of a hydrophobic segment that targets the enzyme to the cytoplasmic face of membranes of the endoplasmic reticulum (ER) (Flint *et al.* 1997). This motif was preceded by a stretch of hydrophilic residues (122) that contained sites for serine phosphorylation (Barford *et al.* 1994).

The three dimensional structure of PTP-1B has been determined by X-ray crystallography (Barford *et al.* 1995) and is shown in Figure 1. PTP-1B is a single domain of 8 α-helices and 12 β-strands. Four central parallel β-strands (β-3, β-12, β-4, β-11) are flanked by antiparallel β-strands on either side and above (β-5, β-6) the central β-sheet. The center is also buried by two α-helices (α-1 and α-2) on one side and a four-helix (α-3, α-4, α-5, α-6) bundle on top. The crevice of the catalytic site is formed by the conserved PTP signature motif (I/V)HCSAG*XX***R**(S/T)G which is found on a loop connecting the COOH-terminal of β-12 and the NH_2-terminal of α-4 (Barford *et al.*

1995). This loop contains the invariant His 214 which is important in stabilizing and optimizing the position of the nucleophilic Cys 215 and most residues whose main chain atoms are required for phosphate-binding (Ser 216-Gly 218 located at the COOH-terminal of β-12 and Ile 219-Arg 221 located at the NH_2-terminal of α-4). Polar groups on neighboring conserved residues such as Ser/Thr 222 also help to orient the phosphate binding loop by forming hydrogen bonds to main chain carbonyl groups on the loop (Gly 218) (Barford *et al.* 1994). In addition, other invariant nonpolar residues create a hydrophobic pocket that binds and stabilizes the phenyl ring of phosphotyrosine (Barford *et al.* 1995).

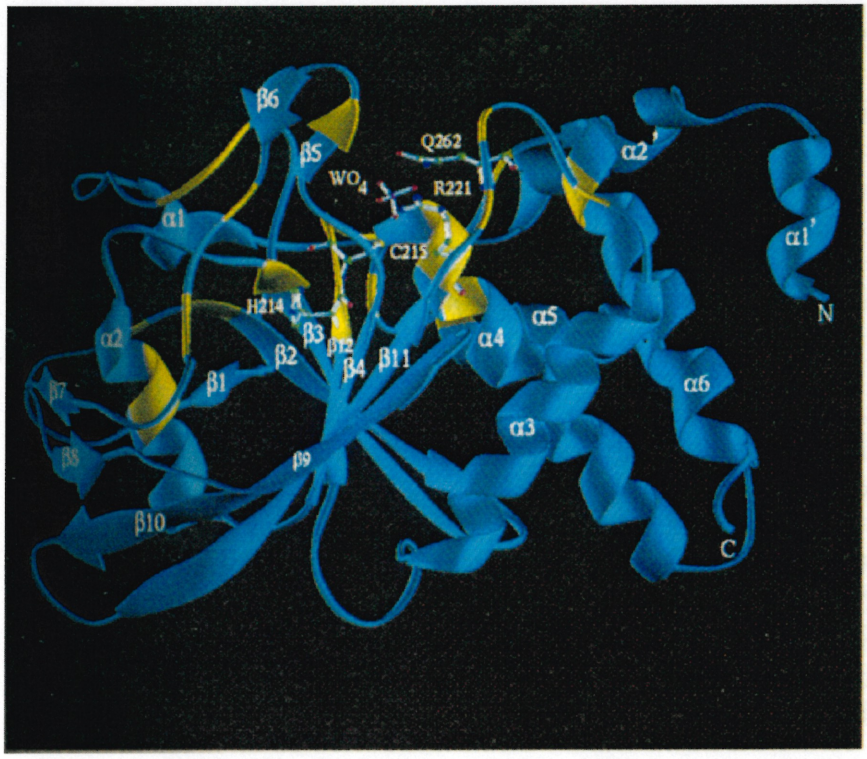


Figure 1. A ribbon representation indicating the secondary structure elements of PTP-1B, the catalytic site, and invariant residues (yellow). The α helices and β strands are labeled. The tungstate ion and the catalytic side chains of His^{214}, Cys^{215} and Arg^{221} are shown.

(*Source: Barford et al. 1994*)

Upon binding of a phosphotyrosine substrate to the catalytic site of PTP-1B a localized conformational change is induced such that a surface loop (residues 179-187) containing Asp 181 and Phe 182 swings into the catalytic site to form the closed loop conformation (Barford *et al.* 1995). Nucleophilic attack on the phosphorous atom of the phosphotyrosine substrate by the S_γ atom of the thiolate ion of Cys 215 forms a thio-phosphate intermediate. This thio-phosphate intermediate is stabilized by the main chain NH groups of Ser 216 to Arg 221 which form 6 hydrogen bonds to the terminal oxygens of the phosphate bound to the catalytic site. The guanidinium side chain of Arg 221 forms two salt bridges with the terminal phosphate oxygens. Phe 182 contributes by forming a hydrophobic pocket that binds the phosphotyrosine of the substrate (Barford *et al.* 1995; Flint *et al.* 1997). After formation of the thio-phosphate intermediate the phenolic oxygen of the tyrosyl leaving group is protonated by the general acid side chain of Asp 181 to facilitate cleavage of the scissile phosphate-oxygen (P-O) bond in the substrate for release of the phosphate-free protein product. The carboxylate side chain of Asp 181 is then suspected to later deprotonate a water molecule that hydrolyzes the enzyme-phosphate intermediate to release free phosphate (Flint *et al.* 1997). Figure 2 illustrates the reaction mechanism of PTP-1B as proposed by Barford *et al.* (1995).

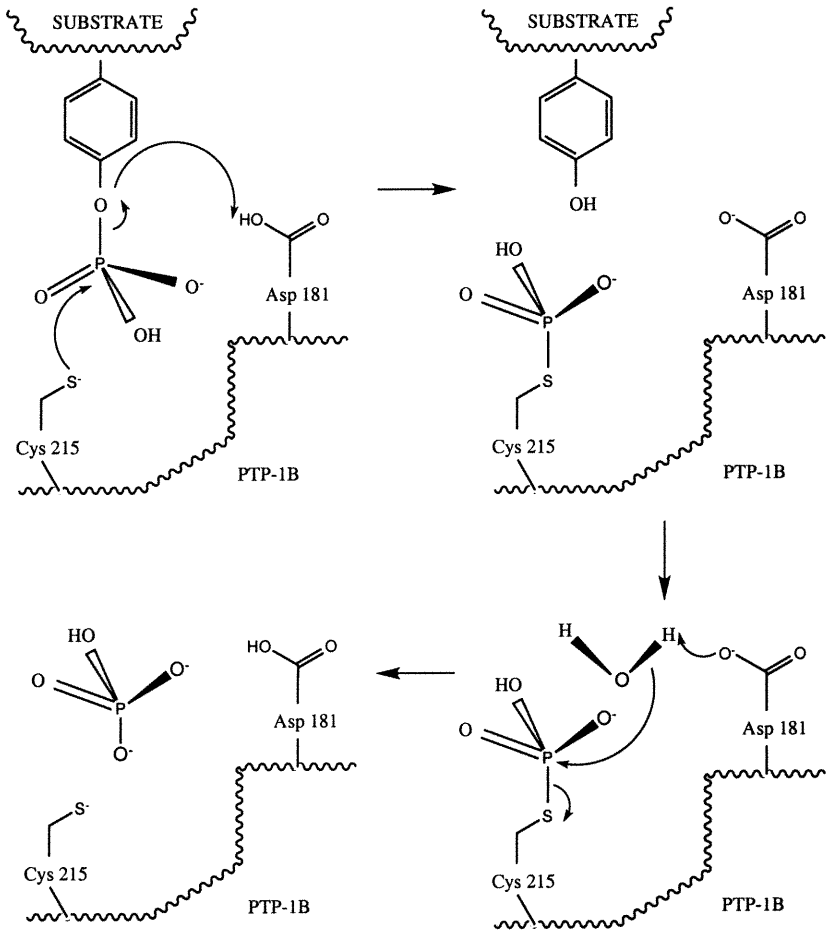


Figure 2. Schematic representation of the reaction steps catalyzed by PTP-1B.

(*Reference: Barford et al. 1995*)

The depth (9Å) and the hydrophobic character of the phosphate binding cleft may explain the specificity of PTPs for phosphotyrosine containing peptides since the smaller phosphoserine/threonine side chains may be too short to reach the recognition site (Barford *et al.* 1994; 1995). The enzyme also displays a subtle preference for multiple acidic side chains N-terminal to the phosphotyrosine residue as found on multiple sites of both human insulin receptor and rat liver insulin receptor substrates (IRS-1) (Kenner *et al.* 1996).

Insulin signaling pathway

One of the most extensively studied phosphatases to date is PTP-1B. Numerous studies of PTP-1B have been performed to elucidate the role and importance of PTP-1B. Initial studies of PTP-1B were suggestive of its role in insulin signaling and recently there has been a direct implication for the role PTP-1B in the insulin signaling pathway. PTP-1B is believed to participate in negative regulation of the insulin pathway by dephosphorylating the activated insulin receptor.

The insulin pathway is crucial in maintaining normal blood glucose homeostasis. It regulates the secretion of insulin by β-pancreatic islet cells and the balanced counter-regulatory hormone secretion of glucagon (Kahn *et al.* 1998; Taylor 1999). Insulin responds to increased glucose levels in the blood by stimulating glucose influx and metabolism from blood into muscle and adipose tissues, thus inhibiting gluconeogenesis by the liver (White and Kahn 1994). In contrast, during starvation, glucose is low which directs insulin-sensitive cells to mobilize stored fuels and promote gluconeogenesis in the liver by exporting the appropriate amino acids and substrates from muscle to liver (Taylor 1999).

Insulin action is mediated by the insulin receptor (IR) which is a transmembrane glycoprotein with intrinsic tyrosine kinase activity. The insulin receptor is a heterotetramer ($\alpha_2\beta_2$) of two αβ-subunits that are each linked to each other by disulfide bonds between the α-subunits (Hubbard *et al.* 1994). The extracellular α-subunits contain the insulin binding site(s) whereas the intracellular β-subunits contain the insulin-regulated tyrosine protein kinase. Insulin binds to the α-subunit of one αβ-subunit resulting in a quaternary structure change that places phosphorylation sites of one β-chain

within reach of the active site of the other β-chain such that intramolecular trans-autophosphorylation can occur at specific tyrosine residues of the adjacent β-chain (Hubbard *et al.* 1994). Phosphorylation of tyrosine residues Tyr 1158, Tyr 1162, and a third tyrosyl residue of the "regulatory domain" activates the intrinsic tyrosine kinase activity of the IR (Ahmad *et al.* 1995). In its unphosphorylated form, Tyr 1162, prevents the binding of both substrate and ATP in the active site (Ahmad *et al.* 1995). Upon activation of the IR with insulin, tyrosine phosphorylation of the principle insulin receptor substrate (IRS-1) on multiple tyrosine residues ensues (White and Kahn 1994; Kahn *et al.* 1998). The IRSs bind to the Src homology 2 (SH2) domains of the regulatory unit, p85, of phosphoinositol (PI) 3'-kinase which in turn recruits the catalytic unit, p110, of PI 3'-kinase (Kahn *et al.* 1998). Upon activation of PI 3'-kinase the insulin-responsive glucose transporter (GLUT4) is translocated from intracellular vesicles to the plasma membrane resulting in the transport of glucose from the blood into the cells for storage. Other processes including glycogen synthesis, protein synthesis, antilipolysis and suppression of hepatic gluconeogenesis also occur in response to this cascade of protein-protein interactions (Kahn *et al.* 1998). The general insulin signal transduction pathway is illustrated in Figure 3.

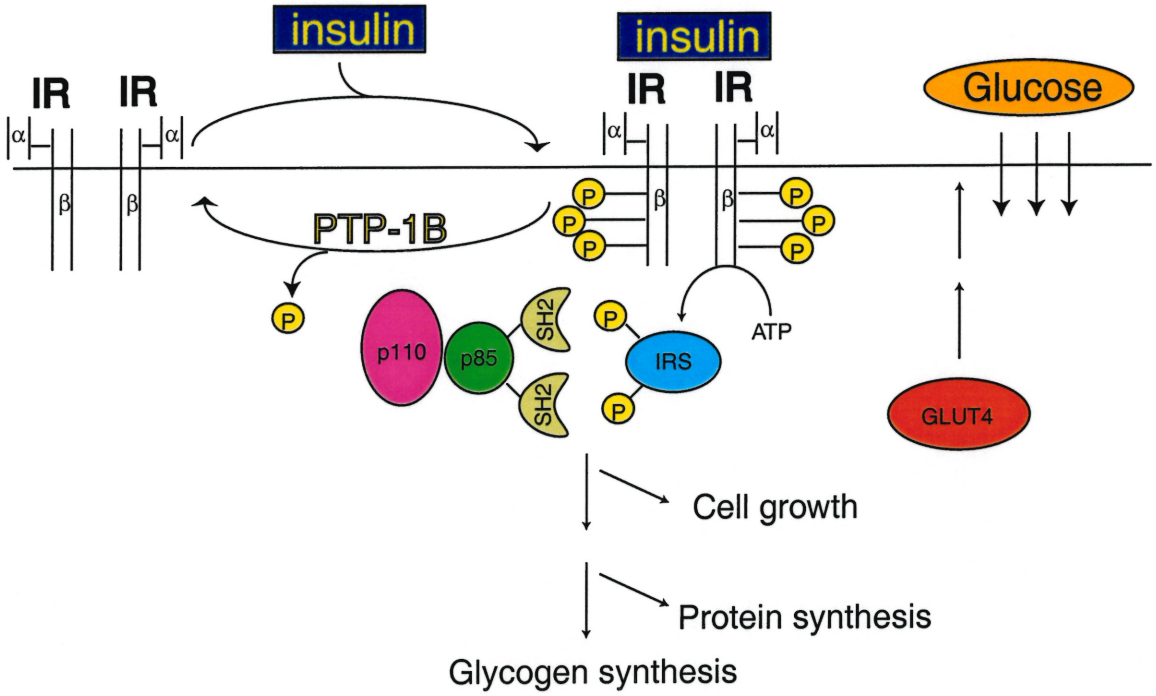


Figure 3. Insulin signaling pathway.

The most common metabolic disease involving dysfunction of the insulin signaling system is diabetes mellitus. Diabetes mellitus is the leading cause of blindness, renal failure and lower limb amputations in adults and a major risk factor for cardiovascular disease and stroke (Taylor 1999). There are two types of diabetes mellitus; an insulin-dependent type-I and non-insulin dependent type-II diabetes mellitus (NIDDM). Type-I diabetes is due to autoimmune destruction of pancreatic β-cells which leads to deficient insulin secretion but is, however, compensated with insulin injections. Type-II diabetes is due to resistance of insulin action and eventually a deficiency in insulin secretion. The onset of NIDDM is initially characterized by an increase in insulin secretion to compensate for the body's resistance to respond to insulin but eventually insulin deficiency develops due to failure of insulin secretion by pancreatic β cells

(Taylor, 1999). Insulin-sensitive tissues such as skeletal muscle and adipocytes develop insulin resistance thus impairing insulin action on glucose uptake and metabolism (Taylor, 1999). Susceptibility to insulin resistance is genetically determined and affected by environmental factors such as diet, physical activity, and age thus >90% of patients are afflicted with type-II diabetes.

Role of PTP-1B in insulin signaling pathway

There has been mounting evidence suggesting the role of PTP-1B in the insulin signaling pathway. Initial studies of purified placental PTP-1B microinjected into *Xenopus* oocytes were shown to block insulin-induced S6 peptide phosphorylation and inhibit insulin-induced oocyte maturation (Cicirelli *et al.* 1990; Tonks *et al.* 1990). Subsequent studies have demonstrated that PTP-1B is an abundant PTP expressed at relatively high levels in insulin-sensitive tissues, such as skeletal muscle, liver and adipose tissue (Goldstein 1993). The studies by Kenner et al (1993) demonstrated incubation of rat L6 skeletal muscle cells with insulin induced a significant time and concentration-dependent increase in PTP activity which was associated with enhanced activity of PTP-1B and expression of PTP-1B mRNA and protein. Overexpression of catalytically active PTP-1B inhibited insulin-stimulated insulin receptor autophosphorylation, tyrosine phosphorylation of IRS proteins and metabolic signaling of glucose incorporation into glycogen (Kenner *et al.* 1996). Studies of various mutant insulin receptor proteins in COS cells by Bandyopadhyay *et al.* (1997) further identified that the interaction of the insulin receptor with PTP-1B is dependent upon the level of

insulin-stimulated receptor autophosphorylation of insulin receptor tyrosine residues Tyr 1146, Tyr 1150 and/or Tyr 1151. In addition, tyrosine phosphorylation of PTP-1B residues Tyr 66, Tyr 152 and Tyr 153 by the activated insulin receptor tyrosine kinase is important for the association of PTP-1B with the insulin receptor (Bandyopadhyay *et al.* 1997).

Recently, PTP-1B has been identified to have a major role in the negative regulation of insulin. Elchebly *et al.* (1999) observed that PTP-1B $^{-/-}$ knockout mice demonstrated slightly lower blood glucose levels and half the level of circulating insulin compared to their PTP-1B $^{+/+}$ littermates. Furthermore, when the PTP-1B $^{-/-}$ knockout mice were placed on a high-fat diet to induce development of type-II diabetes, the researchers observed enhanced insulin sensitivity in muscle and liver tissue and resistance to weight gain. In contrast, PTP-1B $^{+/+}$ littermates rapidly gained weight and became insulin resistant when placed on a high-fat diet. Because PTP-1B acts by dephosphorylating the activated insulin receptor, the increased insulin sensitivity in its absence, in the PTP-1B $^{-/-}$ knockout mice, suggests increased and/or prolonged phosphorylation of the insulin receptor. The resistance to weight gain is inexplicable because knockouts and wild-type mice showed no significant differences in insulin-stimulated receptor phosphorylation in adipose tissue. Consequently, PTP-1B has been identified as a potential therapeutic drug target for the treatment of type-II diabetes and obesity (Elchebly *et al.* 1999).

III. T-cell Protein Tyrosine Phosphatase

Of the one hundred PTPs identified thus far, T-cell protein tyrosine phosphatase displays overall 65% sequence identity to PTP-1B and 74% sequence identity within the conserved catalytic domain (Tiganis *et al.* 1997), refer to Figure 4. Furthermore, the undetermined role of TCPTP introduces a possibility of complicating the development of inhibitors specific to PTP-1B. Because evidence suggests that TCPTP plays an important role in positive proliferation of T- and B-cells, drugs designed to inhibit PTP-1B must be highly specific for PTP-1B.

Studies by You-Ten *et al.* (1997) have found TCPTP to be expressed at all stages of mammalian development and most notably in lymphoid cell lineages. Studies of homozygous $^{-/-}$ TCPTP deficient mice were observed to be deficient in developing immunocompetent T- and B-cells. Furthermore, TCPTPs were found to be critical in maintaining the integrity of the bone marrow microenvironment for the early development of blood cells suggesting its possible role in both hematopoiesis and immunity (You-Ten *et al.* 1997). Human TCPTP is an intracellular nontransmembrane protein that was originally isolated from a human peripheral T-cell cDNA library (Cool *et al.* 1989). This cDNA encodes a 48-kDa protein that contains the conserved catalytic PTP signature motif including the invariant PTP catalytic acid, Asp 182, and a non-catalytic C-terminal segment. Alternative mRNA splicing gives rise to two splice variants; a 48-kDa (TC48) and a 45-kDa (TC45) form of

```
TCPTP: IEREFEELDTQRRWQPLYLEIRNESHDYPHRVAKFPENRNRNRYRDVSPYDHSRVKLQNA 64
       +E+EFE++D    W  +Y +IR+E+ D+P RVAK P+N+NRNRYRDVSP+DHSR+KL
PTP1B: MEKEFEQIDKSGSWAAIYQDIRHEASDFPCRVAKLPKNKNRNRYRDVSPFDHSRIKLHQE 62

TCPTP: ENDYINASLVDIEEAQRSYILTQGPLPNTCCHFWLMVWQQKTKAVVMLNRIVEKESVKCA 124
       +NDYINASL+ +EEAQRSYILTQGPLPNTC HFW MVW+QK++ VVMLNR++EK S+KCA
PTP1B: DNDYINASLIKMEEAQRSYILTQGPLPNTCGHFWEMVWEQKSRGVVMLNRVMEKGSLKCA 122

TCPTP: QYWP-TDDQEMLFKETGFSVKLLSEDVKSYYTVHLLQLENINSGETRTISHFHYTTWPDF 183
       QYWP  +++EM+F++T   + L+SED+KSYYTV  L+LEN+ + ETR I HFHYTTWPDF
PTP1B: QYWPQKEEKEMIFEDTNLKLTLISEDIKSYYTVRQLELENLTTQETREILHFHYTTWPDF 182

TCPTP: GVPESPASFLNFLFKVRESGSLNPDHGPAVIHCSAGIGRSGTFSLVDTCLVLMEKGDD-- 241
       GVPESPASFLNFLFKVRESGSL+P+HGP V+HCSAGIGRSGTF L DTCL+LM+K  D
PTP1B: GVPESPASFLNFLFKVRESGSLSPEHGPVVVHCSAGIGRSGTFCLADTCLLLMDKRKDPS 242

TCPTP: -INIKQVLLNMRKYRMGLIQTPDQLRFSYMAIIEGAKCIKGDSSIQKRWKELSKEDLSPA 300
        ++IK+VLL MRK+RMGLIQT DQLRFSY+A+IEGAK I GDSS+Q +WKELS EDL P
PTP1B: SVDIKKVLLEMRKFRMGLIQTADQLRFSYLAVIEGAKFIMGDSSVQDQWKELSHEDLEPP 302

TCPTP: FDH-----SPNKIMTEKYNG-------NRIGLEEEKLTGDRCT--------------GLS 334
        +H       P K + E +NG        N    ++EE       C                 G+
PTP1B: PEHIPPPPRPPKRILEPHNGKCREFFPNHQWVKEETQEDKDCPIKEEKGSPLNAAPYGIE 362

TCPTP: SKMQDTMEENSESALRKRIREDRKATTAQKVQQMKQRLNENERKRKRWLYWQPILTKMGF 394
       S  QDT  E     +   +R  + A+ A+     + ++  ++           YW+P L  M
PTP1B: SMSQDT--EVRSRVVGGSLRGAQAASPAKGEPSLPEKDEDHALS-----YWKPFLVNMCV 415

TCPTP: MSVILVGAFVGWRLFFQQN 413
        +V+  GA++ +R  F  N
PTP1B: ATVLTAGAYLCYRFLFNSN 434
```

Figure 4. Sequence alignment for T-cell protein tyrosine phosphatase (TCPTP) (48 kDa splice variant) and placental PTP-1B. The middle line is the consensus sequence and + indicates conservation.

human TCPTP which differ in their C-terminal domains (refer to Figure 5). The C-terminal ends vary in size, hydrophobicity, and function in subcellular localization, regulating substrate specificity and enzymatic activity (You-Ten *et al.* 1997; Hao *et al.* 1997). It is the difference in the non-catalytic C-terminal domains of these two splice variants that is suspected to generate functional diversity of TCPTPs. Although the specific roles of these TCPTP splice variants have yet to be determined, studies on their mechanism of localization to subcellular compartments have been reported.

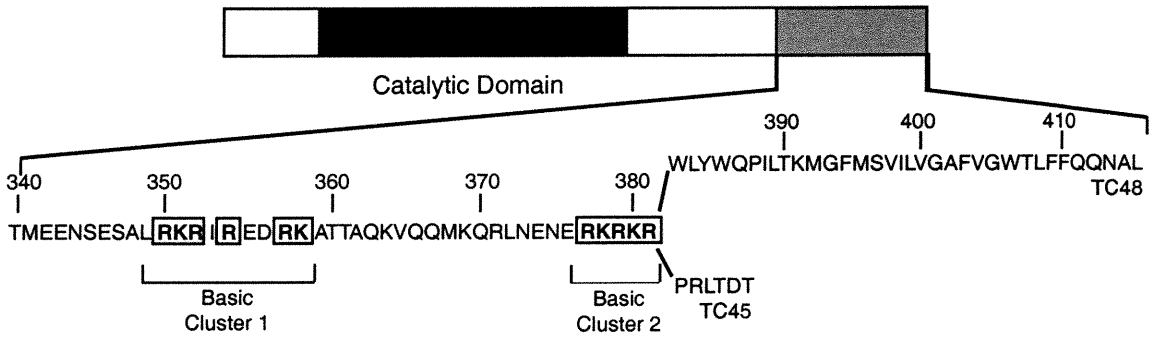


Figure 5. Schematic representation of TCPTP splice variants.

The catalytic region is shown in black, the noncatalytic region is shown in gray. The sequence of the noncatalytic domain (residues 340-415) is shown for both the COOH-termini of splice variant TC45 and TC48. The two basic clusters of the bipartite nuclear localization sequence are boxed.

TC48 contains a 34-residue C-terminal hydrophobic tail which, similar to PTP-1B, localizes the protein to the ER (Cool *et al.* 1990). This C-terminal domain consists of a 19-residue hydrophobic tail and residues 346-358 which consist of basic amino acids (Lorenzen *et al.* 1995). Supportive studies performed by Kamatkar *et al.* (1996) have demonstrated that the 34-residue hydrophobic tail inhibits interaction of TC48 with DNA and prevents its entry into the nucleus. Furthermore, gel filtration studies have demonstrated that TC48 elutes as a tetramer (200 kDa) and only as a monomer if the truncated TC48 lacking the C-terminal 34 amino acids was used (Kamatkar *et al.* 1996). This suggests that the C-terminal tail of TC48 is important for oligomerization and may be capable of interacting with other proteins which may in turn have targeting functions (Kamatkar *et al.* 1996; Tiganis *et al.* 1997).

In contrast, this hydrophobic tail is replaced by a unique hydrophilic 6-residue (PRLTDT) C-terminal end in TC45 and is targeted to the nucleus. Two basic clusters (residues 350-358 and 377-381) have been found to form a bipartite nuclear localization sequence (NLS) which promotes efficient nuclear translocation (Tiganis *et al.* 1997).

These two basic clusters also exist in TC48 which means that the hydrophobic C-terminal tail of TC48 must override the bipartite NLS to permit targeting to the ER (Lorenzen *et al.* 1995). Although TC45 contains a bipartite NLS it has been observed to exit the nucleus following epidermal growth factor (EGF) stimulation which suggests that it can possibly gain access to cytoplasmically localized substrates (Tiganis *et al.* 1998). Studies by Hao *et al.* (1997) have suggested that the non-catalytic C-terminal may very well play a regulatory role in modulating the activity of the catalytic domain and that TC45 may also contain an auto-regulatory site.

Similar to PTP-1B, both TC45 and TC48 and their C-terminal truncated versions bind *p*-nitrophenyl phosphate (pNPP) as substrate (Hao *et al.* 1997; Kamatkar *et al.* 1996) however they differ in enzymatic activity toward specific phosphorylated protein substrates. More specifically, it is the C-terminal regulatory domain which is important in determining interaction with large peptide/protein substrates and in affecting PTP enzymatic activity (Hao *et al.* 1997; Kamatkar *et al.* 1996). Although the two splice variants share identical catalytic domain it is their difference in the non-catalytic C-terminal that differentiates their enzymatic activity, substrate specificity, DNA binding, and subcellular localization (Kamatkar *et al.* 1996).

In addition, affinity testing of the nearly full set of known human PTPs using "substrate-trapping" mutants (D181A) for autophosphorylated insulin receptor kinase (IRK) identified PTP-1B and TCPTP, amongst others, to be specific for the autophosphorylated insulin receptor (Wälchi *et al.* 2000). These same PTPs were further confirmed to dephosphorylate the insulin receptor (Wälchli *et al.* 2000) which suggests a possibility that TCPTP may directly or indirectly affect the insulin signaling pathway.

PTP-1B and TCPTP share a high degree of sequence similarity, a common subcellular localization and have the common ability to dephosphorylate the insulin receptor. Efforts directed at the development of therapeutic agents towards PTP-1B inhibition for the treatment of type-II diabetes must take into account the close similarity between PTP-1B and TCPTP. To aid the discovery of selective inhibitors against PTP-1B, description of the inhibitor-enzyme interactions in structural terms may help identify the significant differences between PTP-1B and TCPTP. These key differences can then be exploited to maximize PTP-1B inhibition while maintaining TCPTP inhibition to a minimum.

Extensive studies of PTP-1B-inhibitor complexes by X-ray crystallography, NMR spectroscopy and molecular modelling have been carried out and have provided a wealth of structural information. Unfortunately, no structural information can be obtained on TCPTP inhibition. This is due to the impossibility in obtaining crystals of TCPTP-inhibitor complexes. The lack of structural information prompted the initiation of an effort to use NMR spectroscopy for the determination of the three dimensional structure of TCPTP and develop methods to describe TCPTP inhibition in structural terms.

NMR spectroscopy has advanced as a sensitive and versatile technique for the structural studies of proteins in solution. Although research on the structural properties and functional role of PTP-1B have progressed, only structural studies of PTP-1B employing NMR techniques has been initiated in our laboratory. By studying both PTP-1B and TCPTP concurrently may help reveal distinct structural differences between TCPTP and PTP-1B that can be exploited in the rational design of selective inhibitors towards PTP-1B.

Chapter 2

NMR Spectroscopy

I. Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a versatile and informative technique employed in the chemical research laboratory to identify and elucidate structures of small molecules. In the past three decades, NMR has progressed as an eminent tool in structural biology and has undergone significant improvements. The NMR techniques developed thus far have provided a means for studying macromolecular solution structures at atomic resolution, for characterization of dynamic properties of macromolecules, such as the internal mobility within a protein or its overall rotational diffusive mobility and for describing properties of partially unfolded proteins (Wagner 1997). Presently, NMR spectroscopy is becoming a useful tool in drug discovery because it permits characterization of very weak interactions between macromolecules and ligands. This application of NMR, termed SAR ('structure-activity relationships') by NMR, as developed by Fesik and coworkers (Shuker *et al.* 1996; Wagner 1997) has enabled the discovery of high-affinity ligands for target proteins.

Even with the improvements in both NMR hardware and NMR techniques, NMR studies of large biomolecules, such as TCPTP, are not trivial. Therefore, it is worthwhile to briefly describe some basic theoretical concepts that are important in understanding the key techniques applied in the study of large proteins by NMR.

Basic NMR Concepts

The NMR active nuclei available in proteins for NMR spectroscopy are ^{13}C, ^{1}H and ^{15}N atoms. These NMR active nuclei are all characterized with a nuclear spin quantum number (*I*) of ½. In the absence of a strong external magnetic field they may be viewed as tiny magnetic bars with random orientations. However, in the presence of a static magnetic field (denoted B_0) these nuclei adopt specific orientations with respect to the magnetic field (refer to Figure 6). The magnetic moment (μ) of the nuclei, which is characterized by its angular momentum, J, and magnetogyric ratio, γ, such that $\mu = \gamma J$, experiences a torque and interacts with the field by precessing about the magnetic field lines in a discrete number of orientations. The number of spin orientations available to the spinning nucleus is dependent upon its nuclear spin quantum number, where a nuclei with $I = ½$ has two spin orientations ($2I+1$) available. Therefore, the nuclear spins of ^{13}C, ^{1}H or ^{15}N atoms may either take up an orientation parallel to the magnetic field, the energetically lower α-state, or antiparallel to the magnetic field, the higher β-state (refer to Figure 6). The rate at which the nucleus precesses is dependent on the field strength and the magnetogyric ratio of the spin and is known as the *Larmor frequency* (ω_0) of the nucleus.

At equilibrium the orientations of the nuclei are distributed such that there is an excess of nuclei in the α state as defined by the Boltzmann distribution:

$$\frac{N_\beta}{N_\alpha} = e^{-\Delta E/kT}$$

where $N_{\alpha,\beta}$ represents the number of nuclei in each spin orientation, k the Boltzmann constant and T the temperature. This small population difference between the possible

spin states may be represented as a collection of spins distributed randomly about the *z*-axis forming a cone parallel to the magnetic field which gives rise to a resultant bulk magnetization vector M_0 (refer to Figure 6).

