Thiol Reactive Chemical Probes for Studying Protein Ion Channel Structure and the Application of Chemical Probes to the Study of U2AF, an Essential RNA Splicing Factor

by

Louise Yuli Foong

A thesis submitted in conformity with the requirements for the degree of

Doctor of Philosophy

Graduate Department of Chemistry

University of Toronto

"Thiol Reactive Chemical Probes for Studying Protein Ion Channel Structure and the Application of Chemical Probes to the Study of U2AF, an Essential RNA Splicing Factor."

by Louise Yuli Foong

Doctor of Philosophy, Graduate Department of Chemistry

University of Toronto

2001

Abstract

In Part I and II, thiol reagents were designed to react with cysteine residues within ion channel proteins. In Part I, a novel thiol reagent, 2-(methylthiosulfonate)ethyl-*N*-(*N*,*N*-dimethylaminoethyl)carbamate (MTSAC), was characterized when linked to the simple ion channel, gramicidin. MTSAC-treated channels showed a pattern of steps in the current recordings on the timescale of the carbamate bond isomerization; the size of the steps was sensitive to the location of the thiol-reactive site in relation to the channel entrance. MTSAC may be useful for establishing proximity to the pore in studies of ion channel protein structure.

In Part II, the synthesis of a group of lidocaine analogs was accomplished. A thiol-reactive lidocaine analog could be used to probe the region surrounding the local anesthetic binding site in cysteine mutants of sodium channels. A group of lidocaine analogs was synthesized, with 2, 3, and 4 carbons between the amino nitrogen of lidocaine and the thiol-reactive group.

In Part III, chemical probes were employed to study the RNA and protein interactions of U2AF, an essential RNA splicing factor. The synthesis of two novel thioester reagents was accomplished, benzophenone-3MPA and biotin-3MPA. EDTA-3MPA, a

thioester nuclease reagent previously used to study DNA-protein interactions, was also synthesized and used in this study. The thioester reagents allowed the attachment of reporter groups to a series of U2AF mutants with cysteine at the N-terminus. U2AF is a heterodimer composed of a large subunit $U2AF^{65}$ and a small subunit $U2AF^{35}$. $U2AF^{65}$ is essential for RNA splicing, containing an RS domain that is necessary for its activity and may be involved in regulation of activity. A series of N-terminal cysteine mutants of $U2AF^{65}$, full-length $U2AF^{65}$ and N-terminal truncations of $U2AF^{65}$ (Δ1-14, Δ1-24, Δ1-64 and Δ1-94) were engineered, cloned and purified. Biotin-3MPA was used to specifically label proteins with N-terminal cysteine mutations. It was used to detect the presence of an N-terminal cysteine, and the successful reaction of that cysteine to other thioester reagents. EDTA-3MPA was used to derivatize the full-length and (Δ1-64) $U2AF^{65}$ mutants. Addition of Fe^{2+} to these proteins generated a site-specific nuclease reagent to footprint the location of the $U2AF^{65}$ RS domain with respect to the pre-mRNA. (EDTA)-$U2AF^{65}$ created a footprint pattern on an RNA substrate, indicating the RS domain was in close proximity to the branchpoint adenosine. Benzophenone-3MPA was reacted with full-length and (Δ1-64) $U2AF^{65}$ mutants. (Benzophenone)-$U2AF^{65}$ and (benzophenone)-(Δ1-64)$U2AF^{65}$ were tested for crosslinking ability to the small subunit $U2AF^{35}$. (Δ1-64)$U2AF^{65}$ appeared to crosslink to $U2AF^{35}$ at a low level.

Acknowledgements

Thank you to Professor G. Andrew Woolley for guidance and support.

Thanks also to Professors Ron Kluger and Robert Batey for useful suggestions.

This research was funded by NSERC and the Ontario Ministry of Education.

Thanks to Ciska Böcker, Daniel Overste, Ayube Reayi and Valentin Zunic for their assistance.

The completion of this thesis would not have been possible without the love and support of the people closest to me. This work is dedicated to my sister Chloë Ting-Ting Foong and my parents. I thank my dear friend Timothy Burrow, whose help and guidance have been invaluable.

Table of Contents

Table of Figures

PART I: THE DEVELOPMENT OF A NOVEL THIOL REAGENT, MTSAC: STUDIES IN A MODEL SYSTEM

PART II: THE SYNTHESIS OF THIOL-REACTIVE ANALOGS OF LIDOCAINE

PART III: PROBING THE FUNCTION OF THE RS DOMAIN OF U2AF^{65}, AN ESSENTIAL RNA SPLICING FACTOR

PART III (CONTINUED):

Table of Abbreviations

PART I:

BES	N,N-bis[2-hydroxyethyl]-2-aminoethanesulfonic acid
DCC	dicyclohexylcarbodiimide
DMAP	dimethylaminopyridine
DMF	dimethylformamide
DTT	dithiothreitol
gram	gramicidin
gram'-SH	gramicidin-(OCO)-CH_2SH
gram''-SH	gramicidin-(OCO)-CH_2CH_2SH
gram'-S-MTSAC	gramicidin-(OCO)-CH_2CH_2S-MTSAC
gram''-S-MTSAC	gramicidin-(OCO)-CH_2S-MTSAC
gram''-S-MTSEA	gramicidin-(OCO)-CH_2CH_2S-MTSEA
MS-EI	mass spectrometry, electron impact
MS-ES	mass spectrometry, electrospray
MS-FAB	mass spectrometry, fast atom bombardment
MTSEA	methylthiosulfonate ethylamine/ammonium
MTSAC	2-(methylthiosulfonate)ethyl-*N*-(*N,N* dimethylaminoethyl)carbamate
THF	tetrahydrofuran
TLC	thin-layer chromatography
TEA	triethylamine
TFA	trifluoroacetic acid

PART II:

DMPA-Cl	2-chloro-N-(2,6-dimethylphenyl)acetamide
DMPA-NHEt	2-(ethylamino)-N-(2,6-dimethylphenyl)acetamide
DMPA-N(Et)(EtOH)	2-((ethyl,2'-hydroxylethyl)amino)-N-(2,6-dimethylphenyl) acetamide
DMPA-N(Et)(EtCl)	2-((ethyl,2'-chloroethyl)amino)-N-(2,6-dimethylphenyl) acetamide
DMPA-N(Et)(PrBr)	2-((ethyl,3'-bromopropyl)amino)-N-(2,6-dimethylphenyl) acetamide
DMPA-N(Et)(PrI)	2-((ethyl,3'-iodopropyl)amino)-N-(2,6-dimethylphenyl) acetamide
DMPA-N(Et)(PrMTS)	2-((ethyl,3'-methylthiosulfonate-propyl)amino)-N-(2,6 -dimethylphenyl)acetamide
DMPA-N(Et)(BuBr)	2-((ethyl,4'-bromobutyl)amino)-N-(2,6-dimethylphenyl) acetamide
EtOH	ethanol
EtOAc	ethyl acetate
$EtNH_2$	ethylamine
$EtNH_3Cl$	ethylamine hydrochloride
MeOH	methanol
MS-CI	mass spectrum, chemical ionization
MTSEA	methylthiosulfonate ethylamine
NaMTS	sodium methylthiosulfonate
NaOAc	sodium acetate
TsCl	*p*-toluenesulfonyl chloride (tosyl choride)
TLC	thin layer chromatography

PART III:

aa	amino acid
ddH_2O	distilled deionized water
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
his-tag	histidine tag
IPTG	isopropylthiogalactoside
MOPS	3-morpholinopropanesulfonic acid
NCBI	National Centre for Biotechnology Information
pre-mRNA	pre-messenger RNA
PIPES	piperazine-1,4-bis(2-ethanesulfonic acid)
PPT	polypyrimidine tract
PAP	poly(A) polymerase
PMSF	phenylmethylsulfonyl fluoride
RRM	RNA recognition motif
snRNA	small nuclear ribonucleic acid
snRNP	small nuclear ribonucleoprotein
Tris	tris(hydroxymethyl)aminomethane
ΔU2AF NE, ΔNE	U2AF-depleted nuclear extract

Part I: The Development of a Novel Thiol Reagent, MTSAC, and Studies in a Model System*

Section A: Introduction to Reagents Used to Study Protein Ion Channels and the Development of MTSAC

Membrane protein ion channels are the subject of intensive study because of their central role in regulation of ion flow across cell membranes. Despite their importance, little is known about their structure and how they function because they reside within a lipid bilayer; this bilayer environment complicates the use of conventional techniques for structural determination such as X-ray crystallography and NMR spectroscopy. Alternative means of probing protein channel structure are therefore required.

A powerful technique for studying ion channel structure is the recording of currents through a membrane containing the channels. Such electrophysiological recordings can provide single-molecule resolution and, when used in conjunction with site-directed mutagenesis, can be used to elucidate structure-function relationships.

For large complex ion channels like the sodium channel, much attention has been directed towards elucidation of the pore region; the pore region is believed to responsible for the ion selectivity of the channel (MacKinnon, 1995). A currently popular method for studying the pore region and the ion pathway through the protein is "scanning-cysteine accessibility mutagenesis" (*e.g.:* Akabas *et al.*, 1994; Sun *et al.*, 1996; Karlin & Akabas, 1996). Amino acids believed to be in or near the pore of the channel are mutated to cysteine (via site directed mutagenesis). The expressed protein is then exposed to a thiol reactive compound and the effect (if any) on channel behaviour is recorded. A number of different thiol reagents are in use. Metal ions such as silver and cadmium which form complexes with thiol groups have been used (Lü & Miller, 1995; Tomaselli *et al.*, 1995;

---

* Part I reprinted with permission from Foong, L.Y; You, S.; Jaikaran, D.C. J.; Zhang, Z.; Zunic, V.; Woolley, G. A. "Development of a novel thiol reagent for probing ion channel structure: Studies in a model system." (1997) *Biochemistry 36*(6), 1343-1348. Copyright 1997 American Chemical Society.

Backx *et al.*, 1992) and methyl thiosulfonate (MTS) reagents are widely used (Akabas *et al.*, 1992). MTS reagents include MTS ethylammonium ($MTSEA^+$), MTS ethyltrimethylammonium ($MTSET^+$) and MTS ethylsulfonate ($MTSES^-$). These reagents covalently link a charged group to the channel. The MTS reagents are employed most commonly in whole cell current measurements to identify changes (usually an overall decrease in current) from wild type behaviour (Akabas *et al.*, 1994; Sun *et al.*, 1996; Stauffer & Karlin, 1994; Kürz *et al.*, 1995). Examples of single channel recording of MTS-modified channels have also been reported (Mindell *et al.*, 1994).

While these approaches have provided new structural insight in a number of cases, they are not without shortcomings. Generally, the reagent must be applied for a relatively long time (minutes) relative to the time frame of protein motion. Because of this exposure time, the reagent may react with even a very minor channel conformation (Lü & Miller, 1995). Attempts to overcome the problem using very brief applications of reagents have been reported (Cheung & Akabas, 1996).

Another basic difficulty with the thiol-labelling approach is the assumption that a change in total current is in fact due to the presence of the thiol reagent in the pore. One cannot rigorously exclude the possibility that binding of a thiol reagent might inactivate a channel even at some distance from the pore region. In the whole cell case, thiol modification may even affect the interaction of other molecules with the channel, that is, the effect of the thiol reagent may be long-range. A demonstrated change in single-channel conductance after application of the reagent would be direct evidence for the presence of the reagent in or near the ion pathway. However, even here, the conductance change may be due to a change in protein structure. For example, Mindell *et al.* (1994) have proposed an alteration of protein structure as an explanation for the observed changes in single channel current after treatment of a cysteine-modified diphtheria toxin channel with $MTSES^-$. A thiol reagent that provided a unique 'signature' in current recordings, when it was present near the pore, but not when it was distant, would allow one to draw more definite conclusions about the location of the reagent.

Carbamate bonds undergo cis-trans thermal isomerization at a frequency of 1-100 Hz at room temperature (Jaikaran & Woolley, 1995; Kessler & Molter, 1976). A thiol reagent incorporating a carbamate bond in addition to a charged amino group would be

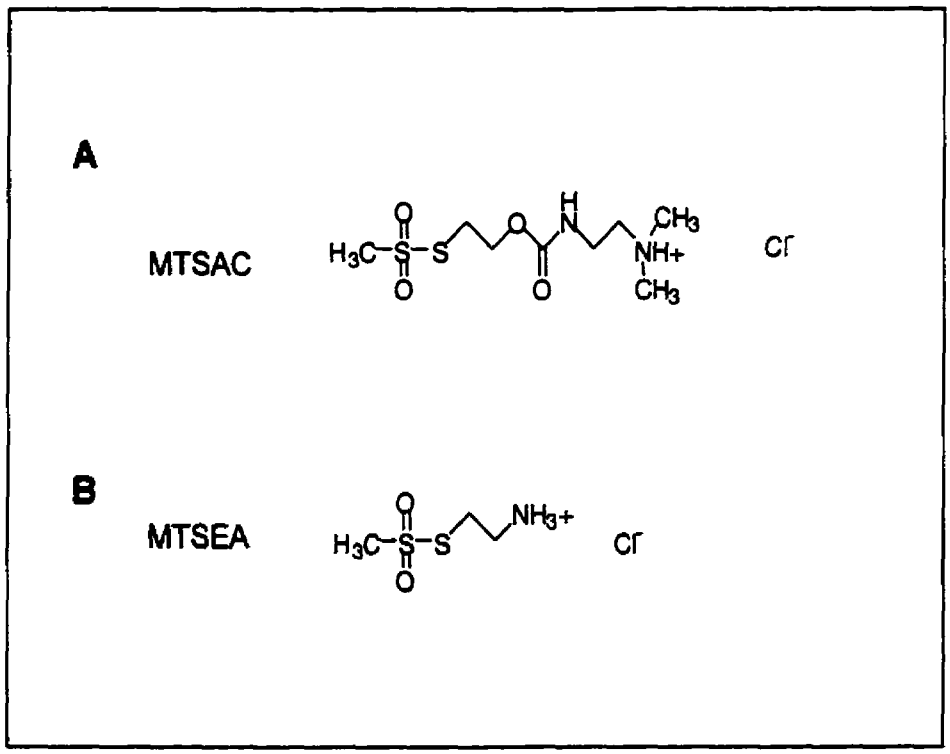


Figure 1: Chemical structures of the thiol reagents of (A) 2-methylthiosulfonate) ethyl-N-(N,N-dimethylamino -ethyl)carbamate (MTSAC) and (B) 2-(methylthiosulfonate) ethylamine (MTSEA).

expected to change its effective shape (*i.e.* its electrostatic charge distribution) as the isomerization occurred. If the reagent were located near the pore in the ion channel protein, this change should alter ion flux, thus acting as an 'internal current switching device' causing current steps or a 'flicker' in the single channel current, on a time frame specified by the activation energy of carbamate bond isomerization (Woolley *et al.*, 1995; Jaikaran & Woolley, 1995). If current steps were seen on the expected timescale when this thiol reagent was added to a cysteine-modified channel, it must be concluded that the thiol group is in fact near the pore (and the coupled reagent is moving freely). A long range effect is very unlikely since any coupling of the carbamate isomerization to protein conformational changes would alter the kinetics of isomerization.

Based on these arguments we designed and synthesized a thiol reagent designated 2-(methylthiosulfonate)ethyl-*N*-(*N*,*N*-dimethylaminoethyl) carbamate (MTSAC (**1**)) (Fig. 1). To test the properties of this reagent, and to compare it to the commonly used MTSEA reagent (Fig. 1), we have investigated its effect on single-channel currents in a structurally well-defined channel, gramicidin (Andersen & Koeppe, 1996; Woolley & Wallace, 1992). We synthesized two different derivatives of the gramicidin channel, bearing thiol groups at different distances from the C-terminal ends. An analysis of single channel recordings obtained with these derivatives after reaction with MTSAC indicates that this reagent can serve as a qualitative indicator of proximity to the pore.

SECTION B: EXPERIMENTAL PROCEDURES

Gramicidin was obtained from Sigma (St. Louis, MO), lipids used for single-channel recording from Nu-Chek-Prep, Inc. (Elysian, MO), and MTSEA from Toronto Research Chemicals (Toronto, ON). All other chemicals were obtained from Aldrich (Milwaukee, WI).

1. Synthesis of 2-(methylthiosulfonate)ethyl-N-(N,N-dimethylaminoethyl) carbamate (MTSAC)

The synthesis of 2-(methylthiosulfonate)ethyl-*N*-(*N,N*-dimethylaminoethyl) carbamate (MTSAC) is outlined in Scheme 1 and was accomplished by Shaochun You and Zhihua Zhang. Sodium sulfide nonahydrate (24 g, 0.2 mol) was dissolved in absolute ethanol (60 mL), cooled to -15°C, and methanesulfonyl chloride (**3**) (12.5 g, 0.11 mol in 20 mL abs. EtOH) added dropwise. The sodium chloride was filtered off and the sodium methylthiosulfonate (**4**) isolated and recrystallized from absolute ethanol to obtain 3.8 g (0.03 mol) pure product (NMR (D_2O): δ3.3 (s, 3H)). 2-Bromoethanol (2.5 g, 20 mmol) was added dropwise to sodium methylthiosulfonate (**4**) (1.34 g, 10 mmol in 15 mL abs. EtOH) and the mixture refluxed for 8 h. Sodium bromide formed and was filtered off, and the filtrate concentrated and extracted with $CHCl_3/H_2O$. The water phase was concentrated and then suspended in 30 mL THF to precipitate remaining NaBr. Concentration of the filtrate gave 2-(methylthiosulfonate)-ethanol (**5**) (0.86 g, 5.5 mmol) (NMR (d-DMSO): δ3.35 (t, 2H), δ3.5 (s, 3H), δ3.85 (t, 2H)). Next *p*-nitrophenyl chloroformate (0.2g, 0.5 mmol) in 1 mL THF was cooled to -10°C. To this, TEA (150 mL) was added over 10 minutes, and a suspension of 2-(methylthiosulfonate)ethanol (**5**) (95 mg, 0.6 mmol) in THF (1 mL) was added within 15 min. while maintaining the temperature. The reaction mixture was stirred overnight at 4°C. The resultant yellow suspension was filtered to isolate a yellow solid which was dissolved in $CHCl_3$ and washed with H_2O. Concentration of the chloroform phase yielded 2-(methylthiosulfonate) ethoxycarbonyl *p*-nitrophenolate (**6**) (70 mg, 0.2 mmol) (TLC, $CHCl_3$/MeOH/H_2O 65:25:4 : R_f = 0.9; NMR: δ3.43 (s, 3H), δ3.55 (t, 2H), δ4.55 (t, 2H), δ7.4 (m, 2H), δ8.3 (m, 2H)). Next, this carbonate compound (**6**) (119 mg, 0.37 mol in 2 mL abs. THF) was cooled to -15°C. To this, *N,N*-dimethylethylenediamine (29.4 mg, 0.33 mmol in 2 mL

THF) was added dropwise. After addition the temperature was maintained at 0°C for 10 min. followed by stirring for an additional 10 min. at room temperature. The reaction mixture was then cooled to 0°C again and HCl (1.8 mL of 1.0 M in Et_2O) was added dropwise. Stirring was continued for a further 30 min. and the oily solid produced isolated, and washed with THF. Dissolution in H_2O followed by freeze drying yielded the final compound, MTSAC (**1**) (77 mg, 0.3 mmol). NMR: δ2.85 (s, 6H), δ3.25 (t, 2H), δ3.45 (t, 2H), δ3.45 (t, 2H), δ3.5 (s, 3H), δ4.35 (t, 2H). High resolution MS-EI, 270.071793 for M^+-HCl ($C_{16}H_{14}O_2S$; calculated 270.073137).

2. Synthesis of Gram-(OCO)-CH_2SH and Gram-(OCO)-CH_2CH_2SH

The C-terminus of gramicidin was derivatized to obtain a terminal thiol group, as outlined in Scheme 2. This synthesis was accomplished by Zhihua Zhang and Valentin Zunic. 9-Fluorene-methylchloride (7) was prepared as detailed by Wawzonek (1956). This compound (7) (600 mg, 2.8 mmol) was then reacted with 3-mercaptopropionic acid or 2-mercaptoacetic acid (3.0 mmol) in diisopropylethylamine (1.5 mL) and THF (10 mL) at room temperature, overnight. The solvent was removed under vacuum, and resuspended in aqueous Na_2CO_3, pH 8-9. The mixture was then extracted with $CHCl_3$. The aqueous layer was acidified and extracted with EtOAc. The EtOAc layer was dried, and the solvent removed to give the product (**8a,b**). For (9-fluorenemethyl)-3-mercaptopropionic acid (**8b**), NMR (d-DMSO): δ2.6 (t, 2H), δ2.75 (m, 2H), δ3.1 (d, 2H), δ4.1 (t, 1H), δ7.35 (m, 4H), δ7.7 (m, 4H); MS-EI, 284 for M^+ ($C_{17}H_{16}O_2S$; calculated 284). For (9-fluorenemethyl)-2-mercaptoacetic acid (**8a**), NMR: δ3.7 (s, 2H), δ4.2 (t, 1H), δ7.35 (m, 4H), δ7.7 (m, 4H); MS-EI, 270 for M^+ ($C_{16}H_{14}O_2S$; calculated 270).

Gramicidin, 40 mg (21.3 mmol) was dissolved in 6 mL dichloromethane with DCC (175 mg, 850 mmol) and DMAP (26 mg, 213 mmol). (9-fluorenemethyl)-3-mercaptopropionic acid (**8b**) (18 mg, 63.7 mmol) was added to the gramicidin/DCC/DMAP solution and stirred overnight. The product (**9b**) was purified by gel-filtration chromatography using lipophilic Sephadex (LH-20) with methanol as the mobile phase. The protected gramicidin derivative (**9b**) (19 mmol, 40 mg) was deprotected by dissolution in 50% piperidine in DMF (1.2 mL) followed by concentration to half volume by evaporation under nitrogen. To this, methanol (1 mL) was added and the solution passed through the LH-20 column again to give a disulfide linked

gramicidin dimer. To reduce the disulfide bond, 2 mL of the gramicidin dimer solution in MeOH was added dropwise to a N_2-saturated DTT solution (18 mg in 1.0 mL MeOH) under a N_2 atmosphere. The desired product, gram-(OCO)-CH_2CH_2SH (**10b**) (gram''-SH) was obtained after 6 h stirring at room temperature and purification by passage through the LH-20 column. TLC ($CHCl_3$/MeOH/H_2O 65:25:4): R_f = 0.8. MS-FAB: 1970 for M^+ ($C_{102}H_{146}N_{20}O_{18}S$; calculated 1971).

To obtain gram-(OCO)-CH_2SH (**10a**) (gram'-SH), the same procedure was followed except that (9-fluorenemethyl)-2-mercaptoacetic acid (**8a**) was substituted for (9-fluorenemethyl)-3-mercaptopropionic acid (**8b**).

3. Reaction of Thiol-derivatized Gramicidin with the Thiol Reagents

The thiol derivatives of gramicidin, gram'-SH (**10a**) and gram''-SH (**10b**), were linked to the two thiol reagents, MTSAC (**1**) and MTSEA (**2**) (3 fold excess to gram^n-SH), by reaction in methanol for 2 h at room temperature (Scheme 2). This synthesis was accomplished by Dominic Jaikaran and Valentin Zunic. For gram''-S-MTSEA, MS-FAB: 2046 for MH^+ ($C_{104}H_{150}N_{21}O_{18}S_2$; calculated 2045). For gram''-S-MTSAC, MS-ES: 2160 for M^+ ($C_{109}H_{159}N_{22}O_{20}S_2$; calculated 2160).

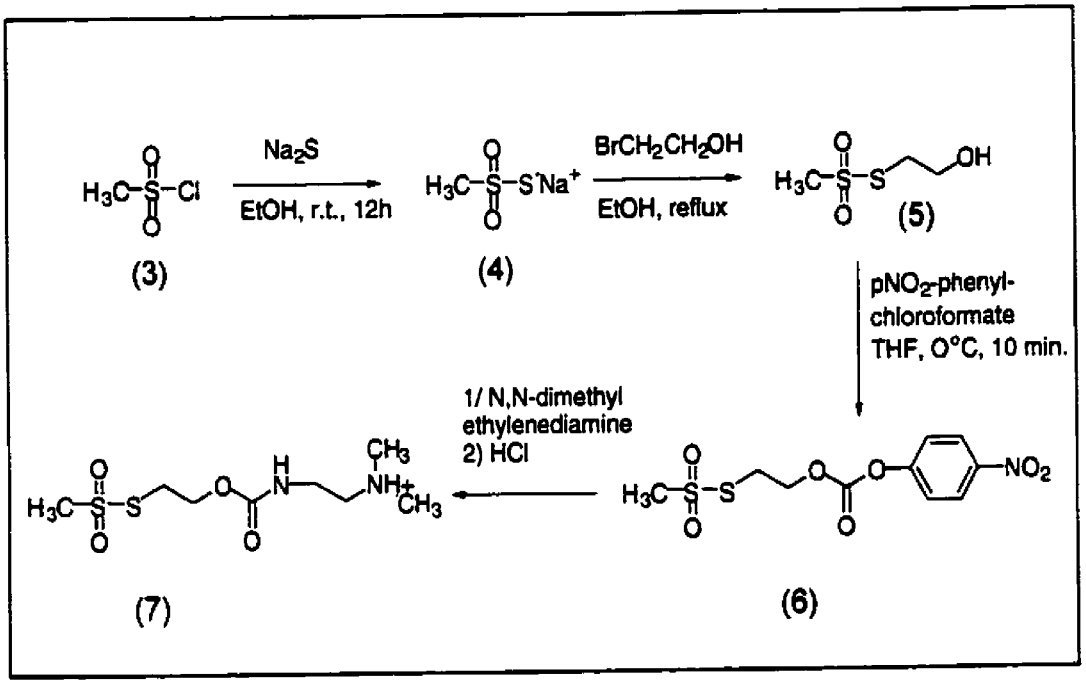
Due to gram'-SH's instability during the deprotection step, an alternative synthesis was also employed where MTSAC (**1**) (60 mL) was coupled with 2-mercaptoacetic acid (3.0 mL) in MeOH (0.4 mL) by stirring for 1 h at room temperature to give the disulfide acetic acid. The disulfide (300 mL) was coupled to gramicidin (14 mg) in DCC (87 mg) / DMAP (14 mg) / CH_2Cl_2 (4.5 mL) to give gram'-S-MTSAC. For gram'-S-MTSAC, MS-ES: 2146 for MH^+ ($C_{108}H_{156}N_{22}O_{20}S_2$; calculated 2145).

Purification of the gramicidin derivatives for single channel measurements was accomplished by HPLC using a reverse phase Rx-C8 column (Zorbax, Rockland Technologies Inc.). A mobile phase consisting of 80% MeOH in H2O, pH 3 (0.1% trifluoroacetic acid, adjusted with TEA) was employed. Retention times were: gram''-S-MTSEA = 12.0 min; gram''-S-MTSAC = 9.2 min.; gram'-S-MTSAC = 8.8 min.

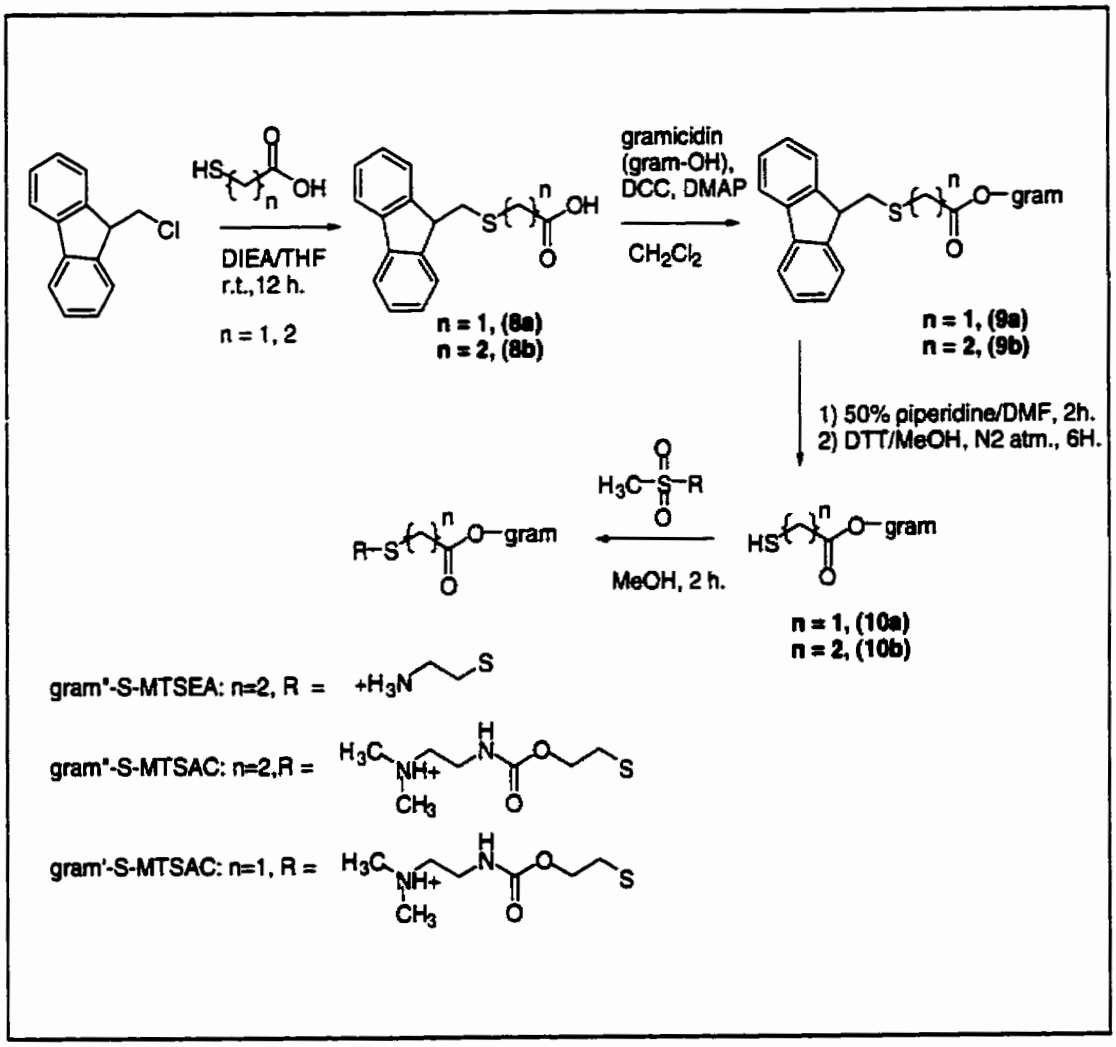
4. Single Channel Measurements

Currents through a glycerol monooleate lipid bilayer containing the gramicidin derivative were recorded and controlled using an Axopatch 1D patch clamp amplifier (Axon Instruments). The lipid bilayer was formed over the opening of a pipette tip as described by Busath & Szabo (1988). The recordings were obtained at room temperature (23°C ± 1°C) with a 200 mV holding potential and filtered at 200 Hz. The aqueous solution used was 1 M CsCl, 5 mM BES, pH 6.3 (adjusted with CsOH). Current recordings were stored and analyzed using Synapse software (Synergistic Research Systems).

Heterodimers of the MTSAC derivatives and unmodified gramicidin were obtained by first adding pure gram^n-S-MTSAC to obtain homodimers, followed by addition of unmodified gramicidin to one compartment only. Heterodimers and homodimers were distinguished by reversing the polarity of the applied voltage during the recording of an open channel. Homodimer currents were of the same magnitude with either voltage whereas heterodimer currents were of different magnitudes.



Scheme 1: Synthesis of MTSAC



Scheme 2: Derivatization of gramicidin with a variable length linker between gramicidin and the thiol reagents, MTSEA and MTSAC

SECTION C: RESULTS AND DISCUSSION

1. Synthesis of a Novel Thiol Reagent, MTSAC

The structure of MTSAC incorporates three functional groups: a methylthiosulfonate (MTS) group at one end of the molecule, an amino group at the other end, and an internal carbamate bond. The MTS group is known to react with thiol groups with high specificity (Kenyon & Bruice, 1977). The terminal charged amino group will affect cation currents electrostatically and the cis-trans isomerization of the internal carbamate bond provides the unique 'current signature' of the reagent. The synthesis of MTSAC was adapted from Bruice & Kenyon (1982) and the synthesis of gramicidin-ethylenediamine (Woolley *et al.*, 1995); the synthetic route is outlined in Scheme 1. Methanesulfonyl chloride (**3**) was converted to the thiosulfonate salt (**4**) by reaction with sodium sulfide. Reaction with 2-bromoethanol afforded 2-(methylthiosulfonate)-ethanol (**5**). This was then reacted with *p*-nitrophenyl chloroformate to give (2-methylthiosulfonate)ethoxycarbonyl *p*-nitrophenolate (**6**). The *p*-nitrophenolate moiety was then displaced by *N,N*-dimethylethylenediamine to give the thiol reagent, MTSAC (**1**).

2. Reaction of Thiol Reagents with Thiol Derivatives of Gramicidin

To test MTSAC, and compare its action with MTSEA (a reagent without a carbamate bond), a model system was required. Gramicidin was chosen—it is structurally well characterized and is popular as a system for studying structure-function relationships in ion channels (Koeppe & Andersen, 1996). Gramicidin forms cation-selective channels in a variety of membranes; channels form through N-terminus to N-terminus dimerization of peptide monomers. The C-termini of the peptides form the entrance and exit of the channel. A variety of C-terminal derivatives of gramicidin have been described; the structure of the channel does not seem to be sensitive to modification at this site (Woolley & Wallace, 1992).

It was noted that the ester bond which forms part of the link between gramicidin and the thiol reagent is prone to hydrolysis. This may be due in part to the presence of the terminal (nucleophilic) amino group: it is possible for the amino nitrogen to make an internal attack on the carbonyl carbon of the ester bond. Hydrolysis of the ester bond was

observed in gram'-SH during the deprotection step, using 50% piperidine in DMF (Scheme 2). In order to avoid this, a simpler and much more efficient synthetic route was proposed and carried out successfully. In this route, the thiol reagent (MTSAC in this case) was coupled directly to 2-mercaptoacetic acid. The resultant disulfide may then be reacted with gramicidin, forming an ester bond linkage between the two molecules. This route avoids the use of the fluorenyl protective group and thus eliminates the need for the deprotection step.

The C-terminus of gramicidin was derivatized by forming esters with two different thiol-carboxylic acids (Scheme 2). In this way, a thiol group is positioned at the channel entrance and exit, close to the ion pathway. Reaction of this thiol group with a thiol reagent would thus place the reagent near the mouth of the channel as shown in Figure 2. The average distance of the thiol group from the channel mouth was varied by using either 2-mercaptoacetic acid or 3-mercaptopropionic acid. It was predicted that the closer the charged amino group of the reagent was to the mouth of the channel (*i.e.* the shorter the derivative), the greater the effect on the ion flux.

3. Single Channel Recordings

Single channel recordings of the gramicidin derivatives were accomplished as described. The experimental conditions described were conditions which gave the best opportunity for observing the greatest difference between the gramicidin derivatives. The use of a relatively concentrated salt solution, 1 M CsCl, allowed peak currents for open channels to be large enough to observe necessary details in the current trace without destabilizing the lipid bilayer. Solutions of peptide obtained from the HPLC purification often required a dilution of up to 2000 times before it could be used. It was found that channel recordings could best be accomplished by adding an excess of peptide to the recording chamber and then washing off the pipette tip at intervals with methanol to reduce the number of channels present as required. Recordings could be obtained quickly by breaking and forming the lipid bilayer immediately after the addition of peptide to the chamber. For glycerol monooleate, a noise level of 0.6-0.7 as measured by the patch clamp amplifier indicated the presence of a stable lipid bilayer.

Figure 3 shows single channel current recordings of unmodified gramicidin (A), gram"-S-MTSEA (B), and gram"-S-MTSAC (C). Currents through each modified

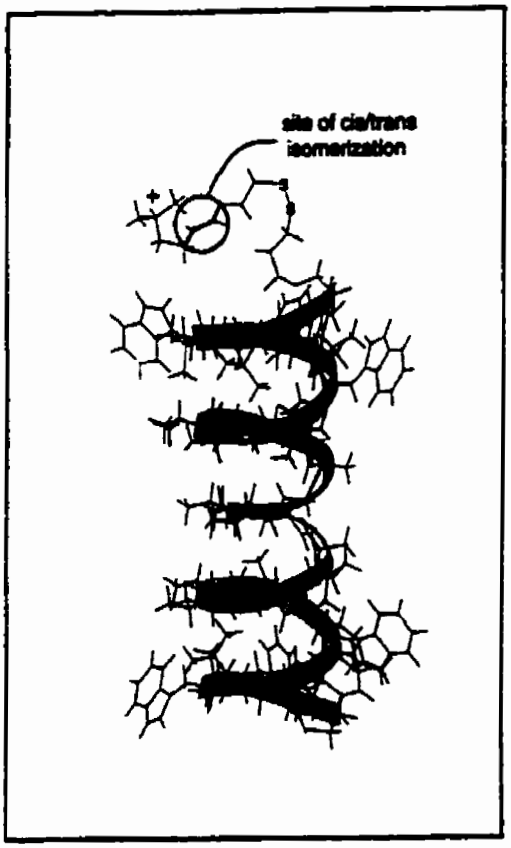


Figure 2: Structure of a gramicidin channel modified at one end with the MTSAC reagent (i.e. a heterodimer channel). The carbamate bond that isomerizes is circled. The channel structure is that reported by Arseniev *et al.* (1985). The modified C terminus is flexible, and only one of many possible conformations is shown.

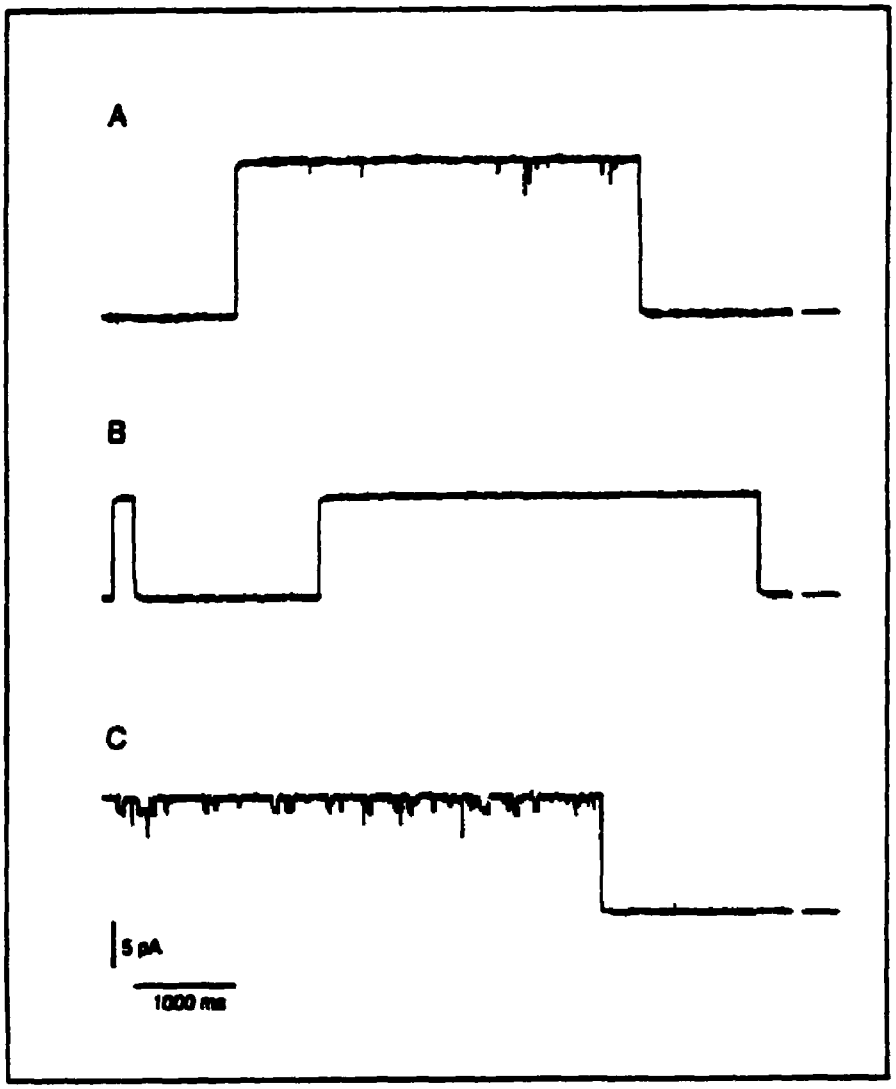


Figure 3: Single channel recordings of gramicidin and derivatives. The baselines (no channel open) are indicated by the horizontal bars: (A) unmodified gramicidin (15.3 ± 0.9 pA, n = 49), (B) gram"-S-MTSEA (10.7 ± 0.4 pA, n = 176), and (C) gram"-S-MTSAC (topmost level 13.5 ± 0.5 pA, n = 133)

channel were smaller than for gramicidin under the same experimental conditions. Thus, the presence of the positively charged thiol reagents at the entrance (and exit) of the channel reduces cation flow through the channel, consistent with previous observations (Mindell *et al.*, 1994; Akabas *et al.*, 1992). Importantly, gram"-S-MTSAC exhibited a flicker in current, as predicted (Fig. 3C). This flicker is not simply noise since it is not present in the baseline current (Fig. 3C, horizontal bar). Gram"-S-MTSEA, which contains a disulfide bond and a positively charged amino group, but not the carbamate group, also does not show current flickers. The disulfide bond does have two preferred rotomers but because the barrier to interconversion is relatively low (7 kcal/mol (Creighton, 1993)), it is too rapid to be detected in these recordings. Thus the current steps in the gram"-S-MTSAC recording can be attributed to thermal cis-trans isomerization of the carbamate bond in the MTSAC reagent.

Previous studies have shown that the trans state of these types of carbamate groups is more stable than the cis state (Kessler & Molter, 1976; Jaikaran & Woolley, 1995). Thus, the longer lived state (the upper current level in Figure 3C) presumably reflects MTSAC reagents in trans conformations. Average lifetimes for the states (38 ms cis/266 ms trans) estimated from recordings of channels in which only one MTSAC reagent was present (see below) are consistent with previous measurements (Jaikaran & Woolley, 1995).

Since each molecule of MTSAC can exist in two states, cis and trans, the number of current levels can in principle increase as 2^N where N is the number of thiol reagents present. Figure 4 shows magnified traces of currents flowing through an open channel with one (panel B) or two (panel C) MTSAC groups. The current observed when two channels are open (each with two MTSAC groups) is shown in panel D. A heterodimer of unmodified gramicidin and gram"-S-MTSAC, which has only one MTSAC group, exhibits two levels of current (Fig. 4B). The size of this step is dependent on the direction of the current flow. For one orientation of the applied voltage, the peak current is 12.9 ± 0.8 pA (n = 31) whereas for the other orientation the peak current is 16 ± 1 pA (n = 30)(which is comparable to unmodified gramicidin (15.3 ± 0.9 pA, n = 59)) and the step size is 0.4 pA. The homodimer of gram"-S-MTSAC shows at least three levels of a possible four. The effect of MTSAC on the open channel current becomes more marked if two (or more) channels are open simultaneously (panel D).

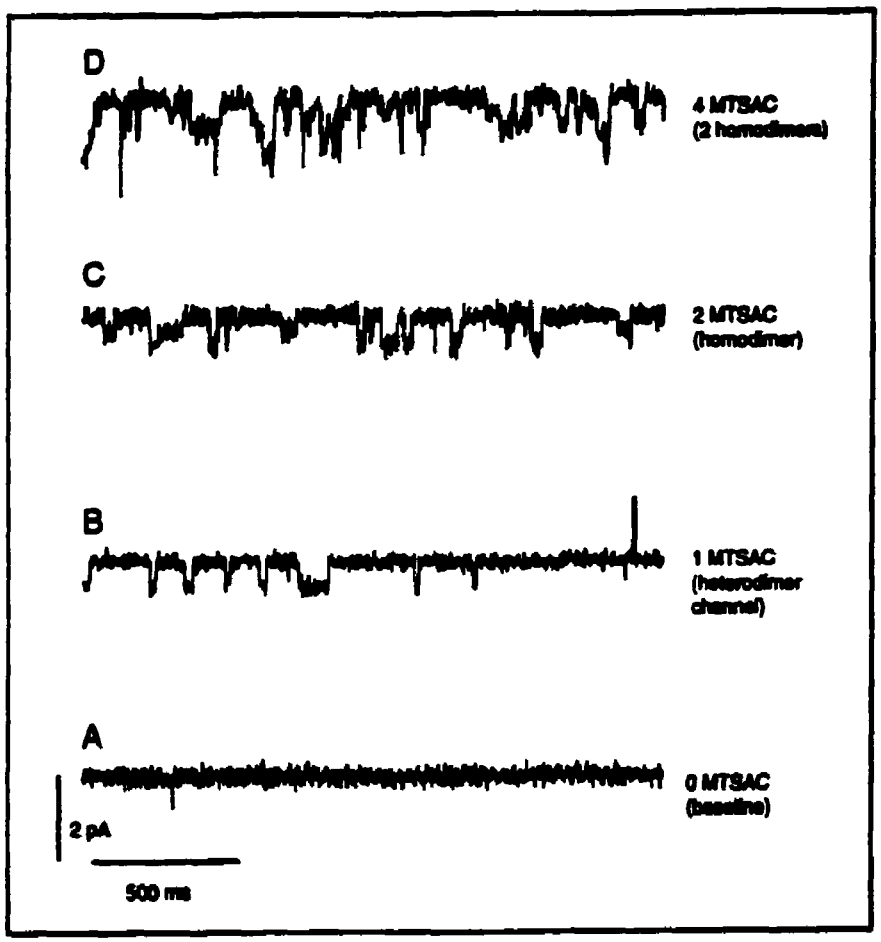


Figure 4: Effects of different numbers of MTSAC reagents on the open channel current. The number of channels open is known from the number of large (-14 pA) steps observed above the baseline. For clarity, the baselines are not shown in B-D and current through the open channel(s) is shown on an expanded scale: (A) baseline current, (B) current through an open gram"-S-MTSAC/gramicidin heterodimer showing two current levels (one step), (C) current through an open gram"-S-MTSAC homodimer (at least three different current levels can be idenitified), and (D) current observed when two gram"-S-MTSAC homodimers are open.

The effect of varying proximity of the MTSAC group to the channel mouth was tested by comparing single channel recordings of gram''-S-MTSAC and gram'-S-MTSAC (Fig. 5). Gram'-S-MTSAC has one fewer methylene unit in the linkage to the terminal amino group than gram''-S-MTSAC so that the charged group should (on average) be closer to the channel mouth. It was expected that gram'-S-MTSAC would show more pronounced current flickers since changes in 'shape' of the MTSAC group would more directly influence ion flux. In Figure 5 two channels are open at the beginning of each record; at the arrow one channel closes, then this channel closes and the baseline is recorded (shown by the horizontal bar). The steps in gram'-S-MTSAC current (B) are more pronounced than the current steps observed with gram''-S-MTSAC (A). The effect is particularly obvious when two channels are open. In the single channel recording of gram'-S-MTSAC, there are 4 current levels apparent, as expected for the homodimer, whereas the fourth level was not clearly apparent in gram''-S-MTSAC. Also, the difference between the top-most and bottom-most levels is 2 pA for gram'-S-MTSAC, as compared to 1 pA for gram''-S-MTSAC. The maximal current was reduced slightly from 13.5 ± 0.5 pA for gram''-S-MTSAC to 12.2 ± 0.4 pA for gram'-S-MTSAC, also indicating that the MTSAC group was having a greater effect on the current. Thus, the effect of MTSAC does indeed appear to be a function of its distance from the pore.

MTSAC behaves essentially as predicted and may prove useful as a new probe of ion channel structure. If it were added to a cysteine-substituted channel of unknown three-dimensional structure, and current steps appeared with lifetimes similar to those shown in figures 3-5, one could conclude that the reagent was interacting directly with the ionic flux. If the reagent instead induced conformational switching in the protein, it is unlikely that the activation barrier for such switching would be the same as that for carbamate isomerization. An increase of even 1 kcal/mol in the barrier would lead to more than a 5-fold increase in state lifetimes. With further development, better reagents may be obtained. For example, placement of an electron-withdrawing group adjacent to the carbamate nitrogen, would be expected to increase the rate of isomerization. This would be desirable for use in a channel with a shorter open lifetime. The charge of the reagent could also be altered to maximize effects on ion flow. Smaller reagents may be useful if site-accessibility proves to be a problem. Such designed thiol reagents should expand the range of techniques available to study the structures of ion channel proteins.

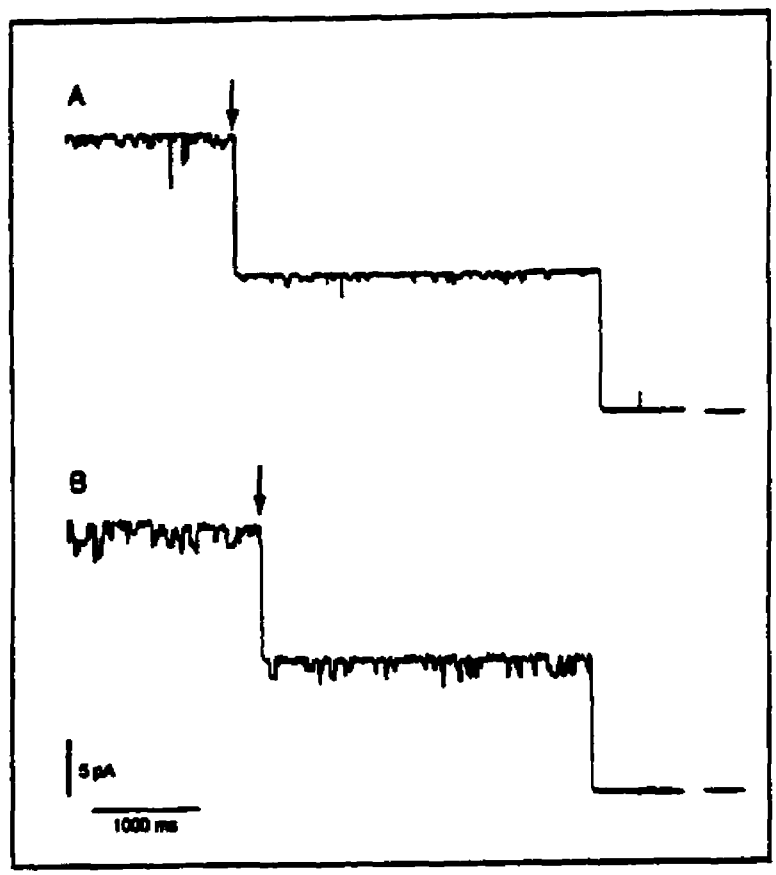


Figure 5: Comparison of single channel records for homodimers of (A) gram"-S-MTSAC and (B) gram'-S-MTSAC. Two channels are open at the start of each record. One channel closes at the point marked with an arrow. The baselines are indicated with horizontal bars. The steps in current are more pronounced with the shorter derivative (B).

REFERENCES

Akabas, M.H. and Karlin, A. "Identification of acetylcholine receptor channel-lining Residues in the M1 Segment of the α-Subunit." (1995) *Biochemistry* **34** (39), 12496-12500.

Akabas, M. H.; Kaufmann, C.; Archdeacon, P.; Karlin, A. "Identification of acetylcholine receptor channel-lining residues in the entire M2 Segment of the α-subunit." (1994) *Neuron* **13** (14), 919-27.

Akabas, M.H.; Stauffer, D. A.; Xu, M.; Karlin, A. "Acetylcholine receptor channel structure probed in cysteine-substituted mutants." (1992) *Science* **258**, 307-10.

Arseniev, A.S.; Barsukov, I.L.; Bystrov, V.F.; Lomize, A.L.; Ovchinikov, Y.A. "^{1}H-NMR study of gramicidin A transmembrane ion channel. Head-to-head right-handed, single-stranded helices." (1985) *FEBS Lett.* **186**, 168-174.

Backx, P. H.; Yue, D. T.; Lawrence, J. H.; Marban, E.; Tomaselli, G. F. "Molecular localization of an ion-binding site within the pore of mammalian sodium channels." (1992) *Science* **257**, 248-251.

Bodansky, M. and Bednare, M. A. "Derivatives of S-9-fluorenylmethyl-L-cysteine." (1982) *Int. J. Peptide Protein Res.* **20**, 437-7.

Bruice, T. W.; Kenyon, G. L. "Novel alkyl alkanethiolsulfonate sulfhydryl reagents. Modification of Derviatives of L-cysteine." (1982) *J. Protein Chem.* **1**, 47-58.

Busath, D. and Szabo, G. "Low conductance gramicidin A channels are head-to-head dimers of beta 6.3-helices." (1988) *Biophys. J.* **53**, 689-695.

Cheung, M. and Akabas, M. H. "Identification of cystic fibrosis transmembrane conductance regulation channel-lining residues in and flanking the M6 membrane-spanning segment." (1996) *Biophys. J.* **70**, 2688-2695.

Creighton, T. E. *Proteins: Structure and Molecular Properties 2nd ed.* New York: W. H. Freeman & Co., 1993.

Hille, B. *Ionic Channels of Excitable Membranes 2nd ed.* Sunderland: Sinauer Associates Inc., 1992.

Jaikaran, D. and Woolley, G.A. "Characterization of thermal isomerization at the single molecule level." (1995) *J. Phys. Chem.* **99**, 13352-13354.

Karlin, A. and Akabas, M. H. "Toward a structural basis for the function of nicotinic acetylcholine receptors and their cousins." (1996) *Neuron* **15**, 1231-1244.

Kenyon, G. L. and Bruice, T. W. "Novel sulfhydryl reagents." (1977) *Methods in Enzymology* **47**, 407-430.

Kessler, H. and Molter, M. "Conformation of protected amino acids. V. Application of lanthanide-shift reagents for conformational studies of tert-butoxycarbonyl alpha-amino acid esters: equilibrium changes and kinetics of the isomerization." (1976) *J. Am. Chem. Soc.* **98**, 5969-5973.

Koeppe, R. E. 2nd and Andersen, O. S. "Engineering the gramicidin channel." (1996) *Annu. Rev. Biophys. Biomol. Struct.*

Kürz, L. L.; Zühlke, R. D.; Zhang, H.-J.; Joho, R. H. "Side-chain accesibilities in the pore of a K^+ channel probed by sulfhydryl-specific reagents after cysteine-scanning mutagenesis." (1995) *Biophys. J.* **68**, 900-905.

Lü, Q. and Miller, C. "Silver as a probe of probe-forming residues in a potassium channel." (1995) *Science* **268**, 304-307.

MacKinnon, R. "Pore loops: an emerging theme in ion channel strucutre." (1995) *Neuron* **14**, 889-892.

Mindell, J. A.; Zhan, H.; Huynh, P. D.; Collier, R. J.; Finkelstein, A. "Reaction of diptheria toxin channels with sulfhydryl-specific reagents: observation of chemical reactions at the single molecule level." (1994) *Proc. Natl. Acad. Sci.* **91**, 5272-5276.

Single-Channel Recording. Sakmann, B., Neher, E., Eds. Plenum Press: New York, 1983.

Stauffer, D. A. and Karlin, A. "Electrostatic potential of the acetylcholine binding sites in the nicotinic receptor probed by reactions of binding-site cysteines with charged methanthiosulfonates." (1994) *Biochemistry* **33**, 6840-6849.

Sun, Z.P.; Akabas, M.H.; Goulding, E.H.; Karlin, A.; Siegelbaum, S.A. "Exposure of residues in the cyclic nucleotide-gated channel pore: P region structure and function in gating." (1996) *Neuron* **16**, 141-149.

Tomaselli, G. F.; Chiamvimonvat, N.; Nuss., H. B.; Balser, J. R.; Perez-Garcia, M. T.; Xu, R. H.; Orias, D. W.; Backx, P. H.; Marban, E. "A mutation in the pore of the sodium channel alters gating." (1995) *Biophys. J.* **68**, 1814-1827.

Wawzonek, S.; Dufek, E. "The acid-catalyzed reaction of 9-fluorenol with 9-alkylidene fluorenes." (1956) *J. Am. Chem. Soc.* **78**, 3530-3533.

Woolley, G.A. and Wallace, B.A. "Model ion channels: gramicidin and alamethicin." (1992) *J. Membr. Biol.* **129**, 109-136.

Woolley, G.A.; Jaikaran, A.S.I.; Zhang, Z.; Peng, S. "Design of regulated ion channels using measurements of cis-trans isomerization in single molecules." (1995) *J. Am. Chem. Soc.* **117**, 4448-4454.

Part II: The Synthesis of Thiol-Reactive Analogs of Lidocaine

SECTION A: INTRODUCTION TO SODIUM ION CHANNELS AND LOCAL ANESTHETICS; DEVELOPMENT OF THIOL REACTIVE ANALOGS OF LIDOCAINE

1. The Sodium Ion Channel

Sodium ion channels are found in the cell membranes of nerve cells and are responsible for the initiation and the propagation of the action potential along the nerve axon. The sodium channel is a voltage-gated channel; a local change in membrane potential causes the channel to open, allowing the flow of sodium ions from the outside of the cell into the cell. Soon after the channel opens, it enters an inactive state for a short period of time (a refractory period) before it returns to the closed state, from which it can be activated by a voltage pulse again (Hille, 1992).

Sodium channels and other voltage-gated protein ion channels are the subject of intense study because of their complex behaviour and their central role in the excitability of nerves. Because of the size of the channel and the fact that it is a membrane protein, conventional techniques of NMR and X-ray crystallography are not readily applied to these systems. Only small fragments of the ion channel may be studied at a time. Most of the current information on the sodium channel has been based on analysis of the primary sequence of the protein (as derived from the cloned genes for the protein) and electrophysiological studies in combination with site-directed mutagenesis.

The sodium channel is composed of a single long peptide sequence which resides within the cell membrane. A hydropathy analysis of the primary amino acid sequence for the sodium channel has allowed prediction of transmembrane folding patterns, so that a predicted "map" of the sodium channel has been proposed, as shown in Fig. 1A (Catterall, 1992). The protein consists of four domains, each domain containing six transmembrane segments. Each domain is linked by a chain of amino acids which form loops on either side of the cell membrane. These loops are heavily glycosylated. The four domains come together to constitute the α-subunit of the sodium channel, which contains all of the major functional structures of the channel. A schematic model of the sodium channel was

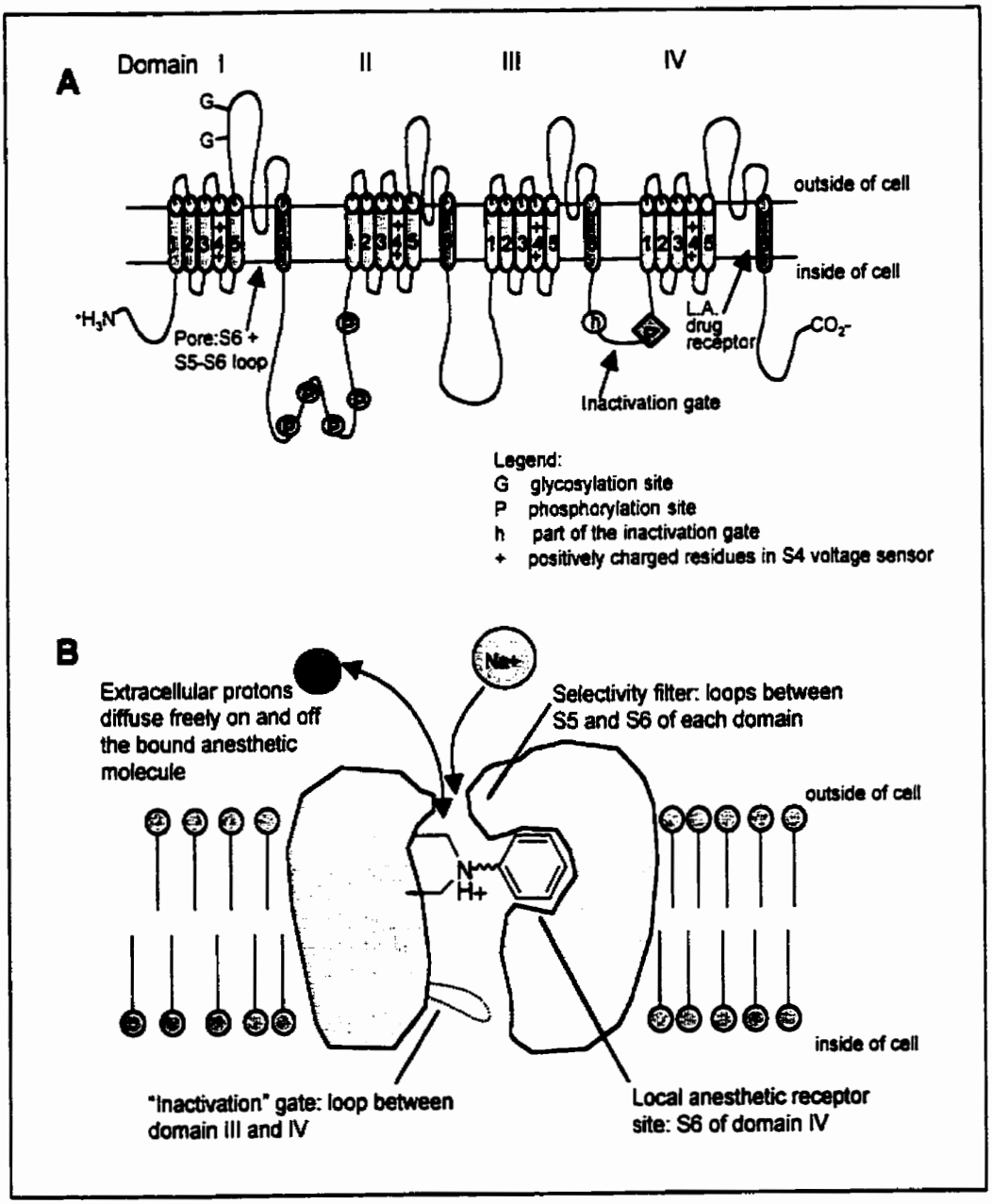


Figure 1: Proposed structure of the sodium channel. (A) The α subunit of the sodium channel is composed of a single polypeptide chain containing 4 homologous domains. Each domain contains 6 transmembrane α-helices. Segment 4 (S4) of each domain contains charged residues which together act as the voltage sensor (yellow). The pore region is defined by S6 of each domain (pink) and the portion of the S5-S6 linker which loops into the cell membrane (adapted from Catterall, 2000). (B) Schematic diagram of the sodium channel blocked by a local anesthetic drug molecule resembling lidocaine (partially adapted from Hille, 1992; Catterall, 1992).

proposed by Hille, as shown in Fig. 1B. There is a putative pore region or vestibule which the sodium ion must pass through in order to pass through the channel; this pore region contains a "selectivity filter" which exhibits an extremely high degree of selectivity for sodium ions (Catterall, 1992). Below this (on the side closer to the inside of the cell) is the inactivation mechanism or "gate" of the channel (Hille, 1992).

The amino acids joining the transmembrane segments 5 and 6 (S5 and S6) of each of the 4 domains have been identified as the putative pore region of the channel (Catterall, 1992). Other important regions of the channel have also been identified. Loops consisting of amino acids joining S5 and S6 behave as a selectivity filter which confers selectivity for sodium ions (*ibid.*). The voltage sensor of the channel has been identified with S4 of each of the 4 domains. The voltage sensor consists of a segment of positively charged amino acid residues ("+" in each of the four S4 segments in Fig. 1A), usually arginine, which are believed to move across the membrane if the voltage across the membrane is depolarized to a more positive potential from its resting potential of -80 mV. This movement is believed to trigger a conformational change which then causes the channel to open (*ibid.*). The loop between domain III and IV is believed to act as the "inactivation gate" of the sodium channel (Bennett *et al.*, 1995; Kellenberger *et al.*, 1997). The inactivation gate is the part of the channel believed to close off the channel shortly after the channel opens, taking the channel into the inactivated state. Deletions and mutations within the III-IV linker either slow down or prevent inactivation of the channel from occurring (Kellenberger *et al.*, 1997).

There has been much research devoted to identifying the amino acids which make up the functional regions of the sodium channel. Much of the functional studies of the sodium channel involve electrophysiological experiments (Catterall, 1992). The genes for sodium channels from various organisms (rat, human) have been cloned and these can be expressed in oocytes from the frog *Xenopus laevis* and the currents from these channels can be studied in whole cell patch clamp experiments, which involves observing the current between the inside of the cell and the outside buffer solution while "clamping" the voltage

at a given potential (Cahalan and Neher, 1992). *X. laevis* oocytes are very large cells (1 mm in diameter) and allow injection of nucleic acid, either DNA or RNA, which the cell will actively express (Soreq and Seidman, 1992). Because of the size of the cells, electrodes may be pushed into the cell so that the current across the cell membrane may be monitored and the voltage controlled ("clamped"). In this manner and by injecting sodium channel genes containing various mutations, much information has been obtained concerning the structure-function relationships within the channel.

In recent years, the sodium channel has been the focus of solution NMR studies. Although the entire sodium channel is too large a molecule to be studied by NMR, fragments of the protein may be used to provide clues as to their individual function within the ion channel. Kevit & coworkers isolated the linker between domains III and IV, which had been identified as the inactivation domain in previous chemical studies (Rohl *et al.*, 1999). The NMR study by Kevit & co-workers found that the domain III-IV linker was found to contain a stable α-helix which may act as the movable barrier to block ion flow through the channel.

2. The Use of Local Anesthetics and Scanning Cysteine Mutagenesis to Study Sodium Ion Channel Structure

Local anesthetics are drugs which bind reversibly to the sodium channel. These drugs bind with higher affinity to the inactivated state of the sodium channel and stabilize the inactive state of the channel (Hille, 1992). Local anesthetics share two common structural features: (1) one or more aromatic rings which confers hydrophobicity and (2) a hydrophilic group which is usually a secondary or tertiary amino group (Catterall & Mackie, 1996). The receptor site of the local anesthetics has been localized to at least two aromatic amino acid residues found in S6 of domain IV, F1764 and Y1771 (Fig. 1B; Ragsdale *et al.*, 1994). The interaction is likely a hydrophobic interaction between the aromatic ring of the local anesthetic drug and the aromatic residues.

Local anesthetics and quaternary alkyl ammonium analogs of the local anesthetics (which are have permanent positive charge in solution) have been used extensively to study structure of the sodium ion channel. It is postulated that there are two pathways for local anesthetics to enter the sodium channel, the hydrophobic pathway and the hydrophilic

pathway (Hille, 1977; Hille 1992). These two pathways were postulated on the basis of observations made with permanently charged quaternary local anesthetics (hydrophilic) and uncharged hydrophobic local anesthetics.

Quaternary analogs of local anesthetics are only effective in blocking the sodium channel when applied to the inside of the cell. Also, these drugs may only bind and unbind when the sodium channel is in the open state. The hydrophilic pathway to the local anesthetic binding site is the pathway through the open channel. It may only be taken by a local anesthetic drug that is protonated.

Hydrophobic local anesthetic drugs may arrive at the local anesthetic binding site by diffusing through the lipid bilayer that makes up the cell membrane or directly through channel wall. In this case, the channel need not be open for the drug to leave the binding site since it may leave by diffusion through the channel. The hydrophobic pathway to the local anesthetic binding site is taken by local anesthetic drugs in the uncharged form.