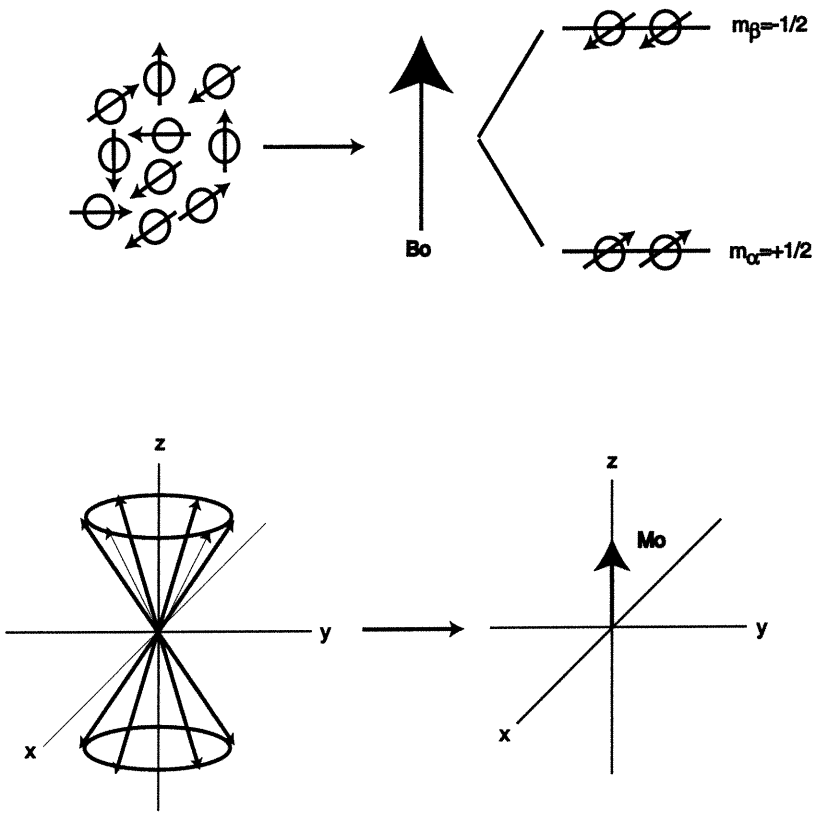


Figure 6. NMR vector model where many like spins are represented by a bulk magnetization vector M_0.

M_0 may be viewed as $N_\alpha - N_\beta$ such that at thermal equilibrium the excess of spins in the α-state (m_α=+½) places M_0 parallel to the *z*-axis.

To obtain an NMR signal this bulk magnetization is perturbed by a second magnetic field (B_1) by applying a radio-frequency (rf) pulse perpendicular to the static field B_o. If the electromagnetic radiation of this rf pulse is oscillating at the Larmor frequency of the nucleus then the bulk magnetization, which is now made up of the small excess of population in the higher energy state, also precesses about this applied magnetic

field. Since this electromagnetic field is applied perpendicular to the bulk magnetization the components of the bulk magnetization will experience a torque such that as it precesses about the applied magnetic field it will also fall from the *z*-axis into the transverse *x*-*y* plane. Therefore, when a 90° ($\pi/2$) pulse is applied, the rf generator is left open long enough such that the bulk magnetization may be brought into the *x*-*y* plane. However, if the generator is left open longer, as in a 180° (π) pulse, the bulk magnetization is brought along the –*z*-axis. The energy (E) involved in generating the NMR signal is related to the frequency (ν) by:

$$\Delta E = h\nu = \frac{h\gamma B_0}{2\pi}$$

$$\text{therefore, } \omega_0 = 2\pi\nu = \gamma B_0$$

Each NMR active nucleus gives rise to a unique precessing frequency which is characteristic of the nucleus type as well as its surrounding chemical environment. Naturally, the system will try to re-establish its equilibrium condition such that the transverse magnetization vector will gradually disappear and simultaneously grow along the *z*-axis. Consequently, the NMR signal decays with time and is detected as a free induction decay (FID). To convert the resulting intensity versus time-domain spectrum into a frequency-domain spectrum the FIDs are processed by Fourier transformation. For convenience, the resulting frequency at which this NMR signal occurs is expressed relative to the strength of the magnetic field as a chemical shift in parts per million (ppm).

The lifetimes of excited nuclear spins are usually long enough to give reasonably narrow NMR resonance lines; however, the natural tendency of these nuclei to return to

their equilibrium state is achieved through various relaxation pathways which renders the NMR signals short-lived.

II. Spin relaxation parameters

When a rf pulse is applied, the bulk magnetization vector moves away from the thermal equilibrium z-axis towards the x-y transverse plane. The system will naturally re-establish the equilibrium conditions by a process termed relaxation. The two different relaxation parameters that describe this phenomena are longitudinal relaxation and transverse relaxation. Bloch has postulated that both these relaxations are first order processes characterized by time constants T_1 (longitudinal) and T_2 (transverse).

Longitudinal relaxation:T_1

Longitudinal relaxation describes the recovery of the magnetization along the z-axis. In other words, the recovery of the thermal equilibrium condition. When the bulk magnetization (M_0) is precessing in the x-y transverse plane, relaxation whereby the bulk magnetization decreases in the transverse plane and returns to the z-axis (refer to Figure 7) may be described by longitudinal relaxation. Longitudinal relaxation, also referred to as spin-lattice relaxation, corresponds to the equilibrium spin populations being re-established since the spins lose their energy to the surroundings. The exponential recovery of the z-magnetization, M_z, is described as:

$$M_z = M_0 (1-e^{-t/T1})$$

where M_0 is the bulk magnetization at thermal equilibrium and T_1 is the time constant for this process. The T_1 relaxation parameter is important to consider in implementing NMR experiments since it determines the repetition time at which a given experiment can be repeated.

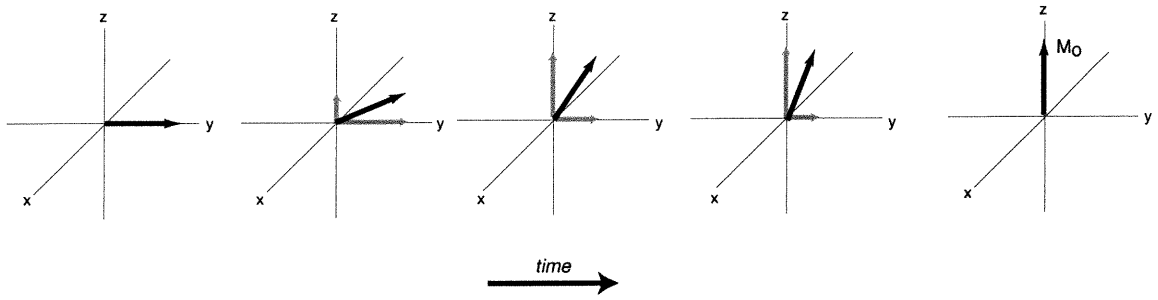


Figure 7. Longitudinal relaxation.
Recovery of M_z magnetization as *x-y* transverse plane signal diminishes.

Transverse relaxation: T_2

Transverse relaxation describes the loss of magnetization in the *x-y* transverse. Following a 90° pulse, in the transverse plane the bulk magnetization results from the summation of many phase coherent microscopic vectors arising from each individual nuclei in a magnetic field. This state is referred to as coherence. Due to magnetic field inhomogeneity across a sample, arising from the hardware and the sample itself, these individual nuclei experience slightly different magnetic fields. As a result, the individual magnetization vectors may start fanning-out which ultimately leads to no net magnetization in the transverse plane (refer to Figure 8). This process of relaxation is referred to as transverse relaxation. The magnetic field differences which lead to a loss of phase coherence in the transverse plane also arise from local magnetic fields generated by intramolecular and intermolecular interactions. Therefore, it is customary to distinguish between the contributions to T_2 as follows:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_{2(\Delta Bo)}}$$

where T_2 refers to the contribution from molecular relaxation processes and $T_{2(\Delta Bo)}$ to that due to experimental factors such as field inhomogeneity. In order to minimize field inhomogeneity it is important to properly shim the magnet to ensure the sample experiences as homogenous a magnetic field as possible.

The T_2 relaxation parameter is important for spectral resolution because T_2^* is inversely proportional to the NMR resonance linewidth at half height:

$$\Delta\nu_{1/2} = \frac{1}{\pi T_2^*}$$

As a result, it is an important relaxation parameter which affects spectral resolution of the NMR experiment. For example, a short T_2^* (fast transverse relaxation) corresponds to a faster loss of transverse phase coherence leading to a greater frequency difference between the individual vectors and thus a broader line in the frequency dimension.

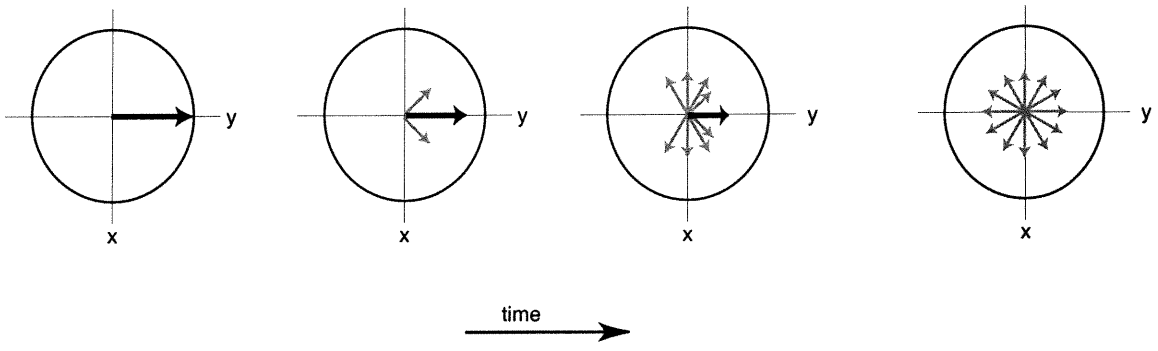


Figure 8. Transverse relaxation.

Local field differences within the sample cause spins to precess with slightly differing frequencies, eventually leading to loss of transverse phase coherence.

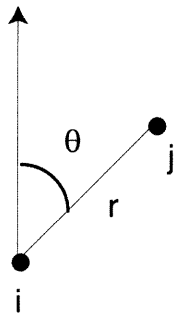
The time constants T_1 and T_2 describe important spin relaxation parameters which may affect the linewidth, resolution and sensitivity of the NMR spectra, and must therefore be considered when designing and implementing NMR experiments for the study of proteins.

III. Relaxation mechanisms

Relaxation occurs mainly through local fluctuating magnetic fields created by various molecular interactions such as dipole-dipole interactions, chemical shift anisotropy and quadrupolar relaxations. For protein NMR the most significant relaxation processes are dipole-dipole and chemical shift anisotropy mechanisms.

Dipole-dipole interaction

In solution, the rapid reorientation of the molecule averages dipole-dipole interactions to zero due to the angular dependence of the dipole-dipole interaction (refer to Figure 9). Consequently, the dipolar interaction is not directly observable however these interactions generate local fields that contribute to relaxation.



$$D_{ij} = \frac{\mu_0 h}{4\pi} \frac{\gamma_i \gamma_j}{r^3} \frac{(3\cos^2\theta - 1)}{2}$$

Figure 9. Dipole-dipole interaction.

Dipole-dipole interaction is the major relaxation mechanism for spin-½ nuclei in solution. When the dipoles of two spin-½ nuclei approach each other their magnetic moments interact depending on their relative orientations. As illustrated in Figure 9 this interaction is related to the product of the magnetogyric ratio (γ) of the two interacting nuclei (i and j), the internuclear distance (r) and the angle (θ) between the internuclear vector and the external magnetic field. The magnetogyric ratio is the intrinsic NMR sensitivity of a given nucleus therefore the greater the sensitivity of a nucleus the greater the magnitude of dipole-dipole interaction contributing to relaxation. The main source of relaxation for protons is due to the interaction between magnetic dipoles of neighboring protons. Furthermore, relaxation of heteronuclei is dominated by dipole-dipole interactions arising from the bound protons.

Chemical shift anisotropy

The chemical shift of an NMR active nucleus arises as a result of the nucleus type and the opposing electron shielding it experiences from the surrounding electron clouds. Some chemical groups such as amides, carboxyls, carbonyls, aromatics, display an unsymmetrical (anisotropic) electronic distribution. Consequently, the chemical shift of a nucleus within such a group will be dependent on the orientation of the group with respect to the magnetic field. The chemical shift anisotropy (CSA) is best described with the chemical shift tensor. The chemical shift tensor describes the extent of electron shielding a given nucleus experiences. The principal components of the chemical shift tensor, σ_{xx}, σ_{yy}, σ_{zz}, are used to describe the three-dimensional nature of electron shielding for a given chemical moiety. Each of the three components represents the

chemical shift (in ppm) of this group as if aligned parallel to the external magnetic field (refer to Figure 10).

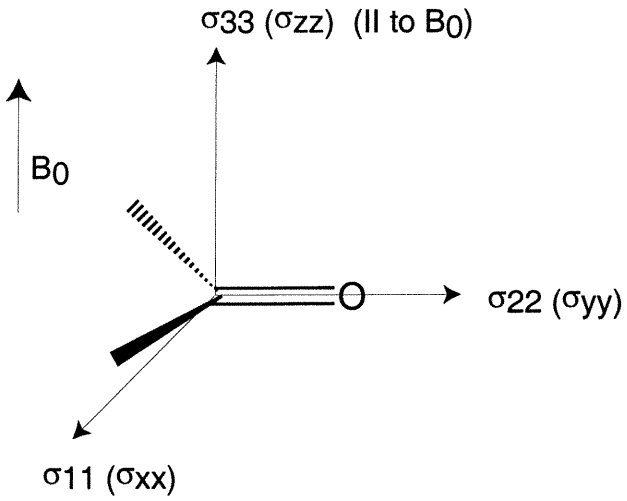


Figure 10. The three principal components of the chemical shift tensor for a carbonyl groups.

In solution, fast tumbling of the molecule averages the components of the chemical shift tensor to its isotropic value $\sigma_0 = 1/3\ (\sigma_{xx} + \sigma_{yy} + \sigma_{zz})$. Nonetheless, the fast molecular reorientation creates fast oscillation of the chemical shifts. These oscillations can be pictured as local fields that can contribute to the relaxation of the NMR signal. It is noteworthy to bear in mind that CSA interactions are directly related to the chemical shift tensor elements expressed in ppm therefore as the external field increases in strength, the CSA interactions become more prominent.

Correlation time (τ_c) and spectral density function

Analogous to exciting a nuclear spin with a magnetic field precessing at its Larmor frequency, nuclear spin relaxation proceeds through small perturbations arising from local magnetic fields. Various local magnetic fields are generated from dipole-

dipole and CSA interactions produced by the motions of molecules. These molecular motions are characterized by the rotational correlation time (τ_c). The correlation time is the average time a molecule takes to rotate through one radian. For a protein, this tumbling rate is related to its molecular weight, molecular shape, surface structure, protein-protein interaction and internal mobility (Wagner 1997). It is the random molecular motions and collisions with other molecules which provide a wide range of frequencies available for relaxation. Knowing that collisions between molecules in liquids occurs at an average time of $10^{-14} - 10^{-12}$ s (Cavanagh *et al.* 1996), frequencies of local fields generated by molecular motions will range from zero to 10^{12} Hz. In other words, dipole-dipole and CSA interactions can be modulated by molecular motions over a wide range of frequencies that will induce relaxation of excited magnetic spins. The local fields produced by the modulation of these interactions can be seen as small magnetic fields capable of perturbing the individual magnetic moments in an analogous fashion as a rf pulse does to the bulk magnetization. Relaxation theory states that the magnitude of these perturbations at a given frequency is related to the correlation time. Therefore, the distribution of frequency at which significant local field perturbations are available for relaxation is defined as the spectral density function $J(\omega)$. The spectral density function describes the power available for relaxation from local magnetic fields as a function of frequency. The spectral density functions, $J(\omega)$, for fast, intermediate and slow molecular tumbling rates, are illustrated schematically in Figure 11A with the corresponding T_1 and T_2 curves in Figure 11B.

Analysis of the spectral density functions (see Figure 11A) helps understand the effect of molecular motion on T_1 and T_2. Small, fast tumbling molecules are

characterized by a short τ_c such that the power available for relaxation is low and over a wide range of frequencies. As a result there is inefficient longitudinal and transverse relaxation which is illustrated in Figure 11B whereby both T_1 and T_2 are long. Due to the long T_2 generally narrow NMR resonance linewidths are observed. For large, slow tumbling molecules such as proteins, most of the power available for relaxation is in the low frequency range. From Figure 11B, the result is that these low frequencies do not promote efficient longitudinal relaxations (long T_1) but contribute significantly to efficient transverse relaxation (short T_2). The shorter T_2 means there is a faster loss of NMR phase coherences therefore generally low intensity and broad NMR signals are observed for biomolecules.

In cases when the correlation time is in the order of the reciprocal of the Larmor frequency ($\tau_c \approx 1/\omega_0$) longitudinal relaxation is most efficient because significant amount of power available for relaxation is available at the resonance frequency (ω_0) such that T_1 reaches a minimum. As a result, recovery of the bulk magnetization to the equilibrium z-axis occurs faster and small broadening of the NMR resonance linewidths are observed.

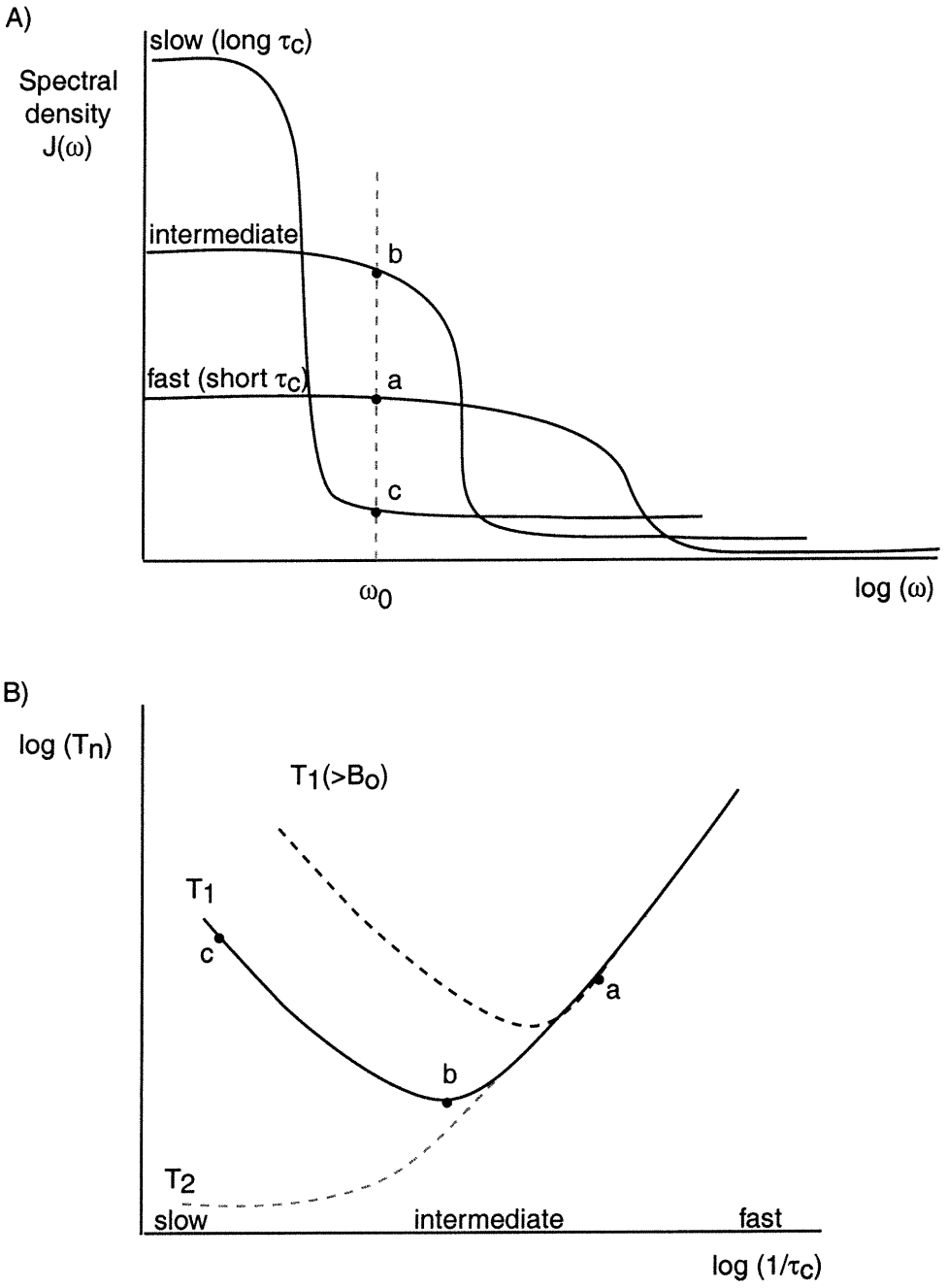


Figure 11. A) Schematic representation of the spectral density as a function of frequency shown for molecules at varying tumbling rates and, B) corresponding T_1 and T_2 curves for spins with a Larmor frequency ω_0 as a function of molecular tumbling rates (inverse τ_c correlation times).

The T_1 curve field dependency is illustrated as the darker dashed curve. ω_0 is field dependent therefore the T_1 minimum occurs for faster motion at higher fields. The dependence of T_2 relaxation on molecular tumbling rates is illustrated as the lighter dashed curve. In contrast to T_1, T_2 relaxation is sensitive to very slow motions.

The limitations of using NMR spectroscopy for the study of large proteins (M.W. ≥25 kDa) is due to the fact that large molecules have long τ_c that results in shorter T_2. Since T_2 directly affects the linewidth, (see equation on p.26) signals are simply lost in the noise and are difficult or impossible to detect. Means of reducing transverse relaxation (increasing T_2) would contribute to better intensity and resolution of NMR signals arising during the study of large proteins. For example, increasing the temperature of a protein solution results in less efficient transverse relaxation because higher temperatures encourage faster protein tumbling rates, characterized by shorter τ_c, thus translating into a longer T_2. In addition, replacement of protons with deuteriums would weaken dipole-dipole interactions due to the smaller magnitude of γ_D compared to γ_H.

IV. Strategy for studying proteins using solution NMR

Determination of the three-dimensional (3D) structure of a protein initially requires efficient expression and purification of high levels of pure monomeric protein at concentrations of at least 1 mM. Subsequently, chemical shift assignment of the polypeptide backbone resonances of the different amino acid types using through-bond and through-space correlation experiments must be obtained. Torsion angle restraints are then derived from three-bond scalar coupling constant data and interproton distance restraints are obtained by measuring through-space NOE interactions belonging to residues that may be far apart in the primary sequence but less than 5 Å apart in space. In addition, residual dipolar couplings which are dependent on the rotational anisotropy relate the local geometric orientation of N-H and C-H internuclear vectors relative to the rotational diffusion axis. Finally, an ensemble of protein structures that satisfy these experimental NMR restraints are calculated using distance geometry, torsion angle dynamics and simulated annealing alogrithms.

In order to obtain complete sequential assignment of the polypeptide backbone various multidimensional, multi-resonance NMR experiments have been developed and implemented. Recall that the NMR vector model may be used to illustrate how a short, nonselective rf pulse applied to excite spin state transitions results in the bulk magnetization vector falling from the z-axis towards the x-y transverse plane. This magnetization continues to precess at its Larmor frequency, undergoing chemical shift evolution, and the NMR signal detected is the chemical shift frequency of this particular nucleus from the carrier frequency (ω_0). If two NMR active nuclei are connected through

a chemical bond they are scalar-coupled, the magnetic spin states of one nuclei will influence the frequency at which the other precesses. For scalar-coupled spin-½ nuclei, instead of observing one unique NMR frequency signal, a doublet of NMR signals, separated by their coupling constant (J), is observed. This doublet is a result of the excited nuclei observing both the α and β spin states of its coupled neighbor such that half of the spin population of the former will precess slightly faster and the other half slightly slower at $\pm J/2$. Using the NMR vector model this may simply be viewed as two vectors, separated by their coupling constant, precessing away from each other in the transverse plane.

The vector model is simple and sufficient in illustrating how the bulk magnetization rotates under the influence of a pulse and precesses under the influence of chemical shift evolution and scalar-coupling evolution. Although it is easy to envision simple NMR experiments using the vector model, it is inadequate for describing the more complex multidimensional, multi-resonance NMR experiments employed in the study of macromolecular proteins. As a result, product operator formalism has been developed to facilitate the description of the spin system state resulting during the application and evolution of the NMR pulse sequence. Refer to Appendix A for a more a complete explanation of the product operator formalism.

Insensitive Nuclei Enhanced by Polarization (INEPT)

To study proteins using solution NMR a brief explanation of the basic INEPT transfer is useful. In proteins, the INEPT pulse sequence forms the basis of most NMR experiments employed for the study of proteins in solution. Recall that the intrinsic

NMR sensitivity of an NMR active nucleus depends on both its natural abundance and its magnetogyric ratio (γ) where the higher the γ, the larger the energy difference between the available spin transition states. As a result the Boltzmann population difference is greater and a higher resonance frequency is observed. Proton has a natural abundance of 99.98% and has one of the largest γ, with the exception of tritium, which makes it a favorable nucleus for detection in high resolution NMR. Most simply, through scalar couplings shared between heteronuclear partners the INEPT transfer enables the high energy bath of nuclei such as ^{1}H to be transferred to the lower-γ nuclei such as ^{13}C or ^{15}N resulting in higher NMR sensitivity of these nuclei. INEPTs permit the transfer of magnetization from one nucleus to another that is scalar coupled. Refer to Appendix B for a more detailed explanation of the INEPT pulse sequence.

Generally, the NMR experiments employed to obtain through-bond correlations are designed to transfer magnetization through heteronuclear coupled nuclei and then back to the more sensitive proton for detection. The most simple of these experiments is the HSQC NMR experiment.

Heteronuclear Single Quantum Coherence (HSQC)

The HSQC experiment is composed of two consecutive INEPT transfers. The first INEPT enables the proton magnetization of the amide proton to be transferred to the amide nitrogen and instead of detecting the ^{15}N chemical shift at the nitrogen frequency this magnetization is transferred back to the amide proton for detection since proton is the more sensitive nucleus. At the end of the second INEPT the ^{1}H chemical shift frequency of the amide protons are detected such that during detection, two chemical shift

frequencies are actually obtained: the ^{15}N chemical shifts of the amide nitrogens obtained during the t_1 period (after the first INEPT) and then the ^{1}H chemical shifts of the amide proton during t_2 (after the second INEPT). After Fourier transformation, the resulting 2D correlation map contains the amide nitrogen chemical shifts along the y-axis and the amide proton chemical shifts along the x-axis. One would expect a single ^{1}HN-^{15}N correlation for each of the backbone amides of the amino acids of the primary sequence, with the exception of prolines which do not contain an amide proton and those amino acids which contain an amide proton in the side chains (e.g. tryptophan, lysine, arginine, histidine, asparagine and glutamine). For simplicity, the NMR signals are observed from the top and represented as a contour map whereby the position of the NMR signal is characteristic of the chemical shifts and the number of contours represent the intensity of the NMR signal (see Figure 23).

V. Challenges in studying proteins by solution NMR spectroscopy

Proteins contain numerous protons which results in a wide NMR spectrum. The main caveat of studying proteins using solution NMR techniques is that with increasing protein molecular weight the escalating proton content results in more signals with increasing NMR resonance signal linewidths. The increasingly large number of protons promotes efficient transverse relaxation through dipole-dipole interactions such that NMR signal linewidths are broadened and hence extensive resonance overlap is observed.

In the past three decades NMR techniques used to study protein in solution have progressed significantly. Initially, protein NMR required the development of more powerful high field instruments with stable radio-frequency generation to increase NMR signal intensity, sensitivity and resolution (Wagner 1997; Kay 1997). Most notably, the advent of pulse field gradients replaced complicated and time-consuming phase cycling methods used to remove unwanted magnetization and suppress unwanted solvent signals, such as H_2O. The development of self-shielded pulsed field gradient probes resulted in improved spectral quality and high sensitivity without artifacts and solvent peaks (Kay and Gardner 1997). Improvements in hardware, techniques for incorporation of ^{13}C, ^{2}H and ^{15}N isotopes as well as intelligent NMR pulse sequences that enable magnetization transfer between through-bond and through-space coupled spins, all have extended the power of NMR spectroscopy for analyzing larger proteins (Kay 1997).

Dimensionality

With the development and application of 2D NMR experiments in the mid 1970s, correlations were permitted to be spread out into two ^{1}H frequency dimensions thereby providing increased spectral resolution. For example, a 2D homonuclear H-H correlation NOE experiment is used to identify backbone connectivities for sequential assignment of up to ~100 residues. Further improvements of 2D NMR methods led to complete structural determination of proteins up to ~100 residues, such as *E. coli* (Dyson *et al.* 1990) and human (Forman-Kay *et al.* 1991) thiroedoxin. However, beyond this limit, 2D methods are insufficient for complete structural determination due to the increased spectral overlap and chemical shift degeneracy. Spectral resolution diminished as the number of resonances increased sharply. Furthermore, magnetization transfer efficiency decreased through the small three-bond ^{1}H-^{1}H J couplings (3 to 12 Hz). Consequently, the NMR signal linewidths became larger than the couplings due to the increasing τ_c of the larger proteins.

For proteins of larger molecular weights (>10 kDa) the problem of spectral resonance overlap was alleviated by isotope incorporation (labeling ^{13}C and ^{15}N) in the protein of interest and by extending the dimensionality of the NMR spectrum through the use of heteronuclear techniques. The advantages of uniform ^{13}C and ^{15}N labeling are two-fold. Firstly, it provides a wide network of large heteronuclear coupling constants by which various correlations can be selectively and efficiently accomplished, and facilitate spreading the data in two, three and even four dimensions. Secondly, the large spectral windows of ^{13}C and ^{15}N offers increased resolution.

Heteronuclear coupling & isotope labeling

With the design and implementation of multidimensional (3D/4D) NMR experiments, the employment of one-bond heteronuclear couplings yield important additional correlation information in the form of ^{13}C and ^{15}N chemical shifts. The larger heteronuclear couplings also permitted extension of the time available for efficient through-bond magnetization transfers to obtain through-bond correlations. The heteronuclear couplings available in proteins are illustrated in Figure 12.

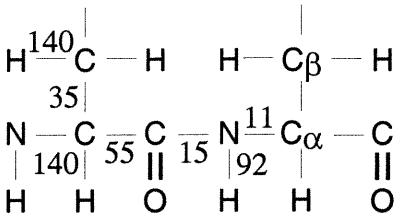


Figure 12. Heteronuclear couplings (in Hz) available in polypeptides.
(*Source: Evans 1995*)

Incorporation of isotopes can be accomplished using bacterial expression systems that employ media using ^{13}C-glucose and ^{15}N-ammonium chloride as the only source of ^{13}C and ^{15}N (Wüthrich 1986). Protein samples obtained in this manner are uniformly (>98%) ^{13}C and ^{15}N labeled. Instead of uniform ^{13}C and ^{15}N labeling proteins in its entirety, substantial simplification of NMR spectra may be attained by selectively labeling with specific ^{15}N-labeled amino acids types. Labeling schemes have also been developed that incorporate exogenous labeled amino acid within a mixture of non-labeled amino acids (Kainosho 1997). Although resonance assignment by selective amino acid

isotope-labeling may be time-consuming it may help to unravel ambiguities in the assignments of signals for large proteins.

Isotope labeling and the availability of heteronuclear coupling stimulated the development of ^{13}C-^{1}H-^{15}N triple resonance NMR experiments and pushed the molecular weight limit for NMR accessible proteins for structure determination from 10 to >20 kDa (Wagner 1997). The power of triple resonance NMR techniques comes from the many magnetization transfers possible that enable the recording of various types of correlations. The magnetization transfers are accomplished by a series of pulses and delays (refer to the discussion of the INEPT sequence). During the delays, any transverse magnetization will relax. For proteins of molecular weights less than 20 kDa, T_2 is long enough that relaxation does not cause a significant problem. But for larger proteins, (molecular weights >20 kDa) the T_2s are in the order of the delay and significant signal losses are obtained making the application of the techniques very difficult, if not impossible.

Although the larger heteronuclear couplings improved the time available for magnetization transfers to occur there is still some loss of signal intensity accompanied with each magnetization transfer. For polypeptides of 250 residues (25 kDa), magnetization transfers are efficient from ^{1}H to ^{13}C through $^{1}J_{cc}$ couplings and back to ^{1}Hs (Clore and Gronenborn 1997). However, for proteins >250 residues (>25 kDa), ^{13}C linewidths quickly approach the $^{1}J_{cc}$ value of ~30-35 Hz. Moreover, efficient dipole-dipole relaxation mechanisms results in broadening of the NMR linewidths. As a result, deuteration has been employed to promote efficient magnetization transfer and reduction of ^{13}C, ^{1}H and ^{15}N signal linewidths.

Deuteration

The application of triple resonance techniques to proteins of molecular weight larger than 25 kDa is limited by a faster transverse relaxation (short T_2) as discussed previously. Deuteration of the protein has been effectively used to decrease transverse relaxation, thereby extending the protein size limit to larger molecular weight ($\geq$ 60 kDa).

The substitution of non-exchangeable protons with deuterons, which possess γ of lower magnitude, significantly reduces the dipolar contribution to relaxation resulting in a slower rate of transverse relaxation (large T_2). For a ^{13}C spin attached to a single proton, replacement of the proton with deuterium diminishes the proton dipolar contribution for transverse relaxation by ~15-fold (Kay and Gardner 1997). Similarly, transverse relaxation of other nuclei other than ^{13}C also decrease upon deuteration (Kay and Gardner 1997). The small scalar coupling of ^{2}H to other nuclei can be readily eliminated by applying broadband deuterium decoupling. Triple-resonance techniques which require long transfer times for transverse ^{13}C magnetization to evolve (in the order of 20 msec) are therefore greatly improved by a combination of high levels of deuteration and ^{2}H decoupling. Deuteration also eliminates the need for resonance assignment of ^{13}C side chain carbon-bound protons. Although deuteration results in significant signal sensitivity gains the actual number of protons available for measuring NOE-derived connectivities for structural determination is limited (Kay 1997; Kainosho 1997).

NOE

The principal source of NMR information used to calculate protein structures is provided by distance constraints obtained from the measurement of through space

interproton nuclear Overhauser enhancement (NOE) interactions. The basis for the NOE is that neighboring protons, within 6 Å, are close enough that when they are perturbed from their equilibrium positions, their magnetization may exchange through a process termed cross-relaxation. The intensity of the NMR resonance results in either a positive or negative NOE. The rate constant for this process is directly related to the distance between the two protons (r^{-6}).

High molecular weight proteins (>25 kDa) require >98% deuteration in order to apply the various NMR techniques needed for complete sequential assignment of the polypeptide backbone atoms. Protons at labile amide sites such as the polypeptide backbone and some sidechain NH positions are re-introduced after protein production by exchange with water during purification. From these, NH-NH distance restraints can be measured but their number is insufficient to determine precise three dimensional structures. Therefore additional distance information is required to supplement NH-NH restraints. For this purpose a labeling strategy has been developed whereby protons of methyl group sidechains are re-introduced into a ^{13}C, ^{2}H, ^{15}N-labeled protein to increase the source of interproton NOE-derived measurements (Gardner and Kay 1997).

Methyl-protonation

Kay and coworkers (Gardner and Kay 1997; Goto *et al.* 1999) developed a strategy using *E. coli* grown in minimal ^{15}N, ^{13}C, ^{2}H media containing ^{13}C glucose as carbon source, ^{15}N ammonium chloride as nitrogen source and supplemented with [3-^{2}H]-U-$^{13}C_5$-α-ketoisovaleric acid and [3,3-$^{2}H_2$]-$^{13}C_4$-α-ketobutyric acid to introduce proton methyl groups of valine (Val), leucine (Leu) and isoleucine (Ile-δ) side chains. These

amino acids are among the most common in the hydrophobic cores of proteins and the packing of these residues is important for stabilizing molecular folds (Gardner and Kay 1997). In addition, protonated methyl groups of Val, Leu and Ile (δ) side chains do not display broad and low intensity NMR signals. The rapid rotation about the methyl 3-fold axis of symmetry decreases the effective τ_c leading to slower dipole-dipole relaxation. As a result, the ^{13}C and ^{1}H linewidths are narrowed due to slower transverse relaxation. NMR experiments have been developed such that the methyl proton and carbon resonances can be correlated to the amide proton for resonance assignment of the methyl groups (Gardner *et al.* 1996). The combination of CH_3-CH_3, CH_3-NH and NH-NH NOE-derived distances result in a more accurate and precise structural determination (Gardner and Kay 1997).

VI. Transverse Relaxation Optimized Spectroscopy

Besides deuteration, spectroscopic techniques have been developed to diminish the observed transverse relaxation of NMR signals in multidimensional experiments. Transverse relaxation optimized spectroscopy (TROSY) is one such method. It is based on the fact that within a system of two scalar coupled ½-spins, such as I and S, the cross correlation relaxation interference between CSA and DD interactions result in different transverse relaxation rates (T_2) of the four individual multiplet components (Guéron *et al.* 1983; Tjandra *et al.* 1996). In other words, the four components of the doublets in a 2D ^{1}H-^{15}N correlation experiment (see Figure 13b) have 4 different relaxation rates due to different contributions from the dipole-dipole (DD) and chemical shift anisotropy (CSA) interactions. The DD and CSA contributions are either additive or they oppose each other resulting, in the latter case, in the narrowing of the resonance as shown in Figure 13b lower right. Because the CSA contributions increase with frequency, hence through magnetic field, it has been predicted that this reduction of transverse relaxation would reach a maximum at ^{1}H-frequencies in the 900-1000 MHz (Pervushin *et al.* 1997). It is also important to note that CSA contributions are only present in amide and aromatic moieties in a protein. Therefore, the trosy principle can only be applicable to the detection of backbone amide since all aromatic side chains are deuterated. In a conventional [^{15}N,^{1}H] correlation experiment, the nuclei are decoupled during the time periods t_1 and t_2, such that a single correlation peak per ^{15}N-^{1}H moiety is detected (Figure 13a). If the same experiment is recorded without decoupling, four cross-peaks of differing linewidths and intensity are observed per ^{15}N-^{1}H moiety due to the different

CSA and DD interactions (refer to Figure 13b). The TROSY pulse sequence selectively detects the one component that has the slowest transverse relaxation rate for both nuclei in the correlation spectrum (Pervushin *et al.* 1997; Pervushin *et al.* 1998) (Figures 13c and 14). All experiments recorded for backbone resonance assignment were using the trosy principle during the detection of amide nitrogens and protons.

.

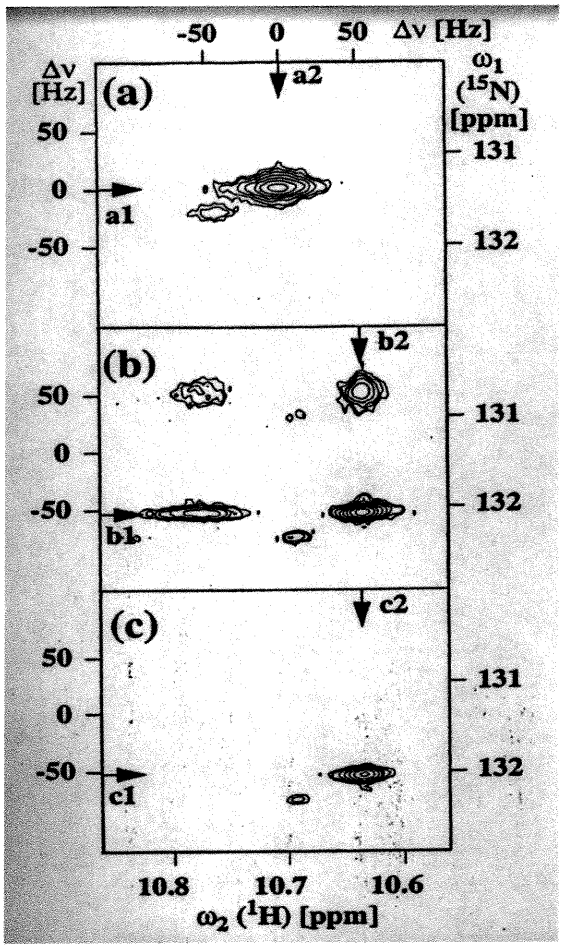


Figure 13. ^{1}HN-^{15}N contour plots of a) conventional broadband decoupled, b) conventional without decoupling and c) TROSY-type ^{1}HN-^{15}N correlation spectra.

(*Source: Pervushin et al. 1997*)

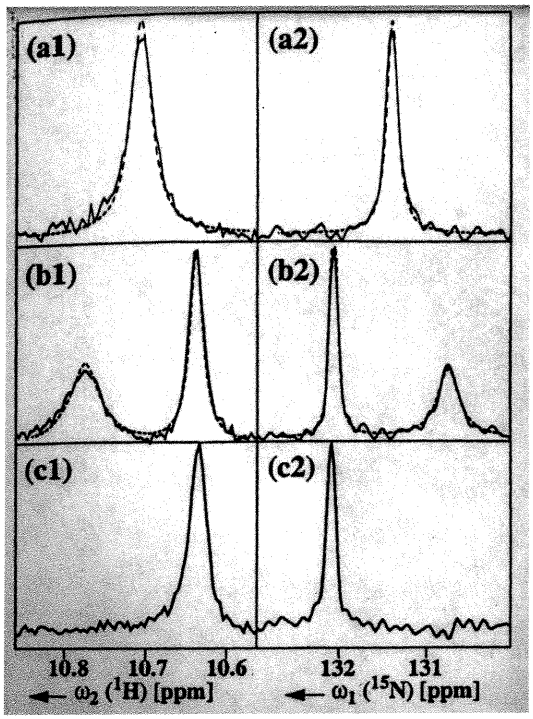


Figure 14. Corresponding ^{15}N and ^{1}H lineshapes of contour plots indicated by black arrows in Fig. 13.

(Source: Pervushin et al. 1997)

Chapter 3

Sample Requirements for Protein Solution NMR

I. Introduction

The major hurdle in applying NMR spectroscopy techniques to the study of proteins in solution is obtaining samples that are well behaved. The protein must be soluble and monomeric at concentrations of at least 1 mM and stable at temperatures of 25-37 °C over long periods of time because data acquisition can take up to 10 days for a single NMR experiment.

To apply triple resonance techniques, the protein of interest must be uniformly enriched with ^{13}C and ^{15}N isotopes. For large proteins, such as TCPTP, deuterium labeling is also required to minimize rapid loss of NMR signals due to dipole-dipole interactions. Bacterial protein expression using minimal media enriched with ^{13}C, ^{2}H and ^{15}N isotopes is an expensive process that decelerates the rate of cell growth. Bacterial expression systems should be efficient: capable of simultaneously withstanding the long, stringent growth conditions and expressing high levels of the isotope-labeled protein.

To obtain NMR concentration of 1-2 mM the protein must be purified and concentrated to high purity. For ^{13}C, ^{2}H and ^{15}N labeled proteins, it is only during purification that deuteriums at labile amide positions can exchange back to protons and permit NMR signal detection through the more sensitive ^{1}H nuclei.

Foremost, a preliminary NMR experiment of the catalytic domain of TCPTP was carried out to assess the feasibility of structural studies. A sample of ^{15}N-labeled TCPTP C216S (1-281) was produced using the pFLAG2 expression system and a 2D ^{1}HN-^{15}N HSQC experiment was collected. A well resolved, homogeneous spectra was obtained which is characteristic of a stable and well-folded protein. These results established the

feasibility of determining the structure of TCPTP by NMR. In contrast, the first 2D ^{1}HN-^{15}N HSQC of the C215S mutant of the catalytic domain of PTP-1B resulted in a non-homogeneous spectra. Unlike TCPTP, protein engineering of PTP-1B C215S (1-321) was required before a well-resolved, homogenous HSQC spectra could even be obtained. Fortunately for TCPTP the next step was to optimize the expression and purification of TCPTP C216S (1-281). Although TCPTP C216S (1-281)-pFLAG2 provided sufficient amounts of protein expression and efficient incorporation of ^{15}N-labeled ammonium chloride, growth of the pFLAG2 vector in the presence of 100% D_2O is inefficient. As a result, the TCPTP C216S (1-281) insert was subcloned into various pET expression systems which are capable of growing more efficiently in the presence of 100% D_2O. These pET translation vectors permitted highly efficient expression, easy subcloning and allowed a variety of poly-His peptide tags to be used for easy detection and purification of the target protein. The TCPTP C216S (1-281) gene was also subcloned into pET translation vectors which did not contain any peptide tags as a precautionary step since the poly-His peptide tag may affect the stability of the protein.

II. Experimental

The pET vectors were selected for the production of the target protein because they offered the potential to achieve the isotope labeling required for structural studies of TCPTP, high levels of soluble protein, ease of purification, and most importantly, cell growth in 100% D_2O with incorporation of ^{13}C and ^{15}N isotopes. The selected pET expression systems are under the control of T7 RNA polymerase. T7 RNA polymerase is highly selective and active that almost all of the cell's resources are converted to target gene expression (Novagen, pET System Manual). Only *E. coli* BL21(DE3) (Stratagene) hosts carry the lacUV5 promoter that is inducible with isopropyl-β-D-thiogalactopyranoside (IPTG) to promote transcription of T7 RNA polymerase. The target protein was cloned into five different pET vectors to evaluate and optimize both expression and purification of the protein.

Genetic Engineering of TCPTP C216S (1-281)

The five pET vectors contained either a poly-His peptide tag at the N- or C-terminus or no tag at all (Novagen, pET System Manual). The poly-His peptide tag is composed of six or ten consecutive histidine residues for fast and versatile affinity purification using a Ni-NTA affinity column. The 10xHis tag provided superior binding and improved purification when impurities were significant, such as in crude cell lysate.

Two different antibiotic selective vectors were selected (Novagen, pET System Manual). The ampicillin selective pET19b vector contained an N-terminal 10xHis tag whereas the pET20b vector contained no peptide tag. Because preliminary bacterial

expression studies demonstrated unregulated protein expression with the pET20b vector, an alternative ampicillin selective vector, pET11a, which contained no peptide tag was also selected. The gene was subcloned into vector pET11a since the polypeptide tags of the other selected vectors may affect the stability of the protein. The FLAG and non-FLAG versions of the target protein were cloned into this vector for comparison purposes. The FLAG octapeptide consisted of Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys and was incorporated since this epitope may improve protein expression. As an alternative, vectors selective for kanamycin were also employed such as pET28b which contained a N-terminal 6xHis tag and pET29b which contained no peptide tag. All pET vectors were purchased from Novagen.

E. coli DH5α harboring the vector pFLAG2 with a TCPTP C216S (1-281) insert was used as the template DNA for the cloning of TCPTP C216S (1-281)-pET constructs. The DNA encoding the TCPTP C216S (1-281) insert was provided by Dr. Ernest Asante-Appiah at Merck Frosst Canada & Co. at Pointe-Claire, Quebec. TCPTP C216S (1-281) possessed an internal *Nde*I site however for proper subcloning of the insert into pET expression systems a *Nde*I site had to be introduced at the 5'-end of TCPTP. As a result the internal *Nde*I site had to be removed from TCPTP by introducing a silent mutation. A silent mutation permits a nucleotide to be mutated such that the actual DNA base is changed however the codon for the specific amino acid is conserved enabling expression of the specific amino acid. At nucleotide position 222 the base thymine was mutated to a cytosine nucleotide in order to alter the unique *Nde*I site while maintaining a codon to express the amino acid tyrosine. Depending on the vector, a TAA stop codon with either a *Sal*I or *Bam*HI restriction site was introduced at the C-terminal. Both the silent

mutation and specific restriction sites were introduced into the TCPTP C216S (1-281)-pET constructs using the general polymerase chain reaction (PCR) protocol as illustrated in Figure 15 (Sambrook *et al.* 1989). The sequence and function of the primers employed for PCR are indicated in Table 1 and Table 2.

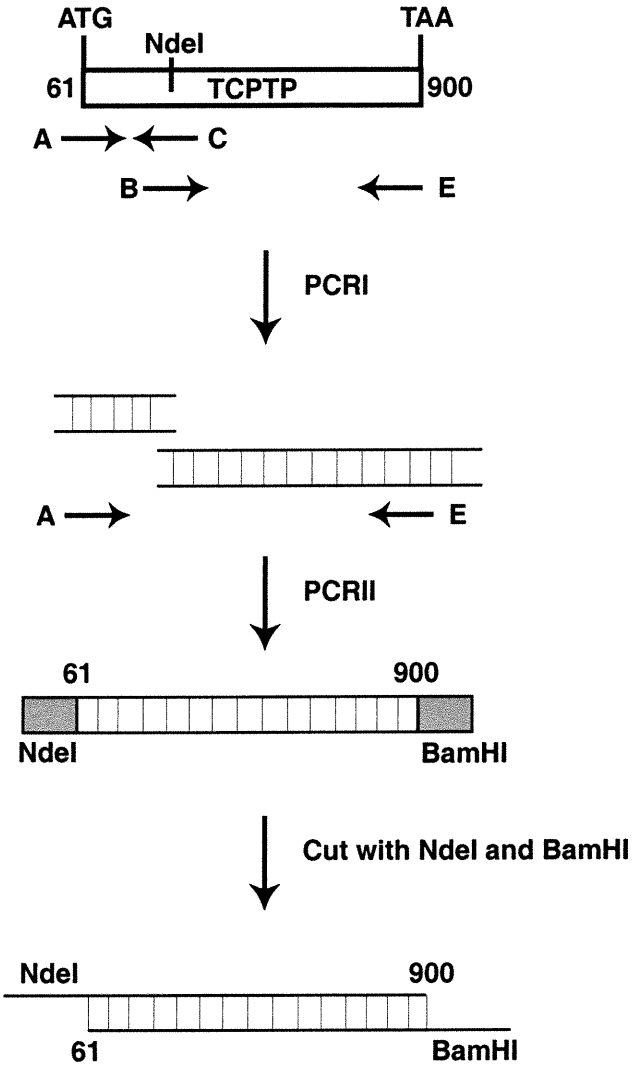


Figure 15. General PCR protocol for the amplification of various TCPTP C216S (1-281) inserts.

Table 1. Primers for PCR of various TCPTP C216S (1-281) inserts.

Primer	Primer sequence
A	5'-CAAGCTTGAATTCCATATGCCCACCACCATCG-3' (sense)
B	5'-ATGTAAGCCCATACGATCACAGTCGTG-3' (sense)
C	5'-CACGACTGTGATCGTATGGGCTTACAT-3' (antisense)
D	5'-ATACTAGAGTCGACTTACTTTATACATTTT-3' (antisense)
E	5'-CGCGGATCCTTACTTTATACATTTTGCTCC-3' (antisense)
F	5'-AGGAGATATCATATGGACTACAAGGACGACGAT-3' (sense)

Table 2. Identity and function of primers for PCR of various TCPTP C216S (1-281) inserts.

Primer	Sequence identity	Function
A	Identical to the starting coding strand of pFLAG2 and nucleotides 61-76 of TCPTP	Introduce *Nde*I site at ATG initiator
B	Identical to the coding strand at nucleotides 208-235 with silent mutation at nucleotide 222 of TCPTP	Remove internal *Nde*I site with silent mutation
C	Complementary to the coding strand at nucleotides 208-235 with silent mutation at nucleotide 222 of TCPTP	Remove internal *Nde*I site with silent mutation
D	Complementary to the coding strand at nucleotides 888-900	Introduce TAA stop codon and *Sal*I site
E	Complementary to the coding strand at nucleotides 883-900 of TCPTP	Introduce TAA stop codon and *Bam*HI site
F	Identical to the starting coding strand for FLAG peptide of pFLAG-2 and nucleotides 100-132 of TCPTP	Introduce *Nde*I site and FLAG peptide at ATG initiator

During the first PCR cycle the internal *Nde*I site of TCPTP C216S (1-281) was removed by introducing a silent mutation at nucleotide 222 of TCPTP. The restriction

sites at the N- and C-termini were introduced using primer A with primer C and primer B with primer E. Two PCR fragments were generated and detected on a 1% (w/v) agarose (FMC BioProducts) gel made up in Tris-borate (TBE) buffer (refer to Figure 16).

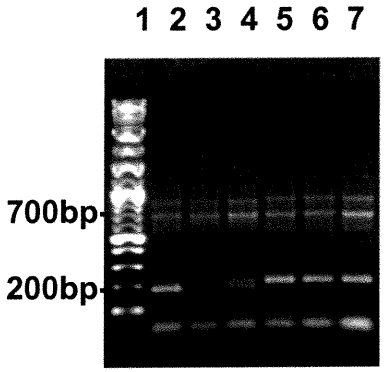
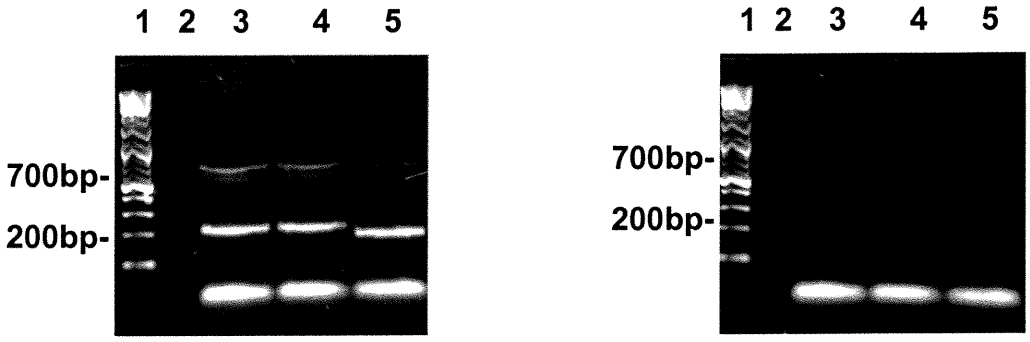


Figure 16. 1% agarose gel of PCR fragments generated to construct TCPTP C216S (1-281) and FLAG-TCPTP C216S (1-281).

Lane 1: 100 bp and 1 kb DNA ladder; lanes 2-4: 190 bp and 704 bp PCR fragments to construct TCPTP C216S (1-281) with varying $MgSO_4$ concentrations (0, 1, 2 mM); and lanes 5-7: 223 bp and 704 bp fragments to construct FLAG-TCPTP C216S (1-281) with varying $MgSO_4$ concentrations (0, 1, 2 mM).

The two products from the first PCR cycle were then analyzed by electrophoresis on a 1% (w/v) low-melt agarose gel (Mandel) made up in TBE buffer with 0.25 μg/mL ethidium bromide (Gibco-BRL, Life Technologies). The expected band sizes were viewed over a UV lamp and evaluated using the DNA 100 bp (100 bp-1.5 kb) and 1 kb ladder (0.5-10 kb) (BioLabs). The correct PCR product lengths were purified directly from the low-melt agarose gel using the protocol supplied with the QIAquick Gel Extraction kit (Qiagene Inc.) (refer to Figure 17).



(a) (b)

Figure 17. Gel extraction of PCR fragments generated to construct TCPTP C216S (1-281) and FLAG-TCPTP C216S (1-281) from 1% low-melt agarose gel (a) before and (b) after gel extraction.

Lane 1: 100 bp and 1 kb DNA ladder; lane 2: empty; lanes 3-4: 223 bp and 704 bp fragments from PCR reaction in Lane #7 of Fig.16; and lane 5: 190 bp and 704 bp fragments from PCR reaction in Lane #2 of Fig.16.

The second PCR cycle was performed to amplify the entire *Nde*I-TCPTP C216S (1-281)-*Bam*HI fragment without the internal *Nde*I site using only primer A and primer E to ligate the two PCR fragments obtained from the first cycle (refer to Figure 18).

Similarly, to amplify the full-length *Nde*I-TCPTP C216S (1-281)-*Sal*I insert by PCR, primers A, B, C and D were used for the first cycling and only primers A and D for the second cycling. To obtain the full-length *Nde*I-FLAG-TCPTP C216S (1-281)-*Bam*HI insert by PCR, primers F, B, C and E were used for the first cycling and only primers F and E were used for the second cycling (refer to Figures 18-20).

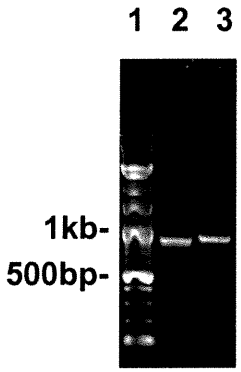


Figure 18. 1% agarose gel of final full length PCR amplified insert of TCPTP C216S (1-281) and FLAG-TCPTP C216S (1-281).

Lane 1: 100 bp and 1 kb DNA ladder; lane 2: expected 866 bp band of full length TCPTP C216S (1-281); and lane 3: expected 901 bp band of full length FLAG-TCPTP C216S (1-281).

The two rounds of PCR each consisted of 25 cycles of denaturation at 95 °C for 45 sec, annealing at 56 °C for 30 sec and extension at 72 °C for 60 sec with only the first cycle employing a longer denaturation step (5 min) and the last cycle employing a longer extension step (5 min). The PCR reactions were carried out on a Biometra T3 Thermocycler and consisted of ~500 ng template DNA, 2 pmol of primer, 0.2 mM dNTPs, 0-2 mM $MgSO_4$ and Deep Vent polymerase (BioLabs) in a total reaction volume of 100 μL.

The final full-length PCR products were then cleaved with the appropriate restriction enzymes in appropriate buffer (as suggested by the supplier): either *Nde*I and *Sal*I or *Nde*I and *Bam*HI, at 37 °C for 1 h. The enzymes were then deactivated at 65 °C for 20 min. The digested TCPTP C216S (1-281) insert was purified using the protocol supplied with the QIAquick PCR purification kit (Qiagene Inc.). Similarly, the pET vectors were digested with the appropriate restriction enzymes: pET11a and pET19b using *Nde*I and *Bam*HI and pET20b, pET28b and pET29b using *Nde*I and *Sal*I. The

digested pET vectors were dephosphorylated with alkaline phosphatase (Boehringer Mannheim) at 37 °C for 1 h to remove 5'-phosphates from the digested ends of the DNA and prevent self-ligation.

The resulting *Nde*I-TCPTP C216S (1-281)-*Sal*I/*Bam*HI fragments were then ligated into the *Nde*I-*Sal*I digested pET20b, pET28b, pET29b vectors and *Nde*I-*Bam*HI digested pET11a and pET19b vectors respectively. Similarly, the *Nde*I-FLAG-TCPTP (1-281)-*Bam*HI fragment was ligated into the pET11a vector digested with *Nde*I and *Bam*HI. All 20 μL ligation reactions employed a 10:1 molar ratio of insert:vector, 1 μL T4 DNA ligase (Boehringer Mannheim) and 1 mM ATP. Half of the ligation reactions were used for transformation into E. coli subcloning efficiency DH5α cells (Gibco-BRL, Life Technologies) or maximum efficiency DH5α cells (Gibco-BRL, Life Technologies) for maximal incorporation of the ligated plasmid. The cells were plated onto solid Luria Broth (LB) agar plates containing the appropriate antibiotic (100 μg/mL of ampicillin for pET11a pET19b, pET20b and 100 μg/mL of kanamycin for pET28b and pET29b) and incubated overnight at 37 °C. Plasmids were isolated and purified using the protocol supplied with the QIAprep Miniprep kit (Qiagene Inc.). Positive inserts were screened by digestion with either *Nde*I/*Sal*I or *Nde*I/*Bam*HI and detected on a 1% (w/v) agarose (FMC BioProducts) gel made up in TBE buffer (refer to Figure 19). All restriction enzymes were purchased from Boehringer Mannheim and antibiotics were purchased from Sigma-Aldrich.

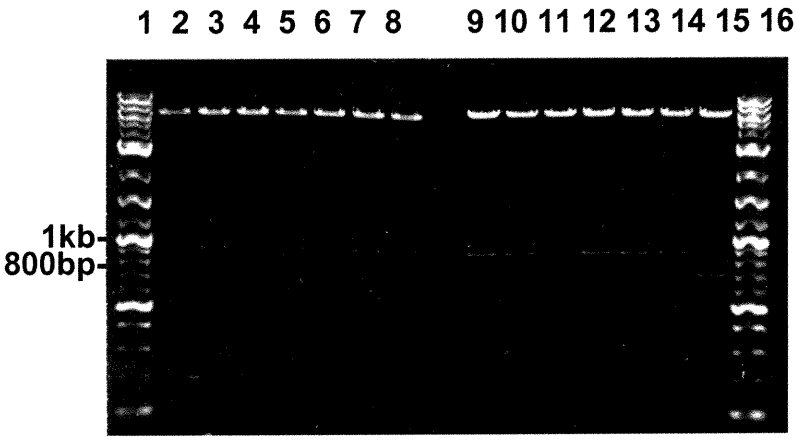


Figure 19. Screening for positive inserts using 1% agarose gel of digested TCPTP C216S (1-281) clones.

Lanes 1 & 16: 100 bp and 1 kb DNA ladder, lanes 2-8: FLAG-TCPTP C216S (1-281)-pET11a clones digested with *Nde*I and *Bam*HI – expect 882 bp band; and lanes 9-15: TCPTP C216S (1-281)-pET19b clones digested with *Nde*I and *Bam*HI – expect 846 bp band. Lanes 2-3, 5, 7-8 contain positive inserts of FLAG-TCPTP C216S (1-281), Lanes 9-10, 12-14 contain positive inserts of TCPTP C216S (1-281).

Sequences of the positive inserts were confirmed by DNA sequencing using the ABI Prism BigDye Terminator Ready Reaction Mix (PE Applied Biosystems). Only ~300-500 ng of the DNA isolated from the Qiagene Minipreps was required for the 20 μL PCR sequencing reactions containing 8 μL of the BigDye Terminator Ready Reaction Mix and 1 μL of the appropriate PCR sequencing primers (5 pmol/μL). The BigDye Terminator Ready Reaction Mix contains AmpliTaq® DNA polymerase, BigDye terminators, dideoxynucleoside triphosphates (dNTPs), magnesium chloride and buffer for the fluorescence-based cycle sequencing reactions. The DNA polymerase was a variant of *Thermus aquaticus* DNA polymerase and contained a point mutation in the active site which resulted in less discrimination against dideoxynucleotides and a second mutation in the amino terminal domain that virtually eliminated the 5'→3' nuclease

activity of AmpliTaq DNA polymerase. The BigDye terminators were a set of dye terminators labeled with novel, high-sensitivity dyes. The A, C, G, T-dye structures contained a fluorescein donor dye each linked to a unique dichlororhodamine acceptor dye. The excitation maximum of each dye label was that of the fluorescein donor and the emission spectrum was that of the dRhodamine acceptor. The fluorescent dyes with the dNTPs were incorporated during the PCR sequencing cycle and emitted a signal at the appropriate wavelength using the excitation energy of the argon laser in the PE Applied Biosystems DNA sequencing instrument during gel electrophoresis. The primers used for the PCR sequencing reactions are tabulated in Table 3. All primers used for DNA sequencing were also purchased from Research Genetics.

Table 3. Primers for PCR sequencing of various TCPTP C216S (1-281) inserts.

Primer	Primer sequence	Identity
T7P	5'-TAATACGACTCACTATAGGGG-3' (sense)	Identical to the T7 promoter coding region of the pET vector
T7T	5'-GCTAGTTATTGCTCAGCGGTG-3' (antisense)	Complementary to the T7 terminator coding region of the pET vector
IM6-1	5'-CACATGCTGCCATTTCTG-3' (sense)	Identical to the coding strand at nucleotides 335-353 of TCPTP
C216SR	5'-CAATGCCTGCACTAGAGTGGAT CACCGC-3' (antisense)	Complementary to nucleotides 694-722 of TCPTP

The PCR sequencing reactions were purified following the protocol supplied with the gel filtration Centri-Sep columns (Princeton Separations). These gel filtration columns have been specifically designed to purify fluorescent reaction mixtures and

provide excellent recovery of DNA fragments >16 base pairs while removing >98% of salts, dNTPs and other low molecular weight compounds. Once the PCR fragments were eluted from the Centri-Sep columns the samples were evaporated under reduced pressure for 1 h. To load the PCR samples onto the sequencing gel, they were resuspended with 4 μL of sample buffer consisting of 200 μL deionized formamide (83%) and 40 μL of 10 mg/mL blue dextran in 50 mM EDTA pH 8.0 (1.7 mg/mL), boiled at 95 °C for 2 min, then left on ice before loading. The samples were loaded onto a denaturing DNA polyacrylamide sequencing gel (10.5 mL of 40% acrylamide, 40 g urea, 32 mL deionized water, 8 mL 10X TBE, 400 μL 10% ammonium persulfate and 40 μL TEMED) with TBE as the running buffer. The gel was allowed to run for 14 h and fluorescent dye emission spectra was detected and collected using the ABI Prism 377 DNA Sequencer program.

DNA preparations of the positive clones were then obtained using the protocol supplied with the Qiagene Maxiprep kit (Qiagene Inc.). All reagents used for DNA sequencing were purchased from Sigma-Aldrich and TEMED was purchased from Bio-Rad Laboratories.

Expression System Selection of TCPTP C216S (1-281)

The most optimal expression system was chosen by assessing the bacterial cell growth rate of the various TCPTP-pET constructs in LB. The identity and amount of the protein expressed during bacterial growth was assessed using Western blot analysis. Two TCPTP C216S (1-281)-pET constructs were chosen based on their high growth rates and

high level of protein expression. The final expression system was chosen based on the efficiency by which pure and soluble protein could be purified.