Since the local anesthetics are amines, the pH of the surrounding solution determines whether or not the drug is present in the charged or uncharged form, and therefore which pathway will be taken to enter and leave the sodium channel. It was found that the pH outside of the cell membrane affected the leakage of local anesthetics from the closed sodium channel while the inner pH had no effect (Hille, 1992). This suggested that only external protons have access to the local anesthetic binding site, that is, protons may pass through a selectivity filter facing the outside of the channel but may not pass through the gate which blocks the channel on the inside. Combining this observation with the observations discussed earlier suggests that the local anesthetic binding site is located in the pore below the selectivity filter and above the inactivation gate (Fig. 1B).

The activity of the local anesthetics combines the effects of utilizing both the hydrophobic and hydrophilic pathways, since the local anesthetic are amines which interconvert between the charged and uncharged forms. The preponderance of one pathway over the other, for a given local anesthetic drug, will depend on the pH of the external environment, the hydrophobicity of the drug, and the pKa of its amino group. In general, the more hydrophobic local anesthetic drugs are more potent and longer-acting. This is likely due to an increased affinity between the local anesthetic binding site and more hydrophobic drugs (Catterall & Mackie, 1996).

One of the most useful methods for studying the structure of the sodium channel as a whole has been the scanning cysteine mutagenesis method described in Part I. In this method, single residues within the sodium channel are mutated to cysteine, and then the residue is reacted with thiol reagents to study the effect (if any) on ion flow through the channel. This method has provided many useful insights into the structure of the pore region of the sodium channel (Kellenberger *et al.*, 1996; Bénitah *et al.*, 1996; Pérez-García *et al.*, 1996). Cysteine mutants of sodium channels have been generated and these mutants were tested with a variety of thiol reagents, including cadmium and the methyl thiosulfonate (MTS) reagents (Akabas *et al.*, 1992; see Part I, Section A). The purpose of using these reagents was to observe the effect on channel behaviour and on ion flow during the channel's opening. Cysteine mutants were tested for sensitivity to cadmium block. Cd^{2+} will bind to the thiol residue of cysteine and block Na^{+} flow through the channel while it is lodged in the pore (Tomaselli *et al.*, 1995). The MTS reagents originally used by Akabas & coworkers (*ibid.*) have been employed to probe the accessibility of various sites on the sodium channel and to identify whether residues were located near the pore region. The advantage of these MTS reagents is their selectivity for reaction with the thiol group of cysteine residues (Kenyon & Bruice, 1977). New types of MTS reagents would be highly useful in the application of scanning cysteine mutagenesis to studying the structure of the sodium channel.

As discussed above, local anesthetics are extremely useful tools in studying the pore region of the sodium channel. They may be used to identify amino acids lining the pore that are near the local anesthetic binding site (Yarov-Yarovoy *et al.*, 2001). They may also be used to identify amino acids which form part of the selectivity filter and inactivation gate, since the mutation of these amino acids will affect the entry of quaternary local anesthetics from the outside of the channel (Sunami *et al.*, 2001). The outer pore region and the selective filter are believed to be composed of the loops joining S5 and S6 of each domain. Mutations of residues within the selectivity filter have been shown to affect local anesthetic binding (Sunami *et al.*, 1997). The distance between the local anesthetic binding site and the selectivity filter is likely small (Balser, 1999). An MTS derivative of a local anesthetic drug would be an extremely useful probe of the sodium channel. In combination with scanning cysteine mutagenesis, an MTS derivative of a local anesthetic could give information on the proximity of various amino acid residues to the local

anesthetic binding site. Amino acids which are believed to form the selectivity filter and the inactivation gate would be of particular interest in such studies.

Backx & coworkers tested a series of novel benzocaine derivatives containing linkers of variable length between the benzocaine molecule and the MTS reactive centre and found that the benzocaine-MTS derivatives in cardiac sodium channels caused "irreversible" blockade (Li *et al.*, 1999). Benzocaine (4-aminobenzoic acid ethyl ester; Merck Index 11th ed.) is a member of the group of local anesthetic drugs. Because the amino group of benzocaine is on the benzene ring, protonation of the amino group is not as favourable as the unprotonated state, which is resonance stabilized by the adjacent phenyl ring. The pKa of benzocaine is 2.5, meaning that benzocaine is present only in the neutral form at physiological pH (CRC Handbook of Chemistry & Physics, 2000). Therefore, benzocaine may only take the hydrophobic pathway to enter the sodium channel (Li *et al.*, 1999). Benzocaine leaves the sodium channel in under 20 milliseconds, which means the duration of its activity is very short compared to other local anesthetic drugs (Hille, 1992). The creation of an MTS derivative of other types of local anesthetic drugs, which have different chemical properties from benzocaine, would be highly useful for probing the pore region of the ion channel near since it would help to identify amino acid residues near the local anesthetic binding site. An MTS derivative of a local anesthetic which enters the channel via the hydrophilic pathway will bind to cysteine residues which are exposed in the pore during the open state of the channel.

3. The Development of Thiol-Reactive Analogs of Lidocaine

Lidocaine is a local anesthetic drug which is widely used in the treatment of cardiac arrhythmia (Catterall & Mackie, 1996). There are side effects associated with its use, primarily due to nonspecific block of all sodium channels throughout the body. Cardiac sodium channels, found in the sinoatrial node (the "pacemaker") which enervates the heart muscle, is structurally different from skeletal sodium channels found in nerves stimulating the skeletal muscles. The notable difference is the presence of a single free cysteine residue found in the pore region of the cardiac sodium channel, in the domain I S5-S6 loop (Chiamvimonvat *et al.*, 1996; Backx *et al.*, 1992). A thiol-reactive version of lidocaine, when reacted to this unique cysteine residue, could be an extremely useful probe of the pore region of the channel in relation to the local anesthetic binding site. Also,

scanning cysteine mutagenesis may be employed so that thiol-reactive lidocaine analogues could be used to study the pore regions of other sodium channels besides the cardiac sodium channel. Residues thought to be in the vicinity to the pore region and/or the lidocaine-binding site can be mutated to cysteine, and an observed irreversible block of the channel would indicate that the lidocaine molecule was covalently attached to the channel.

Lidocaine is more hydrophobic than benzocaine, due to the ethyl groups on the amino nitrogen. The amino group of lidocaine is a tertiary alkyl amine; the pKa of lidocaine is 7.856 (Löfgren, 1948). Therefore, lidocaine will be present in both the charged and uncharged forms *in vivo* and may therefore enter and leave the local anesthetic binding site within the sodium channel by either the hydrophobic or the hydrophilic pathway. Furthermore, due to its increased hydrophobicity, it leaves the sodium channel in the range of a few hundred milliseconds (ms), compared to benzocaine which leaves the sodium channel in less than 20 ms (Hille, 1992). This means that lidocaine is a more potent local anesthetic than benzocaine.

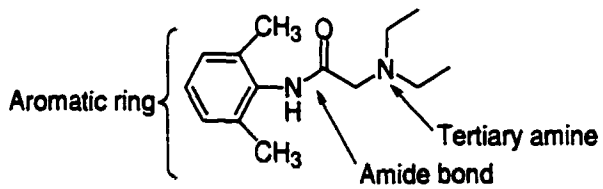


Figure 2: Structure of Lidocaine

Lidocaine is a tertiary amine, with two ethyl groups on the amino nitrogen, and a N-(2,6-dimethylphenyl)acetamide group as the third group (Fig. 2). Lidocaine, as an amine, can be charged or uncharged and thus lidocaine is soluble in either the aqueous environment surrounding the cell or the lipid membrane. The hydrophobic alkyl chains and the aromatic ring allow lidocaine to diffuse from the outside of the cell into the membrane and find its binding site within the channel. The presence of the phenyl group is particularly important: the presence of the phenyl ring correlates with greatly increased local anesthetic activity over amines with simple alkane arms, likely due to a hydrophobic interaction between the phenyl group of lidocaine and the aromatic residues lining the lidocaine receptor site (Zamponi & French, 1994; Ragsdale *et al.*, 1994). It has been shown that changing the length of the alkyl chains on the amino nitrogen does not greatly

affect anesthetic property, *i.e.* the structures of these alkyl chains are not as important to the anesthetic activity as the aromatic group (Liu *et al.*, 1994). This allows the possibility of altering one of the alkyl chains to introduce a thiol-reactive group at the end of the chain without significantly affecting the anesthetic property of the drug.

Lidocaine was first synthesized by Löfgren and Lundqvist in 1946 (Fig. 3A) (Löfgren and Lundqvist, 1948; Astra Pharmaceutical Products, 1960). Using that synthesis as the starting point, a retrosynthesis was proposed to obtain a target compound which resembles lidocaine as closely as possible with the exception of one of the alkyl chains which contains the thiol-reactive group at its terminus (Fig. 3B).

The proposed target compound fulfills two criteria. It retains the required functional groups on the amino nitrogen for local anesthetic ability, the alkyl chains and the N-(2,6-dimethylphenyl)acetamide group. The synthetic method also allows for flexibility in the length and the chemical nature of the carbon chain between the nitrogen and the thiol-reactive group. The reason for this is that the distance between the local anesthetic binding site and the cysteine residue to which the lidocaine group will be anchored is unknown and may be of a dynamic nature, changing with the opening and closing of the channel. The ability of the aromatic ring of the lidocaine analog to find its binding site, once it has reacted with the cysteine, will depend to a large extent on the length of the alkyl chain linker.

This synthetic route entailed the synthesis of halogenated versions of lidocaine. In 1969, Astra chemists made several halogenated versions of lidocaine, by reacting asymmetric secondary amines of the form $NH(CH_3)((CH_2)_x)OH$ with the chloride compound shown after the first step of the synthesis (Fig. 3A) and then converting the alcohol group to a halogen using $SOCl_2$ (Ross *et al.*, 1969). These were found to be long-acting anesthetics (*ibid.*; Stubbins, 1970). These versions of lidocaine are not selective towards which nucleophile will react with them. An MTS version of lidocaine would be highly desirable over these halogenated versions because its selectivity for reaction with thiol groups and its use in scanning cysteine mutagenesis. Backx & coworkers' experiments with MTS derivatives of benzocaine have shown that these MTS reagents are indeed selective for the cysteine residue within the pore of the cardiac sodium channel (Li *et al.*, 1999).

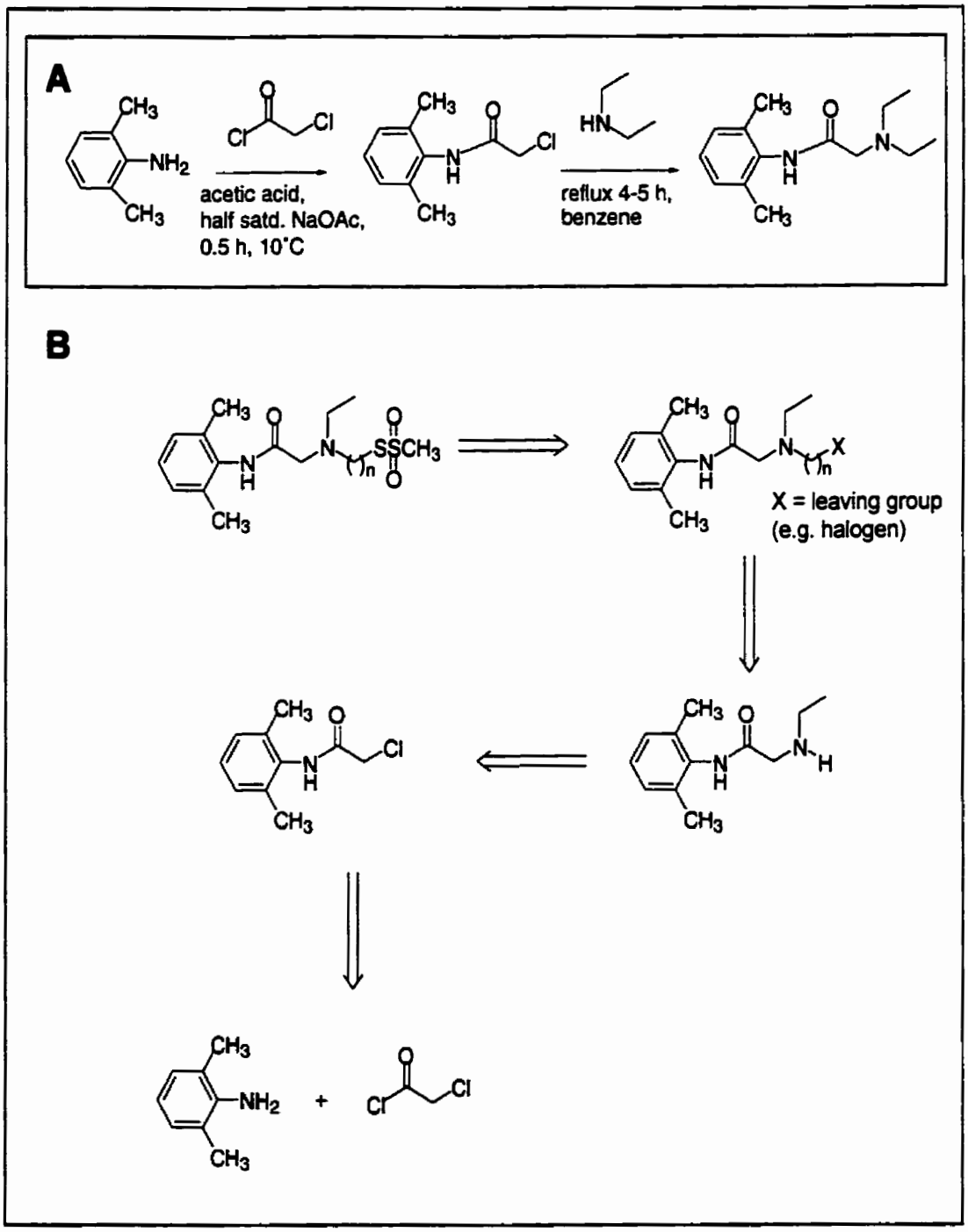


Figure 3: (A) Löfgren and Lundqvist's original synthesis of lidocaine (1948);
(B) Proposed route of synthesis of thiol reactive versions of lidocaine.

A number of obstacles are present in the synthesis strategy centering on the reactivity of the amine. Tertiary amines with three different arms are known to be difficult to synthesize because a linear synthesis route, where one arm of the amine is added at a time, allows for the possibility of multiple reactions at the amino nitrogen (Loudon, 1988). This generally results in a mixture of compounds. In addition, there is the possibility of self-reaction because there is a leaving group and a nucleophile present in the same molecule.

The synthesis was carried out as planned, and a group of lidocaine analogs was synthesized, with a variable alkyl linker containing 2, 3, and 4 carbons between the amino nitrogen and the leaving group. The exchange of halogen for the methylthiosulfonate group was accomplished in the case of the 3-carbon chain compound.

SECTION B: THE SYNTHESIS OF THIOL-REACTIVE LIDOCAINE ANALOGS

All chemicals were obtained from Aldrich (Milwaukee, WI) unless otherwise noted.

1. Synthesis of 2-chloro-N-(2,6-dimethylphenyl)acetamide, DMPA-Cl

The synthesis of DMPA-Cl (**2**) is outlined in Scheme 1. 2,6-Dimethylaniline (**1**) (0.025 mol, 3.1 mL) was added to cold glacial acetic acid (20 mL), and the mixture cooled in an ice bath. To this, chloroacetyl chloride (2.2 mL, 0.0275 mol) was added. A solution of 45% w/v sodium acetate (25 mL) was then added to the mixture (now pH 4.0). The product, a white solid, precipitated out immediately and the whole mixture stirred (simultaneously cooling on ice) for 0.5 h. The precipitate was filtered off and washed repeatedly with mild HCl (e.g. 10% HCl) to remove unreacted 2,6-dimethylaniline, seen as a brown coloration in the washings. A final washing with H_2O removed most acid remaining. Next, the solid was dissolved in chloroform, and extracted with H_2O. The chloroform layer was dried over $MgSO_4$ and the solvent removed under vacuum, to produce a feathery white needle-like solid, which is DMPA-Cl (**2**). The yield of the reaction was approximately 40-50%. TLC ($CHCl_3$/EtOH 3:1; silica): R_f = 0.8. Low resolution MS-EI detects M^+ at 197, 199 (for ^{35}Cl and ^{37}Cl isotopes); high resolution MS-EI obtained a mass of 197.061603 for the ^{35}Cl-containing M^+ ($C_{10}H_{12}NO^{35}Cl$, calculated mass 197.060741). ^{1}H-NMR ($CDCl_3$): δ2.25 (s, 6H), δ4.25 (s, 2H), δ7.1 (m, 3H), δ7.85 (broad s, 1H).

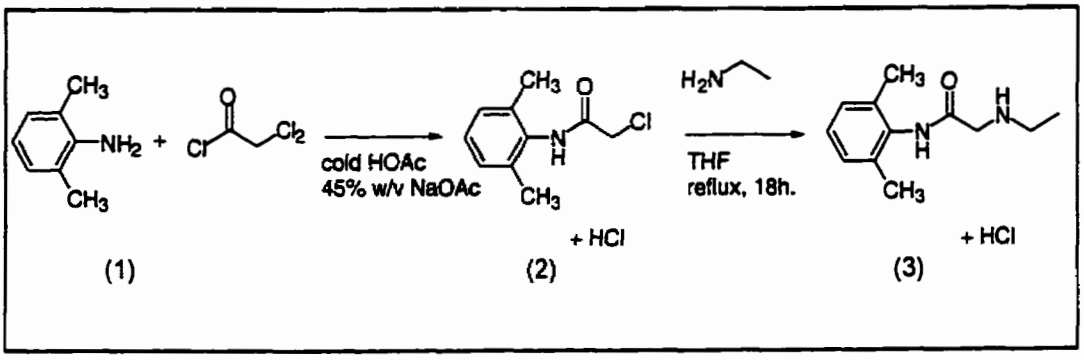
2. Synthesis of 2-(ethylamino)-N-(2,6-dimethylphenyl)acetamide, DMPA-NHEt

The synthesis of DMPA-NHEt (**3**) is also shown in Scheme 1. DMPA-Cl (**2**) was added in 4-5 times molar excess to ethylamine, obtained as a 2 M solution in THF. Amounts used: 28 mL of 2 M NH_2Et in THF (56 mmol); DMPA-Cl, 2.59 g (14.6 mmol). The reaction mixture was refluxed for 18 h. An initial attempt yielded no reaction, due to the fact that ethylamine is extremely volatile. In a regular reflux apparatus, ethylamine evaporated out of the THF solution almost immediately upon heating and escaped to the atmosphere. The condenser must be extremely cold to prevent loss of ethylamine. A mixture of water and ethylene glycol cooled to ≤ 4°C was circulated

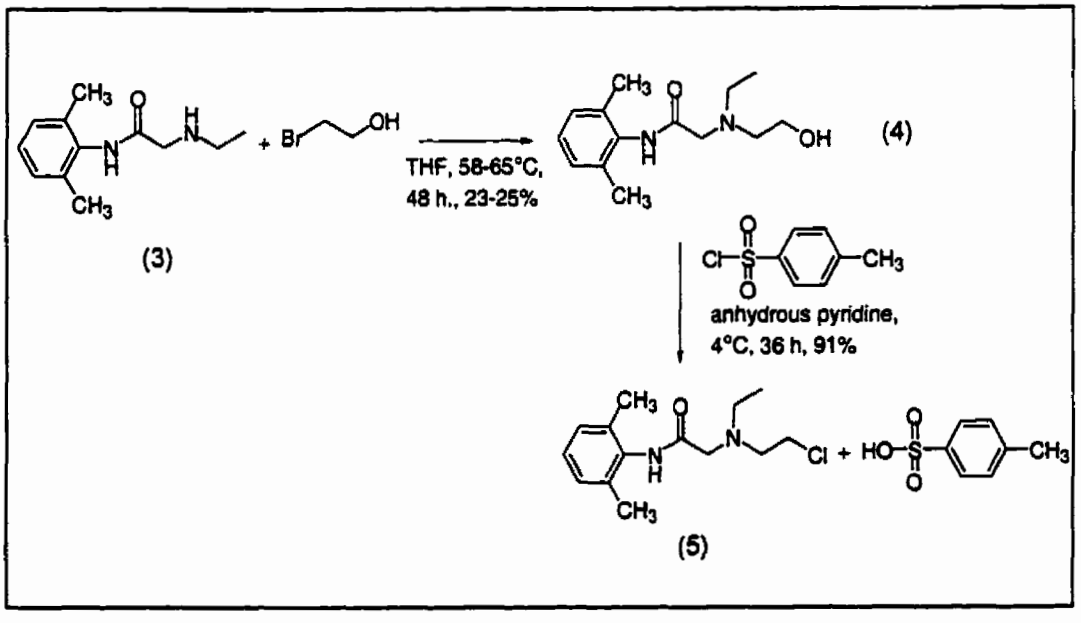
through an insulated condenser, which was also sealed with a stopper. During the reaction, $EtNH_3Cl$ crystals precipitated out of solution. The reaction mixture was then filtered to remove these crystals. THF was removed by vacuum, leaving a greenish-yellow oil which may have solid in it. This was washed with a small amount of strong acid (pH 0-1 HCl), causing the formation of some white solid. The mixture was vacuum filtered to remove the white solid (unreacted DMPA-Cl). The filtrate was then made basic with 2 M NaOH, and the solution evaporated to yield a large amount of white solid. This solid was then washed with diethyl ether, and the resultant diethyl ether solution evaporated under ambient conditions to give pure DMPA-NHEt (**3**) as white flat crystals. The yield of the reaction is 43%. TLC ($CHCl_3$/EtOH 3:1; silica): R_f = 0.4. High resolution MS-EI: 206.142684 for M^+ ($C_{12}H_{18}N_2O$ calculated mass 206.141913). ^{1}H-NMR ($CDCl_3$): δ1.19 (t, 3H), δ2.25 (s, 6H), δ2.8 (q, 2H), δ3.45 (s, 2H), δ7.1 (s, 3H), δ8.83 (broad s, 2H).

3. Synthesis of 2-((ethyl, 2'-hydroxylethyl)amino)-N-(2,6-dimethylphenyl) acetamide, DMPA-N(Et)(EtOH)

The reaction is shown in Scheme 2. DMPA-NHEt (**3**) was gently warmed (58-65°C) with 2-bromoethanol, present in 2-fold excess to the amine, in anhydrous THF for 48 h. Amounts used were: DMPA-NHEt, 300 mg (1.45 mmol); 2-bromoethanol, 206 mL (2.90 mmol). The precipitate formed during the course of the reaction (the hydrobromide salt of DMPA-NHEt) was then filtered off and the filtrate dried under high vacuum (≤0.1 torr) to give a heavy brownish yellow oil. This oily solid was thoroughly washed with H_2O, to give a dry and granular yellowish white solid. This solid, identified as pure DMPA-NH(Et)(EtOH) (**4**) was then dried overnight under high vacuum. The yield of the reaction was 23-25%. The $(DMPA\text{-}NH_2Et)^+Cl^-$ precipitate was collected and dissolved in 2 M NaOH to recover DMPA-NHEt (**2**). The resultant basic solution was evaporated to give a white solid which was washed with ether. Evaporation of the ether solution gives pure DMPA-NHEt (**2**). TLC ($CHCl_3$/EtOH 3:1; silica): R_f = 0.8. Low resolution MS-EI: M^+ is 250. High resolution MS-EI found mass of 250.167819 for M^+ ($C_{14}H_{22}N_2O_2$ calculated mass 250.168128). ^{1}H-NMR ($CDCl_3$): δ1.27 (t, 3H), δ2.22 (s, 6H), δ2.82 (t, 2H), δ2.77 (q, 2H), δ2.82 (t, 2H), δ3.33 (s, 2H), δ7.1 (s, 3H), δ8.8 (broad s, 1H).



Scheme 1: Synthesis of DMPA-Cl and DMPA-NHEt



Scheme 2: Synthesis of DMPA-N(Et)(EtOH) and DMPA-N(Et)(EtCl)

4. Synthesis of 2-((ethyl,2'-chloroethyl)amino)-N-(2,6-dimethylphenyl) acetamide, DMPA-N(Et)(EtCl)

The reaction is shown in Scheme 2. Tosyl chloride was purified according to Fieser & Fieser's method (1967). DMPA-N(Et)(EtOH) (**4**) was dissolved in distilled pyridine and cooled to < 4°C. Tosyl chloride, in 2-fold excess to the alcohol, was added slowly to the solution which turned bright yellow. Amounts used: alcohol, 50 mg (0.20 mmol); pyridine, 1.0 mL (12 mmol); tosyl chloride, 76 mg (0.40 mmol). The reaction mixture was left to stand at 4°C and was complete after approximately 36 h. During this time the solution turned from bright yellow to a dull brownish red. A vacuum filtration was carried out to remove any pyridine hydrochloride crystals. The filtrate was then poured directly on to a mixture of crushed ice and water, and stirred (with scratching) with a glass rod. The mixture was then extracted with diethyl ether, and the ether layer was dried over Na_2SO_4 and reduced *in vacuo* to give a heavy brownish red pungent-smelling oil, identified as pure DMPA-N(Et)(EtCl) (**5**). The yield of the reaction was 91%. TLC ($CHCl_3$/ethyl acetate 3:2; silica): R_f = 0.4 (TsCl, R_f = 0.7; DMPA-N(Et)(EtOH), R_f = 0.1). Low resolution MS-CI: MH^+ is 269, 271 (for ^{35}Cl and ^{37}Cl); the chloride containing fragment was detected at m/z = 120, 122. High resolution MS-CI detected the ^{35}Cl-containing MH^+ at 269.140739 (MH^+ $C_{14}H_{22}N_2O^{35}Cl$, calculated mass 269.142066). H-NMR ($CDCl_3$): δ1.2 (t, 3H), δ2.24 (s, 6H), δ3.0 (t, 2H), δ3.33 (s, 2H), δ3.67 (t, 2H), δ7.1 (m, 3H), δ8.9 (broad s, 1H).

5. Synthesis of 2-((ethyl,3'-bromopropyl)amino)-N-(2,6-dimethylphenyl) acetamide, DMPA-N(Et)(PrBr)

The reaction is as shown in Scheme 3. DMPA-NHEt (**3**) and 1,3-dibromopropane, in 3-fold excess to the amine, were combined in anhydrous THF, so that the amine is present in a concentration of 0.1 M or less. Amounts used were: DMPA-NHEt, 504 mg (2.42 mmol); 1,3-dibromopropane, 741 mL (7.27 mmol); THF 24 mL. This was gently warmed to a temperature well below reflux (55-57°C). After 24 h, another equivalent of 1,3-dibromopropane (247 mL) was added. The precipitate formed during the course of the reaction (the hydrobromide salt of DMPA-NHEt) was filtered off. The filtrate was reduced *in vacuo* to give a heavy yellow oil. A dilute solution of hydrochloric acid was

stirred into the oil. The mixture was then extracted with diethyl ether to remove unreacted 1,3-dibromopropane. The aqueous layer was taken and made basic with 2 M NaOH; the free base precipitated out of solution, and was filtered off directly. The fine white powder was dried *in vacuo*, give pure DMPA-N(Et)(PrBr) (**6**). The yield of the reaction was 33%. TLC ($CHCl_3$/EtOH 3:1; silica): R_f = 0.8. Low resolution MS-CI: M^+ is 326, 328 (for ^{79}Br and ^{81}Br); the bromide containing fragment was detected at m/z = 178, 180. High resolution MS-CI found a mass of 326.098852 for the ^{79}Br-containing compound ($C_{15}H_{23}N_2O^{79}Br$, calculated mass 326.099375). ^{1}H-NMR ($CDCl_3$): δ1.2 (t, 3H), δ2.12 (m, 2H), δ2.25 (s, 6H), δ2.75 (q, 2H), δ2.8 (t, 2H), δ3.37 (s, 2H), δ3.5 (t, 3H), δ7.1 (m, 3H), δ8.72 (weak broad s, 1H).

6. Synthesis of 2-((ethyl,3'-iodopropyl)amino)-N-(2,6-dimethylphenyl)acetamide, DMPA-N(Et)(PrI)

The reaction is shown in Scheme 3. The bromide compound DMPA-N(Et)(PrBr) (**6**) (75 mg; 0.23 mmol) was dissolved in acetone (5 mL), and 3 equivalents of sodium iodide (104 mg, 0.69 mmol) added to the mixture. The reaction was complete after about 0.5 h. The precipitate (NaBr) was filtered off and the acetone evaporated off the filtrate to give a yellow solid. This yellow solid was washed with water, leaving behind a sticky yellow solid. This solid was identified as the iodide derivative, DMPA-N(Et)(PrI) (**7**), with some impurities present. Yield of the reaction is ≥99%. TLC ($CHCl_3$/EtOAc 3:2; silica): R_f = 0.13. Low resolution MS-EI: M^+ is 374, with the iodide containing fragment observed at m/z = 226. No bromide doublets are observed in the mass spectrum, which suggests complete conversion to the iodide derivative. High resolution MS-EI: M^+ has mass of 374.085364 ($C_{15}H_{23}N_2OI$, calculated mass 374.085516). ^{1}H-NMR ($CDCl_3$): δ1.18 (t, 3H), δ2.08 (m, 2H), δ2.22 (s, 6H), δ2.72 (q, 2H), δ2.72 (t, 2H), δ3.24 (t, 2H), δ3.28 (s, 2H), δ7.1 (m, 3H), δ8.72 (weak broad s, 1H).

7. Synthesis of 2-((ethyl,3'-methylthiosulfonate-propyl)amino)-N-(2,6-dimethylphenyl) acetamide, DMPA-N(Et)(PrMTS)

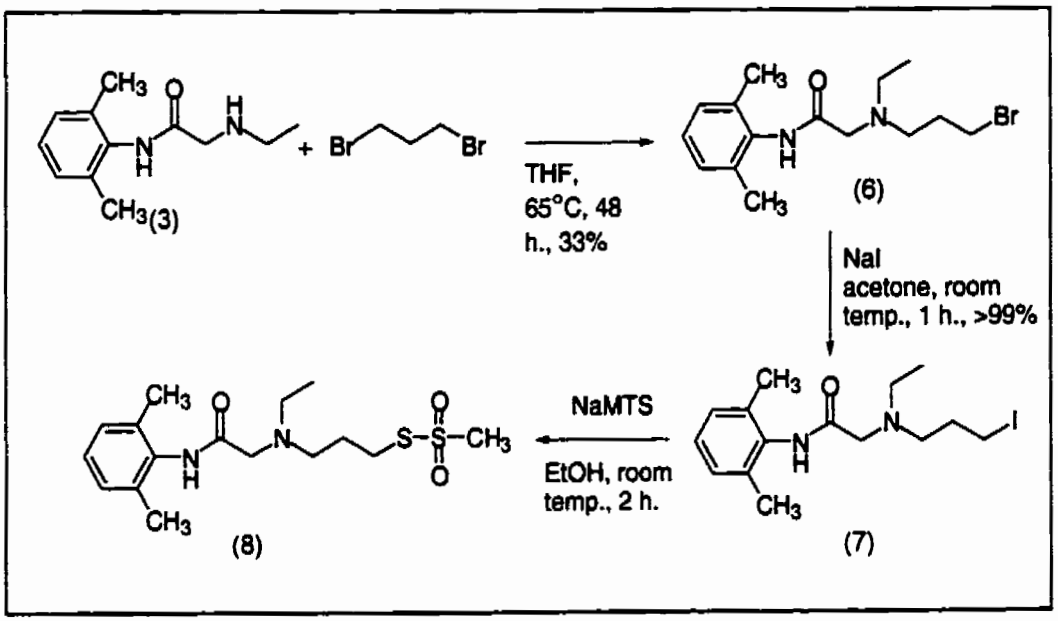
The reaction is shown in Scheme 3. DMPA-N(Et)(PrI) (**7**) (15.2 mg, with some impurity, approximately 0.04 mmol) was dissolved in anhydrous ethanol (5 mL) and sodium methylthiosulfonate (NaMTS), in approximate 3-fold excess (16 mg, 0.12 mmol)

was added. The mixture was stirred at room temperature; the reaction was essentially complete after 1 h. The acetone was removed from the reaction mixture, leaving a sticky yellow solid with a pungent onion-like smell. The solid was washed with chloroform to give a yellow solution with a white solid, which was filtered off. The filtrate was reduced *in vacuo* to give a dark yellow oil. The oil contained a mixture of the methylthiosulfonate derivative, DMPA-N(Et)(PrMTS) (**8**), and the iodide derivative (**7**), with the former in predominance (as suggested by the integration of the peak for the methylene protons adjacent to the iodide in DMPA-N(Et)(PrI) and TLC). TLC ($CHCl_3$/EtOAc 3:2; silica): R_f = 0.25. Low resolution MS-EI: MH^+ detected at 359; MTS-containing fragment at 210; M^+– MTS at 279. High resolution MS-EI: MH^+ has mass of 359.144913 ($C_{16}H_{26}N_2O_3S_2$, calculated average mass 358.513; MH^+ is $C_{16}H_{27}N_2O_3S_2$, calculated mass with most abundant isotopes is 359.146312); M^+–MTS has mass of 279.152382 ($C_{15}H_{23}N_2OS$, calculated mass 279.153108). ^{1}H-NMR ($CDCl_3$): δ1.15 (m, 2H), δ1.40 (t, 3H), δ2.19 (s, 6H), δ2.7 (m, 4H), δ3.16 (q, 2H), δ3.24 (s, 5H), δ7.0 (m, 3H), δ8.63 (weak broad s, 1H).