Growth study of TCPTP C216S (1-281)

One colony from a fresh transformation of BL21(DE3) *E. coli* competent cells harboring the TCPTP C216S (1-281)-pET11a (pET19b, pET20b, pET28b, pET29b) or FLAG-TCPTP C216S (1-281)-pET11a plasmid was inoculated into 3 mL of LB containing the appropriate antibiotic (100 μg/mL ampicillin or kanamycin) and grown overnight at 37 °C with shaking for aeration. For comparison with the pFLAG2 expression vector, one colony from a fresh transformation of BL21 *E. coli* competent cells harboring the human TCPTP C216S (1-281)-pFLAG2 plasmid was inoculated into 3 mL of LB containing ampicillin (100 μg/mL). The overnight cultures were diluted 1:250 (100 μL in 25 mL) in LB containing 100 μg/mL of the appropriate antibiotic and the growth was monitored at OD_{600} at various timed intervals (refer to Figure 20). At an OD_{600} of 0.6-0.7, a 1 mL pre-induction sample of the culture was retained. The remaining culture was induced with a final working solution of 1 mM IPTG (Boehringer Mannheim).

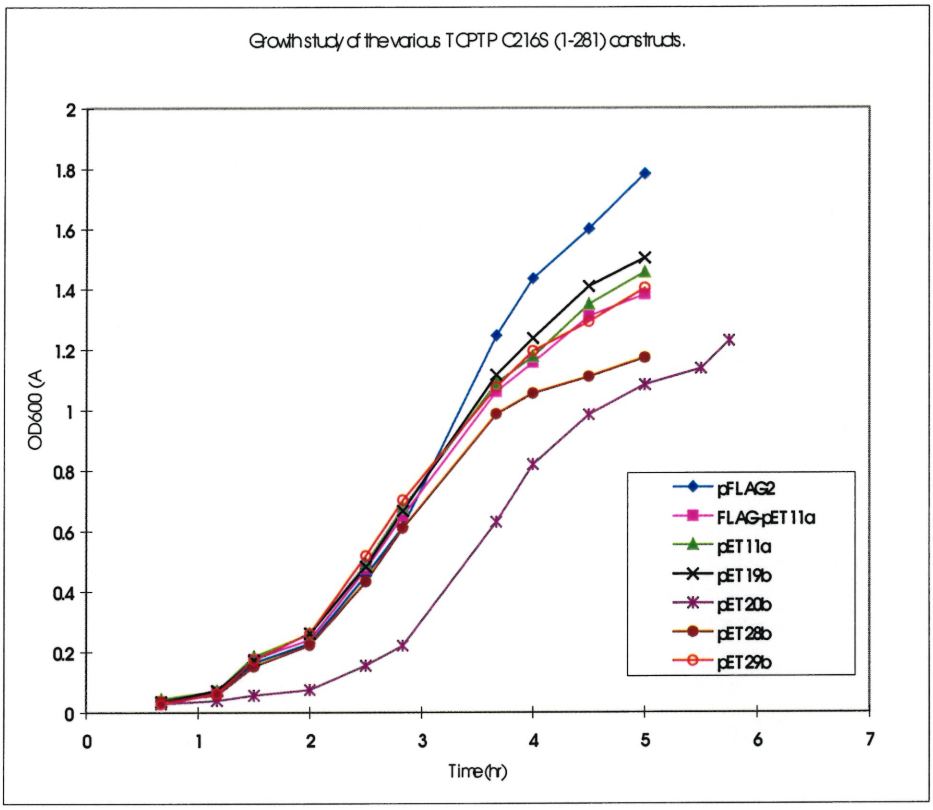
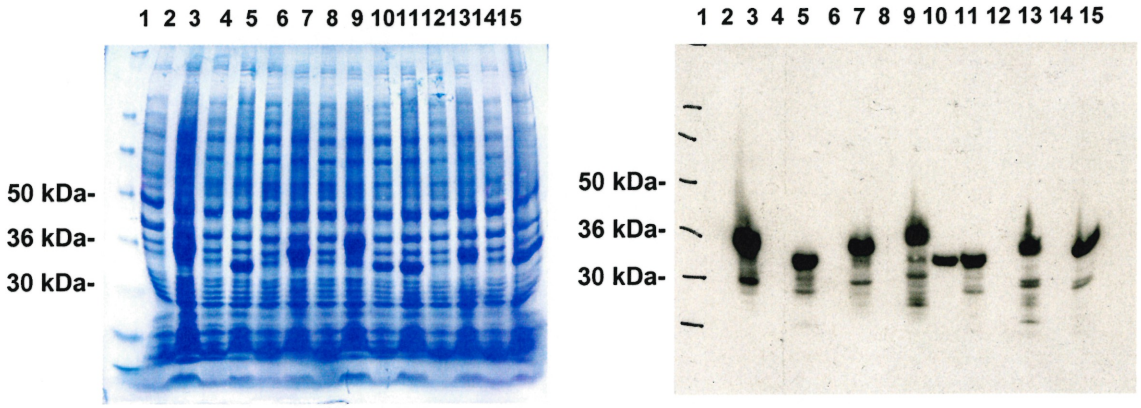


Figure 20. Growth curve study of various TCPTP C216S (1-281)-pET expression systems.

After an induction time of 2 h, 1 mL of the induced cell culture and the 1 mL pre-induction sample were centrifuged separately at 14 000 rpm for 2 min. using an Eppendorf Centrifuge 5415C for 2 min. The supernatant was removed and the pellet resuspended in 100 μL of LB. Each of the pre- and post-induction samples (20 μL) were mixed with 20 μL of 2X reducing loading buffer (2 mL glycerol, 6 mL 10% SDS, 1.5 mL of 1 M Tris-HCl pH 8.0, 1 mL β-mercaptoethanol, 850 μL H_2O, ~1.5 mL >1% bromonphenol blue) and loaded, along with SeeBlue molecular weight markers (Novex), onto two 10-20% SDS polyacrylamide gels with Tris-glycine SDS as the running buffer. The protein bands of one of the gels were detected by Coomassie Blue staining and the

identity of the bands were further confirmed by Western blot analysis of the remaining gel (refer to Figure 21). Similarly, the post-induction samples were diluted accordingly to obtain comparable final OD_{600} readings. These samples were also loaded with 2X reducing loading buffer and SeeBlue molecular weight markers onto two 10-20% SDS polyacrylamide gels to estimate the relative amount of expressed protein (refer to Figure 22). The gels were run at 250 V for 30 min. For Western blot analysis, gels were transferred overnight onto a nitrocellulose membrane (Novex) at 10 V using a transfer buffer (25 mM Tris base, 192 mM glycine, 0.1% SDS, 20% methanol).

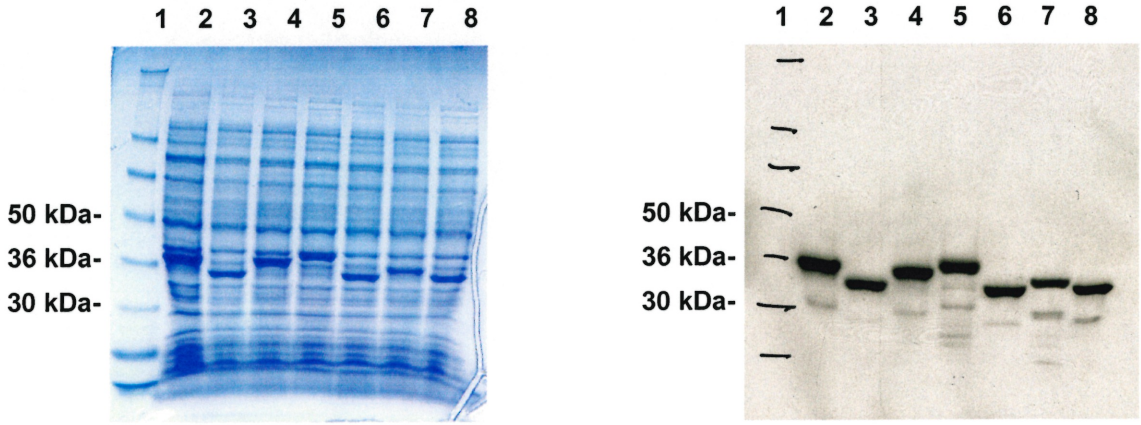
The nitrocellulose membrane was blocked with 5% non-fat dry milk in 0.1% T-TBS (0.1% Tween-20, 20 mM Tris-HCl, 0.5 M NaCl pH 7.4) to prevent non-specific binding. The membrane was then incubated with the primary antibody TC3ED in 1-2% non-fat dry milk and 0.1% T-TBS. The antibody TC3ED recognizes the N-terminal of TCPTP. This primary antibody solution was provided by Wanda Cromlish at Merck Frosst Canada & Co. at Pointe-Claire, Quebec. After three 10 min washes with 0.1% T-TBS, the membrane was incubated with a 1:10 000 dilution of the secondary anti-mouse IgG horseradish peroxidase-linked antibody (Amersham Pharmacia) in 5% non-fat dry milk and 0.1% T-TBS. The membrane was then washed again with three 10 min washes in 0.1% T-TBS before the immunoreactive proteins were visualized by chemiluminescence with the ECL™ Western blotting detection system (Amersham Pharmacia) on Hyperfilm (Amersham Pharmacia) (refer to Figures 21 and 22).



(a) (b)

Figure 21. (a) Coomassie-blue stained gel and (b) western blot analysis of pre- and post-induction samples of various TCPTP C216S (1-281) constructs.

Lane 1: SeeBlue markers; pre- and post induction of TCPTP C216S (1-281) constructs in; lanes 2-3: pFLAG2; lanes 4-5: pET11a; lanes 6-7: pET11a (with N-terminal FLAG); lanes 8-9: pET19b (N-terminal 10xHis); lanes 10-11: pET20b; lanes 12-13: pET28b (N-terminal 6xHis); and lanes 14-15: pET29b respectively.



(a) (b)

Figure 22. (a) Coomassie-blue stained gel and (b) western blot analysis of comparable (normalized $O.D._{600}$) post-induction samples of various TCPTP C216S (1-281) constructs.

Lane 1: SeeBlue markers; comparable post induction of TCPTP C216S (1-281) constructs in; lane 2: pFLAG2; lane 3: pET11a; lanes 4: pET11a (with N-terminal FLAG); lane 5: pET19b (N-terminal 10xHis); lane 6: pET20b; lane 7: pET28b (N-terminal 6xHis); and lane 8: pET29b.

From the growth study curves of the various TCPTP C216S (1-281)-pET expression vectors, TCPTP C216S (1-281)-pET19b demonstrated the growth curve most comparable to that of TCPTP C216S (1-281)-pFLAG2 (refer to Figure 20). The growth curves of the remaining constructs were more or less comparable with the exception of TCPTP C216S (1-281)-pET20b/pET28b which demonstrated a slower growth curve. The significantly slower growth of TCPTP C216S (1-281)-pET20b may be explained by the unregulated expression of the protein due to lack of the lacI repressor protein. Protein expression of the target gene is under the control of T7 RNA polymerase which is only inducible with IPTG. LacI provides additional control of protein expression by acting on both the lac UV5 promoter of the host to repress transcription of T7 RNA polymerase and on the T7 lac promoter in the vector to block transcription of the target gene by any T7 RNA polymerase that may be produced before induction with IPTG.

Positive identification of TCPTP C216S (1-281) protein expression was confirmed using Western blot analysis of the post-induction samples. The human TCPTP C216S (1-281)-pFLAG2 protein and FLAG-TCPTP C216S (1-281)-pET11a were comparable and gave a ~35 kDa band which migrated just above the 36 kDa SeeBlue molecular weight marker. TCPTP C216S (1-281)-pET11a, pET20b and pET29b contained no His-tag and migrated to give an expected ~33.6 kDa band observed just below the 36 kDa SeeBlue molecular weight marker. The TCPTP C216S (1-281)-pET28b protein band migrated slightly above that of pET11a, pET20b and pET29b band, because it contained an N-terminal 6xHis tag. The TCPTP C216S (1-281)-pET19b migrated the slowest because it contained an N-terminal 10xHis tag and was observed above the 36 kDa SeeBlue molecular weight marker. Detection of lower molecular

weight bands on the Western blot may be due to proteolytic cleavage of the TCPTP C216S (1-281) protein resulting in degradation products of the protein. Western blot analysis of the pre-induction samples also demonstrated protein expression by the pET-20b expression system prior to induction with IPTG. In other words, pET-20b did not provide regulated expression of the protein. In terms of band size and intensity (refer to Figure 22), vectors pET-28b and pET-29b resulted in bands that were slightly less intense than the other expression vectors. Only FLAG-TCPTP C216S (1-281)-pET11a and TCPTP C216S (1-281)-pET19b resulted in band sizes and intensity comparable to that of TCPTP C216S (1-281)-pFLAG2. As a result FLAG TCPTP C216S (1-281)-pET11a and TCPTP C216S (1-281)-pET19b were considered as the potential expression systems.

Selection of TCPTP C216S (1-281) Expression System

High levels of protein expression were attained with both FLAG-TCPTP C216S (1-281)-pET11a and TCPTP C216S (1-281)-pET19b constructs. TCPTP C216S (1-281)-pET19b contained a 10xHis tag at the N-terminal which provided a characteristic affinity tag that facilitated purification of the target protein. The final TCPTP construct for NMR studies was chosen after a comparison of the efficiency by which the protein could be purified to high quantity and purity. The two proteins, 2 L cultures of each TCPTP C216S (1-281)-pET19b and FLAG-TCPTP C216S (1-281)-pET11a, were grown using ^{15}N-ammonium chloride as the nitrogen source. Refer below for details on growing ^{15}N-TCPTP C216S (1-281).

The growth rates of the two constructs were similar; however, their purification protocols differed. For 10xHis-TCPTP C216S (1-281)-pET19b the protein was purified

within two chromatography steps using a ~5 mL Ni-NTA affinity column followed by a 5 mL HiTrap Blue™ (Amersham Pharmacia) affinity column. For FLAG-TCPTP C216S (1-281)-pET11a the protein was first purified using a 5 mL HiTrap Blue™ affinity column followed by a 5 mL HiTrap Q (Amersham Pharmacia) anionic exchange column. Detection and visualization of the protein proceeded by electrophoresis on 10-15% SDS Phastgels (Amersham Pharmacia) using Mark12 molecular weight markers (Novex) and staining with Coomassie Blue respectively. The purification protocol for 10xHis-TCPTP C216S (1-281) resulted in more pure protein fractions compared to that of FLAG-TCPTP C216S (1-281). The results and the details of the purification protocols employed for both 10xHis-TCPTP C216S (1-281)-pET19b and FLAG-TCPTP C216S (1-281)-pET11a are found in Appendix C.

To obtain NMR sample conditions for the two respective proteins each of the two purified proteins was first concentrated to ~1 mL by ultrafiltration, at 2000 x g (4°C), using a 5000 MWCO centrifugal filtration device (Millipore). Using the same filtration device the concentrated protein samples were then exchanged with three ~15 mL volumes of 100 mM NaH_2PO_4 pH 7.5 buffer with 5 mM DTT-d_{10} by ultrafiltration. The filtrate after each buffer exchange was evaluated using the Bio-Rad protein assay (Bio Rad Laboratories) to ensure no protein was lost during the concentration and buffer exchange steps. During the last buffer exchange the protein concentration was determined using the calculated extinction coefficient 43 990 $M^{-1}cm^{-1}$. The sample was concentrated to ~0.5 mL to obtain a ~0.4 mM sample. The 2D ^{1}HN-^{15}N HSQC NMR experiments collected on the His-tagged and FLAG-tagged proteins showed no significant differences (Figure 23) which suggested that neither affinity tag interacted

with the catalytic domain. As a result, the His-tagged version of TCPTP C216S (1-281) was selected due to the more efficient purification protocol employing the Ni-NTA affinity column and HiTrap Blue™ affinity column.

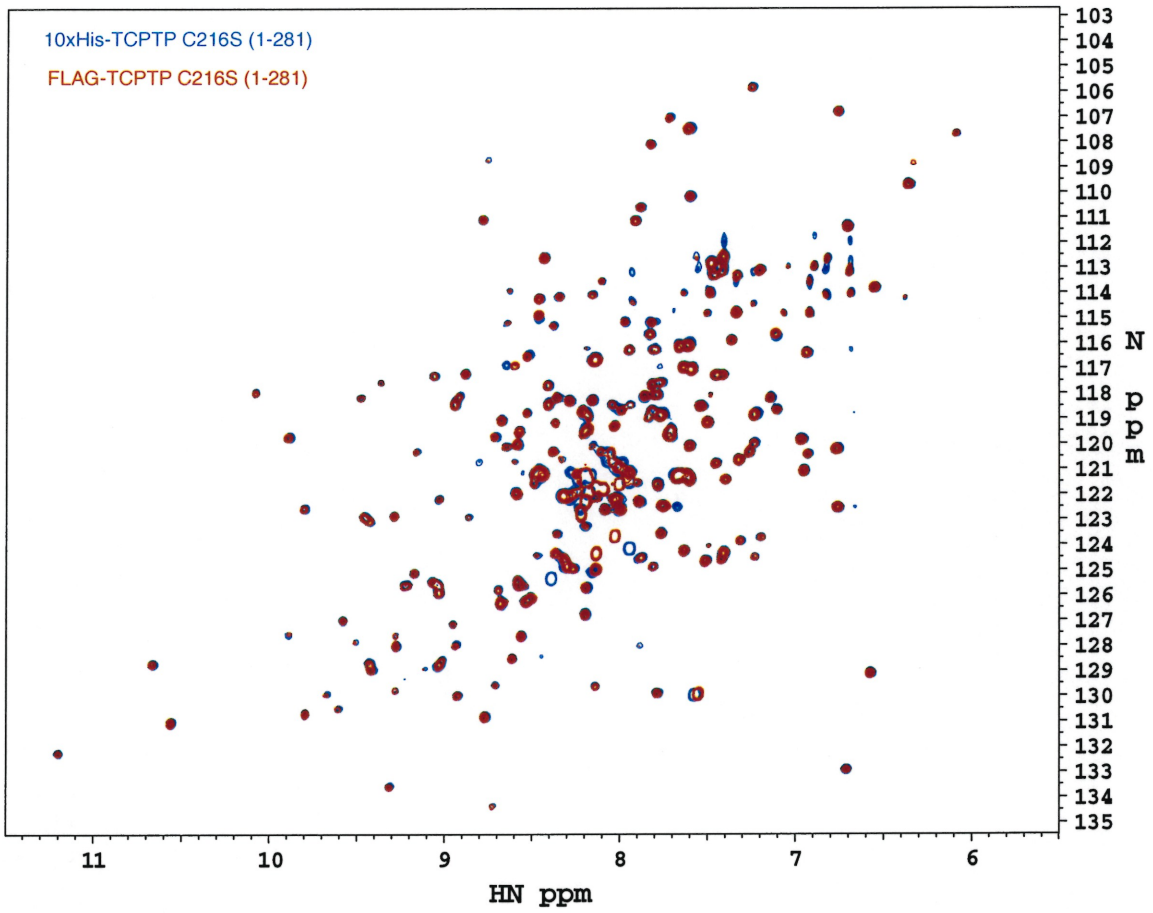


Figure 23. 2D ^{1}HN-^{15}N HSQC correlation of ~0.4 mM 10xHis-TCPTP C216S (1-281)-pET19b (in 100 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10}, 10% D_2O) and ~0.4 mM FLAG-TCPTP C216S (1-281)-pET11a (in100 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10}, 5% D_2O) overlaid for comparison.

Purification of 10xHis-TCPTP C216S (1-281)

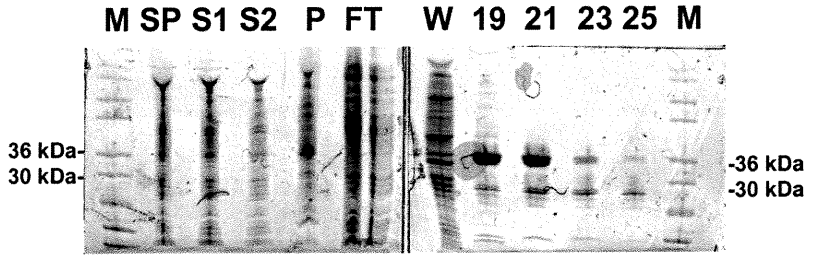
The purification protocol of 10xHis-TCPTP C216 (1-281) was further optimized by comparing two different buffering systems; phosphate buffer and Tris buffer. The incorporation of phosphate buffer during the purification protocol would facilitate amenability of 10xHis-TCPTP C216S (1-281) to NMR sample conditions since the final NMR buffering conditions consist of phosphate buffer.

Both purification protocols employed a Ni-NTA affinity column for the first chromatography step. The second step consisted of an ion-exchange column. In the presence of Tris-HCl buffer (pH 8.0) a MonoQ HR 5/5 anion-exchange column was employed whereas in the presence of phosphate buffer (pH 6.5) a MonoS HR 5/5 cation-exchange column was employed. Detection and visualization of the protein proceeded by electrophoresis on 10-15% SDS Phastgels (Amersham Pharmacia) using Mark12 molecular weight markers (Novex) and staining with Coomassie Blue respectively. In both cases, the first step of purification employed a Ni-NTA column. Although a greater amount of soluble protein was obtained from purification with phosphate buffer rather than Tris buffer during the first purification step, the fractions obtained with phosphate buffer also contained more contaminating proteins. In contrast, a larger amount of pure protein eluted from the final purification step using the MonoQ HR 5/5 anionic exchange column with Tris-HCl buffer rather than the MonoS HR 5/5 cationic exchange column with phosphate buffer. Furthermore, during concentration and buffer exchange by ultrafiltration, a white precipitate was observed for the 10xHis-TCPTP C216S (1-281) protein purified with phosphate buffer. This was probably due to the phosphate buffer being in very close proximity to the isoelectric point of the protein which is ~6.9. As a

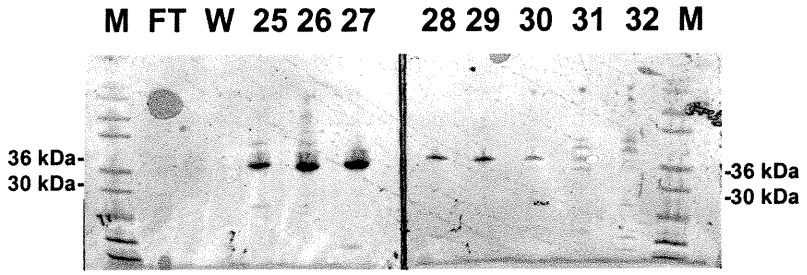
result, the final purification protocol for 10xHis-TCPTP C216S (1-281) was conducted with Tris-HCl buffer (pH 8.0). The results and the details of the purification protocols employed for TCPTP C216S (1-281)-pET19b using Tris-HCl buffer and phosphate buffer are found in Appendix C.

Final Purification Protocol of 10xHisTCPTP C216S (1-281)

The final purification of 10xHis-TCPTP C216S (1-281) was performed using Tris-HCl as the buffering agent first using a 4 mL Ni-NTA affinity column followed by a MonoQ HR 5/5 column; a MonoQ HR 10/10 column was employed for purification of larger cultures. During purification on the Ni-NTA column, the imidazole concentration of the wash step was decreased from 60 mM to 30 mM imidazole to minimize loss of the target protein during the wash step. During purification on the MonoQ HR 5/5 column a 5 column volume wash step using buffer A was inserted and the linear salt gradient to elute the protein was extended to 10 column volumes from 0-1 M NaCl in buffer B. The Coomassie Blue stained gels obtained from this final purification protocol are shown in Figure 24.



(a)



(b)

Figure 24. Final 10xHis-TCPTP C216S (1-281) purification protocol using (a) Ni-NTA column and (b) MonoQ HR 5/5.

M: Mark12 markers, SP: supernatant & pellet (60 mL); S1: supernatant #1 (60 mL); S2: supernatant #2 (40 mL); P: pellet (10 mL); FT: flow-through (100 mL); W: wash (40 mL); numbers denote fraction number (4 mL fractions for Ni-NTA and 2 mL fractions for MonoQ HR 5/5).

Fractions containing 10xHis-TCPTP C216S (1-281) were prepared as NMR samples using the protocol mentioned previously (refer to p.69). Protein concentration determinations using the extinction coefficient overestimated the concentration with an error of ≥ 25%, as established by measurement of 10xHis-TCPTP C216S (1-281) protein concentration by dry weight. A calibration curve of 0.05 mg/mL - 0.5 mg/mL 10xHis-TCPTP C216S (1-281) was constructed using Bio-Rad protein assay (160 μL H_2O, 40 μL Bio-Rad protein assay and 10 μL of diluted protein sample). Readings at 595 nm and 405 nm were used to prepare a calibration curve with a correlation value (r^2) of 0.972 (slope = 0.118, y-int = -0.087).

A 5 L culture was required to produce a 300 μL NMR sample containing 1 mM ^{15}N-^{13}C-^{2}H-TCPTP C216S (1-281). The protein sample was dissolved in 50 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10} and 8% of D_2O for frequency-lock purposes and purged under argon.

Isotope Incorporation

To study a large protein such as TCPTP, isotope labeling with ^{13}C, ^{2}H and ^{15}N are essential. Because the cost for preparing only 1 L of bacterial culture enriched with ^{13}C, ^{2}H and ^{15}N was approximately $2400 for the isotope labels alone, 10xHis-TCPTP C216S (1-281) was expressed with isotope enriched media only after the expression and purification protocols of the protein were established and optimized. In order to obtain 100% ^{13}C, ^{2}H, ^{15}N-labeled TCPTP C216S (1-281) the bacterial cells had to be trained to be grown in minimal media enriched with ^{13}C, ^{2}H, ^{15}N isotopes. As a result the isotopes were incorporated gradually and the culture was serially diluted into minimal media prepared in D_2O in order to obtain 100% deuteration.

Expression of ^{15}N-TCPTP C216S (1-281)-pET19b

One colony from a fresh transformation of BL21(DE3) *E. coli* competent cells harboring the TCPTP C216S (1-281)-pET19b plasmid was inoculated into 3 mL of LB containing ampicillin (100 μg/mL) and grown at 37 °C to an OD_{600} of 0.6-0.8. To obtain ^{15}N labeled protein for NMR experiments, this inoculum was diluted to 1:100 (1 mL in 100 mL) in M9 minimal medium (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L ^{15}N-NH_4Cl) with supplement (0.3% glucose, 10 mg of biotin, 10 mg of thiamine, 100 mg ampicillin, 2 mM $MgSO_4$ and 0.1 mM $CaCl_2$ per litre of M9) and grown overnight with slight agitation at 27 °C to an OD_{600} of 0.6-0.8. This culture was then inoculated into 2 L of ^{15}N-labeled M9 minimal medium containing supplement. The culture was allowed to reach OD_{600} of 0.6, then the temperature was decreased to 27 °C to increase solubility of the protein. Protein synthesis was induced upon addition of a 1 mM IPTG working

solution once the OD_{600} reached 0.7. Cells were harvested after a 2 h induction time, the cell pellet was flash-frozen and stored at –80 °C. A 2 L culture yielded a ~4.9 g cell pellet.

Expression of ^{15}N-^{2}H-TCPTP C216S (1-281)-pET19b

The expression of ^{15}N-^{2}H-TCPTP C216S (1-281)-pET19b was performed using the same protocol for expression of ^{15}N-TCPTP C216S (1-281)-pET19b except the M9 minimal media was prepared in deuterated water. Instead of diluting 1 mL of the starter culture into 100 mL of ^{15}N-M9 with supplement for overnight growth at 27 °C, this solution was allowed to incubate at 37 °C with shaking until an OD_{600} of 0.6-0.8 was reached. The culture was then further diluted to 1:20 (10 mL in 200 mL) in ^{15}N-D_2O-based M9 minimal medium (containing ^{15}N-labeled NH_4Cl) with D_2O-based supplement for overnight growth at 37 °C to an OD_{600} of 0.6-0.8. The overnight culture was then inoculated into 2 L of ^{15}N-D_2O-based M9 minimal media with the remaining D_2O-based supplement. The culture was allowed to reach OD_{600} of 0.6, the temperature was decreased to 27 °C, and protein synthesis was induced upon addition of a 1 mM IPTG working solution once the OD_{600} reached 0.7. Cells were harvested after a 2 h induction time. The cell pellet was flash-frozen and stored at –80 °C. A 2 L culture yielded a ~6 g cell pellet.

Expression of ^{15}N-^{13}C-^{2}H-TCPTP C216S (1-281)-pET19b

The expression of ^{15}N,^{13}C,^{2}H-TCPTP C216S (1-281)-pET19b was performed using the same protocol for expression of ^{15}N-^{2}H-TCPTP C216S (1-281)-pET19b, except

the D_2O-based supplement consisted of 0.2%, instead of 0.3%, of [$^{13}C_6$,$^{2}H_7$]-D-glucose. A 1:100 dilution of the starter culture into 100 mL of ^{15}N-M9 with supplement was grown at 37 °C with shaking until an OD_{600} of 0.6-0.8 was reached. The culture was further diluted 1:20 (50 mL in 1 L) in ^{15}N-D_2O-based M9 minimal medium (containing ^{15}N-labeled NH_4Cl) with ^{13}C-glucose-D_2O supplement for overnight growth at 37 °C to an OD_{600} of 0.6-0.8. The culture was then inoculated into a 10L fermenter containing the remaining 9 L of ^{15}N-D_2O-based M9 minimal media and the remaining ^{13}C-glucose-D_2O supplement. The 10L fermenter was programmed to agitate the media at 400 rpm and an anti-foaming agent (Sigma) was used to prevent foaming of the bacterial cell culture. The culture was allowed to reach OD_{600} of 0.5, then the temperature was decreased to 27 °C. Protein synthesis was induced upon addition of a 1 mM IPTG working solution once the OD_{600} reached 0.6. Cells were harvested after a 2 h induction time. The cell pellet was flash-frozen and stored at –80 °C. A 10 L culture yielded a ~27.4 g cell pellet.

Expression of CH_3(Val, Leu, Ile-δ)-^{15}N-^{13}C-^{2}H-TCPTP C216S (1-281)-pET19b

The expression of CH_3(Val, Leu, Ile-δ)-^{15}N-^{13}C-^{2}H-TCPTP C216S (1-281)-pET19b followed the same protocol used to express ^{15}N-^{13}C-^{2}H-TCPTP C216S (1-281)-pET19b, except that one fifth of the ^{15}N-D_2O-based M9 minimal medium with the [$^{13}C_6$,$^{2}H_7$]glucose-D_2O supplement was retained for later use. Once the final culture had reached an OD_{600} of 0.55, 50 mg/L each of [3,^{2}H]-U-$^{13}C_5$-α-ketoisovaleric acid and [3,3-$^{2}H_2$]-$^{13}C_4$-α-ketobutyric acid was added with ^{15}N-D_2O-based M9 minimal medium supplemented with [$^{13}C_6$,$^{2}H_7$]-D-glucose-D_2O (0.2%), usually 1 hour before induction. The culture was allowed to reach an OD_{600} of 0.55 before the temperature was reduced to

27 °C. Protein synthesis was induced upon addition of a 1 mM IPTG working solution once the OD_{600} reached 0.6. Cells were harvested after a 2 h induction time. The cell pellet flash-frozen and stored at –80 °C. A 2 L culture yielded a ~7 g cell pellet.

The ^{15}N-NH_4Cl was purchased from either Martek Biosciences Corporation or Cambridge Isotope Laboratories (CIL), the [$^{13}C_6$,$^{2}H_7$]glucose from either Martek Biosciences or Isotec Inc., D-glucose-C-d_7 from Isotec Inc., D_2O from CDN Laboratories, α-ketobutyric acid-$^{13}C_4$-3,3-d_2 and α-ketoisovaleric-U-$^{13}C_5$ acid 3-d_1 from CIL.

Multidimensional NMR Experiments

Protein samples obtained from the final purification step were concentrated to ~1 mL. The buffer was exchanged with at least three ~15 mL volumes of 50 mM NaH_2PO_4 pH 7.5 buffer with 5 mM DTT-d_{10} by ultrafiltration, at 2000 x g at 4 °C for 2-3 h, using a 5000 MWCO centrifugal filtration device (Millipore) to minimize the proton content. The filtrate after each buffer exchange was evaluated using the Bio-Rad protein assay (Bio Rad Laboratories) to ensure no protein loss was occurring during concentration and buffer exchange steps. During the final buffer exchange, the protein concentration was determined using the calculated extinction coefficient 43 990 $M^{-1}cm^{-1}$. The protein concentration was corrected in order to account for the 25% overestimation that occurs with use of the calculated extinction coefficient. The sample was concentrated to less than 1 mL and further concentrated using a 3000 MWCO Microcon-3 (Amicon Inc.) by centrifuging at 14000 x g at 4 °C. Usually D_2O was added to the 250-300 μL protein

sample to a final concentration of 5-8%. The sample was then transferred to a 5 mm Shigemi D_2O NMR tube, purged with argon and sealed with parafilm for storage at 4°C until ready for use.

All NMR experiments were performed at 25 °C on a four-channel Varian 600 MHz spectrometer equipped with a pulsed field gradient unit and an actively shielded triple-resonance probe head. A separate radio-frequency channel was used for each of the ^{13}C, ^{1}H, ^{2}H and ^{15}N pulses.

A series of four triple-resonance 3D NMR experiments with ^{2}H-decoupling were used to assign the polypeptide backbone ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and HN chemical shift resonances of $^{15}N,^{13}C,^{2}H$ labeled TCPTP C216S (1-281), namely, HNCA (Yamazaki *et al.* 1994*a*; 1994*b*), HN(CO)CA (Yamazaki *et al.* 1994*a*; 1994*b*), HNCACB (Shan *et al.* 1996) and HN(CO)CACB (Shan *et al.* 1996). The 1 mM triply-labeled protein was made up in 50 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10}, 8% D_2O for frequency lock purposes and purged under argon. A 3D HNCO (Kay *et al.* 1990) experiment was also used to distinguish NMR signals from noise peaks in the four triple-resonance experiments. All of the pulse sequences employed the TROSY principle (Pervushin *et al.* 1997) as implemented by Yang and Kay (1999) to enhance sensitivity in the N and HN dimensions. After each 3D experiment a 2D ^{1}HN-^{15}N HSQC (Müller 1979; Bodenhausen and Reuben 1980) experiment was also performed to ensure that the sample did not degrade.

The 3D HNCA experiment was recorded as a 216* x 64* x 1152* matrix (the * indicates complex points) with acquisition times of 27.0, 20.2, and 64.0 ms in each of t_1,

t_2, and t_3. The corresponding spectral widths in F_1, F_2, and F_3 were 4000.0, 1580.8, and 9000.9 Hz. The data set was acquired with 16 scans per FID and a repetition delay of 1.5 s, giving rise to a total acquisition time of 104 h. For the 3D HN(CO)CA experiment, a matrix of 212* x 64* x 1152* complex points was acquired with acquisition times of 26.5, 20.2 and 64.0 ms in each of t_1, t_2, and t_3. The corresponding spectral widths in F_1, F_2, and F_3 were 4000.0, 1580.8, and 9000.9 Hz. The data set was acquired with 16 scans per FID and a repetition delay of 1.5 s, giving rise to a total acquisition time of 104 h. Both the 3D HNCACB and HN(CO)CACB experiments were acquired with 216* x 64* x 1152* complex matrices and with acquisition times of 25.6, 20.2, and 64.0 ms (t_1, t_2, and t_3). Spectral widths of 4224.3, 1580.8, and 9000.9 Hz in F_1, F_2, and F_3 were employed. Each FID was recorded with 32 scans and with a repetition delay of 1.5 s to give total acquisition times of 9 days. The HNCO experiment was recorded as a 44* x 64* x 1152* complex matrices with acquisition times of 14.7, 20.2, and 64.0 ms in each of t_1, t_2, and t_3. The corresponding spectral widths in F_1, F_2, and F_3 were 1500.0, 1580.8, and 9000.9 Hz. The data set was acquired with 32 scans per FID and a repetition delay of 1.6 s, giving rise to a total acquisition time of 44 h. All the 3D experiments (except HNCO) used constant time (CT) acquisition in t_1 for better resolution in the carbon dimension. The 2D ^{1}HN-^{15}N HSQC experiments were recorded as a 128* x 1088* complex matrix with acquisition times of 40.5 and 60.4 ms (t_1 and t_2) and spectral widths of 1580.0 and 9000.9 Hz (F_1 and F_2). The total measuring time with 8 scans per FID and a repetition delay of 1.6 s was 1 h.

All spectra were processed using NMRPipe/NMRDraw software (Delaglio 1995) and analyzed using the program NMRView (Johnson and Belvins 1994). For the series of four triple-resonance experiments, the HN dimension of the data sets employed a solvent suppression filter (POLY) (Marion 1989) to deconvolute out the residual water signal prior to apodization with a 65°-shifted squared sine-bell window function. The data were subsequently zero filled to twice the size (2048 complex points), Fourier transformed, phased and the data extracted from the 10.5 to 5.5 ppm HN spectrum window. The data was then transposed and the carbon dimensions of all the data sets were apodized with a 65°-shifted squared sine-bell window function followed by zero filling to 256 complex points, Fourier transformation, phasing, and elimination of the imaginary half of the signal. The data were then transposed again and baseline corrected. The size of the nitrogen time domain was doubled via mirror image linear prediction (Zhu and Bax 1990), apodized using a 59°-shifted sine-bell window function, zero filled to 128 complex points, Fourier transformed, phase corrected, and the imaginaries were eliminated. The absorptive part of each of the four 3D data sets from the HNCA, HN(CO)CA, HNCACB and HN(COCA)CB experiments consisted of 256 x 128 x 1024 real points. The HNCO experiment was processed in a similar fashion; however, the absorptive part consisted of 64 x 128 x 1024 real points.

The 2D ^{1}HN-^{15}N HSQC data sets were first processed with a solvent suppression filter in the HN dimension followed by apodization with a 65°-shifted squared sine-bell window function. The data was subsequently zero filled to twice the size (2048 complex points), Fourier transformed, phased and the data extracted from the 11.5 to 5.5 ppm HN spectrum window. The data was then transposed and the size of the nitrogen time

domain was doubled via mirror image linear prediction, apodized using a 59°-shifted sine-bell window function, zero filled to 128 complex points, Fourier transformed, phase corrected and the imaginaries were eliminated. The absorptive part of the HSQC experiment consisted of 64 x 1024 real points.

All the raw NMR data were processed using NMRPipe/NMRDraw software (Delaglio 1995) in order to obtain well phased NMR spectra. The NMR spectra of $^{15}N,^{13}C,^{2}H$-TCPTP C216S (1-281) (50 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10}, 8% D_2O) obtained during the 3D HNCA, HN(CO)CA, HNCACB, and HN(COCA)CB experiments are presented in Figures 25, 26, 27 and 28 respectively. To assign the polypeptide backbone resonances using NMRView (Johnson and Belvins1994) NMR signals were peak picked by adjusting the threshold level to maximize signal detection and minimize noise peaks The NMR spectra of the 3D HNCO experiment was also used to ensure each of the amide proton signals observed in the HNCO (Figure 29) were peak picked in the four 3D NMR experiments as well as for discriminating weak but significant NMR signals from noise peaks. After each experiment a 2D ^{1}HN-^{15}N HSQC (Figure 30) NMR experiment of the same protein was performed to ensure the protein was still viable and did not undergo any degradation. Comparisons of the sequential 2D ^{1}HN-^{15}N HSQCs obtained during acquisition of these five NMR experiments confirmed that 1 mM 10xHis-TCPTP C216S (1-281) is stable at 25 °C.

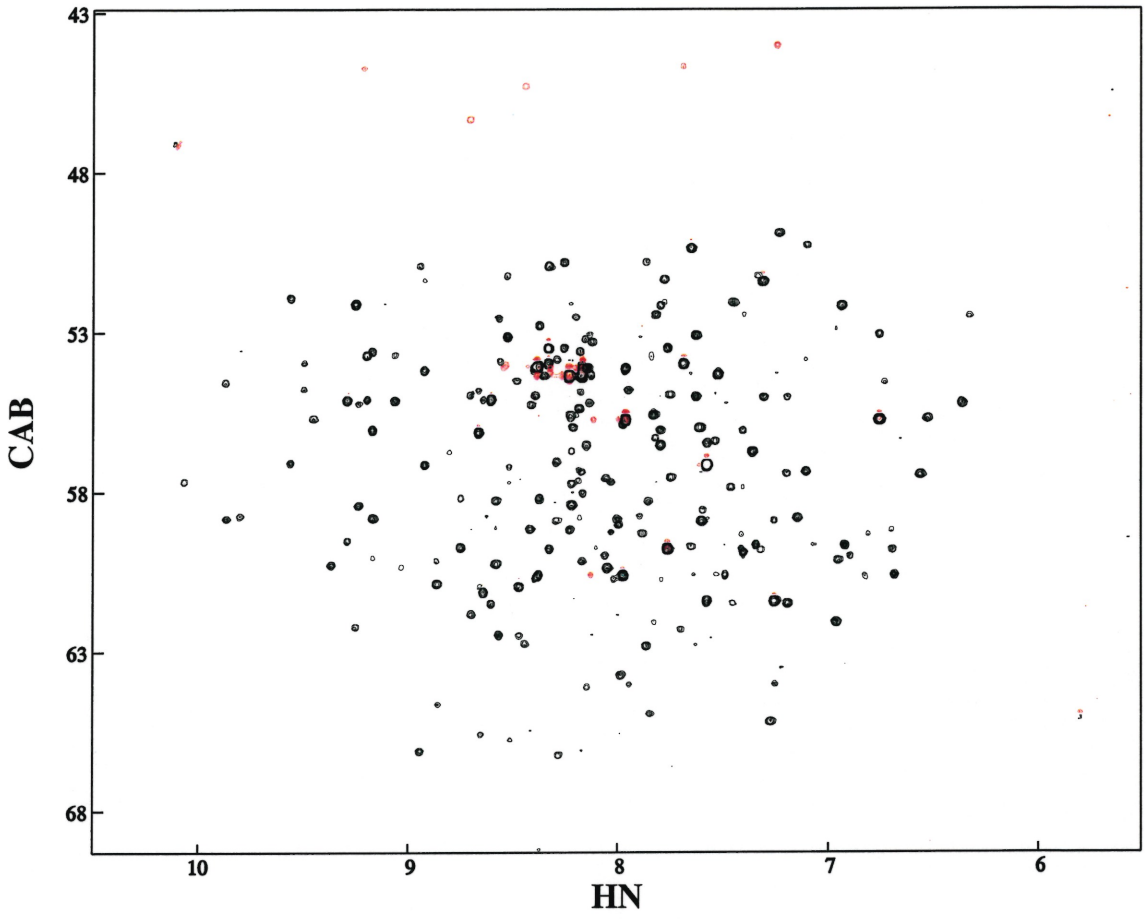


Figure 25. Compressed view of a 3D HNCA cube showing $^{13}C_{\alpha}$-^{1}HN (^{15}N in the *z*-dimension) correlation map of 1 mM ^{15}N,^{13}C,^{2}H-10xHis-TCPTP C216S (1-281) in 50 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10}, 8% D_2O. Signals in black represent positive NMR signals and those in red represent negative NMR signals.

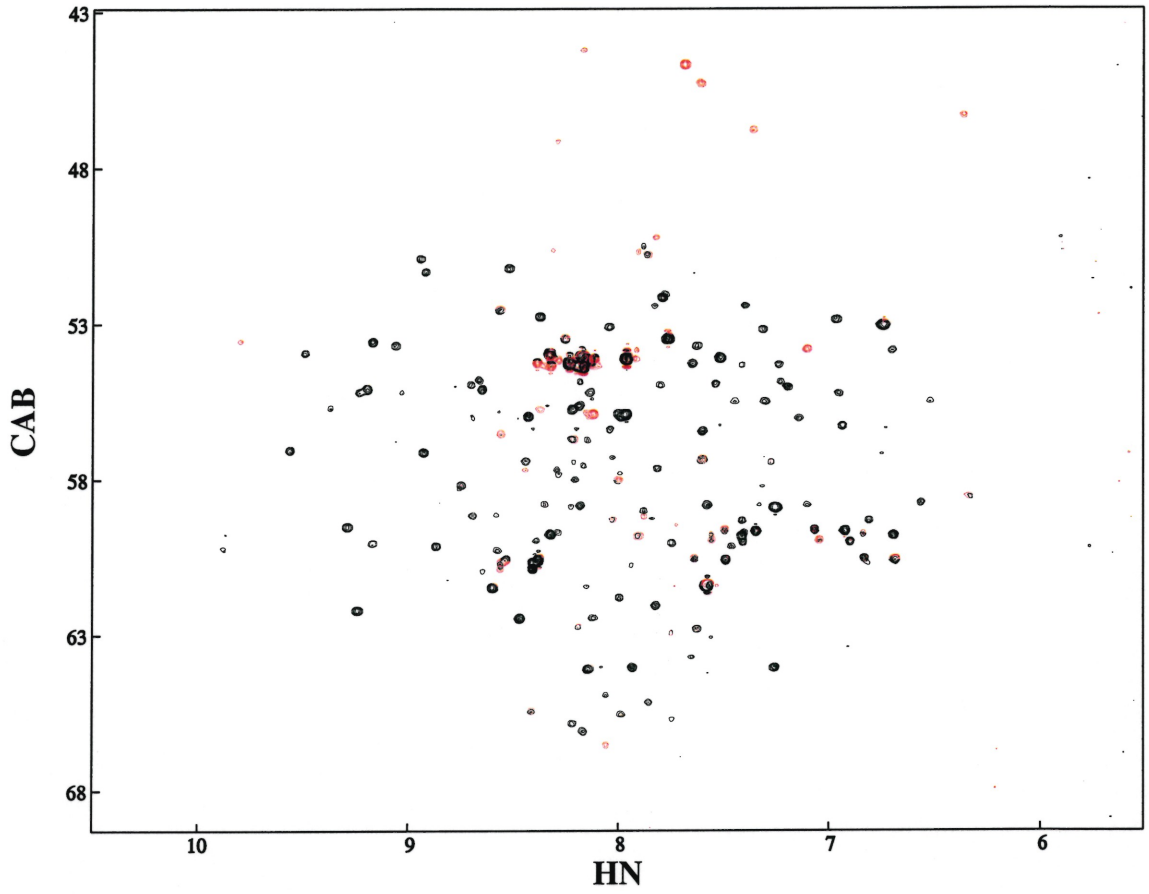


Figure 26. Compressed view of a 3D HN(CO)CA cube showing $^{13}C_{\alpha}$-^{1}HN (^{15}N in the *z*-dimension) correlation map of 1 mM ^{15}N,^{13}C,^{2}H-10xHis-TCPTP C216S (1-281) in 50 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10}, 8% D_2O. Signals in black represent positive NMR signals and those in red represent negative NMR signals.

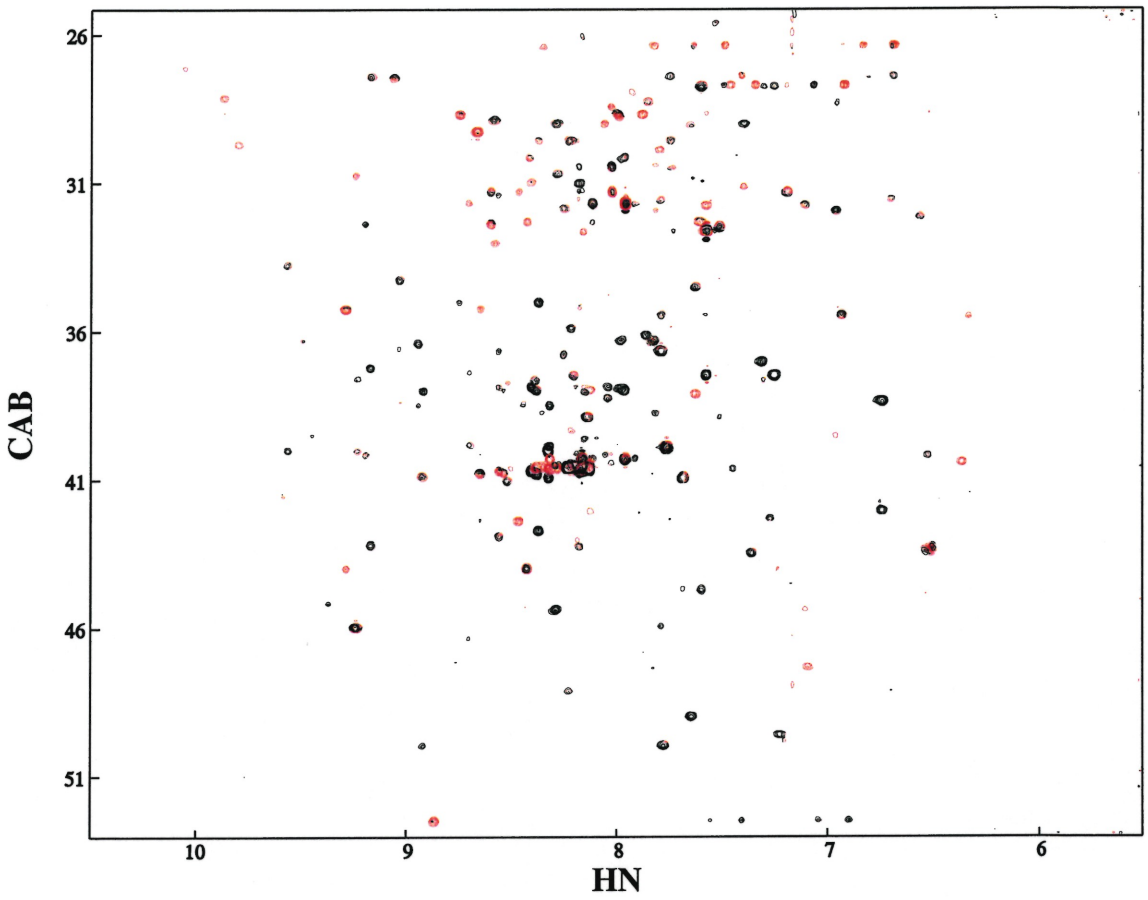


Figure 27. Compressed view of a 3D HNCACB cube showing $^{13}C_{\alpha}$-^{1}HN (^{15}N in the *z*-dimension) correlation map of 1 mM ^{15}N,^{13}C,^{2}H-10xHis-TCPTP C216S (1-281) in 50 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10}, 8% D_2O. Signals in black represent positive NMR signals and those in red represent negative NMR signals.

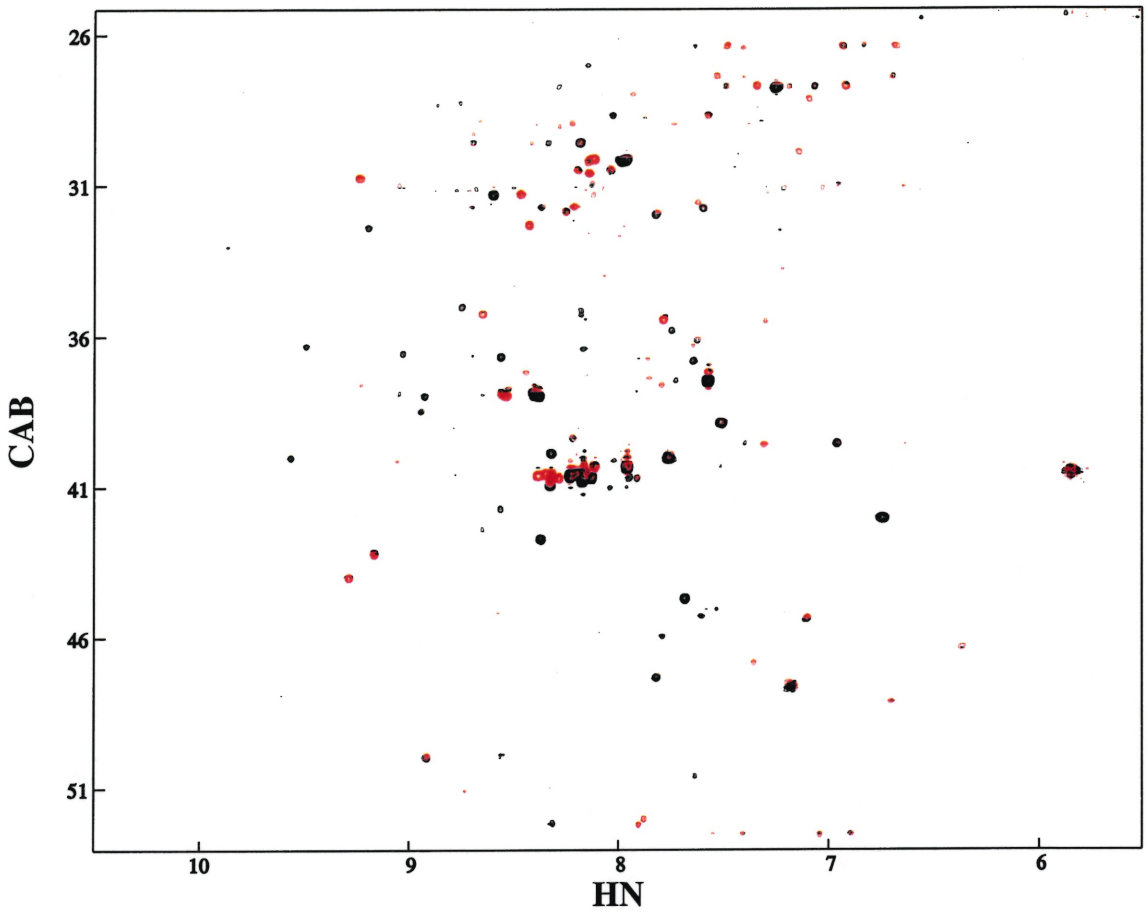


Figure 28. Compressed view of a 3D HN(COCA)CB cube showing $^{13}C_{\alpha}$-^{1}HN (^{15}N in the *z*-dimension) correlation map of 1 mM ^{15}N,^{13}C,^{2}H-10xHis-TCPTP C216S (1-281) in 50 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10}, 8% D_2O. Signals in black represent positive NMR signals and those in red represent negative NMR signals.

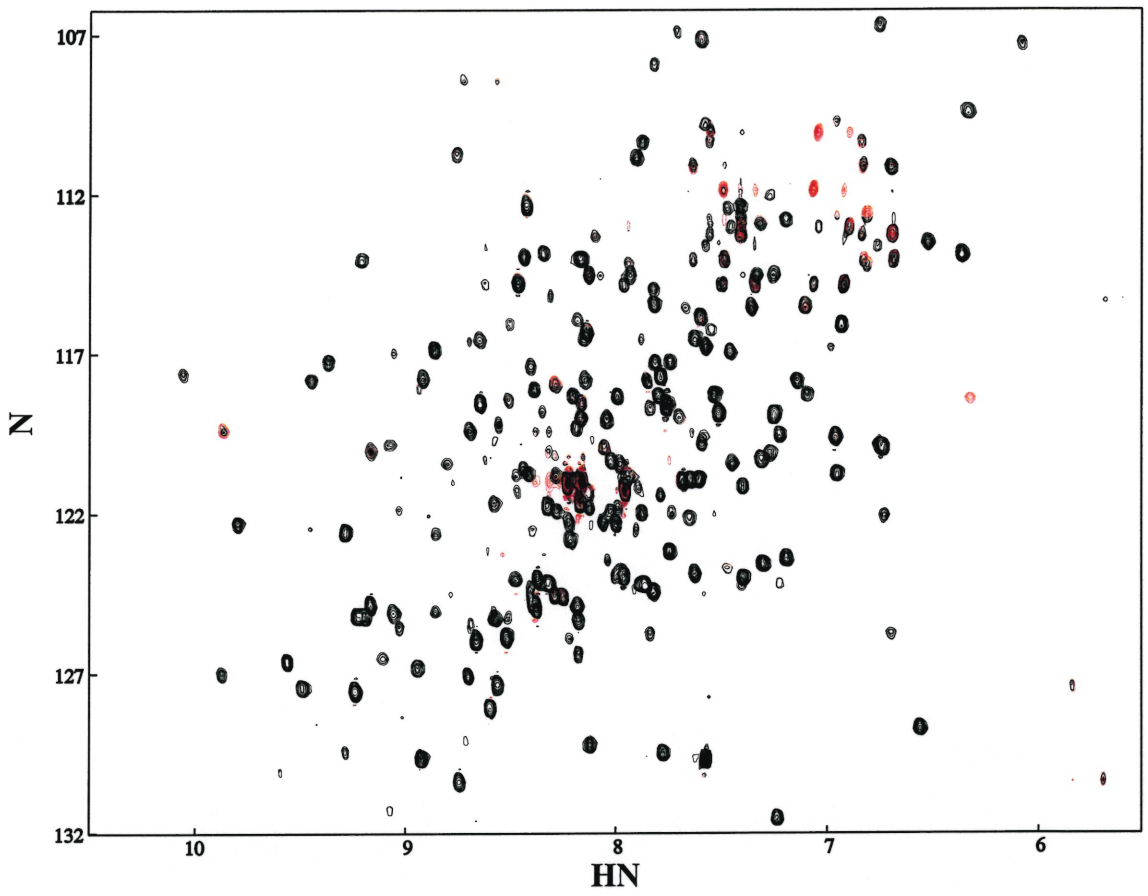


Figure 29. Compressed view of a 3D HNCO cube showing ^{1}HN-^{15}N (^{13}CO in the *z*-dimension) correlation map of 1 mM ^{15}N,^{13}C,^{2}H-10xHis-TCPTP C216S (1-281) in 50 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10}, 8% D_2O. Signals in black represent positive NMR signals and those in red represent negative NMR signals.

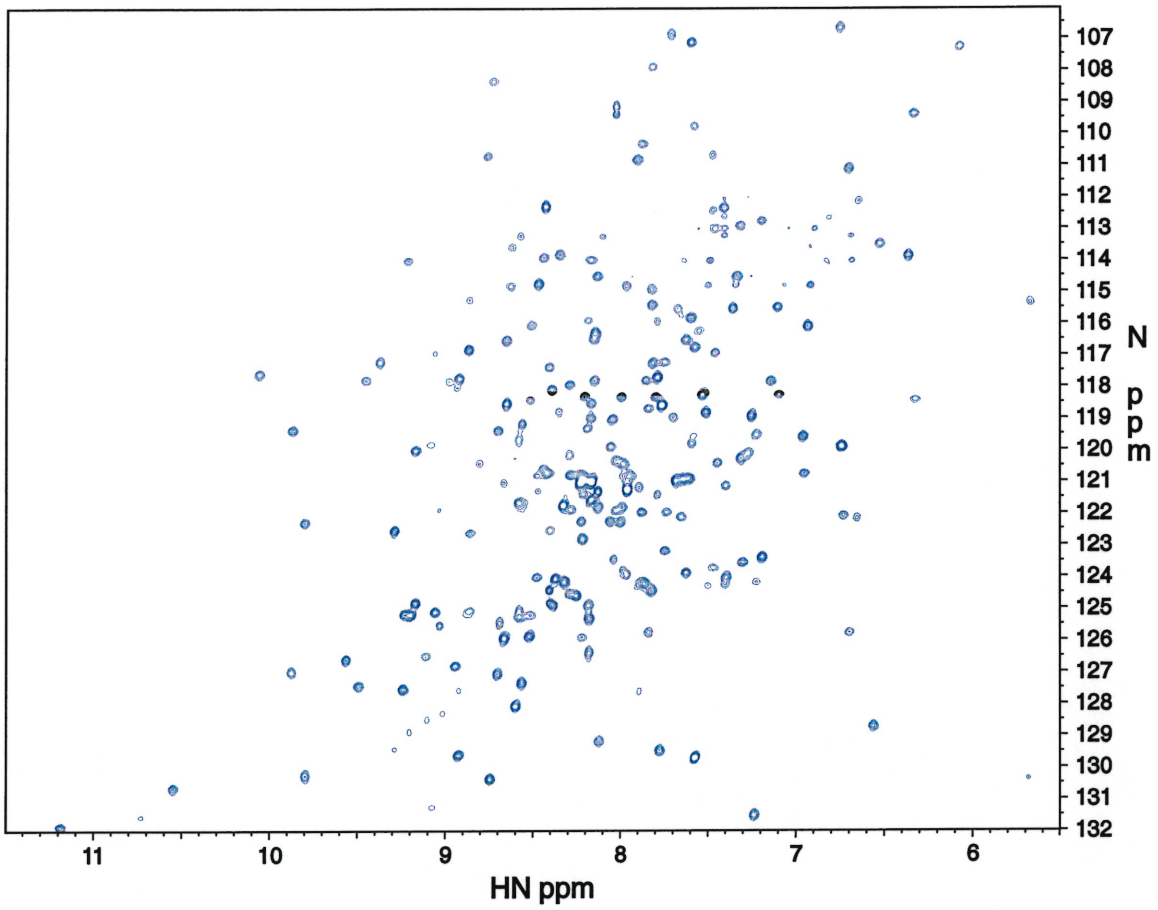


Figure 30. 2D ^{1}HN-^{15}N HSQC of 1 mM ^{15}N,^{13}C,^{2}H-10xHis-TCPTP C216S (1-281) in 50 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10}, 8% D_2O.

A constant time 2D ^{1}H-^{13}C HSQC (Vuister *et al.* 1992; Santoro and King 1992) experiment of CH_3(Val, Leu, Ile-δ)-^{15}N,^{13}C,^{2}H-TCPTP C216S (1-281) was also performed to confirm a successful methyl protonation labeling scheme. The 0.6 mM methyl protonated, triply labeled protein was made up in 50 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10}, 2 mM Pefabloc® (Boehringer Mannheim), 5% D_2O for frequency lock purposes and purged under argon.

The 2D ^{1}H-^{13}C constant time HSQC experiment was acquired with 64* x 1088* complex matrices and with acquisition times of 20.3 and 60.4 ms (t_1 and t_2). Spectral

widths of 1580.0 and 9000.9 Hz in F_1, and F_2 were employed. Each FID was recorded with 16 scans and with a repetition delay of 1.6 s to give total acquisition time of 2 h.

For the 2D ^{1}H-^{13}C constant time HSQC experiment the HN dimension of the data set employed a solvent suppression filter prior to apodization with a 65°-shifted squared sine-bell window function. The data was zero filled to twice the size (2048 complex points), Fourier transformed, phased and the data was extracted from the 4.5 to –1.0 ppm HN spectrum window. The data was then transposed and the nitrogen time domain apodized using a 59°-shifted sine-bell window function, zero filled to 128 complex points, Fourier transformed, phase corrected, window shifted 10 ppm downfield and the imaginaries were eliminated. The absorptive part of the 2D ^{1}H-^{13}C constant time HSQC experiment consisted of 64 x 1024 real points. The 2D ^{1}H-^{13}C constant time HSQC spectra of 0.6 mM CH_3(Val, Leu, Ile-δ)-^{15}N, ^{13}C, ^{2}H-TCPTP C216S (1-281) is shown in Figure 31.

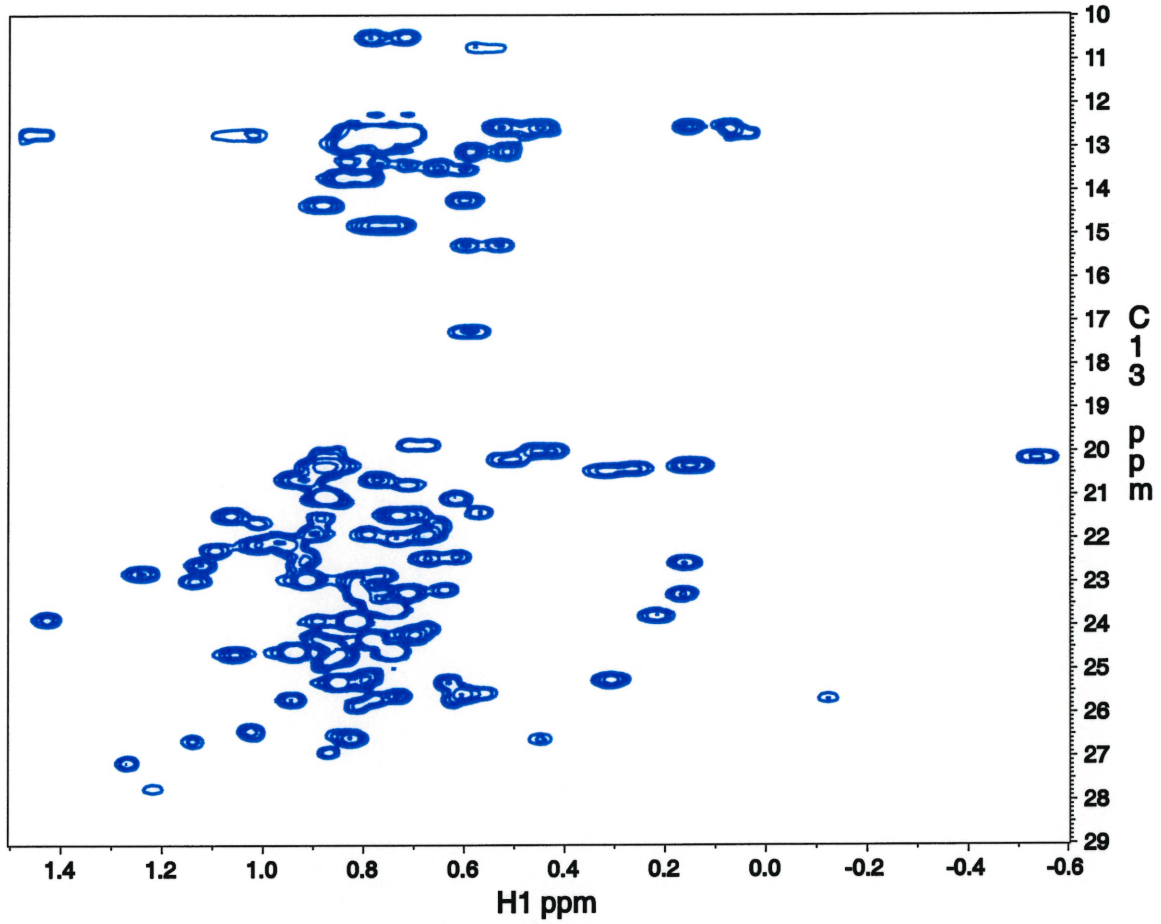


Figure 31. 2D ^{1}H-^{13}C of 0.6 mM CH_3(Val,Leu,Ile-δ)-^{15}N,^{13}C,^{2}H-10xHis-TCPTP C216S (1-281) in 50 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10}, 2 mM Pfabloc, 5% D_2O.