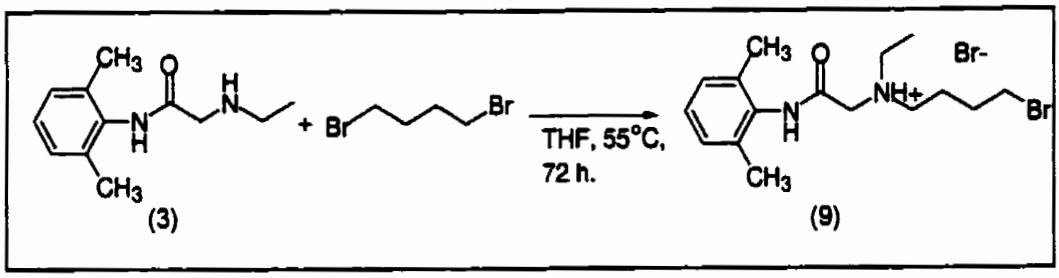
8. Synthesis of 2-((ethyl,4'-bromobutyl)amino)-N-(2,6-dimethylphenyl) acetamide, [DMPA-NH(Et)(BuBr)]$^+$Br$^-$

The reaction is shown in Scheme 4. DMPA-NHEt (**3**) and 1,4-dibromobutane, in 6-fold excess to the amine, were combined in anhydrous THF so that the amine is present in a concentration of 0.1 M or less. Amounts used were: DMPA-NHEt, 115.6 mg (0.560 mmol); 1,4-dibromobutane, 405.8 mL (3.362 mmol); THF (7 mL). This was gently warmed (approximately 55°C). After 24 h, the reaction was essentially complete, and the mixture consisted of a white precipitate suspended in yellow liquid. The precipitate is a mixture of the hydrobromide salts of DMPA-NHEt and the product, [DMPA-NH (Et)(BuBr)]$^+$Br$^-$ (**9**) in an approximate ratio of 1:1. The crude mixture was characterized as follows. TLC (100% MeOH; silica): R_f (product) = 0.11; R_f (DMPA-NHEt) = 0.54. Low resolution MS-EI: M^+ detected at 340/342, bromide-containing fragment detected at 192/194. High resolution MS-EI: M^+ has mass of 340.116348 for the compound containing ^{79}Br, 342.112244 for the ^{81}Br compound ($C_{16}H_{25}N_2O^{79}Br$, calculated mass 340.115025; $C_{16}H_{25}N_2O^{81}Br$, calculated mass 342.112979). The NMR of the crude was compared against the NMR spectrum of pure (DMPA-NH_2Et)$^+X^-$ in D_2O, obtained as

the precipitate in the reaction as described in Section 5 above. ^{1}H-NMR (D_2O) of the desired product, [DMPA-NH(Et)(BuBr)]$^+$Br$^-$: δ1.40 (t, 3H), δ2.16 (s, 6H), δ2.19 (m, 4H), δ3.61 (q, 2H), δ3.72 (m, 2H), δ4.39 (s, 2H), δ7.15 (m, 3H). ^{1}H-NMR (D_2O) of the side product, [DMPA-NH_2Et]$^+$Br$^-$: δ1.29 (t, 3H), δ2.15 (s, 6H), δ3.14 (q, 2H), δ4.10 (s, 2H), δ7.15 (m, 3H).



Scheme 3: Synthesis of DMPA-N(Et)(PrBr), DMPA-N(Et)(PrI), and DMPA-N(Et)(PrMTS)



Scheme 4: Synthesis of [DMPA-NH(Et)(BuBr)]$^{+}$Br^{-}

SECTION C: RESULTS AND DISCUSSION

The synthetic route to obtain the thiol reactive analogs was designed so that the target compounds resembled lidocaine as closely as possible. Thus, the N-(2,6-dimethylphenyl)acetamide group (abbreviated as DMPA) and ethylamino moieties were kept constant, with the second alkyl chain as the variable. The synthesis, as outlined, depended on the successful synthesis of the secondary amine, DMPA-NHEt. From there, one could potentially create a large library of lidocaine analogs. The synthesis of DMPA-NHEt serves as the starting point in the synthesis of a variety of lidocaine analogs, with linking groups of variable length and/or flexibility between the lidocaine group and the thiol-reactive group. It was anticipated that the nucleophilicity of the amine would give rise to several possible side reactions in each step, especially with attempts to react with the dibrominated compounds, and mixtures of compounds would complicate purification procedures. The syntheses of the compounds are discussed below, with pertinent observations on each.

1. Synthesis of DMPA-Cl and DMPA-NHEt

The synthesis of DMPA-Cl was adapted from Löfgren & Lundqvist's original synthesis of lidocaine (1948). The reaction is as outlined in Scheme 1. 2,6-Dimethylaniline (**1**) attacks the acid chloride group of chloroacetylchloride (**2**). To control the reaction, it was necessary to control the degree of protonation of the amine, and this could be achieved with low temperature and an acetate buffer system (pH 4.0). This reaction gave 2-chloro-N-(2,6-dimethylphenyl)acetamide (DMPA-Cl) (**3**) in a yield of 40-50%. In the workup of the reaction, care was taken to completely rinse off all trace of 2,6-dimethylaniline (with dilute HCl) since it would react again in future syntheses. The chloroform/H_2O extraction was a straightforward means of obtaining the compound in pure form.

The synthesis of DMPA-NHEt was perceived to be one of the more difficult reactions to accomplish. There were three problems associated with the reaction as given: ethylamine presented difficulties in handling since it is a gas at room temperature, the product is a secondary amine which (in theory) can react with another molecule of DMPA-Cl, and finally, a molecule of HCl is generated in the course of the reaction so

depending on which amine is the stronger base, either ethylamine or DMPA-NHEt could precipitate out of the solution. The general trend of successful secondary amine syntheses employed a large excess of the halogen to the amine (Spialter & Pappalardo, 1965; Patai, 1968; March, 1988) so this approach was also used.

This reaction was initially attempted with ethylamine hydrochloride and sodium hydroxide was added to obtain the free amine (which is the active nucleophile) in ethanol, but this was not successful due to NaOH's low solubility in ethanol. Moreover, a relatively large amount of NaOH was required to obtain enough free amine. The resultant strongly basic solution tended to hydrolyze the amide bond formed in the previous reaction and negligible amounts of product were formed. The use of an organic base like triethylamine was avoided because of the complication to the purification of the product.

The use of ethylamine dissolved in THF greatly simplified the reaction and the subsequent workup. The product DMPA-NHEt was recovered from the solution, while the "side-product" ethylamine hydrochloride precipitated out of solution during the course of the reaction and could be filtered off directly.

In separating out the product, unreacted DMPA-Cl and the product DMPA-NHEt were separated by protonation/deprotonation of the amine with acid and base. The most expedient method of obtaining the free amine involved evaporation of the basic solution and washing the resultant solid with ether; pure crystals of DMPA-NHEt were obtained by evaporation of the ether at room temperature. The yield of the reaction was 43%, which was considered good since two molecules of ethylamine are consumed for every molecule of DMPA-NHEt produced (meaning a maximum yield of 50%)

DMPA-NHEt is also known as monoethylglycinexylidide (MEGX) and it is the first product of the metabolism of lidocaine by the liver (Foye *et al.*, 1995). It also behaves as a local anesthetic (*ibid.*), likely because it retains the portions of the lidocaine molecule which are involved in its anesthetic action, *i.e.* the aromatic group, the amino portion and one of the alkyl chains.

2. Synthesis of DMPA-N(Et)(EtOH) and DMPA-N(Et)(EtCl)

The sequence of reactions is shown in Scheme 2. DMPA-NHEt (**3**) was reacted with 2-bromoethanol to give DMPA-N(Et)(EtOH) (**4**), as shown in Scheme 2. The

reaction is driven to completion by adding a large excess of the 2-bromoethanol. Because of 2-bromoethanol's reactivity as a halohydrin (as a vicinal halohydrin it is able to form epoxides; Loudon, 1988), a low temperature was employed; it was noted that high temperatures and vigorous refluxing resulted in lowered yields and impurities in the isolated product. $(DMPA\text{-}NH_2Et)^+Br^-$ crystals precipitated out of solution during the course of the reaction. After removal of all solvent *in vacuo*, the product was partially crystallized out of the remaining 2-bromoethanol, present as a thick brown oil. Washing with water removed the oil (composed of 2-bromoethanol), leaving behind pure DMPA-N(Et)(EtOH) as a granular yellowish-white solid. The yield of the reaction was rather low, 23%, and it is believed to be due to 2-bromoethanol's tendency to degrade over the course of the reaction.

A tosylation reaction, according to the procedure given by Fieser & Fieser (1967), was carried out with the alcohol DMPA-N(Et)(EtOH). The aim was to convert the alcohol to the sulfonate ester, which is a good leaving group, and then react it with sodium methylthiosulfonate. The mechanism of this type of reaction is proposed to be nucleophilic catalysis by the solvent, pyridine (Fig. 4) (Loudon, 1988; March, 1985). Pyridine first attacks the *p*-toluenesulfonyl chloride, to form the activated quaternary compound with the pyridine nitrogen bonded to the sulfur. The alcohol then attacks the sulfur centre. The chloride ion which was displaced in the first step of the reaction is then supposed to deprotonate this complex to form the sulfonate ester, but this is not what happened. Instead the chloride actually displaced the entire sulfonyl group to give DMPA-N(Et)(EtCl).

The formation of the chloride derivative was first made apparent in the NMR spectrum, where the doublets for the aromatic protons and the singlet for the *p*-methyl group of the tosyl are not present. The only difference between this NMR spectrum and the NMR spectrum of the alcohol was the definite shift of the triplet representing H_i further downfield from $\delta 3.78$ to $\delta 3.67$ (H_i is the terminal carbon of the alkyl chain being derivatized; it bears the hydroxyl group in the alcohol compound). This indicated that something had happened at this carbon centre. The mass spectrum provided conclusive evidence that the chloride compound was present; the tosylated compound has a much higher molecule weight but is not observed at all. The chlorine isotope distribution is also

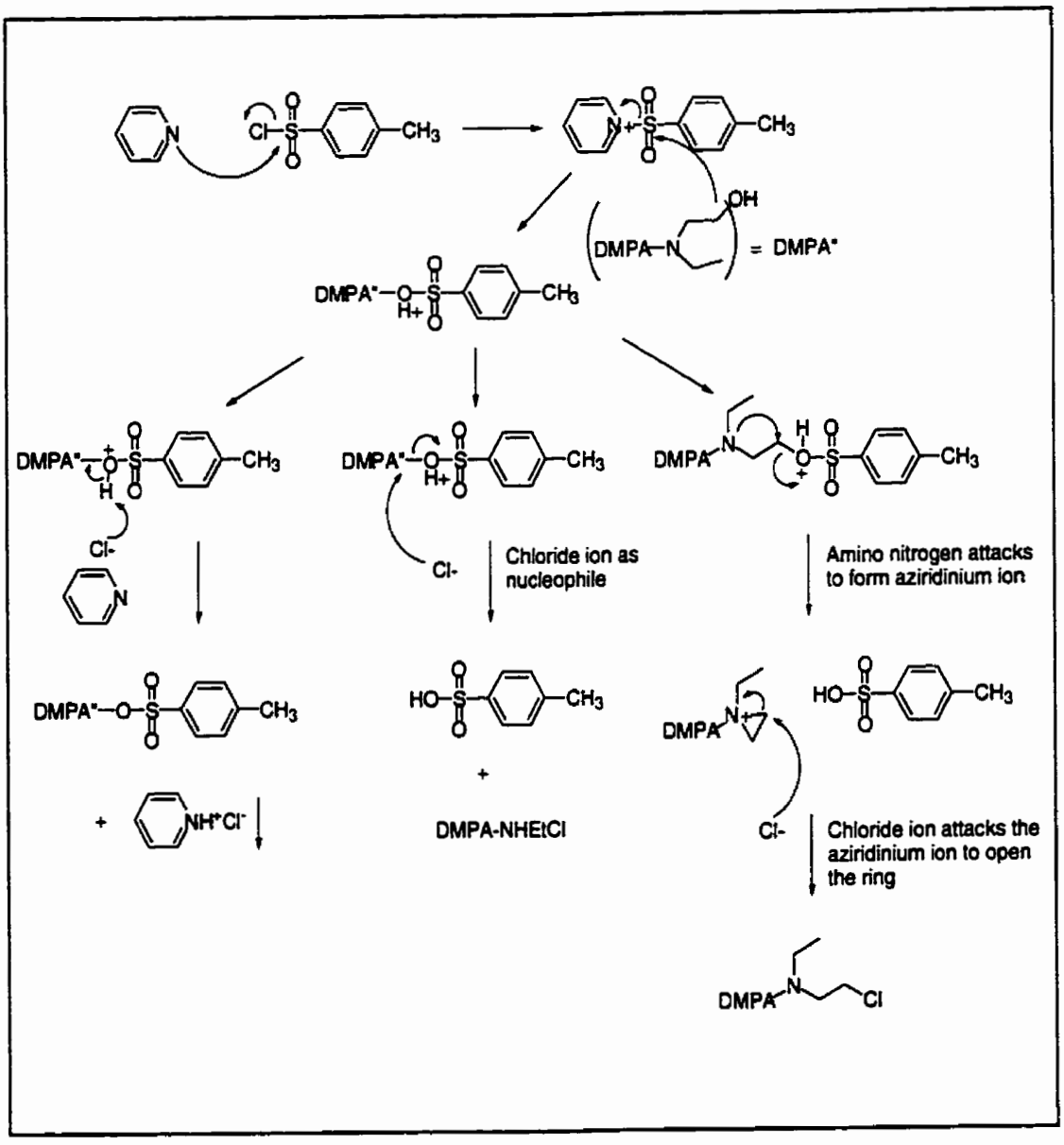


Figure 4: Proposed reaction mechanisms for the tosylation reaction of DMPA-N(Et)(EtOH) which results in the formation of the chloride derivative DMPA-N(Et)(EtCl)

observed in the mass spectrum. This reaction was not expected and it would suggest that the chlorine compound is more stable than the tosylate.

Two possible mechanisms were proposed to account for the formation of the chloride compound. These are outlined in Fig. 4 in the middle and right reaction sequences. In the first case, the chloride ion acts as the nucleophile towards the terminal carbon bonded to the sulfonate oxygen. This forms the product DMPA-N(Et)(EtCl) directly. In the second case, the amino nitrogen performs an intramolecular attack on the terminal carbon to form an aziridinium ion. This type of reaction to form an aziridine ring is not uncommon, as will be discussed below. The aziridine ring is highly activated towards nucleophilic attack and the chloride ion accomplishes this to form the product.

Compounds containing the structure $N\text{-}CH_2CH_2\text{-}Cl$ are known as nitrogen mustards. The compound $N(CH_2CH_2Cl)_2$ is used as a drug in cancer therapy: it functions as a DNA alkylator, preventing cell division. Nitrogen mustards are potent alkylating agents due to their ability to form an aziridine ring (Chabner *et al.*, 1996). The nitrogen attacks the chlorine-bearing carbon to form the aziridinium ion, which is highly activated towards nucleophilic attack (e.g. from bases within DNA) (Fig. 5). Attack at either of the carbons forming the aziridine opens the ring, resulting in a covalent bond between the nucleophile and the nitrogen mustard.

As mentioned before, DMPA-N(Et)(EtCl) has been synthesized previously (Ross *et al.*, 1969). In that synthesis, the authors set out specifically to create the chloride, by reacting the alcohol with $SOCl_2$. The yield is comparable to this reaction above, which was nearly quantitative with a yield of 91%. DMPA-N(Et)(EtCl)'s anesthetic ability was tested on several animal tissues including frog nerve and squid giant axon, and was found to behave as a "long-acting anesthetic" (Stubbins, 1970). This is believed to be due to its reactivity: this compound is likely not very selective in which residue it reacts with and may be reacting with any number of nucleophilic residues, besides cysteine, to be found within or near the pore region. It raises the possibility that conversion to DMPA-N(Et)(EtMTS) would abolish this long-acting anesthetic property if this drug were applied to skeletal sodium channels. As described before, skeletal sodium channels do not have the free cysteine residue and the MTS group would make the compound highly selective for reaction with cysteine. In the cardiac sodium channel, if long-acting block of the channel

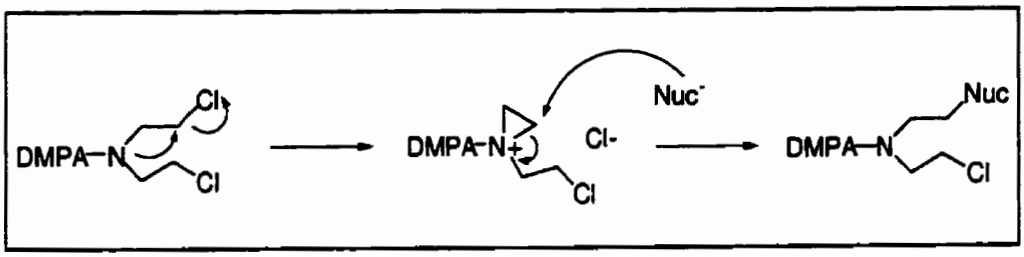


Figure 5: Nitrogen mustard compounds form aziridinium ions which are activated towards nucleophilic attack.

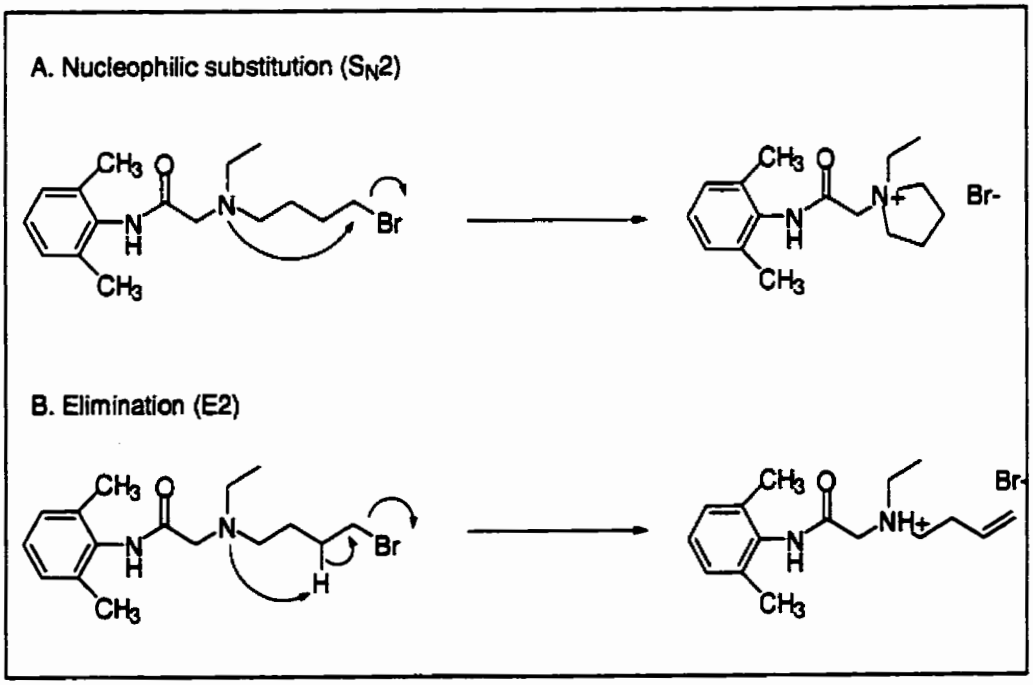


Figure 6: The butyl bromide lidocaine derivative is capable of two types of internal reaction

was observed this would be partial evidence that the compound was indeed reacting with the cysteine residue in question.

3. Synthesis of DMPA-N(Et)(PrBr), DMPA-N(Et)(PrI), DMPA-N(Et)(PrMTS)

DMPA-NHEt (**3**) was reacted with 1,3-dibromopropane, as shown in Scheme 3. To reduce the likelihood of reaction at both ends of the 1,3-dibromopropane, the following measures were taken: DMPA-NHEt was present in low concentration in the THF reaction mixture (0.1 M), the bromine compound was present in a large excess to the amine and finally, a mild reaction temperature was used, 55-57°C (well below THF's boiling point). As with the 2-bromoethanol reaction, $(DMPA\text{-}NH_2Et)^+Br^-$ crystals precipitated out of the solution during the course of the reaction. DMPA-N(Et)(PrBr) (**6**) was isolated using the standard aqueous workup, and in relatively good yield, 33%.

A direct conversion of this bromine derivative to the MTS analog by reaction with NaMTS was not possible; this reaction did not proceed at all even upon heating. Conversion of DMPA-N(Et)(PrBr) to its hydrochloride salt (to block the reactivity of the amine nitrogen) was not successful either. A mixture of the bromide and the chloride derivatives was obtained, with a very small fraction of the MTS derivative present.

In order to obtain the MTS compound, it was necessary to exchange the bromine for a leaving group that is easily exchanged in a subsequent reaction. DMPA-N(Et)(PrBr) was converted to DMPA-N(Et)(PrI) (**7**) by reaction with sodium iodide. This reaction was accomplished under extremely mild conditions, stirring at room temperature for less than 1 hour, and conversion was nearly quantitative. The iodide compound appeared to be more reactive than the bromide compound because it "stuck" to silica TLC plates. In the NMR spectrum, a number of impurity peaks between $\delta 3.5$ and $\delta 5$ were noted, but since the total integration of these peaks was quite low, it was considered unnecessary to purify the compound for the next reaction.

Finally, an exchange reaction between sodium methylthiosulfonate and DMPA-N(Et)(PrI) was carried out to form DMPA-N(Et)(PrMTS) (**8**). This reaction was also accomplished by stirring at room temperature. The NMR spectrum showed the carryover of the impurity from the starting material, DMPA-N(Et)(PrI) and several of the peaks had different chemical shifts from the iodide compound. The peaks representing the

protons from methyl group of MTS and the three methylene groups of the propyl were identified by comparison with the spectrum of MTSEA, $(H_3NCH_2CH_2SS(O)_2CH_3)^+Cl^-$, in which a broad multiplet at δ3.45-δ3.73 was observed (Bruice & Kenyon, 1982). The methyl group introduced (as a part of MTS) is supposed to appear as a singlet in the region of δ3.35 (*ibid.*). This would directly overlap the singlet due to the acetamide methylene group (within the DMPA moiety). The peak at δ3.35 has a much larger integration than if it consisted of the acetamide methylene group alone, so this was taken as evidence for the presence of the MTS group. Also, the "quintet" at δ2.08 from the iodide compound is greatly reduced in size in this NMR spectrum (indicating that very little of the iodide remains).

The mass spectrum showed strong signals for the MTS derivative (at 279) and its signature fragment at 210 and only weak signals were due to the iodide compound (no parent ion of 374 was observed and a small signature fragment peak at 226 was noted).

4. Synthesis of [DMPA-NH(Et)(BuBr)]$^+Br^-$

DMPA-NHEt (**3**) was reacted with 1,4-dibromobutane under similar reaction conditions as the previous reaction with 1,3-dibromopropane as shown in Scheme 4. It was assumed that the product would remain in solution while the hydrobromide salt of DMPA-NHEt precipitate out of solution during the course of the reaction. A workup of the reaction mixture, after filtration to remove the precipitate, found only 1,4-dibromobutane present. The product [DMPA-NH(Et)(BuBr)]$^+Br^-$ (**9**), in this case, co-precipitated out of solution with the DMPA-NHEt. The mixture was identified by TLC (which shows two spots) and NMR, in which the DMPA-NHEt peaks were identified by comparison with an NMR spectrum of pure $(DMPA\text{-}NH_2Et)^+X^-$. The two compounds were present in an approximate 1:1 ratio. Because DMPA-N(Et)(BuBr) forms the salt at relatively the same rate as DMPA-NHEt, the tertiary amine must have a basicity which is very close to that of the secondary amine. This property and the fact that the two compounds have the same solubilities in a number of organic solvents frustrated attempts to separate the two compounds.

Separation attempts with (1) a silica column (100% MeOH, followed by 1.5% HOAc in MeOH) and (2) a mini reverse phase C_{18} column (100% MeOH; 100% H_2O)

were not successful. The separation attempt with the silica column destroyed this compound, suggesting that the bromide compound is highly sensitive to acidic conditions. A debrominated compound was obtained from the separation: no bromide doublets were apparent in the mass spectrum of the eluted compound. This was in contrast to the mass spectrum of the crude mixture, and a parent ion peak of 260 was apparent, corresponding to the mass of the original compound without bromine.

The butyl bromide derivative's stability was a concern from the outset because it is capable of self-reaction. Two possible internal reactions can occur with the butyl bromide compound: it can either undergo a substitution reaction to form a cyclic compound or it can undergo and elimination reaction to form a "3'-butene" derivative (Fig. 6). Which reaction is favoured is not clear because the primary alkyl halide chain favours an S_N2 nucleophilic substitution, while the tertiary amino nitrogen as the base of the reaction suggests an E2 elimination reaction would occur (Loudon, 1988).

An analysis of the mass spectrum of the products helped to identify the reaction that took place. The largest peak in the mass spectrum is of mass 84, which is also the mass of a fragment containing the cyclic structure. The cyclic compound is the product of the S_N2 reaction. Circumstantial evidence also suggests that it is the cyclic product which is obtained. When $[DMPA\text{-}NH(Et)(BuBr)]^+Br^-$ is dissolved in a basic solution (NaOH) to obtain the free amine, the small amount of solid recovered from the solution contained no free amine. This suggests that no conversion is possible. An explanation for this is that the bromide compound reacts with itself in the basic solution, undergoing a substitution reaction to form the cyclic compound. The cyclic compound is a quaternary ammonium ion which is permanently charged and would not be soluble in organic solvents.

The ability of the amino nitrogen to react again, despite it being a tertiary amine, is a distinct possibility considering the reactivity of the nitrogen mustard-like DMPA-N(Et)(EtCl). This butyl derivative's reactivity may also be understood in terms of the amine's ability to form a stable 5-membered ring upon nucleophilic attack on the halogen-bearing carbon. This is in contrast to the previous propyl derivative, DMPA-N(Et)(PrBr). In this case, a similar cyclization reaction would form a highly unstable 4-membered ring (Loudon, 1988). Evidence for the absence of this type of reaction in the propyl derivative

lies in the fact that it is stable to extremes in pH, in contrast to the butyl bromide derivative.

A less harsh method of separation, using reverse-phase chromatography, was employed to try to avoid the self-reaction of the butyl bromide compound. However, the reverse phase C_{18} column separation was of limited success: a separation attempt with 100% methanol as the mobile phase gave an approximate 60% product / 40% DMPA-NHEt split as compared to the crude mixture composed of approximately 50% of each compound. A separation attempt with 100% H_2O and then an increasing gradient of methanol concentration (10%, 20%, etc.) did not effect any separation; all of the mixture eluted with the water in the initial wash.

For the purposes of electrophysiological studies (testing this compound on *X. laevis* oocytes expressing sodium channels), it was considered not absolutely necessary to have the pure compound since only the bromide is reactive. DMPA-NHEt is very similar to lidocaine in its properties and like lidocaine, it may be washed off the channel (Foye *et al.*, 1995). Therefore if a mixture of DMPA-N(Et)(BuBr) and DMPA-NHEt is applied and then washed off after some time, the butyl bromide derivative should remain. If a reduced current is still observed after washing the compound off of the cell, then it should be due solely to the presence of DMPA-N(Et)(BuBr).

5. Future Experiments and Conclusions

A series of halogenated lidocaine analogs was synthesized and the conversion of one of these compounds to the MTS version was accomplished. In future experiments, these compounds may be tested for activity in whole cell patch clamp experiments on sodium channels. These sodium channels may be cardiac sodium channels that contain a unique cysteine residue within the pore region (Backx *et al.*, 1992), or other sodium channels isoforms that do not ordinarily contain cysteine residues within the pore region but have been subjected to site-directed mutagenesis to create cysteine residues.

The halogenated lidocaine analogs should cause irreversible block of the channels, due to the formation of a covalent bond between the lidocaine molecule and the sodium channel, even after the cells are washed free of the drug solution. These halogenated lidocain analogs may react with any nucleophilic residue within the channel. The "propyl-

MTS" derivative of lidocaine, DMPA-N(Et)(PrMTS), should react selectively with cysteine residues. The disulfide bond formed between DMPA-N(Et)(PrMTS) and a cysteine residue may be reduced by the addition of a large excess of a reducing agent such as DTT, thereby removing the blockade of the channel. This is a useful aspect when observing the effect of an MTS lidocaine analog on sodium channels that have been subjected to scanning cysteine mutagenesis. If the cysteine to which DMPA-N(Et)(PrMTS) has reacted is near the local anesthetic binding site, long-lasting blockade of the sodium channel should be observed, which will be removed upon addition of the reducing agent. It is expected that DMPA-N(Et)(PrMTS) should only cause irreversible block in the cardiac sodium channel and not the skeletal sodium channel, which does not have a unique cysteine residue within the outer pore region.

A future synthesis of MTS analogs of lidocaine should focus on an improved synthesis strategy in order to avoid the problem of self-reaction observed in DMPA-N(Et)(EtCl) and [DMPA-NH(Et)(BuBr)]$^+Br^-$. One possible synthetic strategy is the synthesis of MTS-$(CH_2)_n$-Br from Br-$(CH_2)_n$-OH. This could be done by the initial exchange of the bromine for MTS (using NaMTS) and then conversion of the hydroxyl group to bromine with PBr_3. MTS-$(CH_2)_n$-Br may then be reacted with DMPA-NHEt to give DMPA-N(Et)($(CH_2)_n$MTS). This synthesis strategy may avoid the problem of self-reaction and the potential problem of reaction at both ends of the dibrominated alkyl halides used in the original synthesis. This is because the amino group of DMPA-NHEt should reactive selectively with the bromide-bearing carbon, not the MTS group.

In future experiments, new MTS derivatives of lidocaine should be synthesized with different linkers. The linker group may be varied in length and flexibility. When the thiol-reactive lidocaine analog reacts with the pore cysteine residue within the cardiac sodium channel or with an engineered cysteine residue in other sodium channels, the distance between this residue and the local anesthetic binding site is unknown. The length and flexibility of the linker group in the lidocaine analogs may be changed to suit the purposes of the experimenter. The length of the linker group may perhaps be used as an approximate gauge of distance between a given residue within the channel and the local anesthetic binding site. The linker group may also incorporate chemical groups that aid in the solubilization of the lidocaine group in aqueous solution, in which it is normally insoluble.

REFERENCES

Akabas, M. H.; Stauffer, D. A.; Xu, M.; Karlin, A. "Acetylcholine receptor channel structure probed in cysteine-substitution mutants." (1992) *Science* **258**, 307-310.

Astra Pharmaceutical Products. *Xylocaine: Chemistry, Pharmacology and Clinical Applications.* Worcester: 1960.

Backx, P. H.; Yue, D. T.; Lawrence, J. H.; Marban, E.; Tomaselli, G. F. "Molecular localization of an ion-binding site within the pore of mammalian sodium channels." (1992) *Science* **257**, 248-251.

Balser, J. R. "Structure and function of the cardiac sodium channels." (1999) *Cardiovascular Res.* **42**, 327-338.

Bennett, P. B.; Valenzuela, C.; Chen, L.-Q.; Kallen, R. G. "On the molecular nature of the lidocaine receptor of cardiac Na^+ channels." (1995) *Circulation Res.* **77**, 584-592.

Bénitah, J.-P.; Tomaselli, G. F.; Marban, E. "Adjacent pore-lining residues within sodium channels identified by paired cysteine mutagenesis." (1996) *Proc. Natl. Acad. Sci. USA.* **93**, 7392-7396.

Bruice, T. W. and Kenyon, G. L. "Novel alkyl alkanethiolsulfonate sulfhydryl reagents. Modification of derivatives of L-cysteine." (1982) *J. Prot. Chem.* **1**, 47-58.

Cahalan, M. and Neher, E. "[1] Patch Clamp Techniques: An Overview." (1992) *Methods in Enzymology* **207**, 3-14.

Catterall, W. A. "Cellular and molecular biology of voltage-gated sodium channels." (1992) *Physiological Rev.* **72** *(Suppl.)*, S15-S41.

Catterall, W. and Mackie, K. "Local Anesthetics." *Goodman & Gilman's The Pharmacological Basis of Therapeutics, 9th ed.* Eds. Hardman, J. G.; Limbird, L. E.; Molinoff, P. B.; Redden, R. W.; Goodman Gillian, A. New York: McGraw-Hill, 1996.

Chabner, B. A; Allegra, C. J.; Curt, G. A; Calabresi, P. "Antineoplastic Agents." *Goodman & Gilman's The Pharmacological Basis of Therapeutics, 9th ed.* Eds. Hardman, J. G.; Limbird, L. E.; Molinoff, P. B.; Redden, R. W.; Goodman Gillian, A. New York: McGraw-Hill, 1996.

Chiamvimonvat, N., Pérez-García, M. T.; Ranjan, R.; Marban, E.; Tomaselli, G. F. "Depth asymmetries of the pore-lining segments of the Na^+ channel revealed by cysteine mutagenesis." (1996) *Neuron* **16**, 1037-1047.

CRC Handbook of Chemistry & Physics, 81st ed. Ed. D. R. Lide. Boca Raton: CRC Press, 2000.

Fieser, L. F. and Fieser, M. *Reagents for Organic Synthesis, vol. 1.* New York: Wiley, 1967.

Foye, W. O.; Lemke, T. L.; Williams, D. A. *Principles of Medicinal Chemistry 4th ed.* Media: William & Wilkins, 1995.

Hille, B. *Ionic Channels of Excitable Membranes 2nd ed.* Sunderland: Sinauer Associates Inc., 1992.

Kellenberger, S.; Scheuer, T.; Catterall, W. A. "Movement of the Na^+ channel inactivation gate during inactivation." (1996) *J. Biol. Chem.* **271**, 30971-30979.

Kenyon, G. L. and Bruice, T. W. "Novel sulfhydryl reagents." (1977) *Methods in Enzymology* **27**, 407-430.