Chapter 4

Toward the 3D structure determination of TCPTP C216S (1-281)

I. Polypeptide Backbone Resonance Assignment

Structure determination of a protein by NMR spectroscopy consists in three major steps: resonance assignment, measurement of geometrically dependent parameters (chemical shifts, NOE-derived distances, coupling constants, residual dipolar couplings) and calculation of three dimensional models that are consistent with the experimental measurements.

The first step in determining the three dimensional structure of TCPTP C216S (1-281) using multidimensional NMR spectroscopy is the assignment of the polypeptide backbone resonances of a ^{13}C, ^{2}H, ^{15}N-labeled sample of TCPTP C216S (1-281). The triple resonance NMR experiments employed provided direct through-bond correlations of all the NMR active nuclei present in the protein (^{13}C, ^{1}H, ^{15}N). Each NMR signal is characteristic of both the precessing frequency of the specific nucleus type and its local magnetic environment. The principle of resonance assignment is simple. NMR experiments use the network of coupling constants between ^{13}C, ^{1}H and ^{15}N atoms to correlate chemical shifts within a residue (intraresidue). A correlation is a signal that possess two, three or even four chemical shifts. For example, a correlation in a 2D ^{1}H - ^{15}N HSQC experiment will have two chemical shifts: amide proton and amide nitrogen. Also, NMR experiments can correlate chemical shifts of atoms that belong to two adjacent residues (interresidue). These two types of correlation are necessary to identify which type of amino acid the signal belongs, and which residue is next to it.

Assigning each resonance to a specific amino acid of the primary sequence provides information on the type of secondary structure elements (α-helix, β-strand) a

given residue is involved (Wishart *et al.* 1991*a*; Wishart *et al.* 1991*b*). Knowledge of the chemical shifts of all residues can provide useful information in inhibitor binding studies. Since the chemical shift of an atom is very sensitive to its local environment they are easily perturbed by the presence of an inhibitor, or a small conformation change induced upon binding of small molecules. Consequently, chemical shift perturbation studies is a useful tool to map the binding surface of a protein using small molecule probes.

To assign the polypeptide backbone $^{13}C_{\alpha}$, $^{13}C_{\beta}$, ^{1}HN and ^{15}N chemical shift resonances of ^{13}C, ^{2}H, ^{15}N-labeled TCPTP C216S (1-281) a series of four triple-resonance NMR experiments with ^{2}H-decoupling were used (refer to Figure 32). Namely, HNCA, HN(CO)CA, HN(CA)CB and HN(COCA)CB. The HNCA experiment correlates the amide pair ^{1}HN, ^{15}N of residue *i* with its intraresidue $^{13}C_{\alpha}$ ($^{1}HN(i)$; $^{15}N(i)$; $^{13}C_{\alpha}(i)$). The chemical shift of the previous residue in the primary sequence (*i-1*) is also detected giving rise to the interresidue correlation $^{1}HN(i\text{-}1)$; $^{15}N(i\text{-}1)$; $^{13}C_{\alpha}(i\text{-}1)$. Both correlations can be observed because the coupling constants $^{1}J_{N(i),C\alpha(i)}$ and $^{2}J_{N(i),C\alpha(i-1)}$ are usually very similar. Therefore, during the $^{15}N \rightarrow {}^{13}C$ INEPT, magnetization transfer from the amide nitrogen to both $^{13}C_{\alpha}$ intra and interresidues are possible. The HN(CO)CA experiment correlates the intraresidue amide $^{1}HN(i)$ and $^{15}N(i)$ chemical shifts; to the $^{13}C_{\alpha}(i\text{-}1)$ chemical shifts of the previous residue (*i-1*). This correlation is obtained from a magnetization transfer that is selectively executed from the amide to the carbonyl, and to the $^{13}C_{\alpha}$ of the previous residue ($^{15}N(i) \rightarrow {}^{13}CO(i\text{-}1) \rightarrow {}^{13}C_{\alpha}(i\text{-}1)$). As a result the $^{13}C_{\alpha}(i\text{-}1)$ chemical shift obtained from the HN(CO)CA experiment connects the $^{13}C_{\alpha}$ of residue *i-1* to the amide of residue *i* ($^{15}N(i)$). While this interresidue correlation is

redundant to the data obtained in the HNCA experiment, the latter is not sometimes detected or difficult to distinguish from the intraresidue correlation.

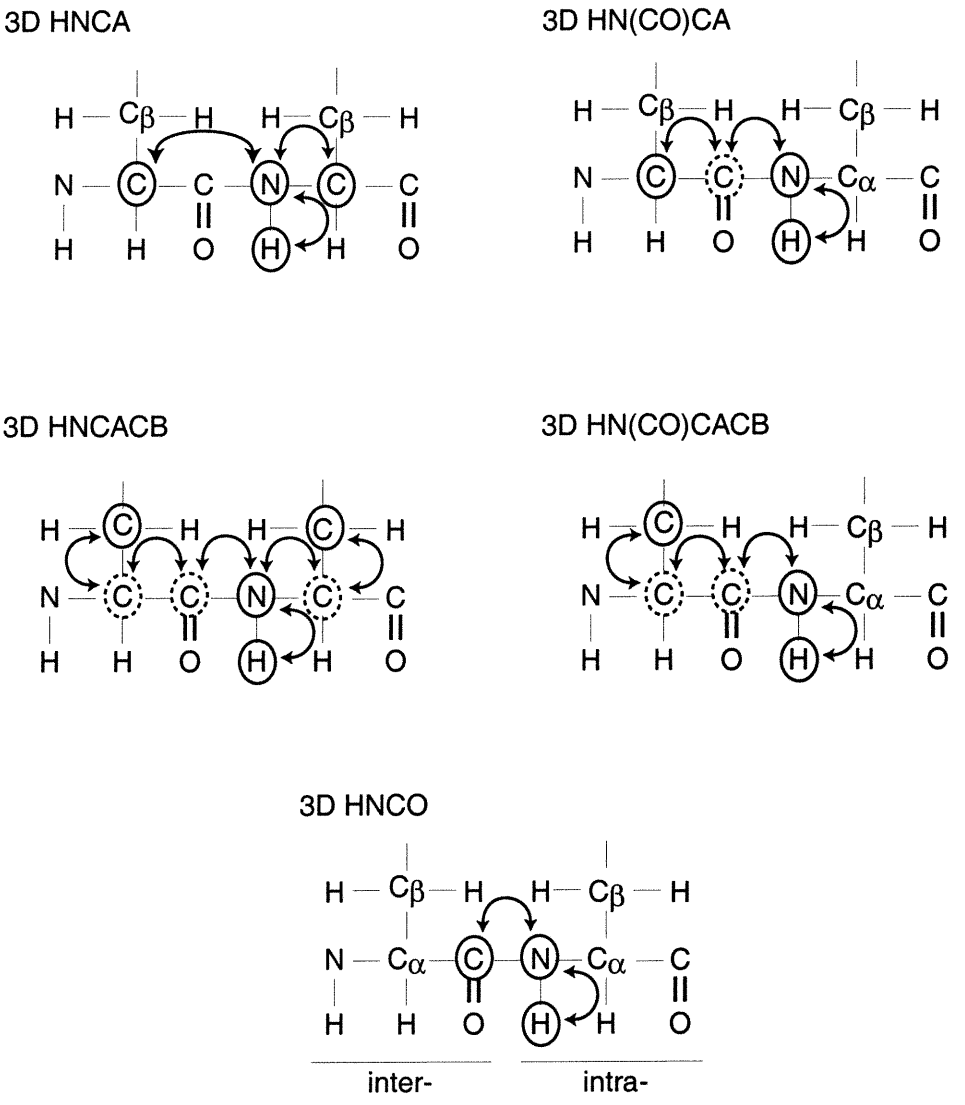


Figure 32. The series of five 3D NMR experiments used to assign the polypeptide backbone resonances.

Magnetization transfer is marked by curved solid lines, and the direction of the transfer is indicated by arrows. Nuclei for which the chemical shift is measured in the three-dimensional experiment are marked by the solid circles.

Two experiments, namely the HN(CA)CB and the HN(COCA)CB, correlate amide ^{1}HN, ^{15}N with $^{13}C_{\beta}$ chemical shifts. The HN(CA)CB is performed in the same fashion as the HNCA but a magnetization transfer (COSY-type) is done from the $^{13}C_{\alpha}$ to the $^{13}C_{\beta}$. For the reasons discussed previously in the HNCA experiment, both $^{13}C_{\beta}$ of residue *i* and the previous residue (*i-1*) are correlated to the amide ^{1}HN(*i*), ^{15}N(*i*) chemical shifts. A fifth experiment, the HNCO, correlates the carbonyl chemical shift to the amide proton nitrogen pair chemical shifts. This experiment has a much higher signal-to-noise ratio compared to the four experiments described before where often, the signal of interest may have an intensity of the order of the noise level. The high sensitivity of the HNCO experiment facilitates the peak detection process by clearly identifying regions where amide peaks were expected. In other words during analysis of the set of 3D experiments the HNCO dataset was used to filter out regions where no peaks were expected to be present.

Before describing in more detail the resonance assignment procedure, it is useful to provide a better insight about the three dimensional nature of the data that were collected. All experiments recorded NMR signals that were modulated with three frequencies, or chemical shifts, (δ_{HN}; δ_{N}; δ_{Cj}) where C_j can be either $C_{\alpha}(i)$, $C_{\alpha}(i-1)$ or $C_{\beta}(i)$, $C_{\beta}(i-1)$, depending of the experiment. Analysis of such dataset in their three dimensional representation is totally impractical. Instead the 3D spectra are represented as a collection of 2D contour maps (similar to the spectrum represented in Figure 23). A 3D spectrum can be compared to a book whereby each page of the book contains a 2D contour map (called a "slice") represented with two chemical shift axis. Typically, the HN (*x*-axis) and Cj (*y*-axis) chemical shifts are plotted on the 2D contour map. The third

axis (*z*-axis), corresponds to the page number of the book and is the amide nitrogen chemical shift. For example, a correlation from the 3D HNCA experiment (δ_{HN}; δ_{CA}; δ_N) will be observed at the intersection corresponding to the chemical shifts of δ_{HN} on the *x*-axis and the δ_{CA} on the y-axis of the 2D contour map located at δ_N in the *z*-axis (the page number in ppm in our book example). All 3D experiments were recorded with 26 ppm spectral window in the nitrogen axis (*z*-axis) with a total of 128 points after data processing. The first 2D contour map was at δ_N = 106 ppm, the second map at 106.20 ppm etc.. and the last contour map at 132 ppm.

The assignment procedure is divided in two major steps; building clusters and matching. The clusters obtained involved taking the large collection of correlations and organizing them on the basis of amide proton-nitrogen pairs so as to assemble all four datasets into one. One amide proton-nitrogen pair, named *k*, has potentially four different correlations as follows: (HN(*k*); N(*k*); $C_\alpha(k)$); (HN(*k*); N(*k*); $C_\alpha(k\text{-}1)$); (HN(*k*); N(*k*); $C_\beta(k)$) and (HN(*k*); N(*k*); $C_\beta(k\text{-}1)$). A cluster is a small sub dataset that assembles all carbon data available regarding the amide proton-nitrogen pair '*k*'. A new contour map for cluster '*k*' is composed of four subspectra or strips around $\delta_{HN(k)}$ and $\delta_{N(k)}$ as shown in Figure 33.

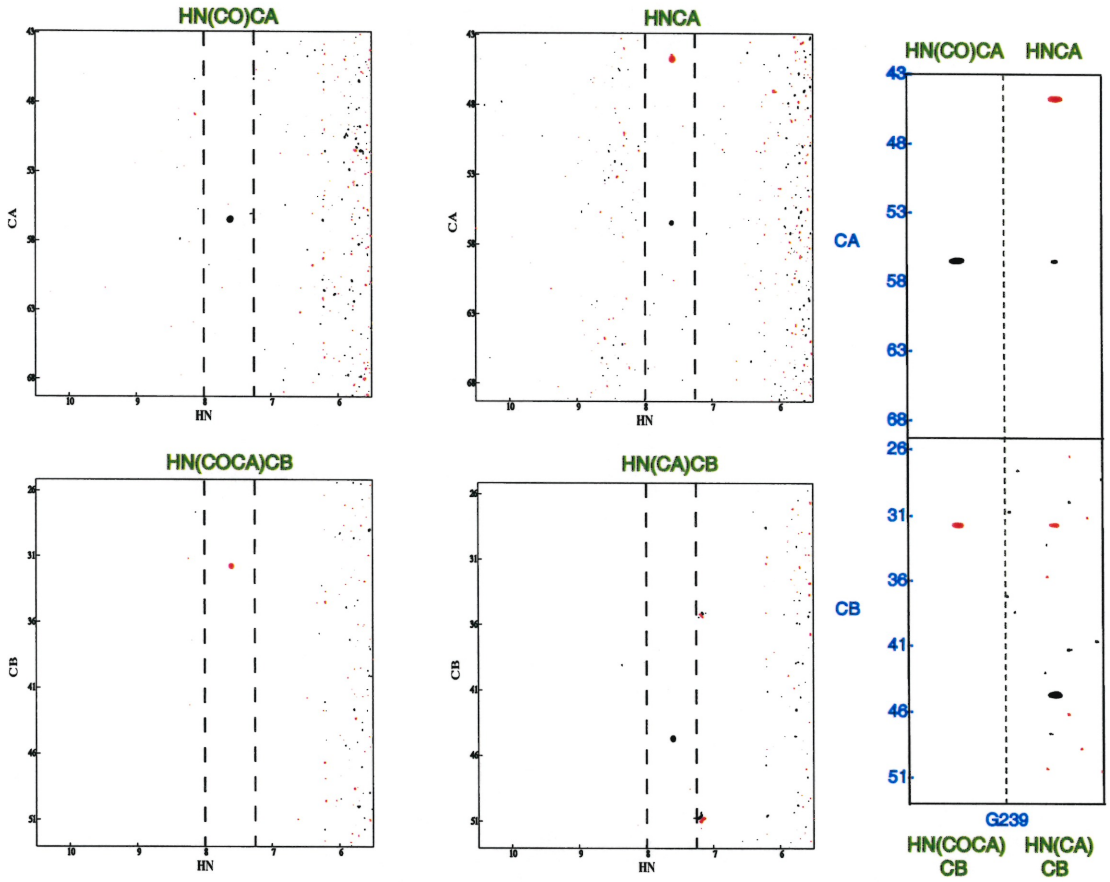


Figure 33. Assemblage of subspectra (strips) from the four 3D NMR experiments in formation of one cluster 'k'.

The chemical shift correlations obtained from the HNCA, HN(CO)CA, HN(CA)CB and HN(COCA)CB experiment at one specific amide nitrogen chemical shift ($\delta_N(k)$) are on one slice of the 3D dataset and represented in the solid box. The specific amide proton chemical shift ($\delta_{HN}(k)$) of the amide nitrogen-proton pair dictates the strip that is obtained from the $\delta_N(k)$ slice. The strips are represented by dotted lines. The strips from each of the four 3D NMR experiments contain chemical shift resoances correlated to one amide nitrogen-proton pair and are organized into one cluster '*k*'. The cluster is represented at the far right.

The organization and the analysis of all spectra could be done manually but it would be highly impractical. NMRView (Johnson and Belvins 1994) is a software written in Tcl/Tk language that was developed for this purpose. It contains a well developed graphical display that allows the user to analyze many spectra simultaneously. This feature is very useful in sorting out which peaks belong to a given HN, N pair of chemical shifts. Scripts for automatic peak detection are also implemented. These features reduce the tedious task of manual peak picking through all 2D contour maps (slices) of all the 3D NMR experiments. Manual editing is only needed when a signal is of lower intensity than the set threshold.

Once the data are reorganized in clusters, the matching step can be performed. The principle is straightforward. A cluster '*k*' contains interresidue correlations ($C_\alpha(i\text{-}1)/C_\beta(i\text{-}1)$) that will match the set of intraresidue correlations ($C_\alpha(i)/C_\beta(i)$) of another cluster. If the matching process would be carried out manually, each cluster (see Figure 33) would be on a single piece of paper. The procedure would consist in finding which cluster contains the interresidue peaks that match perfectly the intraresidue peaks of another cluster and put them adjacent to each other. When two clusters are matched in a sequential way a fragment is formed. Every cluster must then be matched so as to form as long a fragment as possible. Obviously, proline residues do not have an amide proton and this creates breaks in the fragments. When all clusters are matched then the fragments can be matched to the amino acid sequence of the protein. Because of the high diversity of $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts induced by the protein structure, the probability of having two clusters matching the same cluster perfectly is not very high. If it happens, both clusters are left unmatched and the decision is made later. The matching is based on

a comparison of the chemical shifts of the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ of the fragments with the random coil chemical shifts for these same $^{13}C_{\alpha}$, $^{13}C_{\beta}$ of the residue of the sequence. In addition, the sign of the correlation is used in the process. The NMR experiments are recorded in such a way (constant time mode) that the sign of the signal reflects the number of $^{13}C_{\beta}$ bound to the $^{13}C_{\alpha}$ for the $^{13}C_{\alpha}$ experiments or the number of $^{13}C_{\gamma}$ bound to $^{13}C_{\beta}$ for the $^{13}C_{\beta}$ experiments. In the $^{13}C_{\alpha}$ experiments, a positive correlation is observed for all those amino acids that contain a side chain but glycine. The latter is the only amino acid without a side chain. The correlation of glycine $^{13}C_{\alpha}$ has a negative intensity in the HNCA and HN(CO)CA experiments. Similarly, for the $^{13}C_{\beta}(i)$/ $^{13}C_{\beta}(i\text{-}1)$ correlations in the HN(CA)CB and HN(COCA)CB, the number of coupling constants between $^{13}C_{\beta}$ and $^{13}C_{\gamma}$ is observed indirectly by the sign of the NMR signal. For these experiments, side chains containing aromatic or carbonyl groups are actively decoupled during the NMR experiment however amino acids containing long aliphatic side chains could not be decoupled without perturbing the signal for the $^{13}C_{\beta}$ correlations. As a result, a positive NMR signal is observed if $^{13}C_{\beta}$ is coupled to an even number of $^{13}C_{\gamma}$ (e.g. valine) and a negative NMR signal is observed if $^{13}C_{\beta}$ was coupled to an odd number of $^{13}C_{\gamma}$ (e.g. leucine, glutamine). Refer to Table 4 for the random coil chemical shifts and the sign of the signals for the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ correlations of the twenty amino acids. The range of random coil chemical shifts have been obtained using all the data from the repository for biomolecular data derived from NMR spectroscopy (Richarz and Wüthrich 1978). The chemical shift ranges are of a spectral window of 26 ppm therefore aliasing ("folding in") of certain signals in this spectral window have been accounted for.

Table 4 . Random coil chemical shifts for $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ and sign of the correlation for constant time acquisition mode.

Amino Acid	C_{α} random coil chemical shift bounds (ppm)	Sign	C_{β} random coil chemical shift bounds (ppm)	Sign
ALA	55.0 – 50.0	positive	50.0 – 43.0	positive
ARG	60.0 – 55.0	positive	35.0 – 29.0	negative
ASN	55.0 – 50.0	positive	42.0 – 36.0	positive
ASP	57.0 – 51.0	positive	43.0 – 38.0	positive
CYS	58.0 – 53.0	positive	33.0 – 28.0	positive
GLN	58.0 – 53.0	positive	32.0 – 28.0	negative
GLU	60.0 – 53.0	positive	32.0 – 28.0	negative
GLY	49.0 – 41.0	negative	NA	positive*
HIS	58.0 – 54.0	positive	31.0 – 27.0	positive
ILE	64.0 – 57.0	positive	44.0 – 38.0	positive
LEU	58.0 – 52.0	positive	45.0 – 41.0	negative
LYS	60.0 – 53.0	positive	35.0 – 30.0	negative
MET	57.0 – 53.0	positive	37.0 – 32.0	negative
PHE	59.0 – 54.0	positive	42.0 – 38.0	positive
PRO	66.0 – 62.0	positive	34.0 – 31.0	negative
SER	61.0 – 55.0	positive	38.0 – 34.0	positive
THR	64.0 – 58.0	positive	43.0 – 39.0	negative
TRP	58.0 – 53.0	positive	35.0 –25.0	positive
TYR	58.0 – 55.0	positive	43.0 – 38.0	positive
VAL	66.0 – 58.0	positive	36.0 – 31.0	positive

* The C_{α} is detected in the C_{β} experiments at its chemical shift but with a positive intensity.

(*Source: Richaryz and Wüthrich 1978*)

Refer to Figure 34 for an illustration of how the four experiments were used to obtain sequential assignment of $^{13}C_{\alpha}(i)/^{13}C_{\alpha}(i\text{-}1)$, $^{13}C_{\beta}(i)/^{13}C_{\beta}(i\text{-}1)$, $^{1}HN(i)$ and $^{15}N(i)$ chemical shift correlations. This illustration is only a section of one of the best string of matched clusters that was successfully assigned to the primary sequence of TCPTP C216S (1-281).

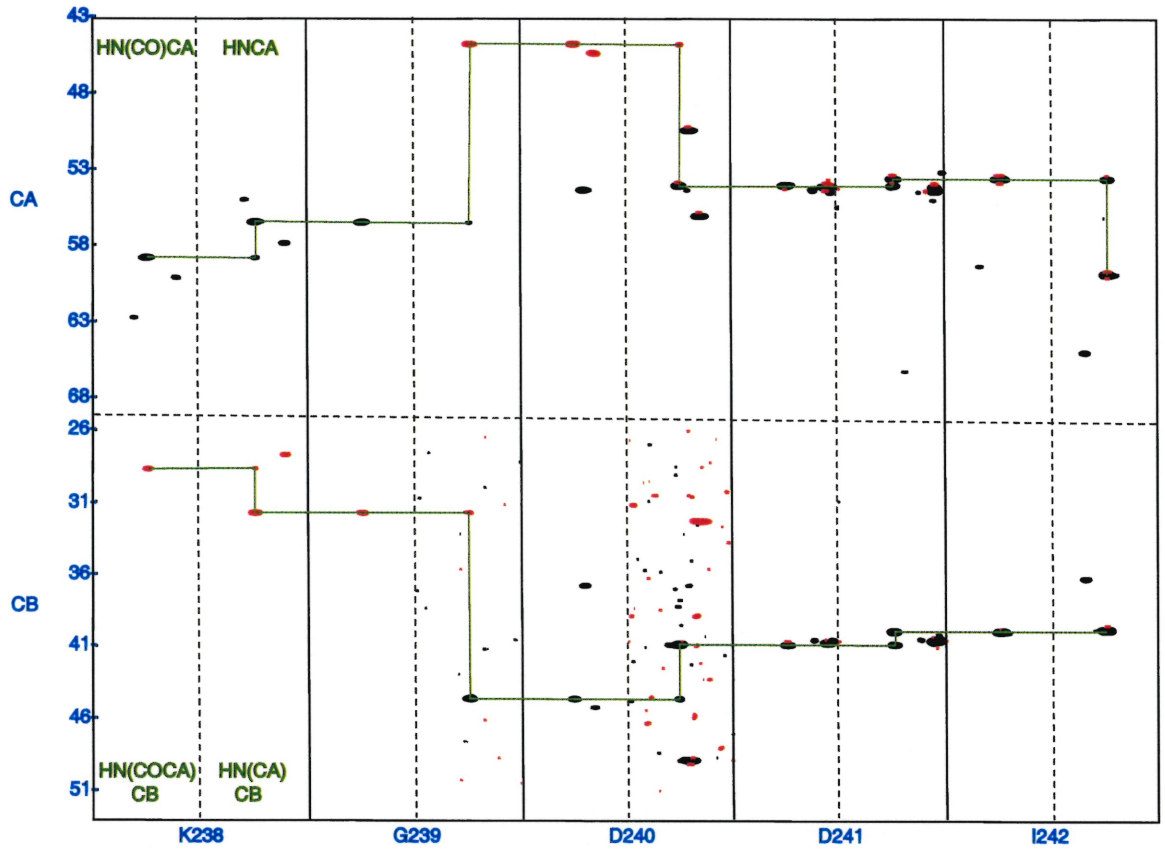


Figure 34. NMR experiments of 1 mM ^{15}N-^{13}C-^{2}H-TCPTP C216S (1-281) in 50 mM NaH_2PO_4 pH 7.5, 5 mM DTT-d_{10} and 5% D_2O. Strip plots of HNCA and HN(CO)CA (upper boxes), HNCACB and HN(COCA)CB (lower boxes) illustrating $^{13}C_\alpha(i)$/ $^{13}C_\alpha(i\text{-}1)$, $^{15}N(i)$, HN(i) and $^{13}C_\beta(i)$/$^{13}C_\beta(i\text{-}1)$, $^{15}N(i)$, HN(i) chemical shift correlations respectively.

The series of four triple resonance multidimensional NMR experiments were performed with a concentration of 0.7 mM on TCPTP C216S (1-281). Assignment of the backbone could not be completed due to the lack of C_β correlations. Many clusters were missing these and made the matching impossible to perform without ambiguity. Also, the C_β chemical shift is very important in assigning a fragment to the primary sequence. With this first set of data, only 42% of the sequence could be assigned. In an attempt to observe more C_β correlations, a new sample was produced by increasing the protein concentration to 1 mM and the suite of 3D experiments were recorded. Although there was some gain in sensitivity there was no significant improvement in the number of polypeptide backbone resonances assigned. The datasets recorded with this more concentrated sample improved the assignment to only 50%. Figure 35 represents the amino acids (in red) of the primary sequence that have been assigned to date. The polypeptide backbone $^{13}C_\alpha$, $^{13}C_\beta$, ^{15}N and ^{1}HN chemical shift resonances of ^{13}C, ^{2}H, ^{15}N-labeled TCPTP C216S (1-281) are tabulated in Table 5.

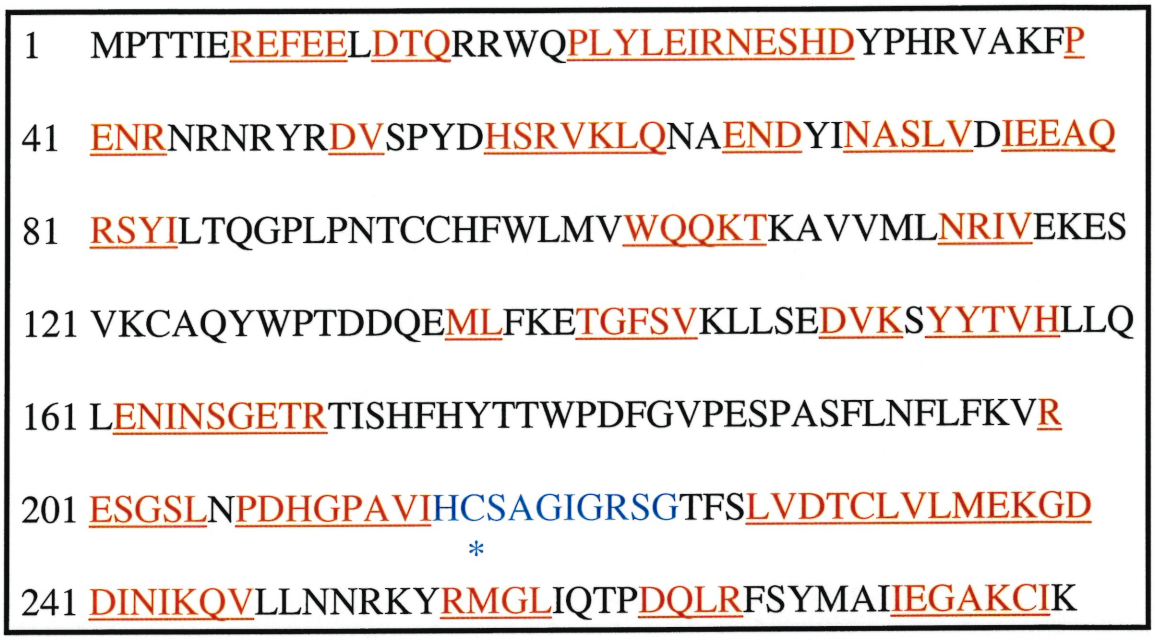


Figure 35. Partial assignment of TCPTP C216S (1-281).

Amino acids successfully assigned are represented in red. The catalytic loop is represented in blue. The asterix is to note that the catalytic cysteine is actually mutated to a serine.

Table 5. Chemical shift resonances of polypeptide backbone $^{13}C_{\alpha}$, $^{13}C_{\beta}$, ^{1}HN and ^{15}N of ^{15}N, ^{13}C, ^{2}H-TCPTP C216S (1-281).

Residue	σ_{HN} (ppm)	σ_{N} (ppm)	$\sigma_{C\alpha}$ (ppm)	$\sigma_{C\beta}$ (ppm)
ARG-7			58.86	28.96
GLU-8	8.23	122.27	59.16	29.57
PHE-9	8.70	119.41	61.81	39.82
GLU-10	8.00	118.40	59.02	28.76
GLU-11	7.88	121.97	59.30	28.69
ASP-13			57.00	
THR-14	8.51	116.07	65.74	40.61
GLN-15	7.75	117.22	54.94	27.37
ARG-16	7.53	118.28		25.61
PRO-20			59.28	28.67
LEU-21	8.03	120.31	57.70	40.43
TYR-22	8.44	120.59	62.75	38.48
LEU-23	8.19	119.38	57.30	40.11
GLU-24	8.04	121.95	59.26	28.45

ILE-25	7.84	118.72	64.94	36.38
ARG-26	8.06	122.30	59.97	29.01
ASN-27	8.39	118.09	54.98	37.64
GLU-28	7.79	118.30	56.05	29.89
SER-29	7.14	117.83	58.83	35.16
HIS-30	8.18	125.27	55.40	30.98
ASP-31	8.12	121.89	53.06	42.04
TYR-32	6.75	120.00		38.36
PRO-40			65.63	30.87
GLU-41	10.06	117.62	57.66	27.14
ASN-42	7.81	117.37	52.43	38.77
ARG-43	7.40	124.03	59.90	29.00
ASP-50			52.86	39.57
VAL-51	6.96	119.55	62.06	31.95
SER-52	7.83	124.49		36.32
HIS-56			58.55	
SER-57	6.34	109.40	55.52	35.52
ARG-58	7.30	123.54	55.03	27.74
VAL-59	7.19	123.44	61.49	31.30
LYS-60	8.60	128.15	55.10	32.39
LEU-61	9.20	125.26	53.70	40.14
GLN-62	9.06	125.12	55.13	27.45
GLU-65			58.82	28.45
ASN-66	7.34	114.55	51.23	37.76
ASP-67	8.52	125.92	53.12	41.02
TYR-68	8.05	119.13		38.25
ASN-70			52.06	35.35
ALA-71	7.78	129.50	51.35	49.96
SER-72	8.92	117.79	57.14	38.00
LEU-73	8.93	129.63	54.20	40.90
VAL-74	9.29	129.43	61.21	31.55
ASP-75				42.72
ILE-76	8.38	124.21	58.20	35.02
GLU-77	8.75	130.39	59.74	28.68
GLU-78	9.87	119.41	58.83	28.14
ALA-79	7.10	118.26	50.25	47.33
GLN-80	7.82	115.47	56.30	26.39
ARG-81	6.94	116.12	52.17	35.46
SER-82	7.79	117.70	56.54	36.67
TYR-83	8.56	119.21	53.89	
ILE-84	9.03	121.89	59.95	38.41
TRP-102			60.04	30.05
GLN-103	9.17	120.12	58.80	27.43
GLN-104	8.35	113.82	54.36	26.44
LYS-105	7.42	112.41	57.46	27.37

THR-106	7.28	120.10	65.18	42.30
ASN-113			52.43	36.21
ARG-114	7.83	114.97	55.50	30.40
ILE-115	7.45	120.43	52.08	40.61
VAL-116	9.12	126.52	58.66	33.77
MET-134			54.87	35.29
LEU-135	8.18	126.41	53.58	43.20
PHE-136	9.17	124.84		37.22
THR-139			59.19	
GLY-140	7.88	110.37	46.81	
PHE-141	7.36	115.52	56.76	43.44
SER-142	9.06	116.98	55.23	36.57
VAL-143	9.03	125.53	60.35	34.26
ASP-149			52.07	37.94
VAL-150	9.05	128.09	63.09	31.17
LYS-151	7.56	127.76	54.17	31.60
TYR-153			55.50	
TYR-154	6.53	113.52	55.71	40.16
THR-155	9.37	117.23	60.28	45.16
VAL-156	8.58	125.29	60.23	33.04
HIS-157	9.87	126.99	54.57	29.59
GLU-162			53.48	31.85
ASN-163	8.26	124.63	50.81	36.77
ILE-164	7.86	124.24	62.80	36.14
ASN-165	7.62	116.45	55.01	
SER-166	7.20	112.79	57.41	37.20
GLY-167	8.44	113.95	45.31	
GLU-168	7.61	120.99	55.97	32.33
THR-169	8.43	112.29	59.52	43.98
ARG-170	9.29	122.59	55.11	35.26
THR-171	8.64	118.52		40.79
ARG-200			59.13	36.84
GLU-201	8.58	121.66	58.25	28.87
SER-202	7.32	112.93	59.82	36.78
GLY-203	7.91	110.86	45.34	45.34
SER-204			61.44	35.45
LEU-205	8.16	116.56	53.20	39.60
ASN-206	7.31	120.27		37.03
PRO-207			64.10	30.61
ASP-208	8.14	116.35	54.10	38.89
HIS-209	7.52	118.95	54.32	32.49
GLY-210	7.24	131.57	44.05	44.00
PRO-211			62.20	30.75
ALA-212	9.25	127.56	52.12	45.95
VAL-213	7.79	121.44		30.15

ILE-214	7.97	123.92	60.60	37.97
HIS-215	8.39	124.89		
LEU-233			57.60	40.34
VAL-234	7.74	121.97	66.56	30.49
LEU-235	8.06	119.91	57.57	40.14
MET-236			58.02	
GLU-237	8.00	122.29	58.84	28.67
LYS-238	7.58	116.78	56.46	31.77
GLY-239	7.60	107.15	44.71	
ASP-240	7.68	120.90	53.98	40.93
ASP-241	8.33	121.69	53.49	40.02
ILE-242	7.76	118.64	59.77	39.90
ASN-243	8.32	124.19	50.92	38.49
ILE-244	8.94	126.81	66.10	36.42
LYS-245	8.17	119.02	60.16	31.26
GLN-246	7.46	116.97	57.84	27.70
VAL-247	8.28	121.90	66.23	30.69
ARG-255			56.05	
MET-256	8.41	122.60	54.98	31.71
GLY-257	8.70	127.10	46.35	
LEU-258	6.37	113.91	55.22	40.42
ASP-263			52.55	41.71
GLN-264	8.57	127.25	62.46	31.35
LEU-265	8.47	114.83	60.94	42.39
ARG-266	8.65	116.60	65.56	40.10
PHE-267	7.97	120.61		
ILE-273			65.16	37.40
GLU-274	7.85	117.85	58.29	28.26
GLY-275	8.76	110.71	47.16	
ALA-276	8.29	124.53	53.85	45.37
LYS-277	7.11	115.53	57.37	31.76
CYS-278	7.60	115.85	58.91	27.76
ILE-279	7.25	118.94	61.42	37.49
LYS-280	7.57	129.70		

Analysis of the second datasets did produce many long fragments as Figure 35 indicates. Many of the short fragments can be assigned at more than one position in the sequence. This is due to the nature of the method. For example, the fragment that was assigned to the tripeptides E65, N66, D67 could also be assigned to tripeptide L193, N194, F195. The first possibility was chosen based on a better chemical shift match for

the C_β of E65. But the values used for this assignment are random coil chemical shifts (Table 4). Possible chemical shift values for a leucine C_β in a protein structure can be as low as 36.3 and for a glutamate C_β as high as 36.5. In other words, the sign of the correlation must match perfectly while the chemical shifts are used as a guide for the assignment. This reiterates the necessity of forming long fragments from matched clusters which can only be accomplished with better (higher intensity) C_β data. In an attempt to understand why it is so difficult to obtain good C_β data, it may be useful to analyze more closely where the long fragments were assigned to the primary sequence and try to correlate this to secondary structure elements. Since no structure is available for TCPTP, the crystal structure of PTP-1B and the sequence alignment of PTP-1B and TCPTP can be used.

Closer analysis of Figure 34 can help understand the difficulty of obtaining C_β data that would lead to a complete backbone assignment. Again, the problem is the measurement of correlations that connect clusters through C_β chemical shifts. It is interesting to note that many correlations, in the C_β experiments, have a high intensity but too many are either very weak or not detected. This suggests that some portion of the protein may behave differently than others. In the absence of any structural information for TCPTP, its high homology with PTP-1B could bring some insights to the problem. Using the sequence alignment (Figure 4) between TCPTP and PTP-1B, and translating the secondary structure elements from the X-ray crystal structure of PTP-1B onto the primary sequence of TCPTP, an interesting observation can be made (Figure 36). Most of the long fragments are assigned to portions of the sequences that are potentially constituting secondary structure elements. Consequently, loops or random coil regions

have fewer assignments made. This observation suggests that conformational heterogeneity or motion of the loops may perturb the chemical shift through broadening which complicate their detection. Also, solvent exposed loops will see their amide protons in faster exchange with the bulk water. Since all experiments rely on amide detection, chemical exchange will also attenuate the NMR signal. This effect increases with pH. All experiments were carried out at pH 7.5 for protein stability reasons.

Figure 36. Secondary structure elements translated from the X-ray crystal structure of PTP-1B onto the TCPTP C216S (1-281) primary sequence by alignment.

Cylinders indicate α-helix and arrows indicate β-strand. Amino acids successfully assigned are represented in red. The catalytic loop is represented in blue. The asterix is to note that the catalytic cysteine is actually mutated to a serine.

In order to test if loop mobility is responsible for missing correlations, an inhibitor was added to the protein in the hope to lock loops around the catalytic site and measure

more correlations, especially in the catalytic loop (see Figure 35). Unfortunately, precipitation of the protein was observed upon addition of the inhibitor. Because it seemed difficult to try stopping movement, increasing it was also tested. NMR studies of PTP-1B demonstrated that increasing the buffer concentration from 50 to 100 mM sodium phosphate allowed for slightly more stable samples. Therefore, the triply labeled TCPTP sample was transferred into 100 mM sodium phosphate, pH 7.5 and the HN(COCA)CB experiment was recorded. A small gain in sensitivity was observed while not sufficient to improve the assignment, it indicated that samples stable at higher temperature (35–40 °C) could yield better NMR data for backbone resonance assignment.

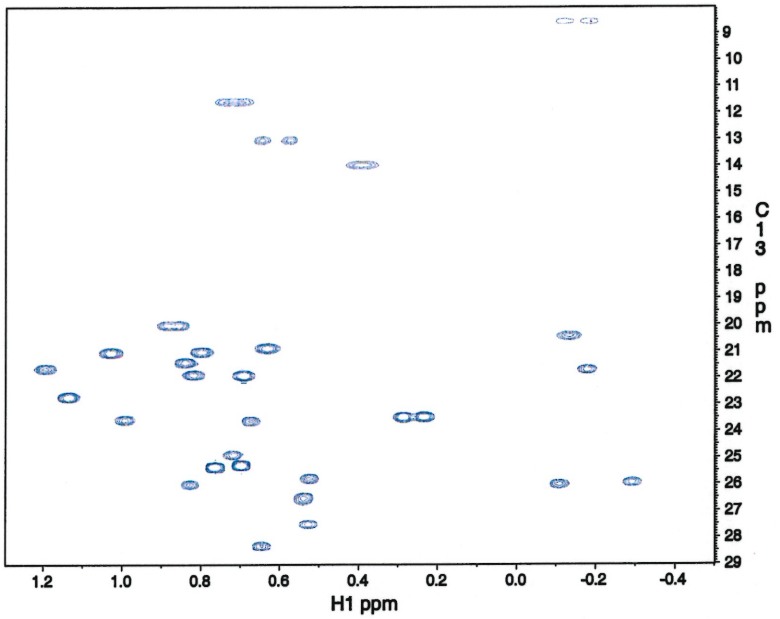
Stability of the TCPTP sample at higher temperatures could be obtained using a strategy that was applied to PTP-1B. This strategy consists in literally attaching the amino and carboxy terminal ends of the protein to a nearby loop with disulfide bridges. The bridges are spontaneously formed upon oxidation of cysteine residues that were introduced by site directed mutagenesis. This approach increased the unfolding temperature by 15 °C as measured by FTIR (Aubin, personal communication). The gain in stability was also observed by NMR where many new correlations were detected when spectra were recorded at 30 °C. The absence of structural data for TCPTP would complicate the introduction of cysteine residues at key locations for disulfide bridge formation. Trial and error based on an homology model could be attempted to build the mutant. Binding assay of known inhibitors could then be used to insure that the binding surface has maintained its wild type conformation. This strategy applied to TCPTP could certainly allow the data collection at temperatures higher than 30 °C.

II. Methyl protonation labeling scheme

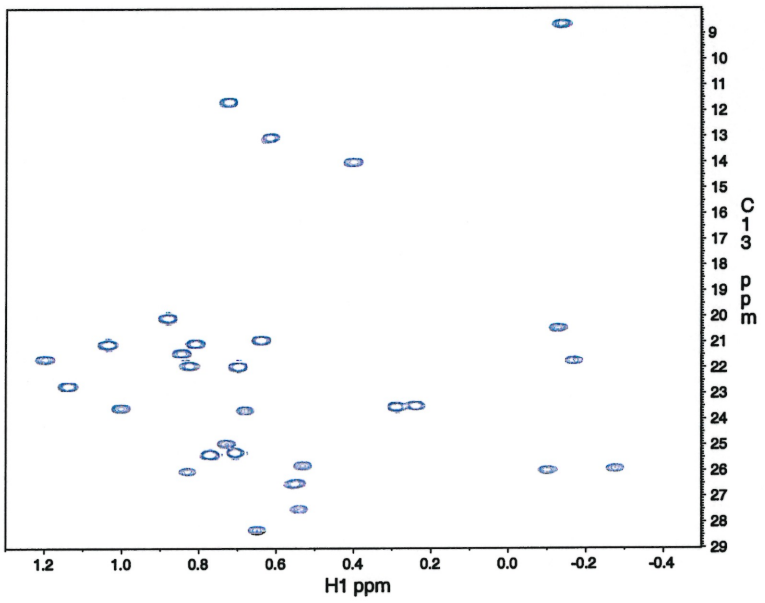
The next step, after performing the resonance assignment of ^{15}N, ^{13}C, ^{2}H-TCPTP C216S (1-281), consists in measuring structural parameters for structure calculation. For a ^{13}C, ^{2}H, ^{15}N-labeled protein all but the exchangeable amide protons are replaced by deuterons, such that only NH-NH distances could be measured. Gardner and Kay (1997) showed that structure calculation based only on these do not provide a sufficient amount of distance constraints to calculate well folded 3D structure. For this reason they have developed a labeling strategy where protons are selectively re-introduced to the methyl groups of Val, Leu and Ile (δ) side chains. Besides the increased amount of interproton distances afforded by the methyl labeling scheme, three more amino acid type can then be unambiguously identified: Val, Leu and Ile. This new type of information can be used to help the resonance assignment of the backbone atoms because the methyl group signal can be correlated to the backbone amide through the network of side chain carbon-13 coupling constants.

Samples labeled in this fashion allow the measurement of NH-NH as well as additional NH-CH_3 and CH_3-CH_3 distance constraints for structural calculations. The 2D ^{1}H-^{13}C HSQC spectra of CH_3(Val, Leu, Ile-δ)-^{15}N, ^{13}C, ^{2}H-TCPTP C216S (1-281) is shown in Figure 31. The 2D spectra confirmed that the methyl side chains of Val, Leu and Ile(δ) were successfully protonated however upon careful analysis of the 2D ^{13}C-^{1}H HSQC spectrum of CH_3(Val, Leu, Ile-δ)-^{15}N, ^{13}C, ^{2}H-TCPTP C216S (1-281) a few NMR signals appeared as doublets.

These doublets were attributed to slow rotation of the isopropyl moieties of Val and Leu side chains and the ethyl moiety of the Ile(δ) side chain. This behavior poses a problem because it decreases both resolution and intensity. To confirm this hypothesis, parallel studies with a known protein stable in the 25 °C to 30 °C range was performed. For this study the carboxy-terminal SH2 domain from phospholipase-Cγ (PLCCSH2) was methyl protonated at the Val, Leu, Ile-δ side chains using the same methyl protonating labeling scheme. This protein is very stable at 30 °C and a 2D ^{1}H-^{13}C HSQC spectrum of CH_3(Val, Leu, Ile-δ)-^{15}N, ^{13}C, ^{2}H-PLCC-SH2 was obtained at both 25 °C and 30 °C. The 2D ^{1}H-^{13}C HSQC spectra of CH_3(Val, Leu, Ile-δ)-^{15}N, ^{13}C, ^{2}H-PLCC-SH2 at 25 °C and 30 °C are shown in Figure 36. As anticipated, doublets or broadening of signals was observed at 25 °C, which confirmed that at the lower temperature some of the isopropyl and ethyl moieties were rotating slowly enough to observe different chemical shifts from the different rotational conformers. However, at 30 °C faster rotation was promoted and one average conformer is observed such that only one sharp signal is detected. The results from the PLCC-SH2 study suggest that the doublets and the broad signals observed in the 2D ^{1}H-^{13}C HSQC spectrum of CH_3(Val, Leu, Ile-δ)-^{15}N, ^{13}C, ^{2}H-TCPTP C216S (1-281) are due to the observation of rotamers. They also indicated that the temperature at which the NMR experiments were acquired was not optimal. Interestingly, these observations strengthen the argument of loop mobility occurring in a motional regime that degrades spectroscopic parameters (linewidth, intensity). This reiterates the need to acquire NMR experiments of ^{15}N, ^{13}C, ^{2}H-TCPTP C216S (1-281) at higher temperatures in order to improve the quality of the NMR data.



(a)



(b)

Figure 37. 2D ^{1}H-^{13}C of 1 mM CH_3(Val, Leu, Ile-δ)-^{15}N, ^{13}C, ^{2}H-PLCγ-SH2 in 100 mM NaH_2PO_4 pH 6.5, 5 mM DTT-d_{10}, 2 mM Pefabloc and 5% D_2O at (a) 25 °C and (b) 30 °C.

III. Conclusion

The work presented in this thesis is the first attempt at studying the catalytic domain of TCPTP using NMR techniques. This project is very challenging due to the large size of the protein. Optimization of both the protein expression system and protein purification protocol was necessary in order to generate high levels of pure ^{13}C, ^{2}H, ^{15}N-TCPTP C216S (1-281) amenable to NMR structural studies. Triple resonance (^{13}C, ^{1}HN, ^{15}N) multidimensional NMR experiments with deuterium decoupling were acquired. The data analysis allowed the assignment of ~50% of the polypeptide backbone resonances. The partial assignment was due to the difficulty in obtaining C_β data necessary for unambiguous connections between residues in the sequence. The results suggest that acquisition of NMR experiments at higher temperatures such as 30 °C – 37 °C are required to improve NMR signals for complete assignment. This was further confirmed with the results obtained using the methyl-protonated (Val, Leu, Ile-δ) triply labeled sample of TCPTP C216S (1-281). The success of the methyl-protonation scheme permits the measurement of NOE-derived interproton distances required later for 3D structural calculations. To improve the stability of TCPTP, FTIR may be used to optimize NMR sample buffer conditions. Furthermore, protein engineering techniques such as the introduction of disulfide bridges as a means to immobilize the loops may be employed.

Already, the partial assignment can be used to study the interaction of small molecules in the binding surface. Based on the perturbation of the chemical shifts induced upon binding a better understanding of where the small molecules are interacting

may be obtained. Such studies performed in parallel on PTP-1B can contribute to the design of selective inhibitors.

References

Ahmad, F., Li, P-M., Meyerovitch, J., and Goldstein, B. J. 1995. Osmotic loading of neutralizing antibodies demonstrates a role for protein-tyrosine phosphatase 1B in negative regulation of the insulin action pathway. The Journal of Biological Chemistry, **270**: 20503-20508.

Bandyopadhyay, D., Kusari, A., Kenner, K. A., Liu, F., Chernoff, J., Gustafson, T. A. and Kusari, J. 1997. Protein-tyrosine phosphatase 1B complexes with the insulin receptor *in vivo* and is tyrosine-phosphorylated in the presence of insulin. The Journal of Biological Chemistry, **272**: 1639-1645.

Barford, D., Flint, A. J. and Tonks, N. K. 1994. Crystal structure of human protein tyrosine phosphatase 1B. Science, **263**: 1397-1403.

Barford, D., Jia, Z., and Tonks, N. K. 1995. Protein tyrosine phosphatases take off. Nature Structural Biology, **2**: 1043-1053.

Bodenhausen, G. and Ruben, D. J. 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chem. Phys. Lett., **69**: 185-189.

Brown-Shimer, S., Johnson, K. A., Lawrence, J. B., Johnson, C. and Bruskin, A. 1990. Molecular cloning and chromosome mapping of the human gene encoding protein phosphotyrosyl phosphatase 1B. Proc. Natl. Acad. Sci. U.S.A., **87**: 5148-5152.

Cavanagh, J., Fairbrother, W. J., Palmer III, A. G. and Skelton, N. J. 1996. Protein NMR Spectroscopy, Academic Press, San Diego.

Chernoff, J., Schievella, A. R., Jost, C. A., Erikson, R. L. and Neel, B. G. 1990. Cloning of a cDNA for a major human protein-tyrosine-phosphatase. Proc. Natl. Acad. Sci. U.S.A., **87**: 2735-2739.

Cicirelli, M.F., Tonks, N. K., Diltz, C. D., Weiel, J. E., Fischer, E. H., and Krebs, E. G. 1990. Microinjection of a protein-tyrosine-phosphatase inhibits insulin action in Xenopus oocytes. Proc. Natl. Acad. Sci. U.S.A., **87**: 5514-5518.

Clore, G. M. and Gronenborn A. M. 1997. NMR structures of proteins and protein complexes beyond 20,000 M_r. Nat. Struc. Biol., NMR supplement: 849-853.

Cool, D. E., Tonks, N. K., Charbonneau, H., Walsh, K. A., Fischer, E. H., and Krebs, E. G. 1989. cDNA isolated from a human T-cell library encodes a member of the protein-tyrosine-phosphatase family. Proc. Natl. Acad. Sci. U.S.A., **86**: 5257-5261.

Cool, D. E., Tonks, N. K., Charbonneau, H., Fischer, E. H., and Krebs, E. G. 1990. Expression of a human T-cell protein-tyrosine-phosphatase in baby hamster kidney cells. Proc. Natl. Acad. Sci. U.S.A., **87**: 7280-7284.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., Bax, A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR, **6:** 277-293.

Dyson, H. J., Gippert, G. P., Case, D. A., Holmgren, A., and Wright, P. E. 1990. Three-dimensional solution structure of the reduced form of Escherichia coli thioredoxin determined by nuclear magnetic resonance spectroscopy. J. Mol. Biol., **29**: 4129-36.

Elchebly, M., Payette, P., Michaliszyn, E., Cromlish, W., Collins, S., Loy, A. L., Normandin, D., Cheng, A., Himms-Hagen, J., Chan, C-C., Ramachandran, C., Gresser, M. J., Tremblay, M. L. and Kennedy, B. P. 1999. Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. Science, **283**: 1544-1548.

Evans, Jeremy N. S. 1995. Biomolecular NMR Spectroscopy, Oxford University Press, Oxford.

Flint, A. J., Tiganis, T., Barford, D. and Tonks, N. K. 1997. Development of "substrate-trapping" mutants to identify physiological substrates of protein tyrosine phosphatases. Proc. Natl. Acad. Sci. U.S.A., **94**: 1680-1685.

Forman-Kay, J. D., Clore, G. M., Wingfield, P. T., and Gronenborn, A. M. 1991. High-resolution three-dimensional structure of reduced recombinant human thioredoxin in solution. Biochemistry, **30**: 2685-98.

Gardner, K. H., Konrat, R., Rosen, M. K. and Kay, L. E. 1996. An (H)C(CO)NH-TOCSY pulse scheme for sequential assignment of protonated methyl groups in otherwise deuterated ^{15}N,^{13}C-labeled proteins. J. Biomol. NMR, **8**: 351-356.

Gardner, K. H. and Kay, L. E. 1997. Production and incorporation of ^{15}N, ^{13}C, ^{2}H (^{1}H-δ1 methyl) isoleucine into proteins for multidimensional NMR studies. J. Am. Chem. Soc., **119**: 7599-7600.

Goldstein, B. J. 1993. Regulation of insulin receptor signaling by protein-tyrosine dephosphorylation. Receptor, **3**: 1-15.

Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C. and Kay, L. E. 1999. A robust and cost-effective method for the production of Val, Leu, Ile (δ1) methyl-protonated ^{15}N-, ^{13}C-, ^{2}H-labeled proteins. J. Biomol. NMR, **13**: 369-374.

Guan, K. L., Haun, R. S., Watson, S. J., Geahlen, R. L. and Dixon, J. E. 1990. Cloning and expression of a protein-tyrosine-phosphatase. Proc. Natl. Acad. Sci. U.S.A., **87**: 1501-1505.

Guéron, M., Leroy, J. L. and Griffey, R. H. 1983. Proton nuclear magnetic relaxation of ^{15}N-labeled nucleic acids via dipolar coupling and chemical shift anisotropy. J. Am. Chem. Soc., **105**: 7262-7266.

Hao, L., Tiganis, T., Tonks, N. K. and Charbonneau, H. 1997. The noncatalytic C-terminal segment of the T cell protein tyrosine phosphatase regulates activity via an intramolecular mechanism. J. Biol. Chem., **272**: 29322-29329.

Hubbard, S. R., Wei, L., Ellis, L. and Hendrickson, W. A. 1994. Crystal structure of the tyrosine kinase domain of the human insulin receptor. Nature, **372**: 748-754.

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

Kahn, B. B. 1998. Type 2 diabetes: When insulin secretion fails to compensate for insulin resistance. Cell, **92**: 593-596.

Kainosho, M. 1997. Isotope labelling of macromolecules for structural determinations. Nat. Struc. Biol., NMR supplement: 858-861.

Kamatkar, S., Radha, V., Nambirajan, S., Reddy, R. S., and Swarup, G. 1996. Two splice variants of a tyrosine phosphatase differ in substrate specificity, DNA binding, and subcellular location. J. Biol. Chem., **271**: 26755-26761.

Kay, L. E., Ikura, M., Tschudin, R. and Bax, A. 1990. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J. Magn. Reson., **89**: 496-514.

Kay, L. E. 1997. NMR methods for the study of protein structure and dynamics. Biochemistry and Cell Biology, **75**: 1-15.

Kay, L. E. and Gardner, K. H. 1997. Solution NMR spectroscopy beyond 25 kDa. Current Opinion in Structural Biology, **7**: 722-731.

Kenner, K. A., Hill, D. E., Olefsky, J. M. and Kusari, J. 1993. Regulation of protein tyrosine phosphatases by insulin and insulin-like growth factor I. J. Biol. Chem., **268**: 25455-25462.

Kenner, K. A., Anyanwu, E., Olefsky, J. M. and Kusari, J. 1996. Protein-tyrosine phosphatase 1B is a negative regulator of insulin- and insulin-like growth factor-I-stimulated signaling. J. Biol. Chem., **271**: 19810-19816.

Lorenzen, J. A., Dadabay, C. Y. and Fischer, E. H. 1995. COOH-terminal sequence motifs target the T cell protein tyrosine phosphatase to the ER and nucleus. J. Cell Biol., **131**: 631-643.

Marion, D., Ikura, M. and Bax, A. 1989. Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time-domain data. J. Magn. Reson., **84**: 425-430.

Markus M. A., Dayie, K. T., Matsudaira, P. and Wagner, G. 1994. Effect of deuteration on the amide proton relaxation rates in proteins. Heteronuclear NMR experiments on Villin 14T. J. Magn. Reson., Series B, **105**: 192-195.

Müller, L. 1979. Sensitivity enhanced detection of weak nuclei using heteronuclear multiple quantum coherence. J. Am. Chem. Soc., **101**: 4481-4484.

Novagen. 1997. pET Sytem Manual, 7th Ed., Novagen Inc., Madison.

Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. 1997. Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc. Natl. Acad. Sci. U.S.A., **94**: 12366-12371.

Pervushin, K. V., Wider, G. and Wüthrich, K. 1998. Single transition-to-single transition polarization transfer (ST2-PT) in [^{15}N,^{1}H]-TROSY. J. Biomol. NMR, **12**: 345-348.

Puius, Y. A., Zhao, Y., Sullivan, M., Lawrence, D. S., Almo, S. C. and Zhang, Z.Y. 1997. Identification of a second aryl phosphate-binding site in protein-tyrosine phosphatase 1B: a paradigm for inhibitor design. Proc. Natl. Acad. Sci. U.S.A., **94**: 13420.

Richarz, R. and Wüthrich, K. 1978. Carbon-13 NMR Chemical Shifts of the Common Amino Acid Residues Measured in Aqueous Solutions of the Linear Tetrapeptides H-Gly-Gly-X-L-Ala-OH. Biopolymers, **17**: 2133-2141.

Sambrook, J., Fritsch, E. F., and Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Santoro, J. and King, G. C. 1992. A constant-time 2D Overbodenhausen experiment for inverse correlation of isotopically enriched species. J. Magn. Reson., **97**: 202-207.

Shan, X., Gardner, K. H., Muhandiram, D. R., Rao, N. S., Arrowsmith, C. H., and Kay, L. E. 1996. Assignment of ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and HN resonances in an ^{15}N,^{13}C,^{2}H labeled 64 kDa trp repressor-operator complex using triple-resonance NMR spectroscopy and ^{2}H-decoupling. J. Am. Chem. Soc., **118**: 6570-6579.

Shuker, S. B., Hajduk, P. J., Meadows, R. P. and Fesik S. W. 1996. Discovering high-affinity ligands for proteins: SAR by NMR. Science, **274**: 1531-1534.

Sørensen, O. W., Eich, G. W., Levitt, M. H., Bodenhausen, G. and Ernst, R. R. 1983. Product operator formalism for the description of NMR pulse experiments. Progress in NMR Spectroscopy, **16**: 163-192.

Taylor, S. I. 1999. Deconstructing type 2 diabetes. Cell, **97**: 9-12.

Tiganis, T., Flint, A. J., Adam, S. A., and Tonks, N. K. 1997. Association of the T-cell protein tyrosine phosphatase with nuclear import factor p97. J. Biol. Chem., **272**: 21548-21557.

Tiganis, T., Bennett, A. M., Ravichadran, K. S. and Tonks, N. K. 1998. Epidermal growth factor receptor and the adaptor protein $p52^{Shc}$ are specific substrates of T-cell protein tyrosine phosphatase. Mol. Cell. Biol., **18**: 1622-1634.

Tonks, N.K., Diltz, C. D. and Fischer, E. H. 1988*a*. Purification of the major protein-tyrosine-phosphatases of human placenta. J. Biol. Chem., **263**: 6722-6730.

Tonks, N.K., Diltz, C. D. and Fischer, E. H. 1988*b*. Characterization of the major protein-tyrosine-phosphatases of human placenta. J. Biol. Chem., **263**: 6731-6737.

Tonks, N. K., Cicirelli, M. F., Diltz, C. D., Krebs, E. G., and Fischer, E. H. 1990. Effect of microinjection of a low-Mr human placenta protein tyrosine phosphatase on induction of meiotic cell division in Xenopus oocytes. Mol. Cell. Biol. **10**: 458-463.

Tonks, N. K., and Neel, B. G. 1996. From form to function: signaling by protein tyrosine phosphatases. Cell, **87**: 365-368.

Tjandra, N., Szabo, A. and Bax, A. 1996. Protein backbone dynamics and ^{15}N chemical shift anisotropy from quantitative measurement of relaxation interference effects. J. Am. Chem. Soc., **118**: 6986-6991.

Vuister, G. W. and Bax, A. 1992. Resolution enhancement and spectral editing of uniformly carbon-13-enriched proteins by homonuclear broadband carbon-13 decoupling. J. Magn. Reson., **98**: 428-435.

Wagner, G. 1997. An account of NMR in structural biology. Nat. Struc. Biol., NMR supplement: 841-844.

Wälchli, S., Curchod, M-L., Gobert, R. P., Arkinstall, S. and van Huijsduijnen, R. H. 2000. Identification of tyrosine phosphatases that dephosphorylate the insulin receptor. J. Biol. Chem., **275**: 9792-9796.

White, M. F., and Kahn, C. R. 1994. The insulin signaling system. J. Biol. Chem., **269**: 1-4.

Wishart, D. S., Sykes, B. D. and Richards, F. M. 1991*a*. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J. Mol. Biol., **222**: 311-333.

Wishart, D. S., Sykes, B. D. and Richards, F. M. 1991*b*. Simple techniques for the quantification of protein secondary structure by 1H NMR spectroscopy. FEBS Lett., **293**: 72-80.

Wüthrich, K. 1986. NMR of Proteins and Nucleic Acids, John Wiley & Sons, New York.

Wüthrich, K. 1998. The second decade – into the third millenium. Nat. Struc. Biol., NMR Supplement: 492-495.

Yamazaki, T., Lee, W., Revington, M., Mattiello, D. L., Dahlquist, F. W., Arrowsmith, C. H. and Kay L. E. 1994*a*. An HNCA pulse scheme for the backbone assignment of ^{15}N,^{13}C,^{2}H-labeled proteins: Application to a 37-kDa *Trp* repressor-DNA complex. J. Am Chem. Soc. **116**: 6464-6465.

Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. and Kay L. E. 1994*b*. A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N, ^{13}C, ^{2}H labeled proteins with high sensitivity. J. Am Chem. Soc. **116**: 11655-11666.

Yang, D. and Kay, L. E. 1999. Improved ^{1}HN-detected triple resonance TROSY-based experiments. J. Biomol. NMR, **13**: 3-9.

You-Ten, K. E., Muise, E. S., Itié, A., Michaliszyn, E., Wagner, J., Jothy, S., Lapp, W. S., and Tremblay, M. L. 1997. Impaired bone marrow microenvironment and immune function in T cell protein tyrosine phosphatase-deficient mice. J. Exp. Med., **186**: 683-693.

Zhang, Zhong-Yin. 1988. Protein-tyrosine phosphatases: biological function, structural characteristics, and mechanism of catalysis, Crit. Rev. Biochem. Mol. Biol., **33**: 1-52.

Zhu, G. and Bax, A. 1990. Improved linear prediction for truncated signals of known phase. J. Magn. Reson., **90**: 405.

Appendix A

Product operator formalism

The basis of product operator formalism is that interactions which the macroscopic magnetization vector under the influence of pulse, chemical shift or scalar coupling Hamiltonians experience is easily described using Cartesian product operators. At equilibrium the bulk magnetization, M_z, may be described as a density matrix operator along the z-axis, I_z. Under the influence of a β pulse from the y-direction (βI_y), I_z, as well as the x-magnetization (I_x) and y-magnetization (I_y) transform according to:

$$I_z \xrightarrow{\beta I_y} I_z \cos\beta + I_x \sin\beta$$

$$I_x \xrightarrow{\beta I_y} I_x \cos\beta - I_z \sin\beta$$

$$I_y \xrightarrow{\beta I_y} I_y$$

The starting state of the spin system is on the left-hand side of the equation and the resulting state of the spin system under the influence of the specific Hamiltonian is described on the right-hand side. The Hamiltonian operator terms above the arrow describe interactions of the spins with external fields as well as with each other and thus the influence under which the system evolves. The Hamiltonian operator which describes how the system evolves under the influence of a pulse, chemical shift and scalar coupling are described by $\beta I_{x,y}$, $\Omega I_z t$ and $2\pi J I_z S_z t$ respectively, where,

(a) $\beta I_{x,y}$ represents a pulse with a flip angle β about an axis in the x, y plane, which forms an angle ϕ with the x-axis,

(b) $\Omega I_z t$ represents the evolution of chemical shift Ω of nucleus I during time t, and

(c) $\pi J 2 I_z S_z t$ represents the evolution of coupling J between nuclei I and S during time t.