Liu, L.; Wendt, D. J.; Grant, A. O. "Relationship between structure and sodium channel blockade by lidocaine and its amino-alkyl derivatives." (1994) *J. Cardiovascular Pharmacology* **24**, 803-812.

Löfgren, N. M. *Studies on Local Anesthetics: Xylocaine, A New Synthetic Drug.* Stockholm: Hoeggstroms, 1948.

Löfgren, N. M. and Lundqvist, B. J. (1948) *Chemical Abstracts*, P6379a.

Loudon, G. M. *Organic Chemistry, 2nd ed.* New York: Benjamin/Cummings Publishing Co. Inc., 1988.

March, J. *Advanced Organic Chemistry, 3rd ed.* New York: John Wiley & Sons, 1985.

Merck Index 11th ed. Rahway, NJ: Merck & Co., 1989.

Nuss, H. B.; Tomaselli, G. F.; Marban, E. "Cardiac sodium channel (hH1) are intrinsically more sensitive to block by lidocaine than are skeletal muscle (m1) channels." (1995) *J. Gen. Physiol.* **106**, 1993-1209.

Pérez-García, M. T.; Chiamvimonvat, N.; Marban, E.; Tomaselli, G. F. "Structure of the sodium channel pore revealed by serial cysteine mutagenesis." (1996) *Proc. Natl. Acad. Sci. USA.* **93**, 300-304.

Patai, Saul. *The Chemistry of the Amino Group.* London: Interscience Publishers, 1968.

Ragsdale, D. S.; McPhee, J. C.; Scheuer, T.; Catterall, W. A. "Molecular determinants of state-dependent block of Na^+ channels by local anesthetics." (1994) *Science* **265**, 1724-1728.

Ragsdale, D. S.; McPhee, J. C.; Scheuer, T.; Catterall, W. A. "Common molecular determinants of local anesthetic, antiarrhythmic and anticonvulsant block of voltage-gated Na^+ channels." (1996) *Proc. Natl. Acad. Sci. USA.* **93**, 9270-9275.

Ross, S. B.; Sandberg, R. W.; Sjoberg, B. O. H.; Akerman, S. B. A. (1969) *Chemical Abstracts* **71**, 70340a.

Qu, Y.; Rogers, J.; Tanada, T.; Scheuer, T.; Catterall, W. "Molecular determinants of drug access to the receptor site for antiarryhthmic drugs in the cardiac Na^+ channel." (1995) *Proc. Natl. Acad. Sci. USA.* **92**, 11839-11843.

Soreq, H. and Seidman, S. "[14] Xenopus oocyte microinjection: from gene to protein." (1992) *Methods in Enzymology* **207**, 225-265.

Spialter, L. and Pappalardo, J. A. *The Acyclic Aliphatic Tertiary Amines.* New York: Macmillan (1965).

Stubbins, J. F. "A new class of ultralong-acting local anesthetics." (1970) *J. Med. Chem.* **13**, 558-559.

Sunami, A.; Dudley Jr., S. C.; Fozzard, H. A. "Sodium channel selectivity filter regulates antiarrhythmic drug binding." (1997) *Proc. Natl. Acad. Sci. USA* **94**, 14126-14131.

Sunami, A.; Glaaser, I. W.; Fozzard, H. A. "Structural and gating changes of the sodium channel induced by mutation of a residue in the upper third of IVS6, creating an external access path for local anesthetics." (2001) *Mol. Pharm.* **59** (4), 684-691.

Tomaselli, G. F.; Chiamvimonvat, N.; Nuss, H. B.; Balser, J. R.; Pérez-García, M. T.; Xu, R. H.; Orias, D. W.; Backx, P. H.; Marban, E. "A mutation in the pore of the sodium channel alters gating." (1995) *Biophys. J.* **68**, 1814-1827.

Yarov-Yarovoy, V.; Brown, J.; Sharp, E. M.; Clare, J. J.; Scheuer, T.; Catterall, W. A. "Molecular determinants of voltage-dependent gating and binding of pore-blocking drugs in transmembrane segment IIIS6 of the Na^+ channel α subunit." (2001) *J. Biol. Chem.* **276** (1), 20-27.

Zamponi, G. W and French, R. J. "Amine blockers of the cytoplasmic mouth of sodium channels: a small structural change can abolish voltage dependence." (1994) *Biophys. J.* **67**, 1015-1027.

Part III: Studying the Function of the RS Domain of U2AF^{65}, an Essential RNA Splicing Factor

SECTION A: RNA SPLICING AND U2AF; A STRATEGY TO ADDRESS UNRESOLVED ISSUES CONCERNING THE ROLE OF THE U2AF^{65} RS DOMAIN IN SPLICEOSOME FORMATION

1. RNA Splicing

In cells, the information required for the cell to live and reproduce is encoded in its DNA sequence in discrete units called genes. In eukaryotes (higher organisms with compartmentalized cellular organelles; including all multicellular organisms), DNA is contained within the cell nucleus. For genes to be expressed into proteins which carry out various functions within the cell, the gene is first transcribed into its pre-messenger RNA (pre-mRNA) counterpart by RNA polymerase. The pre-mRNA of eukaryotes often contains non-coding sequences called introns dispersed between coding sequences called exons. These introns must be removed and the exons spliced together to form the mature RNA which is then exported from the nucleus and translated into protein by the ribosomes. The processing of pre-mRNA, in which introns are excised and exons are spliced together, is termed RNA splicing. The control of RNA splicing is essential since various exons from a single pre-mRNA transcript may be spliced to give rise to different mature RNA products which are translated in turn into different proteins with varying function.

RNA splicing occurs in two chemical steps (Fig. 1). The first step of splicing is a transesterification reaction consisting of a nucleophilic attack of the 2' hydroxyl group of an adenosine within the intron, termed the branch-point adenosine, on the 5' phosphodiester bond on the G which occurs 3' to exon 1. This forms the lariat intermediate, containing exon 2 and the intron. A second transesterification reaction, consisting of the nucleophilic attack of the 3' hydroxyl group at the end of exon 1 on the phosphodiester bond at the beginning of exon 2, joins the two exons and release the intron as a lariat byproduct.

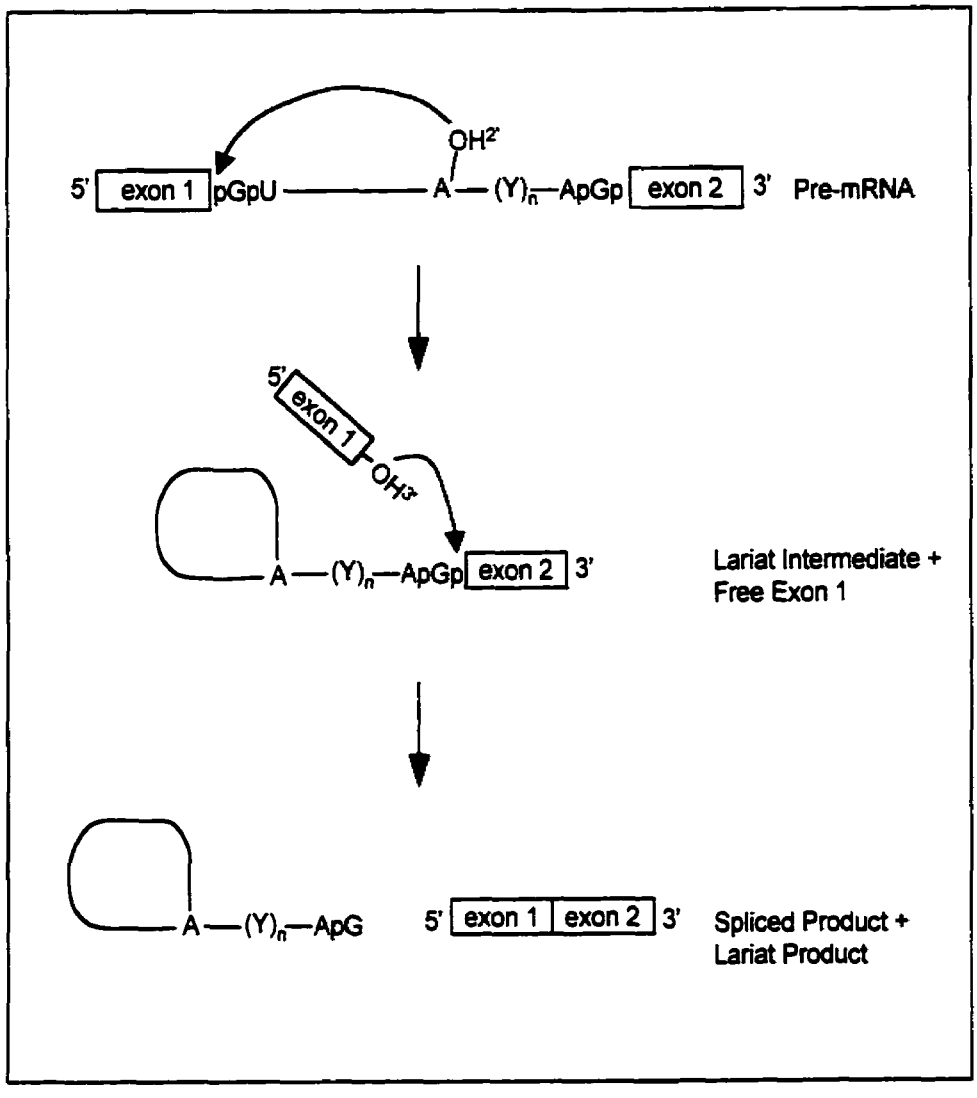


Figure 1: Pre-mRNA splicing occurs via two transesterification reactions. The conserved sequences of the itron are GU at the 5' splice site, the branchpoint adenosine, the polypyrimidine tact (denoted as $(Y)_n$) and AG at the 3' splice site. The first transesterification reaction occurs when the 2' hydroxyl of the branchpoint adenosine attacks the phosphodiester bond (p) 5' to the G at the 5' splice site. This forms the lariat intermediate and frees exon 1. The free 3' hydroxyl at the end of exon 1 attacks the phosphodiester bond 5' to the beginning of exon 2 to form the spliced exon 1 - exon 2 product and the intron as a lariat product (Moore *et al.*, 1993).

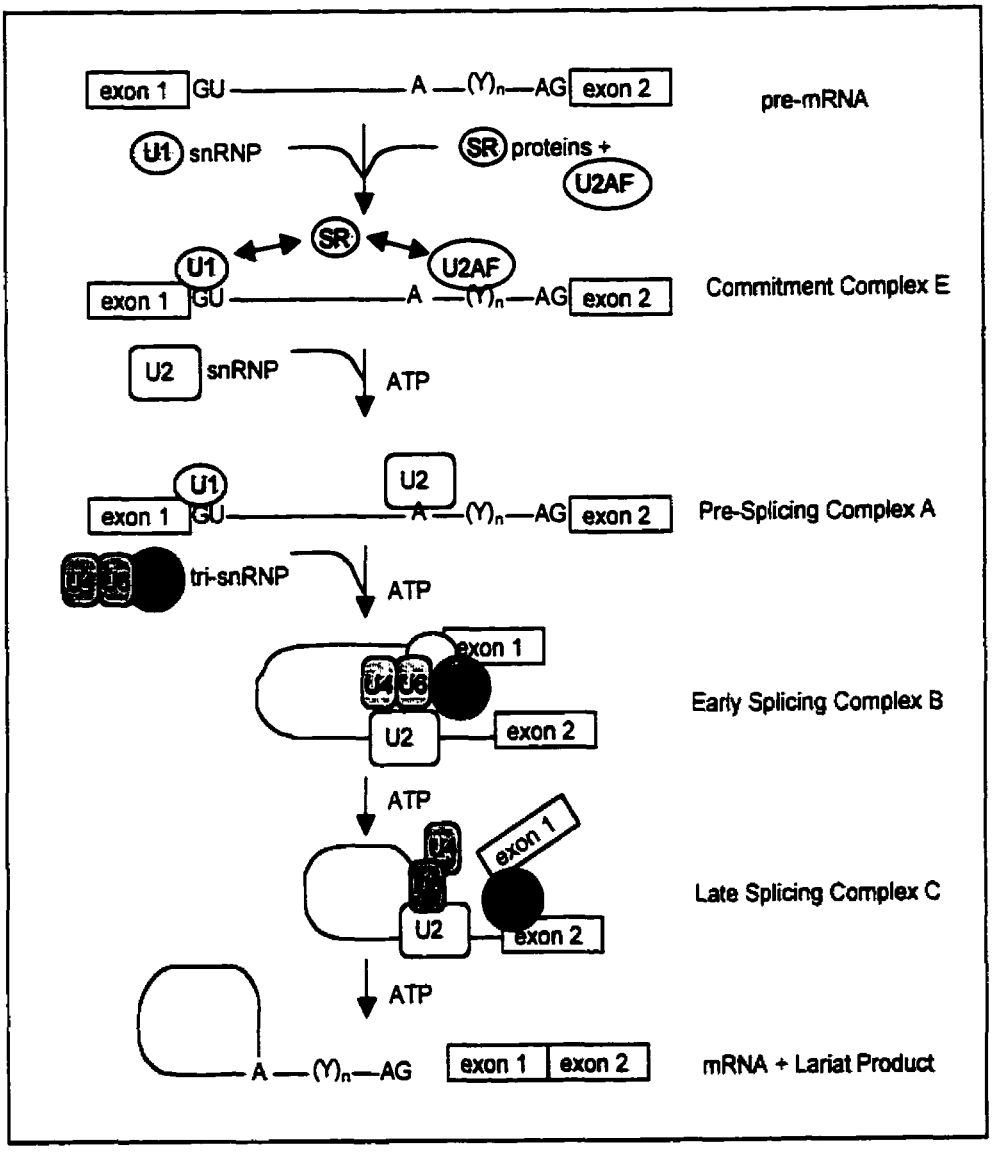


Figure 2: Spliceosome assembly occurs in a dynamic fashion, with snRNPs and other splicing factors entering and leaving the the spliceosome at different points during the course of the two-part splicing reaction. The various complexes which have been established in mammals are labeled E (early), A, B, and C (adapted from Kramer, 1996). In yeast introns, there is no typical polypyrimidine tract or U2AF splicing factor.

The process of RNA splicing is controlled by a large multiple-protein complex called the spliceosome (reviewed in Moore *et al.*, 1993). Studies on pre-mRNA and the spliceosome have been established in yeast (*Saccharomyces cerevisiae*) and homologs identified in mammals. Differences between the yeast and the mammalian systems are given below. Within the pre-mRNA, the introns are defined by conserved (consensus) sequences. The presence of these conserved sequences is essential to the accuracy of RNA splicing and their role will be further explained below. The 5' splice site, at the 5' end of an intron, is AG|GUAUGU in yeast and AG|GURAGU in mammalians (R = purine). The 3' end of the intron, contains three conserved elements, the branchpoint region, a pyrimidine-rich sequence called the polypyrimidine tract, and the 3' splice site. The spacing between these three elements and the length of the polypyrimidine tract is variable, so that the branchpoint region may be up to 100 nucleotides from the 3' splice site. The branchpoint region is UACUAAC in yeast and YNYURAC in mammals (A is the branchpoint adenosine, Y = pyrimidine). Yeast introns do not contain the typical polypyrimidine tracts as found in other species, instead containing a short uridine rich tract which is necessary for the second catalytic step of splicing but appears to be dispensable for initial spliceosome assembly (Rymond & Rosbash, 1985; Kramer, 1996).

Accurate RNA splicing requires a group of small nuclear ribonucleoproteins (snRNPs), which recognize conserved sequences within the intron, and other splicing factors which include the SR family of proteins (defined and discussed in Section A.2). Fig.2 shows a schematic depiction of various complexes (commitment complex in yeast, E in mammalian; A,B,C) which were first characterized biochemically in yeast (Kramer, 1996; Moore *et al.*, 1993). These complexes are composed of different snRNPs and splicing factors, which form during the course of RNA splicing. The spliceosome complexes have been shown to be conserved amongst the eukaryotes. The snRNPs contain small nuclear RNAs (snRNAs), portions of which base pair (via Watson-Crick base-pairing) to portions of the intron, specifically the 5' and 3' ends of the intron and the branchpoint region. The known snRNPs involved in RNA splicing are designated U1, U2, U4/U6 (associated together) and U5. The formation of the commitment complex, the first step to formation of the

spliceosome and RNA splicing, involves U1 snRNP binding to the GU at the 5' end of the intron. Also in the early (E) complex, another splicing factor, U2AF (U2 snRNP auxiliary factor) binds to the polypyrimidine tract within the intron. U2AF recruits U2 snRNP to the branchpoint adenosine, a crucial step in RNA splicing; this forms the A splicing complex. In yeast, as stated earlier, typical polypyrimidine tracts do not occur; instead, it appears that another splicing factor, called Mud2p, recognizes the branch site (Abovich *et al.*, 1994). The binding of U2 snRNP to the pre-mRNA to the pre-mRNA involves the base pairing of its snRNA to the branchpoint region of the pre-mRNA. The snRNa of U2 snRNP is missing the base which should base pair to the branchpoint adenosine in the pre-mRNA so that the branchpoint adenosine is bulged out of the RNA duplex (Query *et al.*, 1994). The branchpoint adenosine is thus activated for the first transesterification reaction. U4/U6 snRNPs and U5 snRNP enter the spliceosome complex together as a tri-snRNP, with U5 snRNP interacting with the 3' end of exon 1 and the 5' end of exon 2, bringing the two exons together to bring about the second transesterification reaction.

2. U2AF, an Essential RNA Splicing Factor

U2AF is an essential splicing factor in the formation of the mammalian spliceosome since it recognizes the polypyrimidine tract within the intron, and recruits U2 snRNP to the branchpoint region of the intron. Without U2AF, U2 snRNP does not bind to the branchpoint region of the intron and splicing does not occur. Human U2AF was first identified by Green & coworkers in 1989 (Zamore & Green, 1989). U2AF homologs are also found in *Schizosaccharomyces pombe* (Potashkin *et al.*, 1993), *Caenorhabditis elegans* (Zorio *et al.*, 1997) and *Drosophila melanogaster* (Kanaar *et al.*, 1993). In yeast, the splicing factor Mud2p is the putative homolog of U2AF. As stated earlier, yeast introns do not contain a typical polyprimidine tract. Instead, Mud2p appears to interact with the branch site (Abovich *et al.*, 1994).

U2AF is a heterodimer, consisting of a large subunit with apparent mass of 65 kDa ($U2AF^{65}$) and a 35 kDa subunit ($U2AF^{35}$) (Fig. 3). $U2AF^{65}$ binds the polypyrimidine tract and recruits U2 snRNP, whereas $U2AF^{35}$ interacts with the 3' splice site and may also

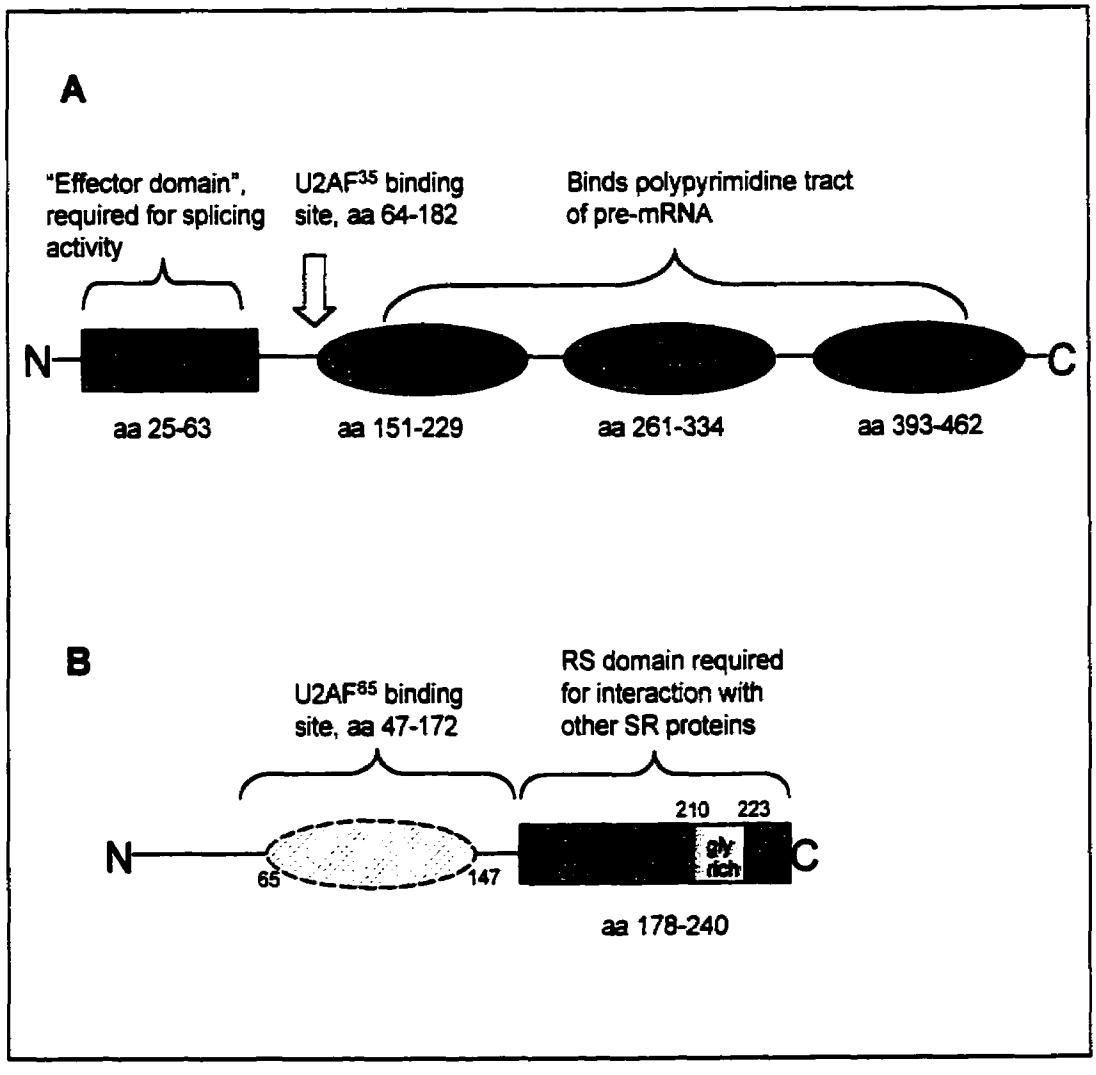


Figure 3: (A) Domain structure of U2AF^{65} (475 aa total) showing the RS domain near the N-terminus and three RNA binding domains (RRMs) towards the C-terminus (Zamore *et al.*, 1992; Zhang *et al.*, 1992). (B) Domain structure of U2AF^{35} (240 aa total) showing the RS domain near the C-terminus and a glycine-rich domain within the RS domain (Zhang *et al.*, 1992; Zuo & Maniatis, 1996). Within the U2AF^{65} binding site is a possible RNA binding domain (shaded area, aa 65-147) which aligns with the RNA binding domain of the yeast splicing factor Mud2p (NCBI Conserved Domain Database accession no. smart00361). Mud2p is the putative homolog of mammalian U2AF^{65} (Abovich *et al.*, 1994).

interact with other splicing factors in regulatory splicing events (Moore, 2000; Zuo & Maniatis, 1996). U2AF^{65} is indispensable for splicing whereas U2AF^{35} may be dispensable, depending on the sequence of the intron within the pre-mRNA substrate. The importance of U2AF^{65} to splicing has made it the focus of studies to identify the nature of its activity and interactions with other splicing factors.

(a) The Structure and Function of the Large Subunit U2AF^{65}

The larger subunit, U2AF^{65}, shares a basic domain structure that is somewhat similar to the SR proteins. U2AF^{65} has been termed an "SR-related protein" (Blencowe *et al.*, 1999). The SR proteins are a highly conserved family of proteins involved in RNA splicing: they are involved in alternative splicing of the pre-mRNA through selection of different 5' splice sites (reviewed in Kramer, 1996; Manley & Tacke, 1996). This is accomplished through bridges of protein-protein interactions between the SR proteins and enhancer elements within the exons, to the snRNPs bound to the intron. In the SR proteins, one or more consensus RNA recognition motifs (RRMs or RNA binding domains, RBDs) are found towards the N-terminus, and an arginine-serine rich (RS) domain is found towards the C-terminus. In U2AF^{65}, the order of the domains is different, with an RS-rich domain found at the N-terminus of the protein and three RRMs towards the C-terminus (Zamore, 1992). The domain structure of U2AF^{65} is shown in Fig. 3A. The first two RRMs of U2AF^{65} are required for binding of the U2AF heterodimer to the polypyrimidine tract.

The role of the RS domain of U2AF^{65} is not well defined. Green & coworkers discovered that a deletion mutant of U2AF^{65} missing the RS domain could not support splicing on a variety of pre-mRNA substrates, suggesting that it plays an important role in U2AF activity (Zamore *et al.*, 1992). Green & coworkers found that the RS deletion mutant of U2AF^{65} could not bind as efficiently to a pre-mRNA substrate as the full-length U2AF^{65} (Lee *et al.*, 1993), suggesting the RS domain may be involved in RNA binding. Further investigation by Green & coworkers found that the U2AF^{65} RS domain contacted the branchpoint region of the intron, suggesting a role in recognition of the branch point region (Valcárcel *et al.*, 1996). It is possible that the U2AF^{65} RS domain, due to its overall positive

charge, may be necessary for stabilizing RNA-RNA interactions within the spliceosome or for inducing conformational changes in the pre-mRNA (Moore *et al.*, 1993). The concentration of basic amino acid residues in the RS domain may neutralize the negatively charged phosphodiester backbone of the RNA, allowing either RNA-RNA interactions or the manipulation of RNA conformation. Substitution of the wild type U2AF^{65} RS domain with peptides containing a net positive charge (*e.g.* $(RS)_n$, $(RG)_7$) was sufficient to restore splicing activity, suggesting that the RS domain may stabilize the interaction between U2 snRNP and the branch point by charge shielding of the RNA phosphodiester backbone (Valcárcel *et al.*, 1996).

The SR proteins are known to be regulated through phosphorylation of the RS domain by SR protein-specific kinases (reviewed in Stojdl & Bell, 1999). Phosphorylated SR proteins have been shown to have enhanced protein-protein interaction activity and are required in early spliceosome complex formation (Xiao & Manley, 1997; Tacke *et al.*, 1997). Conversely, dephosphorylation of SR proteins may be necessary for the completion of splicing to occur (Mermoud *et al.*, 1994). Phosphorylation of the SR proteins appears to reduce nonspecific binding to RNA, suggesting that the RS domain in these proteins affects their RNA binding activity (Xiao & Manley, 1998). Phosphorylation of the RS domain in SR proteins also affects the intranuclear distribution of the SR proteins, regulating the recruitment of the SR proteins to sites of transcription (Misteli & Spector, 1997; Yeakley *et al.*, 1999). Phosphorylation/dephosphorylation of the RS domain in SR proteins is likely a regulatory mechanism in RNA splicing, affecting RNA and protein interaction activity. Given the regulatory function of the RS domains in SR proteins, it is possible that U2AF^{65} activity may also be regulated by phosphorylation. Reed & coworkers postulated that the phosphorylation of U2AF^{65} while it is present in the E complex, destabilizes its binding to pre-mRNA, so that U2AF^{65} leaves the spliceosome in the transition from E to A complex (Champion-Arnaud *et al.*, 1995). Phosphorylation of the U2AF^{65} RS domain may also affect its protein-protein interaction activity, similar to the effect of phosphorylation in the SR proteins. With these questions in hand, it is clear that more study must be done on the role of U2AF^{65}'s RS domain since it plays such an important role in RNA splicing. However,

U2AF is a member of a large multi-protein complex which changes composition over time. Studying the interactions of U2AF with other members of the spliceosome presents challenges which this work is meant to address.

(b) The Role of the Small Subunit U2AF^{35}

U2AF^{35} was purified so that it could be used in subsequent experiments with the chemically derivatized U2AF^{65} mutants. U2AF^{35} is the small subunit of the U2AF heterodimer. The affinity of binding between the two U2AF subunits is extremely high, such that they may only be separated under stringent conditions (Zamore & Green, 1991; Wu & Maniatis, 1993). U2AF^{35} was initially characterized by Green & coworkers in 1992 (Zhang *et al.*, 1992). The domain structure of U2AF^{35} is shown in Fig. 3B. U2AF^{35}, like U2AF^{65}, has been termed an "SR-related protein" (Blencowe *et al.*, 1999). It contains an RS domain near its C-terminus, between amino acids 188-240. The RS domain contains a run of 12 consecutive glycines between amino acids 210-223 which are likely involved in protein-protein interactions (Wang *et al.*, 1997). Through biochemical assays, the region of interaction between the two subunits was identified as amino acids 47-172 in U2AF^{35} and amino acids 64-182 in U2AF^{65} (Zhang *et al.*, 1992). Previous biochemical studies have not identified an RNA binding domain in U2AF^{35}. However, the sequence between amino acids 65-147 may contain a possible RNA binding domain; the given sequence aligns with the RNA binding domain of the yeast splicing factor Mud2p, the putative homolog of U2AF^{65} (NCBI Conserved Domain Database, accession no. smart00361). An inspection of the sequence between amino acids 65-147 of U2AF^{35} shows that it does not contain the consensus RNA binding motifs which have been identified in a large number of RNA binding proteins (Nagai *et al.*, 1995; Conte *et al.*, 2000).

The role of U2AF^{35} was not well defined for some time. At first it was thought that U2AF^{35} was not an essential splicing factor because U2AF^{65} was able to reconstitute splicing in U2AF-depleted nuclear extract in the apparent absence of U2AF^{35} (Zamore & Green, 1991). However, Wu & Maniatis (1993) observed that U2AF^{35}, not U2AF^{65}, interacted with SR proteins (SC35 and SF2/ASF); they proposed that U2AF^{35} formed a bridge of protein

interaction between U2AF^{65}, bound to the 3' splice site, and U1-70K, part of U1 snRNP bound to the 5' splice site. Further research by Zuo and Maniatis (1996) showed that U2AF^{35} was an essential factor for splicing. The RS domain of U2AF^{35} was required for U2AF^{35} to interact with the SR proteins. Wu & Maniatis (1993) postulated that Zamore & Green's (1991) previous result was due to incomplete removal of U2AF^{35} from the nuclear extract or possibly, the presence of another protein in the extract whose activity is redundant with U2AF^{35}. The cumulative data strongly suggested that U2AF^{35} was an essential splicing factor and that it also played a role in exonic enhancer-dependent splicing through its ability to interact with other SR proteins (Wu & Maniatis, 1993; Zuo & Maniatis, 1996).