Therefore, for a magnetization in the transverse plane, precessing under the influence of chemical shift evolution, the product operator is described as (refer to Figure 38),

$$I_x \xrightarrow{\Omega t I_z} I_x \cos \Omega t + I_y \sin \Omega t$$

$$I_y \xrightarrow{\Omega t I_z} I_y \cos \Omega t - I_x \sin \Omega t$$

$$I_z \xrightarrow{\Omega t I_z} I_z$$

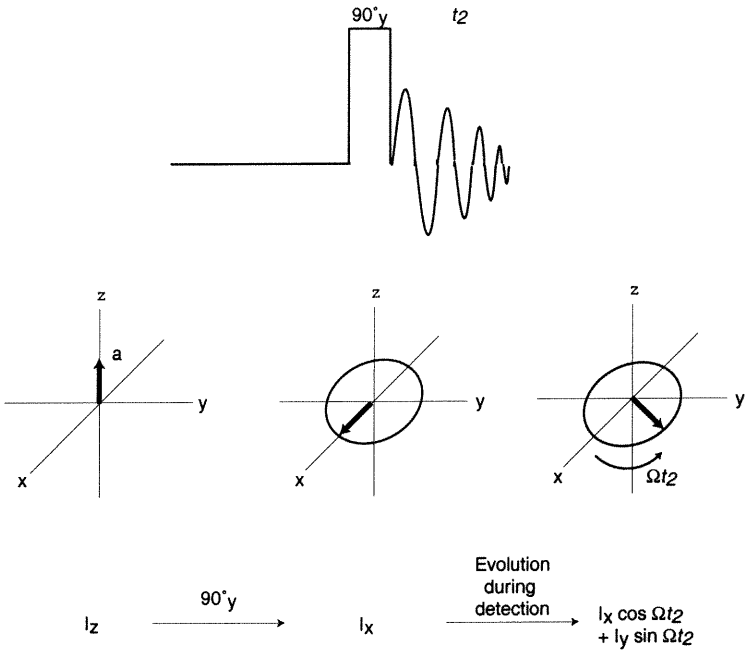


Figure 38. Evolution of magnetization using the vector model and product operator formalism.

In terms of product operator formalism spin-spin scalar coupling between two nuclei I and S may be described according to,

$$I_x \xrightarrow{\pi J 2 I_z S_z t} I_x \cos \pi J_{IS} t + 2 I_y S_z \sin \pi J_{IS} t$$

$$I_y \xrightarrow{\pi J 2 I_z S_z t} I_y \cos \pi J_{IS} t - 2 I_x S_z \sin \pi J_{IS} t$$

$$2 I_x S_z \xrightarrow{\pi J 2 I_z S_z t} I_x S_z \cos \pi J_{IS} t + I_y \sin \pi J_{IS} t$$

$$2 I_y S_z \xrightarrow{\pi J 2 I_z S_z t} I_y S_z \cos \pi J_{IS} t - I_x \sin \pi J_{IS} t$$

Using the vector model, the spin-spin scalar coupling between nuclei I and S may be viewed as two vectors for spin I, precessing at different angular velocities corresponding to either the α state or the β state of spin S. As illustrated in Figure 39 the *x*-magnetization of spin I is modulated by cos $\pi J_{IS}t$ and the *y*-magnetization of spin I is modulated by sin $\pi J_{IS}t$, where the *y*-magnetization for spin I points to +y for those I spins which observe the S spin in the α state and similarly spin I points to –y for those I spins which observe the S spin in the β state. The I_y magnetization may be described as $I_y(S^{\alpha}-S^{\beta})$ however this anti-phase magnetization is not observable since the positive and negative components cancel exactly. The difference in populations of the α and β levels $(S^{\alpha}-S^{\beta})$ is given by $2S_z$ therefore $I_y(S^{\alpha}-S^{\beta})$ is described as $2I_yS_z$. As a result, precession of scalar-coupled nuclei is described by both the in-phase and anti-phase magnetization. For a more complete explanation of the product operator formalism please refer to Sørensen *et al.* (1983).

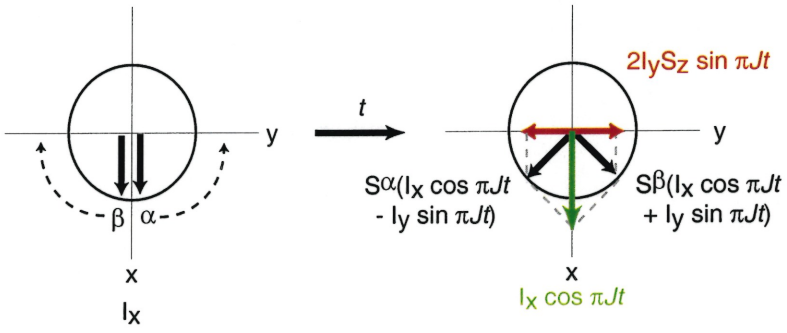


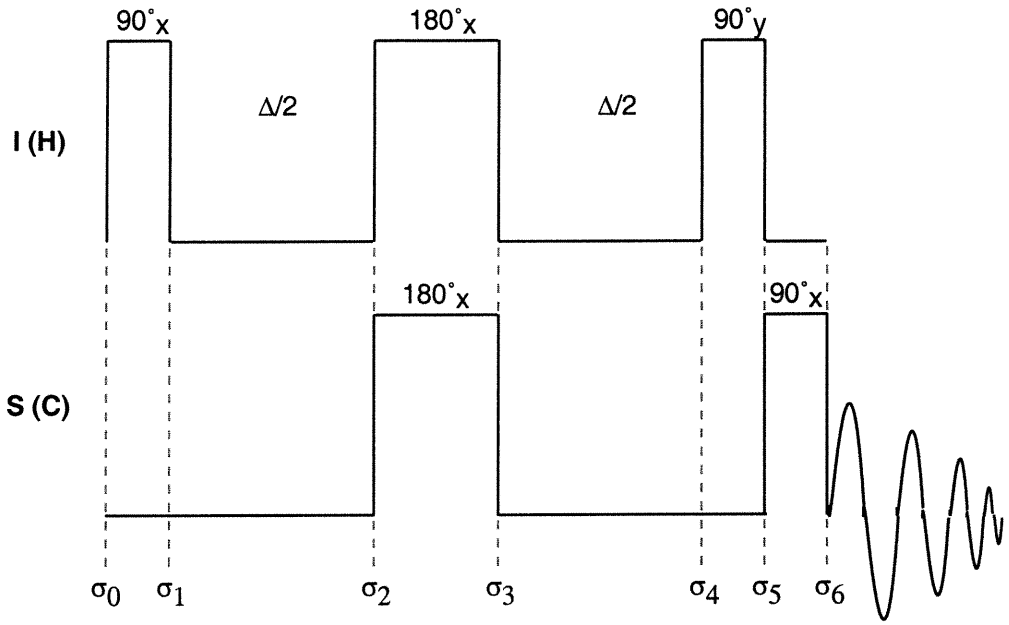
Figure 39. Evolution of scalar coupling using the vector model and product operator formalism.

Appendix B

Insensitive Nuclei Enhanced by Polarization Transfer (INEPT)

The INEPT (Insensitive Nuclei Enhanced by Polarisation Transfer) sequence was developed to enable low-γ nuclei, such as ^{13}C, to be enhanced through polarization transfer from a high-γ nuclei such as ^{1}H. If a heteronuclear couple such as ^{1}H-^{13}C is excited with a 90°(H) pulse to bring the proton magnetization onto the *x-y* transverse plane one would observe a ^{1}H doublet NMR signal. The two halves of the doublet arise from the magnetization of the proton spin being coupled to its partner, ^{13}C, which may be in one of two orientations (α or β). Similarly, if the carbon magnetization can be brought into the transverse plane for detection two halves of a ^{13}C doublet NMR signal would also be observed. The only difference is that the proton doublet intensity is 4x greater than that of the carbon doublet since $\gamma_{1H} = \frac{1}{4}\gamma_{13C}$. Through scalar couplings shared between heteronuclear partners the INEPT permits the high energy sink of ^{1}H to be transferred to the lower-γ nuclei such as ^{13}C or ^{15}N resulting in higher NMR sensitivity of these nuclei. The INEPT pulse sequence is illustrated in Figure 40 with the corresponding product operator formalism and NMR vector model to represent the state of the spin system as it proceeds through the NMR pulse sequence.

a)



b) $\sigma_0 \approx I_z + (\gamma_s/\gamma_I)S_z$

$\downarrow 90°_x$ (H)

$\sigma_1 \approx -I_y + (\gamma_s/\gamma_I)S_z$

$\downarrow \pi J_{IS}\Delta/2$

$\sigma_2 \approx -I_y \cos \pi J_{IS}\Delta/2 + 2I_xS_z \sin\pi J_{IS}\Delta/2 + (\gamma_s/\gamma_I)S_z$

$\downarrow 180°_x$ (H)

$\downarrow 180°_x$ (c)

$\sigma_3 \approx I_y \cos \pi J_{IS}\Delta/2 - 2I_xS_z \sin\pi J_{IS}\Delta/2 - (\gamma_s/\gamma_I)S_z$

$\downarrow \pi J_{IS}\Delta/2$

$\sigma_4 \approx I_y \cos \pi J_{IS}\Delta/2 \cos\pi J_{IS}\Delta/2 - 2I_xS_z \cos \pi J_{IS}\Delta \sin\pi J_{IS}\Delta/2$
$- 2I_xS_z \sin \pi J_{IS}\Delta/2 \cos\pi J_{IS}\Delta/2 - I_y \sin \pi J_{IS}\Delta/2 \sin\pi J_{IS}\Delta/2$
$- (\gamma_s/\gamma_I)S_z$

$\approx I_y (\cos^2 \pi J_{IS}\Delta/2 - \sin^2 \pi J_{IS}\Delta/2)$
$- 2I_xS_z (2\cos \pi J_{IS}\Delta/2 \sin \pi J_{IS}\Delta/2)$
$- (\gamma_s/\gamma_I)S_z$

since $\cos AB = \cos^2 A - \sin^2 B$ and $\sin AB = 2 \cos A \sin B$

$\approx I_y (\cos 2\pi J_{IS}\Delta/2)$
$- 2I_xS_z (\sin 2\pi J_{IS}\Delta/2)$
$- (\gamma_s/\gamma_I)S_z$

since $\Delta = 1/2J_{IS}$

$\approx - 2I_xS_z - (\gamma_s/\gamma_I)S_z$

$\downarrow 90°_{+/-y}$ (H)

$\sigma_5 \approx - 2I_zS_z - (\gamma_s/\gamma_I)S_z$ (H-$90°_y$) or $- 2I_zS_z - (\gamma_s/\gamma_I)S_z$ (H-$90°_{-y}$)

$\downarrow 90°_x$ (C)

$\sigma_6 \approx - 2I_zS_y + (\gamma_s/\gamma_I)S_z$ or $2I_zS_y + (\gamma_s/\gamma_I)S_z$

c)

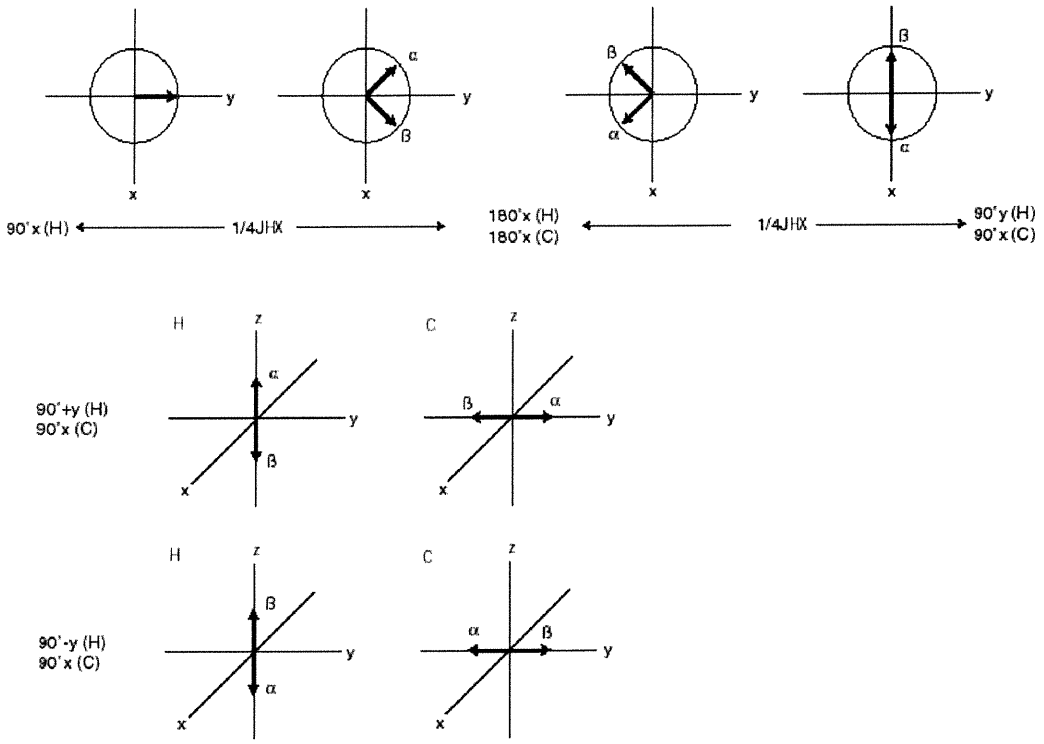


Figure 40. a) INEPT pulse sequence, b) INEPT described by product operator formalism and, c) evolution of proton vectors during the INEPT pulse sequence.

Consider spin-½ scalar coupled pair such as ^{1}H-^{13}C represented by I and S respectively. The INEPT pulse sequence is characterized by an initial $90°_x$ excitation of the high-γ ^{1}H spin. This pulse excites all the protons such that the proton magnetization is brought into the transverse plane (σ_1) evolving under the influence of chemical shift as well as heteronuclear coupling to the ^{13}C spin during the $\Delta/2$ period (σ_2). Recall that the in-phase magnetization of ^{1}H scalar coupled to ^{13}C is represented by I_y product operator however the ^{1}H anti-phase magnetization is represented by $2I_xS_z$ product operator. After this $\Delta/2$ period, a $180°_x$ (H) pulse is applied to refocus chemical shift evolution of the

proton magnetization during the second Δ/2 period rendering ^{1}H chemical shift negligible. However, a simultaneous $180°_{x}$ (C) pulse ensures that the proton magnetization continues to evolve under the influence of heteronuclear coupling to ^{13}C (σ_3). The reason is that the $180°_{x}$ (C) pulse inverts the ^{13}C magnetization from +z-axis to -z-axis such the proton vector's sense of precession is inverted. The faster moving vector of the ^{1}H spin that was previously coupled to $^{13}C_{\beta}$ is now the slower moving vector and is coupled to $^{13}C_{\alpha}$ and vice versa. The two halves of the doublet therefore continue to dephase from each other such that the *J*-coupling evolves during the second Δ/2 period. This is more obvious with the use of product operator where in σ_4 the individual terms from σ_3 each undergo heteronuclear coupling evolution. After a total period of Δ of 1/2J the two proton vectors are antiphase along ±x, as described at the end of σ_4, such that the subsequent $90°_{y}$ (H) pulse aligns these vectors along the ±z-axis (σ_5). This corresponds to the selective polarization inversion (SPI) of one-half of the proton doublet. The simultaneous 90°x (C) pulse enables the newly created population differences to produce the enhanced spectrum to be observed in the transverse plane (σ_6).

Due to SPI of one of the two proton spin transitions only one of the two halves of the ^{1}H doublet is inverted however the signal intensities are unchanged since there is no change in the population difference across each of the two ^{1}H transitions. In contrast, both the ^{13}C-spin population differences have been affected by this SPI, due to scalar coupling, such that the relative signal intensities would be -3:5. This ratio arises from combination of both the natural ^{13}C magnetization (1:1) and the magnetization transferred from the proton by selective polarization transfer (-4:4).

If the experiment is repeated with the phase of the last 90° (H) pulse at -y, the antiphase proton vectors that were -z, in the previous experiment, now become +z and vice versa. The resulting ^{13}C-spin doublets are likewise inverted such that the relative intensities are now (5:-3). Subtraction of the two experiments by inverting the receiver phase causes the polarisation transfer contribution to add and equalize the intensity of the two lines (-8:8) but cancels the natural magnetization lines that were not generated by polarization transfer. As a result detection of the ^{13}C magnetization results in a 4-fold increase in NMR signal intensity since the difference between the two scans results in a magnetization of $4I_zS_y$.

Appendix C

Purification of 10xHis-TCPTP C216S (1-281)-pET19b

E. coli cells containing 10xHis-TCPTP C216S (1-281) were lysed with 20 mM Tris-HCl pH 8.0, 5 mM β-mercaptoethanol, 5 mM benzamidine and Complete™ protease inhibitor (Boehringer Mannheim). The general lysis procedure always consisted of resuspending the frozen cells with 4 mL lysis buffer/g cell pellet and lysing the cells by two passes through a French Press at 18 000 psi. The cell lysate was then centrifuged at 31 000 x g for 30 min at 4°C and the supernatant was retained. The pellet was resuspended in lysis buffer and centrifuged for 10 min to obtain a second supernatant. After the centrifugation step, the pellet was resuspended in lysis buffer. The amount of soluble 10xHis-TCPTP C216S (1-281) protein obtained from the lysis procedure was assessed by taking aliquots of the lysate, obtained after the French Press, after the first and second centrifugation steps and from the final resuspended pellet. Examination of the aliquots proceeded by electrophoresis on 10-15% SDS Phastgels (Amersham Pharmacia) using Mark12 molecular weight markers (Novex). The bands were visualized by staining with Coomassie Blue (refer to Figure 41). Considering the varying volumes from which the aliquots were obtained for gel electrophoresis the relative ratio of soluble to total protein content is almost comparable (refer to lanes S1 and SP of Figure 41 respectively). The second supernatant (lane S2 of Figure 41) fraction also contained some soluble protein therefore the two centrifugation steps were necessary. Insoluble protein remaining in the pellet (refer to lane P of Figure 41) appears prominent

however it should be noted that this sample was 5x more concentrated in comparison to the other samples.

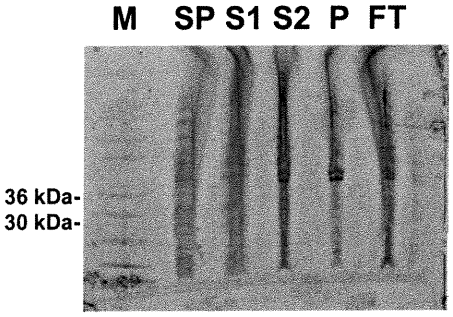
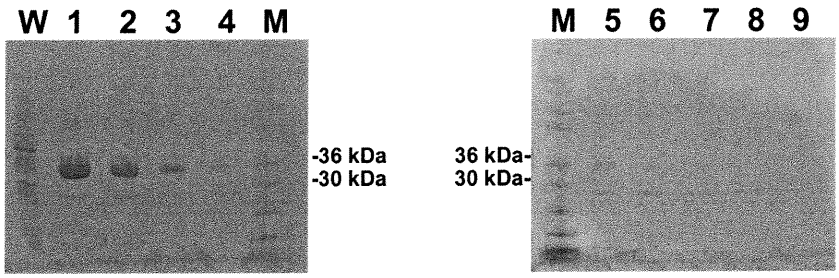


Figure 41. Lysis of 10xHis-TCPTP C216S (1-281)-peT19b.
M: Mark12 markers, SP: supernatant & pellet (25mL); S1: supernatant #1 (25 mL); S2: supernatant #2 (10 mL); P: pellet (5 mL); FT: flow-through (35 mL).

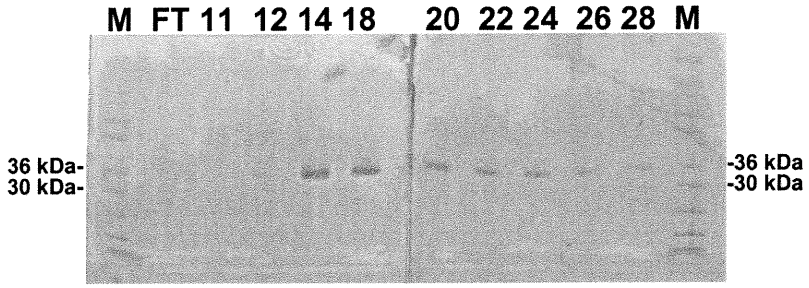
Purification of the collective supernatants (S1 and S2) for 10xHis-TCPTP C216S (1-281) consisted of two chromatography steps. The supernatants were first loaded onto a ~5 mL Ni-NTA (Qiagene Inc.) affinity column equilibrated with 20 mM Tris-HCl pH 8.0, 5 mM β-mercaptoethanol and 60 mM imidazole (buffer A). The column was washed with ten column volumes of buffer A and the protein eluted within five column volumes of buffer B (buffer A containing 250 mM imidazole). Fractions containing the 10xHis-TCPTP C216S (1-281) protein were always detected by electrophoresis on 10-15% SDS Phastgels (Amersham Pharmacia) and the bands were visualized by staining with Coomassie Blue (refer to Figure 42a).

Fractions of the protein obtained from the Ni-NTA column were further purified using a 5 mL HiTrap Blue™ (Amersham Pharmacia) affinity column equilibrated with 20 mM Tris-HCl pH 7.5, 5 mM DTT and 0.1 mM EDTA (buffer A). The column was

washed with four column volumes of buffer A followed by approximately three column volumes of 100 mM NaCl in buffer B (buffer A with 2 M NaCl). The protein was eluted in a ten column volume of 500 mM NaCl in buffer B. The fractions containing 10xHis-TCPTP C216S (1-281) (refer to Figure 42b) were pooled and concentrated to obtain a NMR sample.



(a)



(b)

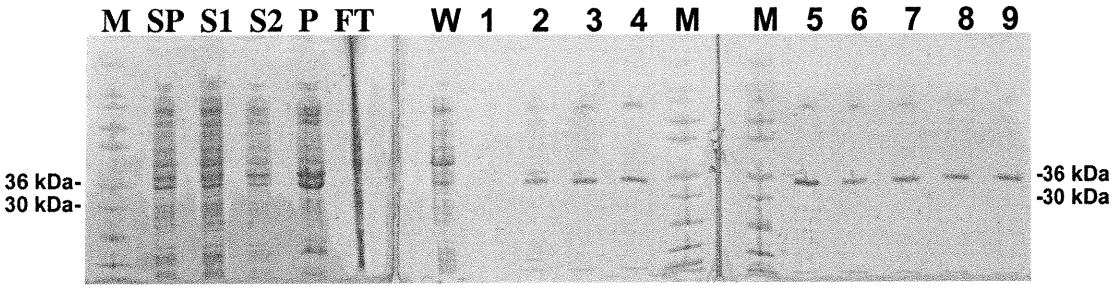
Figure 42. Purification of 10xHis-TCPTP C216S (1-281) using (a) Ni-NTA column and (b) HiTrap Blue™.

M: Mark12 markers; numbers denote fraction number (5 mL fractions for Ni-NTA column and 2.5 mL fractions for HiTrap Blue™).

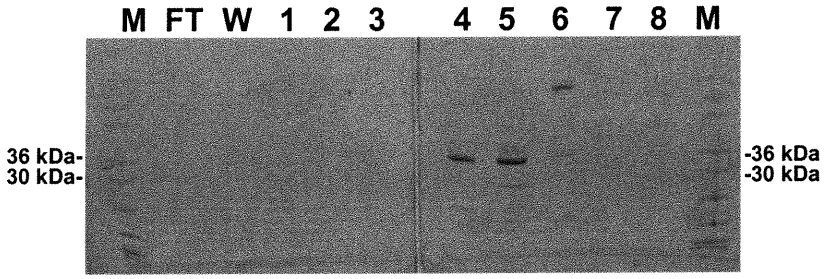
Purification of FLAG-TCPTP C216S (1-281)-pET11a

The purification protocol for FLAG-TCPTP C216S (1-281) was obtained from Dr. Ernest Asante-Appiah at Merck Frosst Canada & Co. at Pointe-Claire, Quebec. The cell pellet was lysed following the general lysis procedure (refer to p.IX) except the lysis buffer consisted of 20 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 5 mM DTT and Complete™ protease inhibitor (Figure 43a).

The protein was first purified using the above mentioned protocol employing a 5 mL HiTrap Blue™ affinity column except the column was washed with ten column volumes of buffer A and the protein was immediately eluted over a ten column volume of 100 mM NaCl in buffer B (Figure 43a). Fractions containing FLAG-TCPTP C216S (1-281) protein were then purified using the same buffers with a 5 mL HiTrap Q (Amersham Pharmacia) anionic exchange column. The column was washed with five column volumes of buffer A and the protein eluted at ~200 mM NaCl during a linear ten column volume salt gradient from 100 mM-500 mM NaCl in buffer B (Figure 43b).



(a)



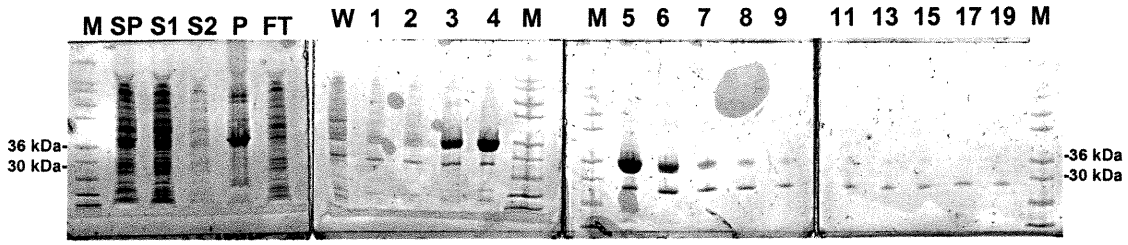
(b)

Figure 43. Purification of FLAG-TCPTP C216S (1-281) using (a) HiTrap Blue™ and (b) HiTrap Q.

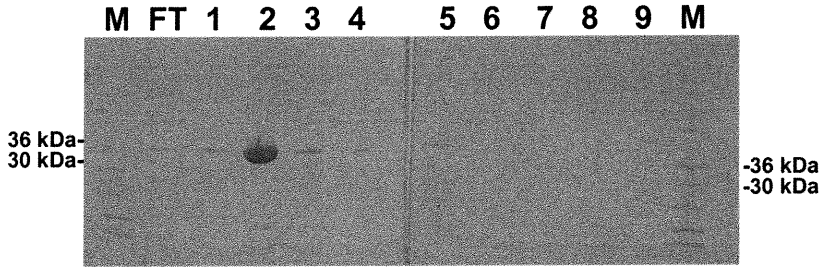
M: Mark12 markers, SP: supernatant & pellet (30mL); S1: supernatant #1 (30 mL); S2: supernatant #2 (10 mL); P: pellet (5 mL); FT: flow-through (40 mL); W: wash; numbers denote fraction number (5 mL fractions for HiTrap Blue™ and 5 mL fractions for HiTrap Q).

Purification of 10xHis-TCPTP C216S (1-281)-pET19b using Tris-HCl vs phosphate buffer

A frozen cell pellet of 10xHis-TCPTP C216S (1-281) was lysed following the general lysis procedure (refer to p.IX) except the lysis buffer consisted of 20 mM Tris-HCl pH 8.0, 5 mM β-mercaptoethanol, 1 mM benzamidine and 1% Triton X-100. The Complete™ protease inhibitor was omitted from the lysis buffer since it was determined to contain EDTA which can bind to the Ni^{2+} sites of the Ni-NTA affinity column. The detergent Triton X-100 was included in an attempt to solubilize more protein from the cell pellet. The supernatants (Figure 44a) were purified using the described Ni-NTA affinity column (1-2mL) purification protocol; however, buffer A consisted of 20 mM Tris-HCl pH 8.0, 5 mM β-mercaptoethanol and 1 mM benzamidine. Furthermore, the column was first washed with 5 column volumes of buffer A, then with 5 column volumes of buffer A with 60 mM imidazole. The protein was eluted within 10 column volumes of buffer B (buffer A with 250 mM imidazole). To scavenge any free Ni^{2+} metal ions and ensure the protein remained reduced 10 mM EDTA and10 mM DTT were added to each fraction respectively. Fractions containing 10xHis-TCPTP C216S (1-281) (Figure 44a) were further purified using a MonoQ HR 5/5 (Amersham Pharmacia) anionic exchange column equilibrated with 20 mM Tris-HCl pH 7.5, 5 mM EDTA and 5 mM DTT (buffer A) instead of the HiTrap Blue™ affinity column. The protein was eluted at ~180 mM NaCl during a linear 6 column volume salt gradient from 100 mM - 1 M NaCl in buffer B (buffer A with 2 M NaCl). Fractions containing 10xHis-TCPTP C216S (1-281) (Figure 44b) were prepared as NMR samples by the previously mentioned protocol (refer to p.69), except only 50 mM instead of 100 mM NaH_2PO_4 pH 7.5 buffer was used with 5 mM DTT-d_{10}.



(a)



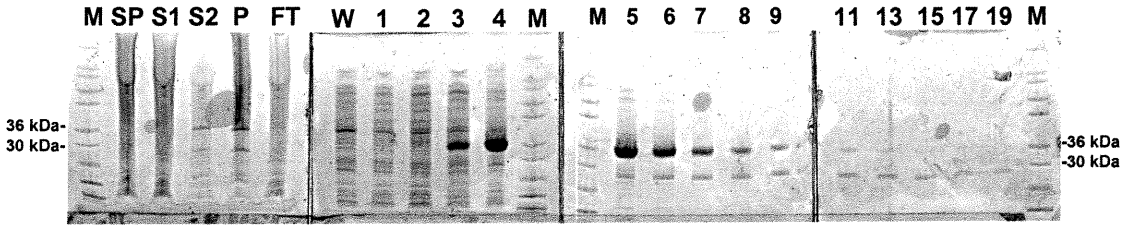
(b)

Figure 44. Purification of 10xHis-TCPTP C216S (1-281) with Tris-HCl buffer using (a) Ni-NTA affinity column and (b) MonoQ HR 5/5.

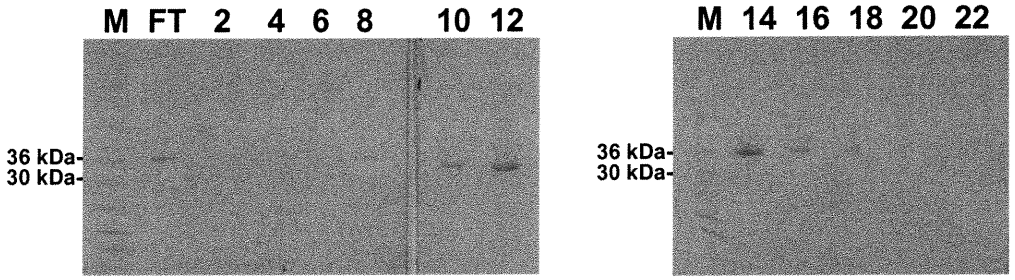
M: Mark12 markers, SP: supernatant & pellet (12mL); S1: supernatant #1 (12 mL); S2: supernatant #2 (5 mL); P: pellet (5 mL); FT: flow-through (17 mL); W: wash (5mL); numbers denote fraction number (1 mL fractions for Ni-NTA and 1 mL fractions for MonoQ HR 5/5).

For comparison purposes, 10xHis-TCPTP C216S (1-281) was similarly purified, except 100 mM NaH_2PO_4 pH 8.0 was used instead of 20 mM Tris-HCl pH 8.0. The protein was purified using the same Ni-NTA protocol mentioned above followed by an ion-exchange column (Figure 45). Instead of using a MonoQ HR 5/5 anionic exchange

column, a MonoS HR 5/5 (Amersham Pharmacia) cationic exchange column was used in the second chromatography step. Consequently, fractions containing 10xHis-TCPTP C216S (1-281) were pooled together and diluted 1:5 with 50 mM NaH_2PO_4 pH 4.6 (monobasic) to adjust the pH from 8.0 to 6.5. The same salt gradient employed for the MonoQ HR 5/5 was also used for the MonoS HR 5/5 column and the protein was observed to elute at ~400 mM NaCl. Fractions containing 10xHis-TCPTP C216S (1-281) were prepared as NMR samples using the previously mentioned protocol (refer to p.69).



(a)



(b)

Figure 45. Purification of 10xHis-TCPTP C216S (1-281) with phosphate buffer using (a) Ni-NTA affinity column and (b) MonoS HR 5/5.

M: Mark12 markers, SP: supernatant & pellet (12mL); S1: supernatant #1 (12 mL); S2: supernatant #2 (5 mL); P: pellet (5 mL); FT: flow-through (17 mL); W: wash (5mL); numbers denote fraction number (1 mL fractions for Ni-NTA and 1 mL fractions for MonoS HR 5/5).