Recently, it has been found that U2AF^{35} plays a very important role in RNA splicing: it is responsible for the recognition of the YAG sequence at the 3' splice site (Wu *et al.*, 1999; Zorio & Blumenthal, 1999; Merendino *et al.*, 1999). The 3' end of the (mammalian) intron and the 3' splice site are highly conserved in sequence, featuring a YAG (Y is a pyrimidine) at position –1 to –3 relative to the 3' splice site and the polypyrimidine tract beginning at position –5, extending 10 or more nucleotides back (Burge *et al.*, 1999). Through an *in vitro* selection experiment, it was observed U2AF^{65} bound to pyrimidine rich sequences only, with no specificity for a polypyrimidine tract followed by YAG (Singh *et al.*, 1995). A second *in vitro* selection experiment with the U2AF heterodimer found that it bound with highest affinity to a substrate that was almost identical to the mammalian 3' splice site as described above, a polypyrimidine tract followed by YAG (Wu *et al.*, 1999). This result suggested that U2AF^{35} activity was responsible for the recognition of the 3' splice site. It was observed earlier that mammalian introns may contain either long or short/weakened polypyrimidine (PPT) tracts (Reed, 1989). An intron with a short or weak PPT is "AG-dependent", requiring the presence of the 3' splice site AG for spliceosome assembly and formation of the lariat intermediate. An intron containing a long PPT is "AG-independent", meaning the same splicing process can occur in the absence of the 3' splice site AG. U2AF^{65} alone can rescue splicing of AG-independent introns in U2AF-depleted extract but it cannot rescue splicing of AG-dependent introns (Zuo & Maniatis, 1996; Wu *et al.*, 1999). This phenomenon may be explained by the activity of U2AF^{35}. In AG-independent introns with a

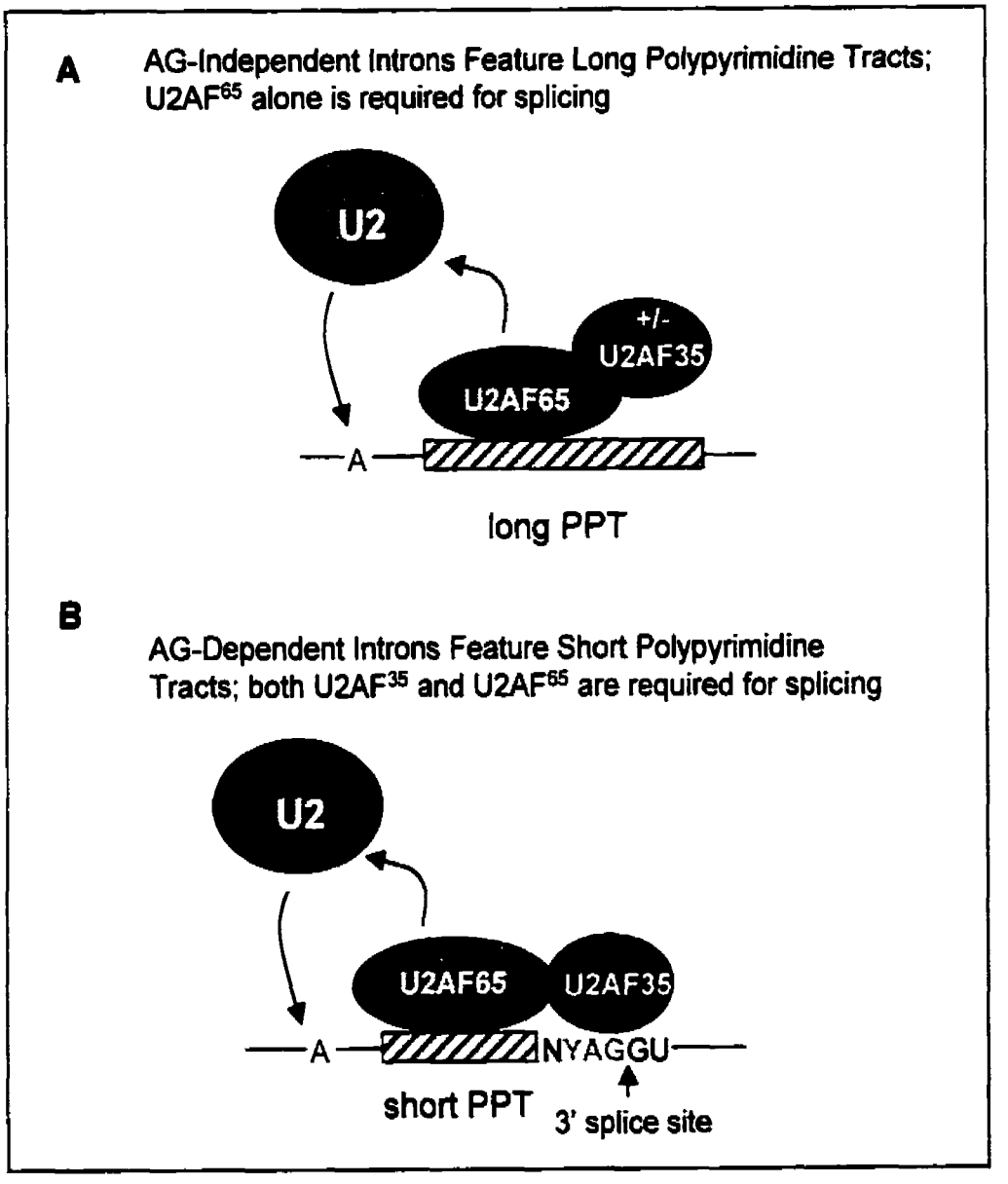


Figure 4: U2AF^{35} recognizes the 3' splice site. The 3' splice site features a highly conserved YAG sequence (orange), where Y is a pyrimidine. (A) The AG-independent intron features a long polypyrimidine tract (PPT) to which U2AF^{65} can bind tightly enough without additional binding of U2AF^{35} required. U2AF^{65} can recruit U2 snRNP to the branchpoint and the presence of U2AF^{35} is not required. (B) The AG-dependent intron has a short/weak PPT so that the additional binding of U2AF^{35} to the 3' splice site is required to stabilize the U2AF heterodimer binding to the PPT. In this case both subunits of U2AF are necessary for splicing to occur (adapted from Moore, 2000).

strong PPT, U2AF^{65} binding to the PPT is strong enough that the presence of U2AF^{35} is not absolutely necessary, and is thus dispensable from splicing (Fig. 4A). In the case of the AG-dependent intron with a short/weak PPT, U2AF^{65} binding to the intron is weak and the additional interaction between U2AF^{35} and the 3' splice site AG is required to maintain U2AF heterodimer binding (Fig. 4B). The binding specifity of U2AF^{35} for the 3' splice site YAG to be 2 nucleotides removed from the end of the PPT may be exploited to allow for alternative splicing. In the case of *male-specific-lethal*-2 (*msl*-2) transcription in *Drosophila*, an unusually long stretch of nucleotides between the PPT and the 3' splice site YAG weakens the interaction between U2AF^{35} and the 3' AG (Merendino *et al.*, 1999). This weakens the overall binding affinity of the U2AF heterodimer with the PPT. Another PPT-binding protein (in this case, *Sex-lethal* protein in *Drosophila*) may then compete with U2AF^{65} for binding to the PPT, thus preventing U2AF from recruiting U2 snRNP to the branch point which in turn prevents splicing. This is an important process in the control of alternative splicing of introns.

3. Methods of Studying Protein Function and Interactions Within a Dynamic Multi-Protein System: Adaptation of the Chemical Ligation Strategy to Attach Chemical Probes

(a) Strategies to Study Protein Structure and Function: Chemoselective Chemical Ligation

There are numerous methods to study the structure and/or function of proteins. Spectroscopic methods such as NMR or X-ray crystallography are more easily applied to small proteins or simple protein complexes. It has only been recently that these spectroscopic methods have been advanced to a stage where they have been applied successfully to the study of large proteins and multi-protein complexes (*e.g.* the ribosome, reviewed in Puglisi *et al.*, 2000; the potassium ion channel, Doyle *et al.*, 1998). For large and complex multi-protein systems which undergo changes over a course of time, such as the spliceosome, different methods must be used to study the protein interactions which are occurring. Chemical reagents that create an identifiable signature within the protein system are useful in

this regard: they may be added to the protein system at various times to track changes in interactions. An example of a group of chemical reagents that are commonly used in this regard are protein cross-linkers, which link interacting proteins with one another. Other examples include compounds that bind to and alter the behaviour of a particular protein, thus affecting the behaviour of the overall system. To study specific interactions, a useful method is to attach the chemical probe in a site-specific manner to a particular protein and observe the signals created by that probe. One way to do this is to use Schultz's method to insert unnatural amino acids at a specific position within the protein (Ellman *et al.*, 1991). Although the method accurately delivers a chemical probe to a specific position within a protein, the relatively low efficiency of charging the tRNA and *in vitro* protein translation can prove to be obstacles in obtaining adequate amounts of protein (Barrett *et al.*, 1999).

Another method to site-specifically derivatize a protein with a chemical probe is to react the nucleophilic residue of an amino acid (*e.g.* cysteine) within the protein in question to a chemical probe. For this method to be useful, the reaction should be site-specific to a given amino acid within the protein to be studied and the yield of the reaction should be high. Chemoselective chemical ligations of peptides provide an ideal starting point for this methodology. Stepwise solid phase synthesis of peptides is limited to a maximum of approximately 100 amino acids (Merrifield, 1965; Gutte & Merrifield, 1969); most proteins of interest to researchers are considerably larger than this. Research has focused on the ligation of two or more peptides, which may be synthesized by solid phase or through recombinant expression, to form large functional proteins (reviewed in Dawson & Kent, 2000; Tam *et al.*, 1999). The common theme of the various chemical ligation strategies is the reaction of a peptide containing a nucleophilic amino acid residue at its N-terminus to the C-terminus of a second peptide, which has been activated with a good leaving group. Two major chemical ligation methods are the thioester-cysteine ("native") ligation by Kent & coworkers (Dawson *et al.*, 1994) and the imine ligation method by Tam & coworkers (Liu & Tam, 1994). Both allow the ligation of two peptides in a single pot reaction and protecting groups are not needed.

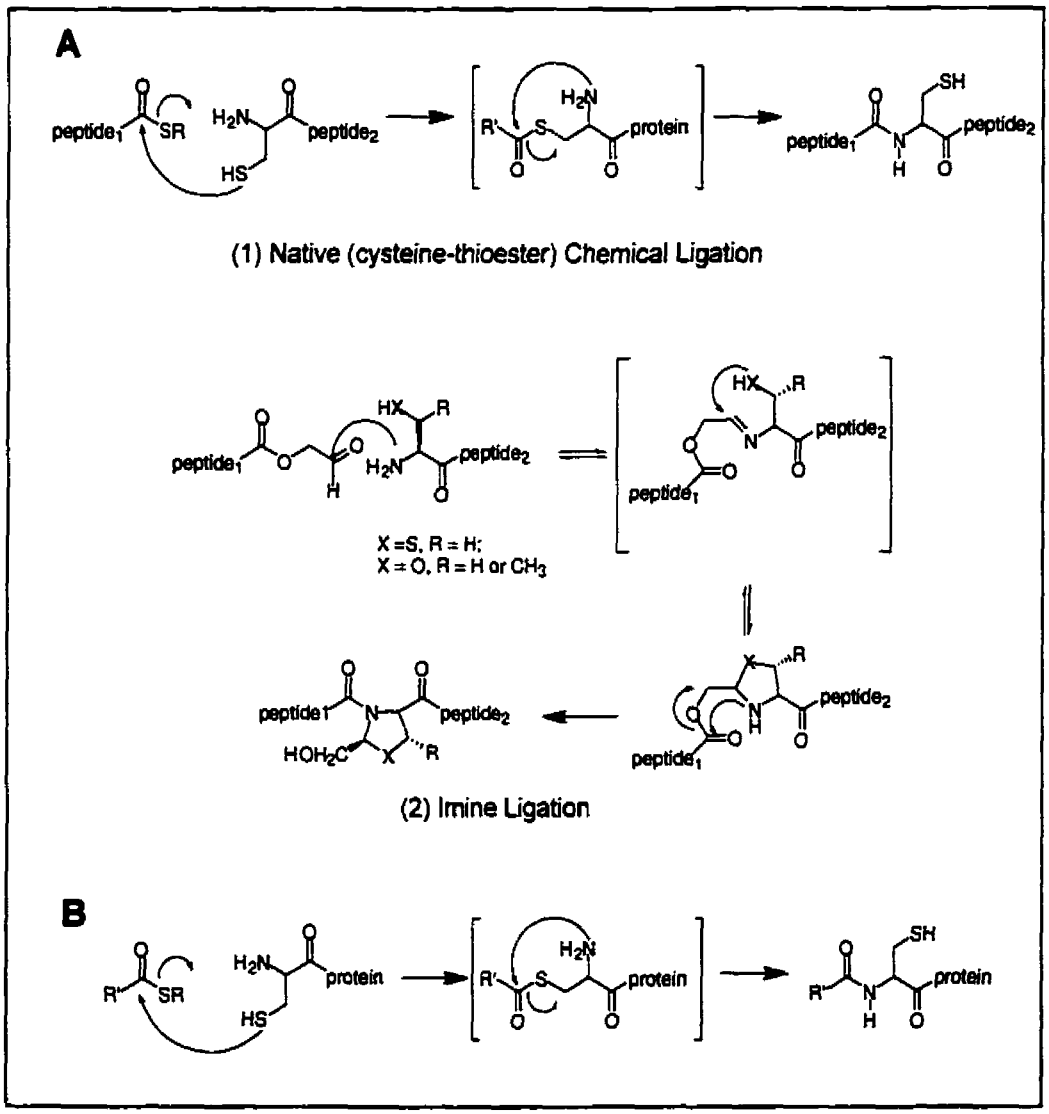


Figure 5: (A) Two major examples of chemoselective peptide ligation methods. (1) Native chemical ligation consists of the reaction of a peptide with a thioester at the C-terminus with a peptide with cysteine at the N-terminus. The first step of this reaction is a transthioesterification. The presence of the amino group causes an internal rearrangement to form an amide bond (Dawson *et al.*, 1994). (2) Imine ligation consists of the reaction of a peptide with an aldehyde at the C-terminus with a peptide with a thiol or hydroxy residue (from cysteine, serine or threonine) at its N-terminus. This reaction forms a proline-like oxazolidine or thiazolidine link between the two peptides (Liu & Tam, 1994). (B) Application of the native chemical ligation reaction to attach a chemical probe/reporter group to the N-terminus of a protein. A thioester derivative of a reporter group (R') is reacted with a protein with an N-terminal cysteine.

Kent & coworkers used the reaction of thioesters with cysteine in their approach to chemical ligation of peptides (Dawson *et al.*, 1994). The reaction of a thioester with cysteine to form an amide bond was first observed by Wieland & Schneider (1953). Reaction of a thioester with thiols forms a new thioester; this reaction is reversible, such that thiols may exchange with the thioester (Fig. 5A). The thiol residue of cysteine may thus react with a thioester reagent to form an intermediate thioester. This reaction is reversible for all cysteines within a protein except if the cysteine is the first amino acid at the N-terminus. In this case, the amino group of this N-terminal cysteine is positioned for an intramolecular attack on this intermediate to form an amide. This second reaction is rapid and irreversible since the amide product is the thermodynamically favoured product. Thus, the ligated product contains a peptide (amide) bond at the ligation site. Kent & coworkers initially used this reaction to join two solid-phase synthesized peptides to form a large, functional protein (Dawson *et al.*, 1994). The first peptide ends with a thioester at its C-terminus and the second peptide begins with a cysteine at its N-terminus. The N-terminal cysteine is essential since the amino group is free for nucleophilic attack on the thioester formed from exchange of the cysteine side chain with the thioester reagent. Internal cysteines may undergo the first reaction with the thioester reagent but because they do not have free amino groups, the reaction is reversible. Besides the selectivity of this reaction, another advantage is that the ligation reaction may be carried out in aqueous buffer, at neutral pH (Dawson & Kent, 2000). Denaturants such as urea may be added to the buffer to unfold the peptides and increase the rate of reaction. Muir & coworkers have further developed the thioester-cysteine strategy to allow the ligation of recombinantly expressed proteins to chemically synthesized proteins with a C-terminal thioester group (Muir *et al.*, 1998).

Besides the thioester-cysteine strategy for joining two peptides, Tam & coworkers have developed another chemical ligation method, the imine ligation, which relies upon the reaction between a peptide with serine, threonine, or cysteine at the N-terminus with another peptide with a C-terminus activated with an aldehyde group (Liu & Tam, 1994; reviewed in Tam *et al.*, 2000). Like the thioester-cysteine reaction, the imine ligation reaction occurs in two steps. The first step is the formation of an imine between the N-

terminal amino group and the C-terminal aldehyde. The nucleophilic thiol or hydroxyl residue is then positioned for an internal attack on the imine. The result of this reaction is the formation of an oxazilidine or thiazolidine ring linking the two peptides (Fig. 5A). The advantage of Tam's methodology is that the N-terminal residue is not limited to cysteine, as with Kent's thioester ligation. The choice of amino acids allows some flexibility in the application of chemical ligation to studying the protein in question. However, the hydroxyl residues of serine and threonine are weak nucleophiles and the ligation reaction must take place in organic solvent when the N-terminal residue is serine or threonine (Tam *et al.*, 2000). The reactivity of the aldehyde at the C-terminus of the second peptide in Tam's ligation method may also be of some concern, so the imine ligation must be carried out in acidic solution to protonate the nucleophilic residues within the peptides.

As stated earlier the chemical ligation methods currently in use for ligating peptides may be adapted to the ligation of chemical probes (reporter groups) to a protein, to study its interactions within a biological system. It is not necessary for the first partner in the chemical ligation reaction to be a peptide with an activated C-terminus. The same reactions may be employed to attach a chemical probe through the N-terminal cysteine of the protein in a site-specific fashion. Of the two methods described above, the thioester-cysteine method is the preferred strategy for the covalent attachment of a chemical probe to a protein (Fig. 5B). The advantage of the thioester-cysteine ligation method over the imine ligation is that the bond that is formed is an amide bond, which is the naturally occuring bond between amino acids within a protein. The oxazilidine or thiazolidine ring, formed when using the imine ligation method, is similar to a proline ring. This would mean that a chemical probe, if attached using this method, would be locked in a certain position with respect to the protein. When ligating a chemical probe to a protein, the length and flexibility of the link between the protein and the chemical probe will affect the range of the observed signal from the probe. It is preferable to have a flexible linker since there may be a chance that a rigid protein-probe link, such as with an oxazilidine or thiazolidine ring, would direct the chemical probe towards the interior of the attached protein, thus obscuring the observed signal from the probe. Verdine & co-workers have employed the thioester-cysteine ligation method to attach

a thioester derivative of EDTA to a (DNA-binding) transcription factor (Erlanson *et al.*, 1996). The thioester reagent and N-terminal cysteine method is extremely useful for investigating protein interactions which are occuring at or near its N-terminus, since it allows the site-specific attachment of chemical probes at the N-terminus.

(b) Application of the Thioester-Cysteine Chemical Ligation Strategy to U2AF^{65}

Considering that the RS domain of U2AF^{65} begins almost immediately at the N-terminus, it is an ideal target for the thioester-cysteine reagent strategy. This strategy relies on the engineering of a cysteine residue at the N-terminus of the protein, so that it may react with a chemical probe containing a thioester functional group, thus forming an amide bond between the N-terminus of the protein and the chemical probe. Besides attaching chemical probes to the N-terminus of the wild type U2AF^{65}, N-terminal truncation mutants of U2AF^{65} could be engineered, all beginning with cysteine. The effect of this would be to place the chemical probe at various locations in U2AF^{65}, at the beginning and the end of the RS domain, or within the RS domain. Two types of chemical probes are to be attached to the U2AF^{65} RS domain. The first type of chemical probe would be an activatable nuclease reagent, such as Fe^{2+}-EDTA, which would detect protein-RNA interactions and identify the location of such interactions on the pre-mRNA. The second reagent would be benzophenone, a protein cross-linker that may be used to detect protein-protein interactions. These reagents are discussed in further detail below.

(c) Development of Thioester Reagents Containing Reporter Groups

Two chemical probes were required to study the RNA-protein interactions and the protein-protein interactions of U2AF^{65}. A third chemical probe was developed as a tool to identify N-terminal cysteines in the U2AF^{65} mutants. The following reagents, with descriptions of their chemistry and use, were employed in the study of U2AF^{65}.

The combination of Fe^{2+} and H_2O_2 was first observed to oxidize organic compounds by Fenton in 1894 and is commonly known as Fenton's reagent (Walling, 1957). Fe^{2+} (in aqueous solution) in the presence of hydrogen peroxide, will oxidize to Fe^{3+} and also generate

hydroxyl radicals (Fig. 6). Furthermore, Fe^{3+} may be readily reduced *in situ* by the addition of ascorbic acid to the reaction, thus cycling Fe^{2+} back into another oxidation reaction with H_2O_2 and generating more hydroxyl radicals (Martell, 1982). Fe^{2+}/H_2O_2 may be used as an activatable nuclease reagent. The addition of H_2O_2 to a reaction mixture containing Fe^{2+} will produce hydroxyl radicals. These radicals will initiate cleavage of the phosphodiester backbone of any nearby nucleic acid, usually beginning with abstraction of hydrogen from either the 4' or 5' position of the sugar ring (Pogelzelski & Tullius, 1998). EDTA is employed as a carrier for Fe^{2+} in these reactions. The hydroxyl radicals that are released during the course of the reaction will only cleave portions of the nucleic acid strand which are exposed to the solvent.

In order to study a protein-nucleic acid interaction, Fe^{2+}-EDTA may be free in solution (cleaving all exposed portions of the nucleic acid, Fig. 6A; Tullius & Dombroski, 1986) or tethered to the protein in question (cleaving only nucleic acid in close proximity, Fig. 6B; Erlanson *et al.*, 1996). Following cleavage of a nucleic acid substrate by Fe^{2+}-EDTA/H_2O_2, the resultant fragments may be analyzed by polyacrylamide gel electrophoresis to provide a distinctive "footprint" of the protein-nucleic acid interaction, as shown in Fig. 6C. The nucleic acid substrate is labeled at one end with a radioisotope (^{32}P), commonly the 5' end. After hydroxyl radical cleavage of the nucleic acid substrate, the fragments are separated by electrophoresis and the shorter fragments (observed by autoradiography) will contain the 5' end label and run near the bottom of the gel. This gives information on where the cleavages occur with respect to the 5' end of the substrate. The two types of experiments that may be carried out with Fe^{2+}-EDTA are discussed in detail below.

When a protein is bound to the nucleic acid strand, that portion is protected from hydroxyl radical initiated cleavage. When Fe^{2+}-EDTA is free in solution and H_2O_2 is added to initiate hydroxyl radical production, a protection footprint is observed, where no substrate fragments will be obtained containing the region of the substrate to which the protein was bound (Fig. 5C, panel A). This technique is referred to as hydroxyl radical (protection) footprinting and is conceptually similar to the use of enzyme nucleases (*e.g.* DNase I) to footprint the binding of protein on DNA. Tullius & coworkers used the Fe^{2+}/H_2O_2 system (added exogenously) to study protein-DNA complexes (Tullius & Dombroski, 1986; Dixon *et al.*, 1991). Fe^{2+}-EDTA has been used in several instances to study the interactions of transcription factors with DNA (Zorbas *et al.*, 1989). The same technique has been

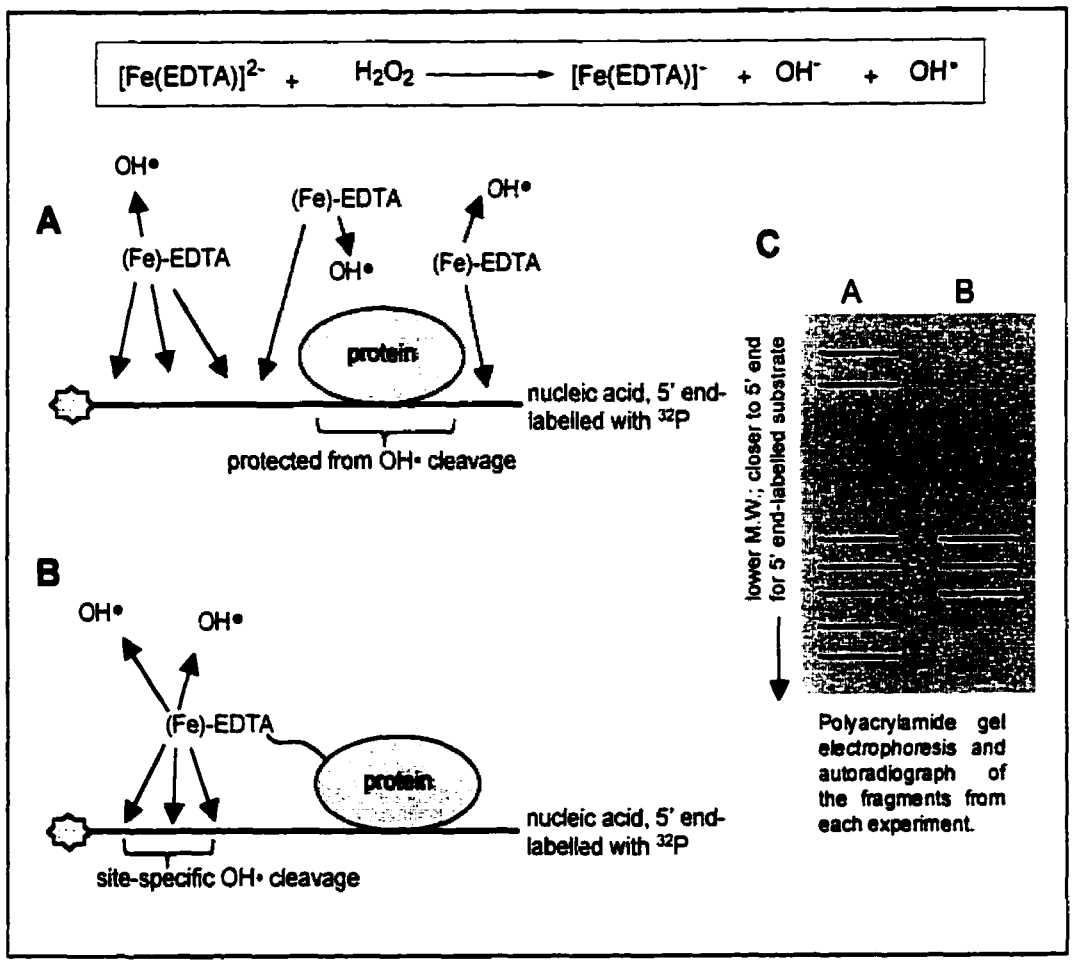


Figure 6: The reaction of Fe^{2+} with H_2O_2 produces hydroxyl radicals, which initiate cleavage of the phosphodiester backbone of DNA/RNA by abstraction of hydrogen atoms from the sugar moiety. This reaction can be used to footprint the location of a nucleic-acid binding protein on its nucleic acid substrate. (A) Protection footprinting: (Fe)-EDTA and H_2O_2 are added exogenously. Cleavage occurs at positions all along the nucleic acid sequence except the position where the protein is bound. (B) (Fe)-EDTA is tethered to the protein to produce site-specific hydroxyl radical initiated cleavage. (C) Separation of the fragments obtained from (A) and (B) by polyacrylamide gel electrophoresis & detection by autoradiography. In (B), site-specific cleavage gives information on the location of the domain to which the (Fe)-EDTA is tethered.

employed to study RNA splicing complexes in which Fe^{2+}-EDTA reagent was added exogenously (Wang & Padgett, 1989). It has been recently used to study the interaction of the RNA binding domain of U1A protein with RNA (Beck *et al.*, 1998). There are a few major advantages of the hydroxyl radical footprinting technique over enzymatic nuclease footprinting. The hydroxyl radical is extremely small and uncharged, so it may penetrate into areas not accessible to an enzyme, particularly in RNA which often has a secondary structure. Also, hydroxyl radicals may cleave at any base; enzyme nucleases are often base-specific. Another advantage of the Fe^{2+}-EDTA reagent is that the generation of hydroxyl radicals from the oxidation of Fe^{2+} may be carefully controlled by the respective amounts of H_2O_2 and ascorbic acid added.

EDTA, when tethered to a specific location within a protein, acts as a means of localizing the nuclease activity of Fe^{2+} to that particular location. When Fe^{2+}-EDTA is tethered to the nucleic acid-binding protein, causing site-specific cleavages on the nucleic acid substrate as shown in Fig. 5B, only a few fragments of the substrate will be seen. In the example given in Fig. 5B, the cleavages occur near the 5' end of a 5' ^{32}P-end labeled substrate, giving rise to short fragments containing the 5' end (Fig. 5C, panel B). The location of these cleavages are determined by the length of the fragments obtained, thus giving information on the location of the chemical probe and the protein domain to which it is attached, with respect to the 5' end of the substrate. The tethering of Fe^{2+} ion with EDTA to a nucleic acid binding protein, to be used as a site-specific nuclease, has been employed in a few cases (Erlanson *et al.*, 1996; Hall & Fox, 1999). As a thioester reagent, Fe^{2+}-EDTA would act as a site-specific nuclease, tethered to the N-terminus of the protein. EDTA-3MPA, a thioester derivative of EDTA, was synthesized by Verdine & coworkers and used as a site-specific nuclease reagent when tethered to a DNA-binding protein (Erlanson *et al.*, 1996). EDTA-3MPA was chosen as a nuclease reagent to be tethered to $U2AF^{65}$ to study the RS domain positioning with respect to the pre-mRNA.

Benzophenone and other aryl carbonyl derivatives have been widely used as protein cross-linkers, in order to identify protein-protein interactions (Dormán & Prestwich, 1994). Benzophenone is a photoactivated reagent: exposure to radiation of ~350 nm excites the

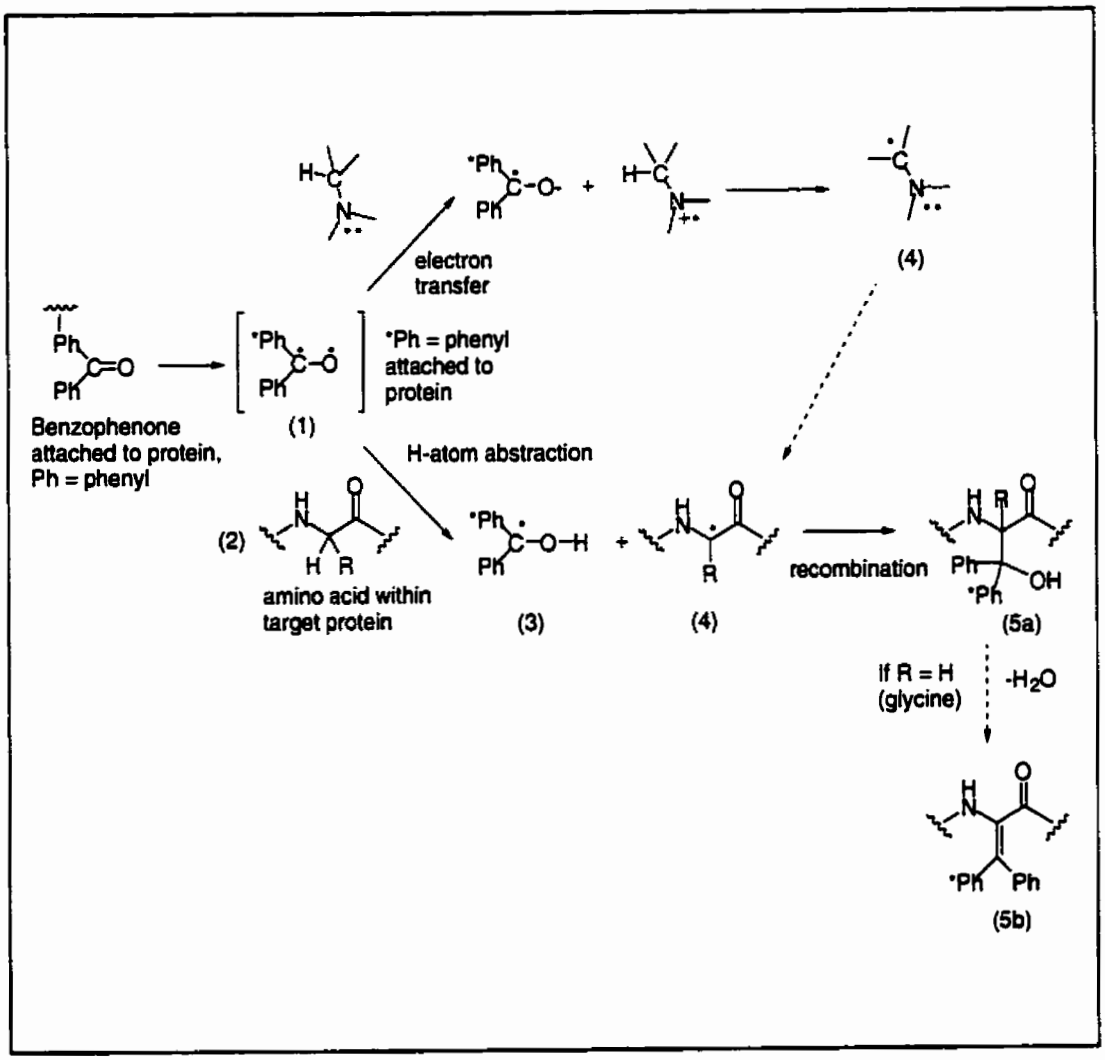


Figure 7: The use of benzophenone photochemistry in protein crosslinking. Excitation of the carbonyl group of benzophenone with UV radiation (~350 nm) causes an n→π* transition to form a "diradical"-like triplet state (1). The excited state of benzophenone may then follow two possible reaction pathways, both leading to the formation of a covalent bond between benzophenone (attached to the protein of interest) and the interacting protein (adapted from Dormán & Prestwich, 1994).

carbonyl group to form a "diradicaloid triplet state" (1, Fig. 7), so that the oxygen becomes weakly electrophilic. The oxygen will abstract the α-hydrogen from a nearby amino acid residue, to produce the benzophenone radical (3) and amino acid ketyl radical (4). These recombine to form a new C-C bond between the two groups. Alternatively, if an amine is adjacent to the excited benzophenone, the electrophilic oxygen will abstract an electron from the nitrogen to form a nitrogen radical (upper panel). The nitrogen radical will abstract hydrogen from an adjacent carbon to form a new alkyl radical (4) which then recombines with the benzophenone radical (3) in the same manner as the ketyl radical. The application of benzophenone as a protein cross-linker is to tether it to a protein and then attempt to cross-link the benzophenone-derivatized protein to any other biomolecule with which it interacts. Benzophenone is highly useful since it may be activated by a relatively long wavelength of 350 nm, which does not harm protein or nucleic acid. Also, it does not require a radical scavenger when it relaxes from the excited state, so that it may be excited many times over. This increases the efficiency of the cross-linking reaction as there is the opportunity for the benzophenone to come within close proximity of the interacting protein, while the proteins are tumbling in solution. The specificity of the benzophenone reagent is dictated by the length and rigidity of the linker between the benzophenone group and the protein. The range of benzophenone's activity as a cross-linker is within 3 Å (Dormán & Prestwich, 1994), so that only close protein-protein interactions may be detected, given a short and rigid linker. A thioester derivative of benzophenone may therefore be used to detect protein-protein interactions occurring at the N-terminus of a protein.

In addition to these compounds a thioester form of biotin was synthesized, to be used as a marker for the presence of the N-terminal cysteine and to detect the reaction of that cysteine to another thioester reagent. Biotin is commonly used to label proteins in biochemical assays; it is often used in conjunction with streptavidin because the very high affinity of biotin-streptavidin binding. Streptavidin based reagents (*e.g.* resin bound streptavidin) and antibodies against biotin are readily available to detect the presence of the biotin label. For example, the presence of biotin may be detected with an antibody against biotin in an immunoblot (Western blot). The thioester derivative of biotin, would thus be

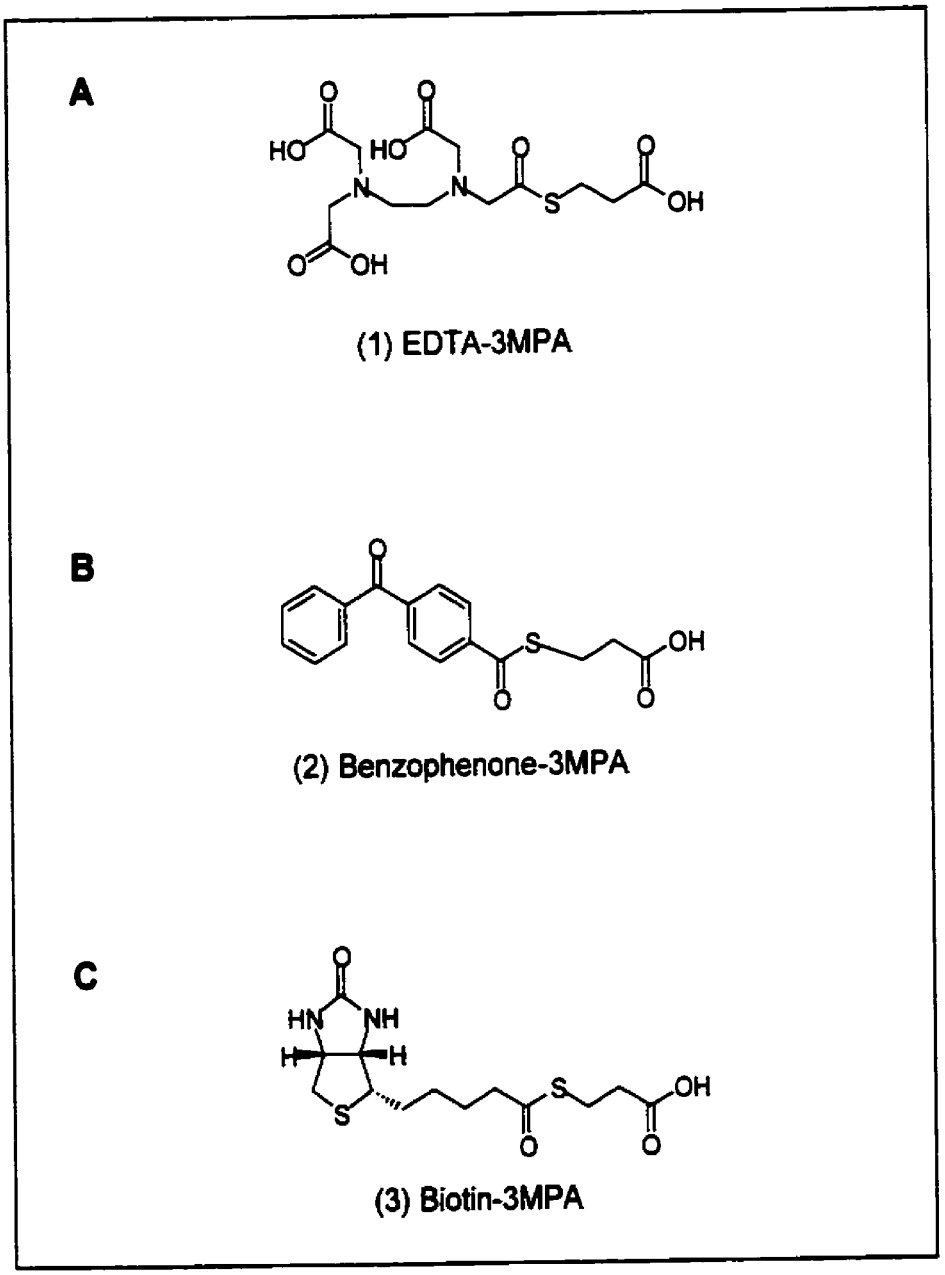


Figure 8: Chemical structures of the thioester reagents (A) EDTA-3MPA, (B) benzophenone-3MPA and (C) biotin-3MPA.

useful as a marker to identify the presence and reactivity of a cysteine at the N-terminus of a protein.

The synthesis of a group of thioester derivatives of chemical probes or labels (in the case of biotin) was carried out, with the purpose of using these reagents to derivatize proteins to which a cysteine residue has been engineered into the start of the N-terminus. 3-Mercaptopropionic acid was employed in all cases to create the thioester derivatives, which are EDTA-3MPA (**1**), benzophenone-3MPA (**2**), and biotin-3MPA (**3**), summarized in Fig. 8.

(d) Development of a Series of U2AF^{65} N-terminal Cysteine Mutants

U2AF^{65}, as with all other naturally occurring proteins, begins with methionine at the N-terminus. This is due to the start codon in all genes corresponding to methionine. The thioester reagent strategy depends on the presence of an N-terminal cysteine in the protein, so this cysteine residue must be engineered into the beginning of the protein sequence. To create an N-terminal cysteine in U2AF^{65}, a protease recognition site in inserted before a cysteine residue that is immediately followed by the rest of the protein sequence. The selected protease must cleave the protein at the end of its recognition site to produce a protein with cysteine at the N-terminus. Factor X_a was chosen because it cleaves at the C-terminal end of the sequence IEGR. Insertion into the pET (Novagen) plasmid expression vector, transformation of the vector into *E.coli*, and subsequent expression produces a protein which will have a histidine tag (his-tag; composed of six consecutive histidines) from the pET expression vector, followed by the protease site, the cysteine residue, and the U2AF^{65} sequence (Fig. 9A). Treatment of the his-tagged protein with Factor Xa removes all of the N-terminal peptide portion up to and including the protease site. This gives a free cysteine residue at the N-terminus of the protein. This protein may then be derivatized with a thioester reagent. A series of mutants were produced, all of which contain at their N-termini the Factor Xa protease recognition site followed by cysteine. These mutants include the full-length U2AF^{65} and a series of N-terminal truncations of U2AF^{65}: Δ1-14, Δ1-24, Δ1-64, Δ1-94 (Fig. 9B). The Δ1-14 truncation places the cysteine at the beginning of the RS domain,

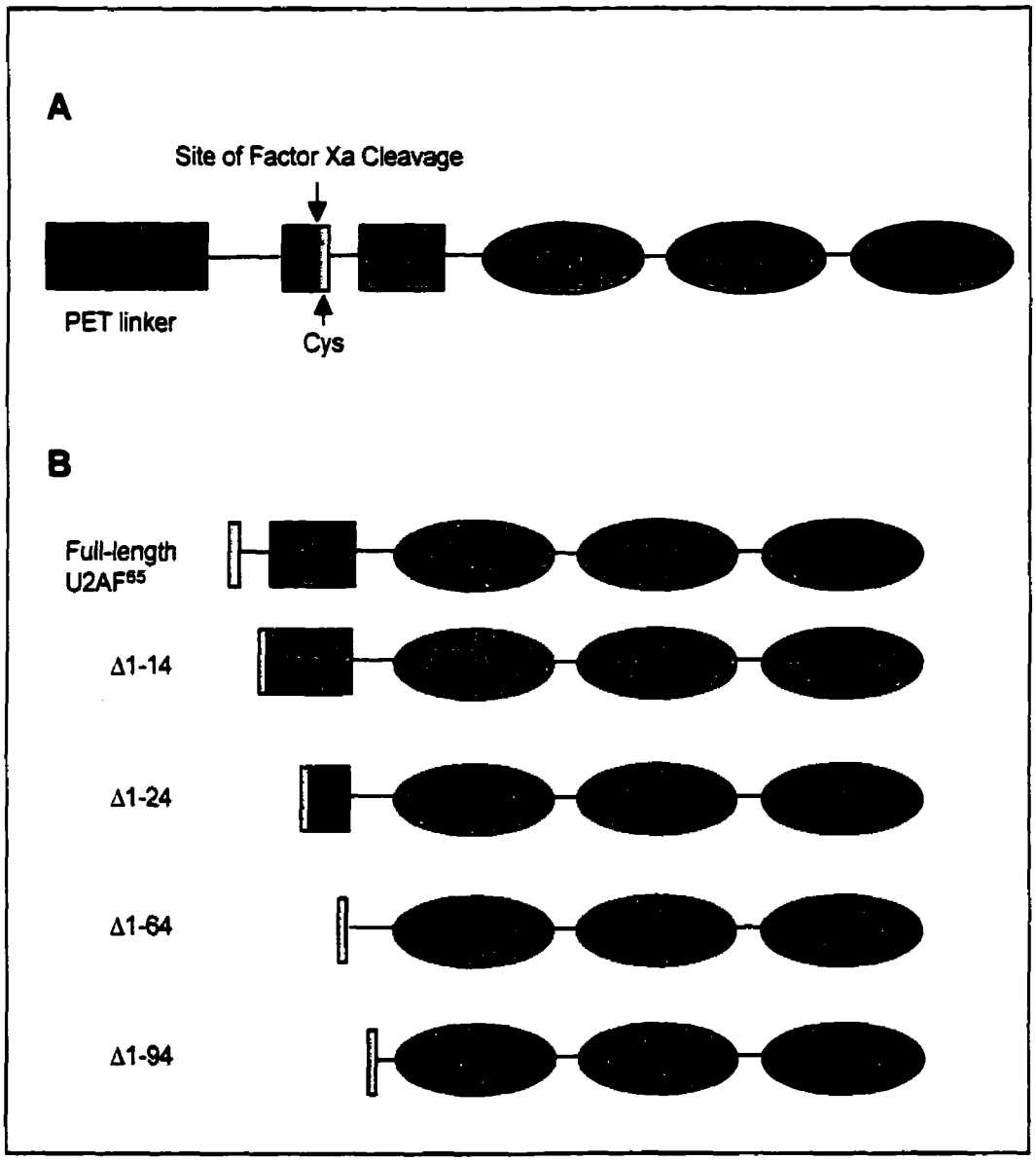


Figure 9: Creation of the N-terminal cysteine U2AF^{65} mutants. (A) The domain structure of the mutant U2AF^{65} after expression from transfected *E. coli* (see Section A.3(d) for details). The Factor Xa protease cleavage site is shown in red. Upon cleavage by Factor Xa, all of the peptide sequence preceding the cysteine will be removed and the U2AF^{65} protein will have cysteine (yellow) at the N-terminus. (B) The N-terminal cysteine U2AF^{65} mutants after cleavage by Factor Xa: full-length U2AF^{65} and proposed N-terminal truncation mutants of U2AF^{65}.

the Δ1-24 truncation places the cysteine within the RS domain, and the Δ1-64 and Δ1-94 mutants are missing the RS domain entirely.

4. Experiments with the U2AF^{65} N-terminal Cysteine Mutants

(a) Assay to Determine the Presence of N-terminal Cysteines in the U2AF^{65} Mutants; Activity of the U2AF^{65} N-terminal Cysteine Mutants

A series of U2AF^{65} mutants beginning with cysteine was created, so that they might be chemically derivatized with thioester derivatives of reporter groups such as Fe^{2+}-EDTA (a nuclease) and benzophenone (a protein-protein cross-linking reagent). The group of U2AF^{65} mutants that could be expressed and purified in adequate quantities included full length U2AF^{65} and the N-terminal truncated versions of U2AF^{65}, Δ1-14, Δ1-64, and Δ1-94. The U2AF^{65} cysteine mutants were cleaved with Factor Xa protease to expose the N-terminal cysteine. The mutants were then reacted with biotin-3MPA to identify the presence and reactivity of the N-terminal cysteine. The biotin-3MPA reagent is a thioester which will react with an N-terminal cysteine, thus labeling the protein with biotin. The biotin label may then be detected by either a streptavidin-based assay or an anti-biotin antibody. An immunoblot (Western blot) assay using anti-biotin antibody was chosen because of its convenience and sensitivity. It was tested on one of the N-terminal cysteine mutants of U2AF^{65}, Δ1-64. The assay was also used to detect successful reaction of the N-terminal cysteine with other thioester reagents and as a qualitative means of estimating the extent of the reaction. Biotin-3MPA is shown here to be a novel and useful reagent for detecting the presence of N-terminal cysteines in proteins, and detecting the successful reaction of the cysteine with other thioester reagents.

The U2AF^{65} N-terminal cysteine mutants were tested in two ways to ascertain their viability and activity before continuing to experiments with the chemical probes. The first test was to see whether the U2AF^{65} proteins were able to bind to a short RNA containing the intron sequence, *i.e.* the branch point region, the polypyrimidine tract and the 3' splice site. U2AF^{35} was added to observe if it was able to bind to the U2AF^{65} cysteine mutants. These

two observations were to make certain that the U2AF^{65} mutants were able to bind to the polypyrimidine tract within the intron and able to form a heterodimer with U2AF^{35}. The second test was to see whether the U2AF^{65} mutants were able to reconstitute RNA splicing when added to a U2AF-depleted nuclear extract. Green & coworkers observed that the presence of the RS domain in U2AF^{65} was necessary for splicing to occur (Zamore *et al.*, 1992) so it was expected the SR deletion mutants (Δ1-64, Δ1-94) would not reconstitute splicing. The ability to reconstitute splicing in U2AF-depleted nuclear extract was considered to be proof of the activity of the protein.

(b) Hydroxyl Radical Footprinting of EDTA-Derivatized U2AF^{65} on pre-mRNA

The attachment of EDTA to RS domain of U2AF^{65} was to serve as a marker for its location with respect to the pre-mRNA. The attachment of a nuclease to the RS domain of U2AF^{65} should allow one to identify the position of the RS domain with respect to the pre-mRNA, in particular to the branch point region or the 3' splice site. Green & coworkers observed that, upon binding U2AF^{65} to a pre-mRNA sequence, the RS domain of U2AF^{65} could be cross-linked to a photoactivatible 4-thiouridine base immediately 3' to the branchpoint adenosine (Valcárcel *et al.*, 1996). From this observation, it was suggested that U2AF^{65}'s RS domain may play some role in recognition of the branchpoint region.

U2AF^{65} interacts with other splicing factors during commitment complex formation. The third RNA binding domain of U2AF^{65} (RRM3) is required for high affinity binding to the polypyrimidine tract (Zamore *et al.*, 1992); RRM3 also contacts a branchpoint-binding protein, SF1/mBBP (Berglund *et al.*, 1998; Rain *et al*, 1998). SF1/mBBP is believed to be a component of the early (E) complex, a factor that recognizes the branchpoint region early on and along with U2AF, facilitates the recruitment of U2 snRNP to the site. SF1/mBBP has been shown to contact U2AF^{65}'s RRM3 and the two proteins bind cooperatively to the pre-mRNA (Berglund *et al.*, 1998; Rain *et al.*, 1998). This raises the possibility that the pre-mRNA may be looped around the U2AF protein so that all of the above mentioned contacts are possible (Fig. 10). It is possible that U2AF plays another important role in early spliceosome formation besides U2 snRNP recruitment, that it may be configuring the secondary structure of the pre-mRNA through its RNA and protein contacts. The attachment of the Fe^{2+}-EDTA nuclease right on the N-terminus of the RS domain and

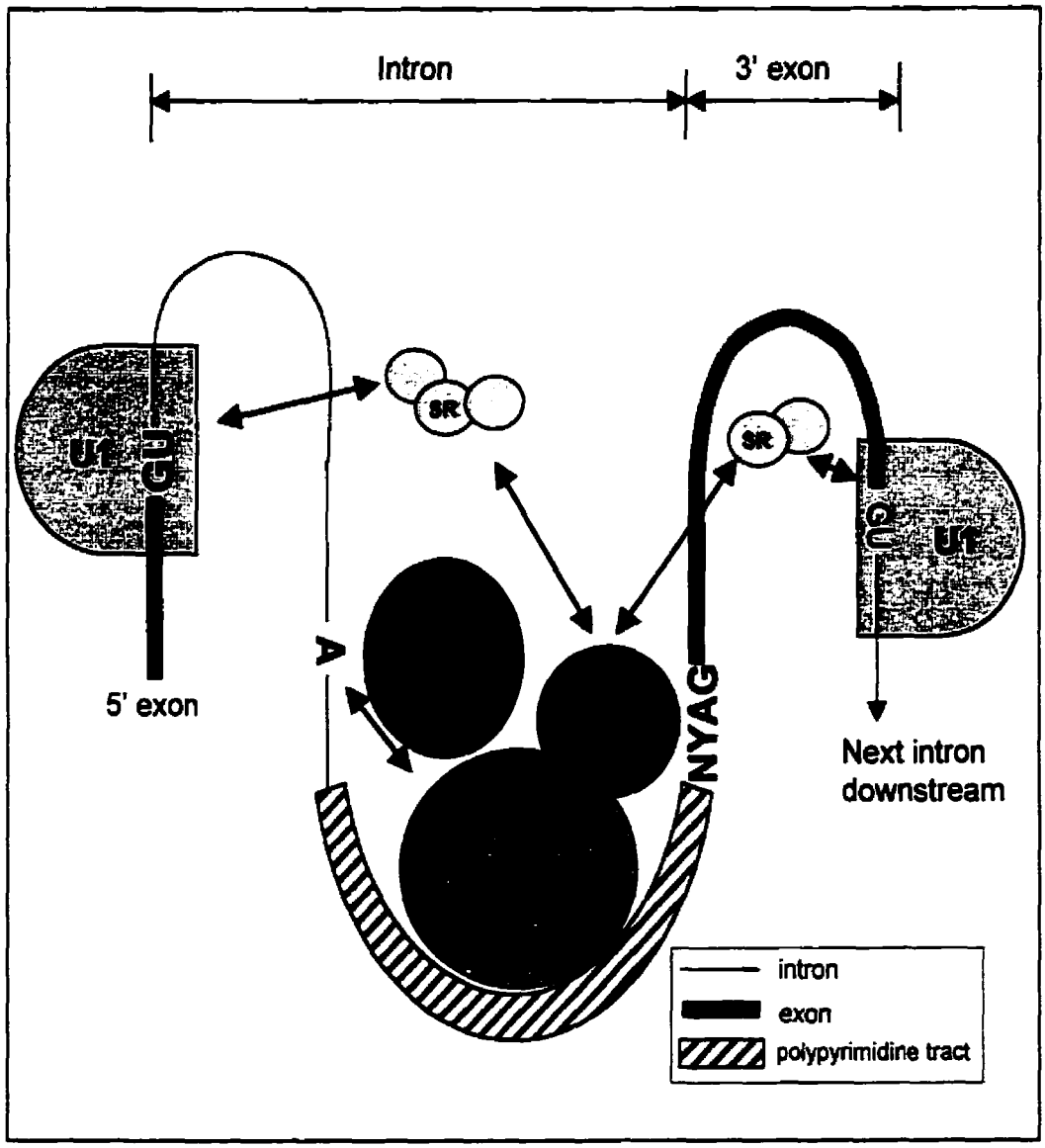


Figure 10: Summary of major U2AF^{65}-U2AF^{35} contacts during early spliceosome complex formation. U2AF^{35} contacts the 3' splice site. Also, U2AF^{35} forms part of a bridge of SR proteins across the intron to U1 snRNP bound to the 5' splice site of the upstream exon and across the downstream exon to the U1 snRNP bound to the 5' splice site of the next intron (Wu & Maniatis, 1993; Zuo & Maniatis, 1996). U2AF^{65} binds to the polypyrimidine tract. Its RS domain contacts the branchpoint adenosine (Valcárcel *et al.*, 1996). The C-terminus of U2AF^{65} contacts SF1/mBBP which contacts the branchpoint adenosine (Rain *et al.*, 1998). To accomodate all of these contacts, the pre-mRNA is shown looped around the various proteins.

subsequent activation of the nuclease to cleave nearby RNA, should identify the location of the RS domain with respect to the pre-mRNA. In the case of the deletion mutant Δ1-14, the placement of the Fe^{2+}-EDTA nuclease is much closer to the high concentration of arginine and serine residues between amino acids 23-64 of the $U2AF^{65}$ sequence. The RS domain may have a dual RNA and protein interaction function and the Δ1-14 mutant may be useful in identifying such interactions.

(c) Cross-linking of Benzophenone-Derivatized $U2AF^{65}$ Constructs in a Minimal System

Attachment of benzophenone to the N-terminus of the $U2AF^{65}$ cysteine mutants was to allow the identification of proteins that are interacting with the RS domain. Such proteins may be regulating the function of U2AF in the spliceosome. The specificity of the benzophenone cross-linking activity is highly dependent on the flexibility and length of the linker between the benzophenone group and the rest of the protein. The longer and more flexible the linker is, the more likely the benzophenone group will cross-link with another protein. A possible disadvantage of a long, flexible linker is the creation of cross-links to proteins which do not actually interact with $U2AF^{65}$'s RS domain and affect $U2AF^{65}$ activity. In this case, benzophenone-3MPA has a very short and inflexible linker (an amide bond) and will therefore only cross-link proteins in the immediate vicinity of the N-terminus of the $U2AF^{65}$ proteins. Because of the shortness and inflexibility of the linker, it is possible that the benzophenone group is too far removed to cross-link to an interacting protein, particularly if the other protein is interacting with a portion of $U2AF^{65}$ that is removed from the N-terminus. The cross-linking activity of benzophenone-derivatized $U2AF^{65}$ cysteine mutants was tested by their ability to cross-link to $U2AF^{35}$. Amino acids 64-182 of $U2AF^{65}$ are believed to interact with $U2AF^{35}$. This section of amino acids follows directly after the RS domain and includes a portion of the first RNA binding domain, which is found in amino acids 151-229 (Zhang *et al.*, 1992; Zamore *et al.*, 1992). Therefore, it was thought that full-length $U2AF^{65}$ would likely not cross-link to $U2AF^{35}$ because the benzophenone group is likely too far from the site of interaction. It was also thought possible (Δ1-64)$U2AF^{65}$ could cross-link with $U2AF^{35}$, since the benzophenone group would be placed immediately at the start of the $U2AF^{65}$ site of interaction with $U2AF^{35}$. This assay may be useful in determining the activity of the benzophenone group and also give a estimate of the specificity and

therefore the usefulness of the benzophenone-3MPA reagent. If it was found to be too specific then the benzophenone reagent could be redesigned to have a long, flexible linker group between the benzophenone group and the thioester portion.

$U2AF^{65}$ is also known to interact with other splicing factors besides SF1/mBBP. UAP56 (56 kDa $hU2AF^{65}$ associated protein) is a putative RNA helicase that is believed to interact with amino acids 138-183 in $U2AF^{65}$, which form part of RRM1 (Fleckner *et al.*, 1997; Ito *et al.*, 1999). UAP56 is required for U2 snRNP binding to the branchpoint (Fleckner *et al.*, 1997). $U2AF^{65}$ also interacts with SAP 155, a component of U2 snRNP; this interaction occurs through amino acids 334-475, which includes $U2AF^{65}$'s RRM3 (Gozani *et al.*, 1998). Reed & coworkers postulated that the interaction between $U2AF^{65}$ and SAP155 is responsible for the recruitment of U2 snRNP to the branchpoint region (*ibid.*). Recently, it was found that the N-terminus of $U2AF^{65}$ interacts with poly(A) polymerase (Vagner *et al.*, 2000). Poly(A) polymerase (PAP) adds a series of adenosine bases to the 3' end of pre-mRNA. PAP binds to a conserved poly(A) signal near the 3' end of the pre-mRNA. There is evidence that RNA splicing and polyadenylation by PAP are coupled (Gunderson *et al.*, 1997). Mattaj & coworkers found that the carboxy terminal domain of PAP interacts with amino acids 17-47 of $U2AF^{65}$ (Vagner *et al.*, 2000); this sequence forms the major portion of the RS domain. A protein representing PAP's carboxy terminal domain was tethered downstream to an intron in the substrate RNA, thereby stimulating $U2AF^{65}$ binding to the intron and splicing of that intron. This suggests that selection of introns for splicing is regulated in part by the presence of an adjacent poly A signal, to which PAP binds and then recruits U2AF to the intron. The RS domain of U2AF therefore might serve as an effector domain, as earlier suggested by Kramer (1996). Benzophenone derivatization at the beginning and end of the $U2AF^{65}$ RS domain (using the following $U2AF^{65}$ mutants: full-length or Δ1-14; Δ1-64 or Δ1-94) could identify interactions with UAP56 and PAP, both of which are believed to interact with regions of $U2AF^{65}$ towards its N-terminus. Cross-linking would likely not be observed with proteins that interact with C-terminal regions of $U2AF^{65}$, such as SF1/mBBP and SAP 155, since the benzophenone tether is very short and rigid.

SECTION B: EXPERIMENTAL PROCEDURE

Unless noted otherwise, all chemical reagents were obtained from Sigma-Aldrich. Acetonitrile used throughout was dried over calcium hydride and distilled before use. All nucleic acid precipitations refer to the standard phenol/chloroform/ethanol precipitation procedure described by Maniatis *et al.*, 1982. Adjustments to standard protocols are mentioned as required.

1. Synthesis of EDTA-3MPA

EDTA-3MPA (**1**) was synthesized and purified as described by Erlanson *et al.* (1996). The synthesis is shown in Scheme 1. Low resolution MS-electrospray: $(M-H)^-$ is 379 ($C_{13}H_{20}O_9N_2S$, calculated average mass 380). The EDTA-3MPA solid was dissolved in distilled water to make a 185 mM stock solution and was stored at –20°C.

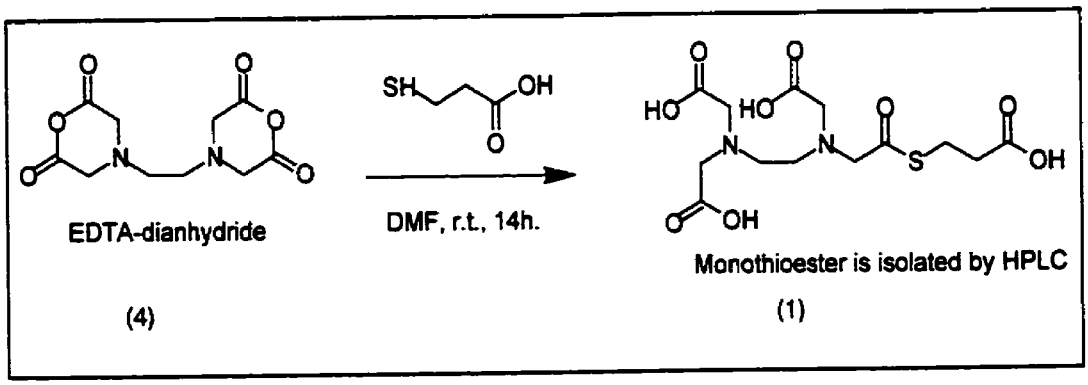
2. Synthesis of Benzophenone-3MPA

The synthesis of benzophenone-3MPA (**2**) is outlined in Scheme 2. 4-Benzoylbenzoic acid (**5**) (5 g, 22 mmol) was dissolved in neat thionyl chloride (20 mL) and heated with stirring until the solid dissolved. The flask was allowed to cool to room temperature and then the flask contents were dried *in vacuo*. The solid was then evaporated from CH_3CN twice. This gave 4-benzoyl benzoic acid chloride (**6**). 4-Benzoylbenzoic acid chloride (approx. 22 mmol) was dissolved in acetonitrile, and diisopropylethylamine (10 mL, 58 mmol) was added. The mixture turned orange at this point. *p*-Nitrophenol (8 g, 58 mmol) was dissolved in 10 mL CH_3CN, and added dropwise to the reaction mixture while stirring. Immediately upon addition of *p*-nitrophenol, a white precipitate formed in the solution. The precipitate, benzophenone *p*-nitrophenyl ester (**7**) (5.6 g, 16 mmol; 73% yield), was filtered off and washed with CH_3CN. It was relatively pure, according to NMR analysis. ^{1}H-NMR (d-DMSO): δ7.60 (m, 2H, ArH), δ7.67 (m, 2H, ArH), δ7.72 (m, 1H, ArH), δ7.78 (m, 2H, ArH), δ7.92 (m, 2H, ArH), δ8.32 (m, 2H, ArH), δ8.38 (m, 2H, ArH). Benzophenone *p*-nitrophenyl ester was dissolved in DMF and 3-mercaptopropionic acid added. The mixture was stirred at room temperature overnight. Purification of the

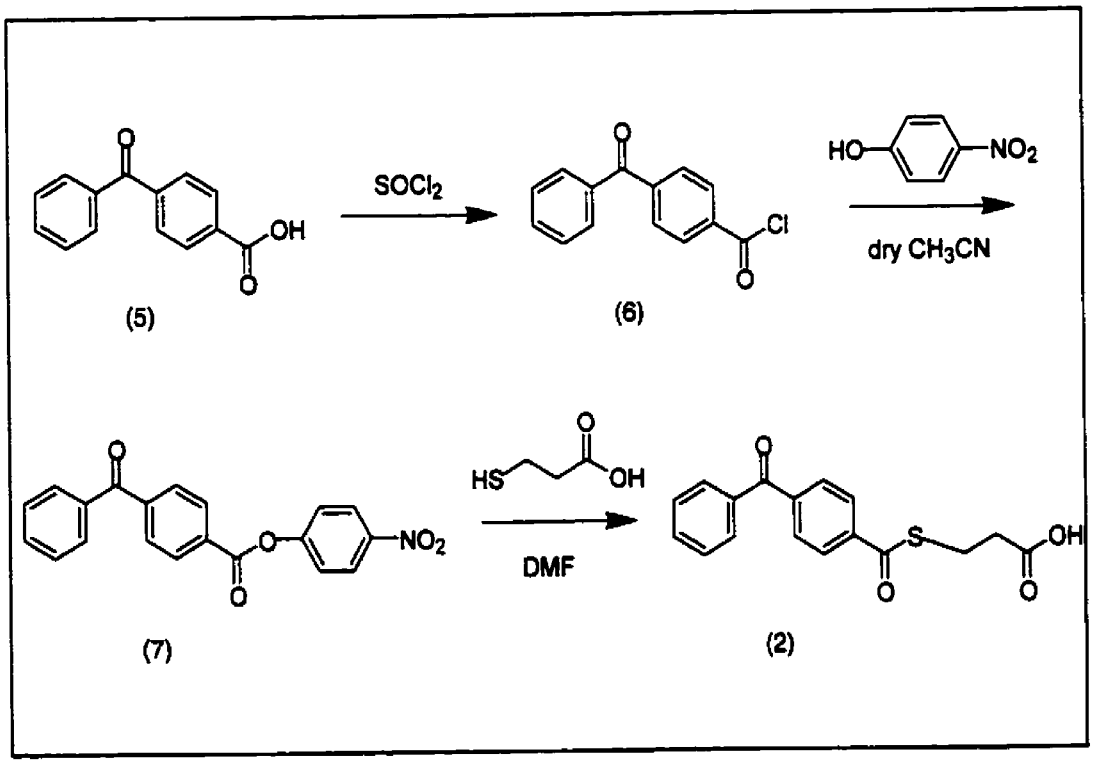
reaction mixture by column chromatography (dichloromethane over silica, 30-60 mesh) separated the thioester benzophenone-3MPA (**2**) from the starting compounds. Benzophenone-3MPA is a a pale yellow fine powdery solid, m.p. 137-138°C.High resolution MS-EI: M^+ is 314.062443 ($C_{17}H_{14}O_4S$, calculated mass 314.061281) ^{1}H-NMR (d-$CDCl_3$): δ2.83 (t, 2H, CH_2), δ3.36 (t, 2H, CH_2), δ7.50 (m, 2H, ArH), δ7.62 (m, 1H, ArH), δ7.79 (m, 2H, ArH), δ7.85 (m, 2H, ArH), δ8.05 (m, 2H, ArH), δ9-10 (broad s, carboxylic acid H). For experimental use, benzophenone-3MPA was dissolved in absolute ethanol to make a 30 mM solution which could be stored for a week at 4°C.

3. Synthesis of Biotin-3MPA

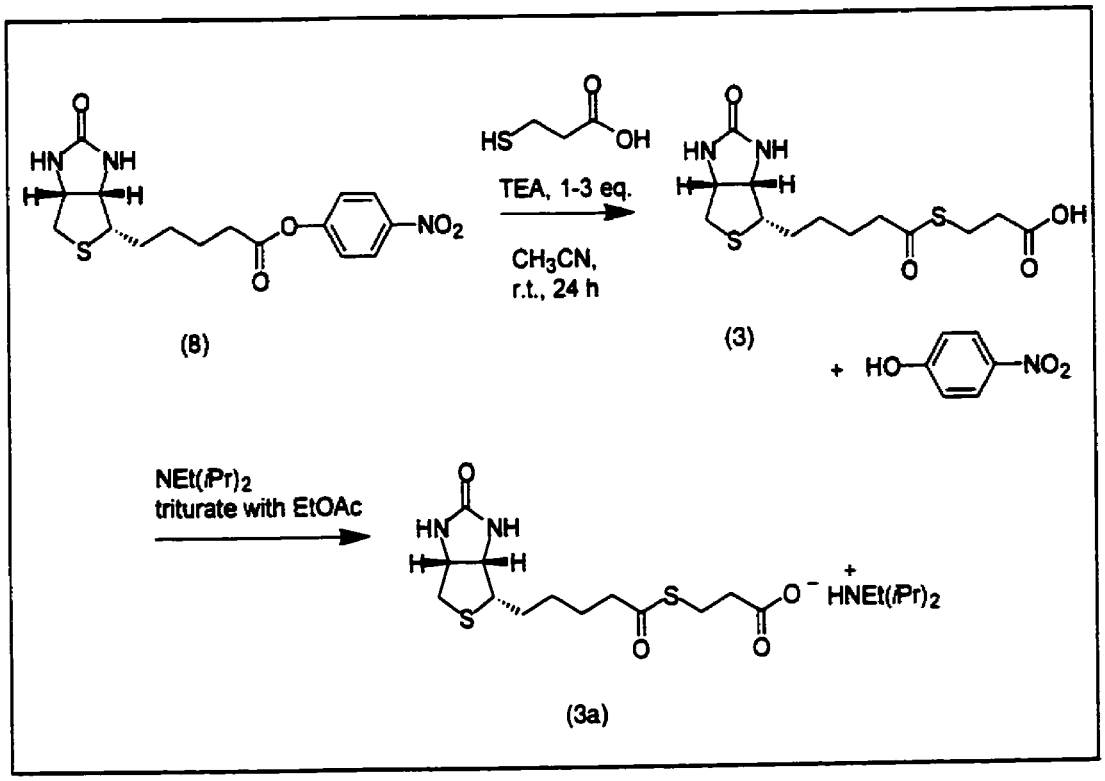
The synthesis of biotin-3MPA (**3**) is as outlined in Scheme 3. The starting compound, *d*-biotin-(*p*-nitrophenyl ester) (**8**) (200 mg, 0.547 mmol) was reacted with 3-mercaptopropionic acid (48 µL, 0.547 mm) in DMF (0.5 mL). Triethylamine (222 µL, 1.6 mmol) was also added. The mixture was stirred overnight at room temperature to form the product, biotin-3MPA (**3**). Diisopropylethylamine was added to biotin-3MPA to form the ammonium salt (**3a**), which was precipitated from solution by triturating the reaction mixture with ethyl acetate. The product (110 mg, 0.331 mmol; 60% yield) was a fine white powder. It was stored at –20°C. Yield of the reaction was 60%. TLC of biotin-3MPA (5% MeOH/CH_2Cl_2): R_f = 0.3. High resolution MS-FAB: MH^+ is 333.0906 ($C_{13}H_{20}O_4N_2S_2$, calculated mass 333.0943). ^{1}H-NMR (d-DMSO): δ1.32 (m, 2H, CH_2), δ1.44 (m, 2H, CH_2), δ1.56 (m, 2H, CH_2), δ2.57 (m, 5H, heterocycle CH_2, $2CH_2$), δ2.86 (d, 2H, CH_2), δ2.96 (t, 2H, CH_2), δ3.08 (m, 1H, heterocycle CH), δ4.11 (m, 1H, heterocycle bridge H), δ4.38 (m, 1H, heterocycle bridge H), δ6.38 (s, 1H, NH), δ6.46 (s, 1H, NH), δ12.25 (broad s, 1H, acid H). For experimental use, biotin-3MPA was dissolved in PIPES buffer (50 mM PIPES, 0.1 M KCl, 10% glycerol) and 10% w/v DMSO, to form an 8 mM stock solution which was stored at –20°C.



Scheme 1: Synthesis of EDTA-3MPA (Erlanson *et al.*, 1996)



Scheme 2: Synthesis of benzophenone-3MPA



Scheme 3: Synthesis of biotin-3MPA

4. Cloning of U2AF^{65} N-terminal Cysteine Mutants

Human U2AF^{65} inserted in pGEX plasmid (hU2AF^{65}/pGEX) was a kind gift from Michael Green (University of Massachusetts, Worcester, MA) and was used as the starting point for cloning the U2AF^{65} N-terminal cysteine mutants. The human U2AF^{65} peptide sequence is stored in the NCBI Peptide database, accession no. P26368. Creation of the full-length, Δ1-14, Δ1-24 and Δ1-64 mutant U2AF^{65} genes (carried in pET plasmid vector) was carried out by Ayube Reayi.

(a) Generation of PCR Primers

PCR primers were synthesized on a Perkin-Elmer Applied Biosystems Oligonucleotide Synthesizer 381A, using the supplied 0.2 μM CE (cyanoethyl) synthesis program. DNA oligonucleotides were synthesized with the final trityl group still attached ("trityl on" synthesis), and were purified manually with PE Applied Biosystems OPC columns, using the manufacturer supplied instructions. The sequences of the PCR primers are given below.

Basic codon interpretation for 5' primers:

5' d(CGC GCG | GAA TTC ATC GAA GGT CGT | TGT | ... ...)3'

CG "clamp" Factor Xa site cys start of U2AF^{65} sequence (5 codons)

U2AF^{65} (full length): 5' d(CGC GCG GAA TTC ATC GAA GGT CGT TGT ATC CAT ATG TCG GAC) 3'

Δ1-14: 5' d(CGC GCG GAA TTC ATC GAA GGT CGT TGT GGT AAA CAA GAG CGG GAC) 3'

Δ1-23: 5' d(CGC GCG GAA TTC ATC GAA GGT CGT TGT GGT CGG CAT CGG AAG CGC) 3'

Δ1-64: 5' d(CGC GCG GAA TTC ATC GAA GGT CGT TGT TTG ACC AGA GGC GCT) 3'

Δ1-94: 5' d(CGC GCG GAA TTC ATC GAA GGT CGT TGT GGT CCA CCC CCA GGC TTT) 3'

3' primer: 5' d(GCG CCC GGA AGC TTT TAC CAG AAG TCC CGG CGG TGA TA) 3'

(b) Insertion of Engineered U2AF^{65} Gene into Expression Vector

PCR was carried out using the hU2AF^{65}/pGEX plasmid as template and the above 5' primers and the 3' primer to generate new U2AF^{65} genes containing the Factor Xa site and

cysteine preceding the actual U2AF^{65} sequence. The PCR reaction mixture followed the standard protocol. The PCR thermocycler program was set to a denaturing temperature of 97°C (2 min), an extending temperature of 70°C (5 min), and an annealing temperature starting at 55°C (1 min) in the first cycle, and dropping by 1°C every subsequent cycle until it reached 50°C, at which point the cycle was repeated 30 times.

(c) Ligation of U2AF^{65} Gene Constructs into Plasmid Vectors

PCR products were precipitated (Maniatis *et al.*, 1982), resuspended in ddH_2O and overdigested with 3 units each of *Eco* RI and *Hind* III enzymes (both New England Biolabs), 37°C, overnight, total volume 25 µL. After digestion the PCR products were precipitated once more.

pET plasmid (Novagen), carrying kanamycin resistance, was also digested with Eco RI and *Hind* III at 37°C overnight, and then treated with alkaline phosphatase (Boerhinger Mannheim)at 37°C for 30 min., precipitated and resuspended in ddH_2O. Amounts of plasmid used depended on stock concentration of pET plasmid available; 100 ng of plasmid per digestion reaction was the usual amount.

Digested PCR products (the insert) and cut pET plasmid were ligated together with T4 DNA ligase (10 U/µL, USB). The insert to plasmid ratio was at least 100:1; for example 1.0 pmol of insert and 0.1 pmol of plasmid. Reactions were 20 µL in volume, 10 U of T4 DNA ligase, incubated at 16°C overnight. Following the ligation reaction, the mixture was precipitated to isolate the ligation product and resuspended in 10 µL (*i.e.* 2-fold concentration).

The ligation products were transfected into Novablue *E. coli* cells (Novagen). A modified version of Novagen's transformation protocol was used, using cell stock (20 µL), ligation product (5 µL) and SOC medium (80 µL, Novagen). The cell stock was placed in an incubator-shaker for 1 h, 37°C. After incubation, the cell stock was spread on LB kanamycin agar plates to select Novablue cells containing the pET vector (Maniatis *et al.*, 1982; growth media from Difco; antibiotic from Sigma-Aldrich). Minipreps (Qiagen) of plasmids from isolated Novablue colonies were digested with *Eco* RI and *Hind* III (1 U each) to identify successful insertion of the engineered U2AF^{65} gene. Successfully ligated plasmids were then transfected into BL21 pLysS *E. coli* strain (Novagen), which also carries chloramphenicol

resistance, for large scale culture preparation and protein expresssion. The same protocol was used as above, except with 5 µL of plasmid stock, 80 µL of BL21 cells, and 250 µL of SOC medium.

5. Cell Culture Preparation

(a) Test Induction of Protein Synthesis

The transfected cell stocks were used to grow 10 mL test cultures. Small samples were taken from each stock, placed in 10 mL of fresh LB-kanamycyin+chloramphenicol broth (Maniatis *et al.*, 1982) and grown overnight in an incubator-shaker set at 37°C. The culture was then diluted 100-fold into fresh growth media and allowed to grow until the culture had reached an $O.D._{590}$ of 0.4-0.6 (as compared to a blank sample of LB medium). A small (*e.g.* 10 µL) sample was taken from the culture. At this point, IPTG was added to final concentration of 1 mM and the culture allowed to incubate for a further 3-4 h. Another small sample (*e.g.* 10 µL) was taken from the culture. The two samples ("before" and "after" induction) were each suspended in 20 µL SDS loading dye and run on a 16% 200:1 SDS PAGE gel. A heavy band at the correct molecular weight, as compared to molecular weight markers (New England Biolabs) was indicative of successful induction. Of the set of $U2AF^{65}$ mutants created, Δ1-24 was the only one which did not express in an appreciable amount.

(b) Large Scale Culture Preparation

The same procedure as in (a) was followed except the 10 mL starter culture was diluted into 1 L of fresh growth media. The amount of IPTG added was adjusted accordingly. The culture was concentrated by centrifugation in a Sorvall UltraCentrifuge (15 min., 12 000 rpm), and the cell pellets thus obtained were stored either temporarily (1-2 days) at –20°C or for longer periods at –80°C.

6. Purification of Protein from Cell Cultures

(a) Denaturing Preparation of (Δ1-64)$U2AF^{65}$, (Δ1-94)$U2AF^{65}$ and (Δ1-14) $U2AF^{65}$

The cell pellet was thawed on ice and suspended in Buffer Gn (6 M guanidine hydrochloride, 100 mM sodium phosphate, 100 mM KCl, 10 mM Tris, 10% glycerol, 1

mM methyl methanethiolsulfonate, pH 8 (Erlanson *et al.*, 1996). The mixture was stirred at 4°C for 45-60 min. PMSF was added to a final concentration of 0.5 mM during the course of stirring. The mixture was then centrifuged to separate insoluble cellular debris (Sorvall Ultracentrifuge, 30 min., 10 000 rpm). The supernatant was taken and processed in the following manner.

For Δ1-64 and Δ1-94:

Imidazole (1 M stock solution) was added to a final concentration of 10 mM to the supernatant. The supernatant was bound to Ni^{+2}-NTA resin (Qiagen) pre-equilibrated in Buffer Gn/10 mM imidazole, 1 mL of resin per 5 mL of lysate, by gentle shaking on a rotary shaker for 1 h at 4°C. The resin was washed twice with fresh Buffer Gn containing 10 mM imidazole and then dialysed to refold the protein. The resin was placed in dialysis tubing presoaked in buffer (12-14 000 MWCO, Spectropor). The starting dialysis buffer was 6 M urea, 100 mM sodium phosphate, 100 mM KCl, 10% glycerol, pH 7.6. The starting buffer was diluted by 2 M increments with diluent buffer containing all of the above except urea, every 2 hours until no urea was present. The final buffer was PIPES buffer (50 mM PIPES, 100 mM KCl, 10% glycerol, 1 mM DTT, pH 7.6). The resin was left to dialyse overnight. All dialysis was done at 4°C. After dialysis, a large amount of precipitated protein was present with the resin. This was partially removed by washing the resin with fresh PIPES buffer containing 10 mM imidazole several times. The resin was next washed with PIPES buffer containing 20 mM imidazole and then the protein was eluted with 100 mM imidazole.

For Δ1-14:

The supernatant was bound to Fast Flow SP-Sepharose (Amersham Pharmacia Biotech) pre-equilibrated in Buffer Gn. The resin was washed three times with fresh Buffer Gn. The resin was then placed in dialysis tubing presoaked in buffer (12-14 000 MWCO, Spectropor), and dialysed to remove denaturant. The dialysis steps and urea buffers were the same as for Δ1-64, except that the stepwise dilution of the starting buffer was halted when a concentration of 1 M urea was reached (*i.e.* the resin was allowed to dialyse into 1 M urea overnight). The resin was loaded into a column and protein was eluted with a stepwise increase of KCl: 0.2 M, 0.4 M, 0.6 M, and 0.8 M. The remaining buffer composition was 50 mM PIPES pH 7.6, 10% glycerol. For each KCl concentration, 3 x 10 mL fractions were

collected. Δ1-14 begins to elute in the 0.4 M KCl fractions and elutes completely in the 0.6 M and the first 0.8 M KCl fractions. The 0.6 M fractions were diluted partially with the 0.4 M fractions (*e.g.* in 3:1 ratio) and dialysed against PIPES buffer (as described for Δ1-64) overnight. Some of the protein precipitated out of solution, forming a flocculent beige solid; this was removed by centrifuging and removal of the supernatant. The supernatant contained a high concentration of protein, usually at least 1 mg/mL. The protein was flash frozen for further use.

(b) Native Preparation of Full-length U2AF^{65}

This protocol was adapted from the protocol given by Qiagen for purification of his-tagged proteins from Ni^{2+}-NTA resin (Qiagen). The cell pellet from a 1 L culture was thawed for 15 min. on ice. Lysis buffer, 10 mL (50 mM KH_2PO_4 pH 8.0, 300 mM KCl, 10 mM imidazole, 0.1% Tween, 10% glycerol) was added to the pellet. Lysozyme (10 mg, USB) was added to the mixture, which was incubated on ice for 30 min. The mixture was then sonicated (Virsonic) on ice, using medium power; six 10 s. bursts were carried out with a 10 s. cooling period between each burst. The lysate mixture was then centrifuged (Sorvall Ultracentrifuge, 10 000 rpm, 30 min.) and the supernatant removed. This cleared lysate was then bound to Ni^{2+}-NTA resin slurry pre-equilibrated with lysis buffer (1 mL resin/4 mL lysate), and gently mixed on a rotary shaker at 4°C for 60 min. The mixture was then loaded on to a column and the unbound lysate collected. The resin was then washed with 3 x 5 mL of wash buffer (50 mM KH_2PO_4 pH 8.0, 300 mM KCl, 20 mM imidazole, 0.1% Tween, 10% glycerol). The protein was eluted with 5 x 1 mL elution buffer (50 mM KH_2PO_4 pH 8.0, 300 mM KCl, 250 mM imidazole, 0.1% Tween, 10% glycerol). All of the U2AF^{65} eluted in the elution buffer fractions. The buffer was exchanged using a size-exclusion resin column (Sephadex G-25 PD-10 column, Amersham Pharmacia Biotech), into MOPS buffer (50 mM MOPS pH 7.6, 100 mM KCl, 10% glycerol, 1 mM DTT).

(c) Denaturating Preparation of U2AF^{35}

His-tagged human U2AF^{35} was expressed from Bdv 325 *E. coli* cell stock carrying human U2AF^{65}/hisU2AF^{35} (kanamycin, ampicillin resistance) plasmid. This cell stock was a kind gift from Donald Rio (UC Berkeley, California). The procedure for expression was as described in Section (5). A cell pellet from a 1 L culture was prepared as described earlier.

The cell pellet was thawed on ice for 15 min. and then was lysed with 15 mL denaturing lysis buffer (6 M guanidine HCl, 0.1 M KH_2PO_4, 0.1 M KCl, 0.01 M Tris, 10% glycerol, pH 8.0). PMSF (300 mM stock solution in ethanol) was added to a final concentration of 0.5 mM. The mixture was stirred for 1 h. at room temperature and then centrifuged (Sorvall UltraCentrifuge, 10 000 rpm, 30 min.) to pellet the insoluble cellular debris. The supernatant was bound to 4 mL Ni^{2+}-NTA resin (Qiagen) pre-equilibrated in denaturing lysis buffer, by gentle mixing at room temperature on a rotary shaker for 1 h. The resin mixture was washed twice with fresh denaturing lysis buffer containing 10 mM imidazole. The protein was refolded on the resin by dialysis, starting with 6 M urea dialysis buffer (6 M urea, 0.1 M KH_2PO_4, 0.1 M KCl, 0.01 M Tris, 10% glycerol) and the same step-wise decrease in urea concentration as described earlier in part (a). The final buffer was 0.1 M KH_2PO_4, 0.01 M Tris, 10 mM imidazole, 10% glycerol, pH 7.6. The resin was loaded on to a column and the unbound protein was collected. The resin was washed with 3 x 5 mL of wash buffer (0.1 M KH_2PO_4, 0.1 M KCl, 0.01 M Tris, 20 mM imidazole, 10% glycerol, pH 7.6). $U2AF^{35}$ was eluted in 3 x 5 mL of elution buffer (0.1 M KH_2PO_4, 0.1 M KCl, 0.01 M Tris, 250 mM imidazole, 10% glycerol, pH 7.6). This elution buffer was exchanged using a size-exclusion resin column (Sephadex G-25 PD-10 column, Amersham Pharmacia Biotech), for MOPS buffer. The concentration of $U2AF^{35}$ thus obtained was 0.5 mg/mL.

7. Factor Xa Proteolysis of the $U2AF^{65}$ Mutants

The concentration of the various $U2AF^{65}$ proteins was determined by Bradford assay (BioRad), using bovine serum albumin (New England Biolabs) as the concentration standard. Protein samples taken from –80°C storage were thawed on ice and then allowed to warm to room temperature. Factor Xa (New England Biolabs) was tested for activity in a time course reaction, checking the level of proteolysis every hour for 4 hours, to ascertain how much protease was required. This test reaction started with 50 µL of protein, at least 0.5 mg/mL. Calcium chloride was added to a final concentration of 2 mM. Factor Xa was then added in a ratio of 1 µg protease : 50 µg target protein. Samples taken every hour after the addition of Factor Xa were analyzed by SDS-PAGE (16% 200:1 SDS-polyacrylamide resolving gel).

In the case of Δ1-14, the Factor Xa proteolysis reaction was controlled by the addition of urea to reduce the activity of Factor Xa. A concentrated urea solution (6 M, pH 7.6) was added to the protein sample to a final concentration of 250 mM before addition of $CaCl_2$ and Factor Xa as described above. For Δ1-14, the reaction was usually halted at 20-30 minutes.

After doing this initial test reaction, the correct amount of Factor Xa to be added to a large sample of protein was determined. The activity of the Factor Xa tended to vary with each stock obtained from the manufacturer; the usual protease/protein ratio varied between 1:100 and 1:500. In the case of Δ1-64 and Δ1-94, the protease reaction could be carried to nearly 100% completion. For full length U2AF^{65} and Δ1-14, nonspecific cleavage of the proteins occurred so the reaction was halted at ~60% completion to maximize the amount of correctly cleaved product. The protease reaction was halted by the addition of a protease inhibitor cocktail: dansyl-EGR-chloromethyl ketone (Calbiochem) was added to 9 μM final concentration, pepstatin A added to 1.3 μg/μL and leupeptin added to 11 μg/mL (both, Boerhinger Mannheim). The protein samples were then derivatized immediately following proteolysis.

8. Further Purification of Cleaved and/or Derivatized Full-length U2AF^{65} and (Δ1-14) U2AF^{65}

As mentioned earlier, the full length U2AF^{65} and Δ1-14 were subject to nonspecific degradation when cleaved by Factor Xa. The main cleavage product (with the N-terminal cysteine) was purified from the nonspecifically cleaved products with the use of SP-sepharose ion exchange resin. An FPLC pump was used (Amersham Pharmacia Biotech), and a 1 mL HiTrap SP-sepharose column. Buffer A composition was either 50 mM PIPES pH 7.6 (for Δ1-14) or 50 mM MOPS pH 7.6 (for full length U2AF^{65}), and 10% glycerol, 1 mM DTT, 0.1 M KCl; Buffer B was the same composition except with 1.0 M KCl. The pump was set at a flow rate of 0.5 mL/min. A sample of protein (3-5 mg total protein, 2-5 mL in volume) was injected on to the column and washed on for 10 minutes before the start of the gradient. The gradient was programmed as follows: 0 to 54% B, 12 min.; 54% B, 5 min.; 54 to 65%

B, 8 min.; 65% B 5 min.; 65 to 100% B, 10 min.; 100% B, 5 min. The steps in the gradient represent the following KCl concentrations: 54% B is 0.59 M KCl, 65 % B is 0.69 M KCl. Fractions of 1 mL were collected for the duration of the gradient and analyzed by SDS-PAGE. A purification of Δ1-14-(EDTA) is shown in Fig. 11B (Section C), with the corresponding gradient program. Non-specifically cleaved proteins eluted early at the start of the gradient; uncleaved protein generally eluted around 0.6 M KCl, with the correctly cleaved protein eluting at the start of the second step of the gradient (0.69 M KCl). The yield of FPLC-purified Δ1-14 was approximately 0.2 mg/mL, depending on the concentration of the starting crude sample. For unknown reasons, the yield from the FPLC purification of full-length U2AF^{65} was not as high as Δ1-14, approximately 0.05-0.1 mg/mL, although the starting concentrations of samples were about the same.

9. Derivatization of Factor Xa-cleaved U2AF^{65} Mutant Proteins with Thioester Reagents

Dithiothreitol (BioShop) was added to the protein samples to a final concentration of 0.1 M. Biotin-3MPA, EDTA-3MPA and benzophenone-3MPA were dissolved in the appropriate solvents as described in Sections B.1-3. To derivatize the protein with either biotin-3MPA or EDTA-3MPA, the thioester reagent was added to a final concentration of 1 mM and the protein solution allowed to stand overnight at 4°C. To derivatize with benzophenone-3MPA, the benzophenone-3MPA (15 mM in ethanol) was added to a final concentration of 500 μM. The protein solution turned pale yellow and precipitate formed within the first 10 minutes of addition. The protein solution was allowed to stand at room temperature for 1 h. and then stood overnight at 4°C. The precipitate was removed by centrifuging, and the supernatant recovered.

10. The Biotin-3MPA Immunoblot Assay

The U2AF^{65} mutants, full-length U2AF^{65}, Δ1-14, and Δ1-64, were cleaved with Factor Xa to expose the N-terminal cysteine, as described in Section B.7. Three samples of each protein, containing approximately 20 ng of protein (not including the smaller weight

degradation products present), were taken. One sample was reacted with biotin-3MPA (final concentration 1 mM), with added 0.1 M DTT, overnight at 4°C. The next two protein samples were reacted with either EDTA-3MPA (final concentration 1mM) or benzophenone-3MPA (final concentration 500 μM), with added 0.1 M DTT, room temperature for 6 h. Biotin-3MPA (final concentration 1 mM) was then added to both samples and the reaction continued overnight at 4°C. As a control, a test protein sample with no N-terminal cysteine (*i.e.* full-length U2AF^{65}, Δ1-14, Δ1-64 which have not been cleaved by Factor Xa) was also subjected to the same reaction conditions with biotin-3MPA. To halt the reactions and analyze the products, each set of samples were suspended in SDS loading dye and separated on SDS polyacrylamide gel (16% 200:1). The proteins were then transferred on to nitrocellulose membrane. Mouse anti-biotin antibody (Sigma-Aldrich Biochemicals) was used to detect biotin. Rabbit anti-mouse antibody conjugated with horseradish peroxidase (HRP) and luminol reagent (both, New England Biolabs) were used to detect the binding of mouse anti-biotin. The SDS-PAGE and corresponding immunoblot for full-length U2AF^{65} are shown in Fig. 12 (Section C).

11. Native Gel Mobility Shift Assay

(a) Preparation of RNA Substrates

The selected RNA sequence was based on the PIP85.b sequence, an optimized splicing substrate (Query *et al.*, 1994). The sequence contains the branchpoint region, the polypyrimidine tract and the first 15 bases of the 3' exon. The mutant features a random purine-rich sequence in place of the polypyrimidine tract. Complementary DNA oligonucleotides with a T7 promoter sequence at the 3' end (TAT AGT GAG TCG TAT TA) were synthesized on a PE Applied Biosystems Oligonucleotide Synthesizer 381 A (trityl-on, 0.2 μmol synthesis). The DNA oligonucleotides were manually purified with PE Applied Biosystems OPC cartridges, according to the manufacturer's instructions.

Wild type DNA sequence (LF1):

5' d(TAT GTT GTG TAG GAC | CTG AGG GAA AAA GAG AGA AGA AGC CAG TCA GCA CCC | TAT AGT GAG TCG TAT TA) 3'

Mutant DNA sequence (LF2):

5' d(TAT GTT GTG TTAG GAC | CTG AGT TGC ATT GCA TGT CCG TCC CAG TCA GCA CCC | TAT AGT GAG TCG TAT TA) 3'

T7 promoter complementary sequence oligonucleotide:

5' d(TAA TAC GAC TCA CTA TA) 3'

RNA was synthesized *in vitro* with T7 RNA polymerase. An oligonucleotide stock mixture was first prepared, consisting of 20 μM DNA template (LF1/LF2), 5 μM T7 promoter complementary sequence, in 10 μM Tris pH 7.5, 10 μM $MgCl_2$. Before using this mixture in the *in vitro* transcription, an aliquot was taken and heated for 3-5 min. at 95°C and then cooled to room temperature for at least 10 min. The *in vitro* transcription buffer consisted of 40 mM Tris pH 7.5, 5 mM DTT, 1 μM spermidine, 0.01% Triton-X, 20 mM $MgCl_2$. The transcription reaction consisted of nucleotide mixture (3 mM in A, C, G, U; all Pharmacia), oligonucleotide mixture as described above (0.2 μM in the template), and T7 RNA polymerase (1 U/50 μL reaction; Boeringher Mannheim). The *in vitro* transcription was incubated overnight at 37°C. The mixture was precipitated to isolate the RNA. The RNA oligonucleotides were purified on a 15% denaturing (8 M urea) 39:1 polyacrylamide gel (1.5 h at 25 mA). The band in the gel containing RNA was identified by UV absorption; the band was cut out and crushed, then suspended in 250 μL of phenol/chloroform solution and 400 μL of 0.3 M sodium acetate pH 5.6. The gel mixture was extracted for 4 h at 37°C and then the gel was separated from the solution using a spin column. The filtered solution was then subjected to the standard precipitation procedure. The RNA thus obtained was resuspended in 50 μL ddH_2O. Before using an aliquot of RNA (~50 pmol) was treated with alkaline phosphatase (37°C, 1 h) and precipitated to remove the RNA. RNA oligonucleotides were phosphorylated with [γ-^{32}P]ATP and T4 polynucleotide kinase (New England Biolabs), for 15 min. at 37°C. The RNA was then gel purified again (same procedure as above).

LF1 RNA sequence:

5' GGG UGC UGA CUG GCU UCU UCU CUC UUU UUC CCU CAG GUC CUA CAC AAC AUA 3'

LF2 RNA sequence:

5' GGG UGC UGA CUG GGA CGG ACA UGC AAU GCA ACU CAG GUC CUA CAC AAC AUA 3'

(b) Gel-shift Assay

The conditions for this assay were partially adapted from the gel-shift assay by Berglund *et al.*, 1998. The proteins (full-length $U2AF^{65}$, Δ1-14, Δ1-64, Δ1-94; $U2AF^{35}$; SF1/mBBP), added in increasing concentrations, were incubated with ~1 nM ^{32}P-labeled RNA (LF1 or LF2) in binding buffer for 60 min. at room temperature. The binding buffer was 25 mM Tris pH 7.5, 25 mM NaCl, 0.25 mg/mL tRNA, 1 mM EDTA. A 0.5x TBE 6% (85:1) native polyacrylamide gel and 0.5x TBE running buffer were chilled to 4°C for at least 2 h. before running the gel. The protein-RNA complexes were separated on the gel, 5 h. at 100 V, at 4°C. The gel was dried on to a sheet of filter paper on a BioRad slab gel drier and exposed overnight on a Phosphorimager screen, then scanned for the exposed image on a Phosphorimager scanner.

12. Splicing Reconstitution in U2AF-Depleted Nuclear Extract

(a) Preparation of RNA Substrates

The pPIP85.b plasmid was cleaved with *Hind* III restriction endonuclease (New England Biolabs). Also, DNA template could be derived from the PCR of pPIP85.b plasmid, between M13F 5' primer and CQ27a 3' primer, corresponding to the 3' exon.

A mutant pre-mRNA substrate, containing purines in place of the polypyrimidine tract within the intron, was generated according to the following procedure. A mutant RNA oligonucleotide was generated, based on the PIP85.b sequence. The following DNA oligonucleotide PCR primer was synthesized:

Primer "C":

5' d(CGC GCG CCT CGA GGG TGC TGA CTG GGA CGG ACA TGC AAT GCA ACT CAG GTC CTA CAC) 3'

PCR was carried out using "C" and CQ27.a primers and pPIP85.b as the template. The resultant DNA, dubbed C27, was precipitated and digested with restriction endonucleases *Xho* I and *Hind* III (overnight, 37°C; both New England Biolabs). At the

same time, pPIP85.b was also digested with *Xho* I and *Hind* III (overnight, 37°C), then treated with alkaline phosphatase (1 h, 37°C; Boerhinger Mannheim), and precipitated. The digested DNA (from PCR) was ligated into the cut pPIP85.b plasmid using T4 DNA ligase (overnight, 16°C; high concentration 10 U/μL, USB). This ligation reaction was precipitated and the ligation reaction product was used to transform *E. coli* Novablue competent cells (Novagen), following the supplier's transformation protocol. The transformed cultures were incubated for 16 h. on LB-ampicillin/agar plates. Colonies that appeared after the incubation period were then grown in 5 mL LB-ampicillin starter cultures (overnight, 37°C). The plasmids were isolated from these starter cultures using the Qiagen Plasmid Mini-Prep kit. The plasmids were then subjected to digestion with *Bsu* 36I (3 h, 37°C). They were then analyzed on a 1.5% agarose/TBE gel, stained with ethidium bromide and visualized under UV radiation. The successful insertion of the C27 oligonucleotide into the pPIP85.b plasmid was apparent when the plasmid was impervious to *Bsu* 36I since it does not contain the *Bsu* 36I restriction site; the wild type pPIP85.b plasmid contains the *Bsu* 36I restriction site and the cleaved plasmid migrates further into the gel than an intact plasmid. The successfully cloned plasmid containing the C27 sequence was as a template in PCR with M13F and CQ27.a as primers. The resultant DNA, to be used as a template for RNA transcription, was called C27.

Complementary ^{32}P-body-labeled RNA was synthesized *in vitro* using T7 RNA polymerase. The 50 μL reaction mixture contained 10 mM DTT, (A,C,G) nucleotides 0.5 mM in each, 6 μM UTP, 1 unit RNasin, GpppG cap analog (New England Biolabs), and T7 RNA polymerase reaction buffer 1 x. To this, ~0.1 μg of DNA template (PIP85.b or C27) was added, with [α-^{32}P]-UTP and 1 unit T7 RNA polymerase. The reaction was incubated for 2 h at 37°C, then precipitated to isolate the RNA, which was resuspended in ddH_2O (50 μL).

(b) Nuclear Extract Preparation

Nuclear extract was prepared from HeLa cell culture as described by Dignam *et al.* (1983) by Andrew MacMillan. U2AF-depleted extract was also prepared by Andrew MacMillan and Patrick McCaw, using the following procedure (MacMillan *et al.*, 1997). The nuclear extract was dialyzed into 1 M KCl/Buffer D (20 mM HEPES pH 7.9, 20% glycerol, 0.2 mM EDTA, 0.05% Nonidet P-40, 0.5 mM). The extract was passed over a

poly(U)-Sepharose 4B column (Pharmacia) then dialyzed into 0.1M KCl/Buffer D. This gave nuclear extract depleted of poly(U)-binding proteins, including U2AF. The column was washed with 1 M KCl/Buffer D and then with 2 M KCl/Buffer D. The 2M KCl fraction contains various splicing factors which must be added back to the depleted extract for efficient splicing to occur in splicing reconstitution. The 2 M KCl fraction is stored separately from the poly(U)-binding protein depleted nuclear extract, and is added back to this extract to form ΔU2AF NE as used in the splicing reconstitution (see below).

(c) Splicing Reconstitution Assay

Splicing reactions (25 μL) were carried out as described by Grabowski *et al.* (1984), using 40% nuclear extract, incubated at 30°C. The 2 M KCl fraction (as described above) is added to the splicing reaction mixture in a volume ratio of 1 (2 M KCl) to 12.5 (splicing reaction mixture). Experiments consisted of the whole nuclear extract (control), ΔU2AF NE (control), ΔU2AF NE + wild type $U2AF^{65}$ (control; a gift from Patrick McCaw; purified as described in MacMillan *et al.*, 1997), and ΔU2AF NE + ($U2AF^{65}$ cysteine mutant protein; at least 500 ng). After the reactions were completed, they were precipitated and the RNA pellets resuspended in denaturing loading dye containing bromphenol blue and xylene cyanole (0.1% each). The RNA was resolved on a 15% denaturing (8 M urea) polyacrylamide gel (run time: 3 h at 75 W). The gel was exposed overnight on a Phosphor Screen (Molecular Dynamics). The screen was scanned on a Molecular Dynamics Storm 860 Phosphorimager, using ImageQuant 5.0 software.

13. Hydroxyl Radical Footprinting of (EDTA)-(full-length)$U2AF^{65}$ and (EDTA)-(Δ1-64)$U2AF^{65}$

(a) Preparation of RNA Substrates

RNA oligonucleotides (LF1/LF2; PIP85.b / C27) were 5' end-labeled with ^{32}P using the T4 polynucleotide kinase as described in Section 1(a). In the case of LF1/LF2, the RNA was purified by running on to an 8% denaturing polyacrylamide gel (1.5 h, 25 mA) and then extraction from the gel, as described in 1.(a). For PIP85.b and C27, these were purified by running on to a 15% denaturing polyacrylamide gel (2 h, 20 mA) followed by gel extraction.

The activity of the RNA was measured by liquid scintillation using a Beckmann scintillation counter.

(b) Partial Digestion of RNA Substrate with RNase T1 or RNase A to Produce a Base/Molecular Weight Ladder

A sample of the RNA substrate LF2 (1 µL of at least 100 000 cpm/µL) was diluted into 45 µL of reaction buffer (18 mM citric acid, 0.9 mM Na_2EDTA, 6.3 M urea, 13.5 mg/mL tRNA, pH 5.0) + 4 µL of dH_2O. The sample was then heated to 55°C for at least 5 min. 1 µL of RNase T1 (1 unit/µL) or RNase A (0.1 µg/µL) was added to the sample and the sample heated for a further 30 s. The reaction was quenched by the addition of phenol/chloroform solution and the sample was then precipitated in the usual manner to obtain the RNA.

(c) Hydroxyl Radical Footprinting of EDTA-Derivatized U2AF Cysteine Mutants on pre-mRNA

All reactions were carried out in reaction buffer, 20 mM HEPES pH 7.5, 100 mM KCl. Protein (2 µL of full-length $U2AF^{65}$/Δ1-14/Δ1-64; added $U2AF^{35}$ in some cases) was added to 18 µL of reaction buffer, diluting the protein 10 fold: ~1.7 µM for full-length $U2AF^{65}$, ~1 µM for Δ1-64 and ~0.3 µM for Δ1-14. The dilution brought the final glycerol concentration to 1%. RNA (LF1/LF2; PIP85.b/ C27) was added (1 µL of at least 100 000 cpm/µL stock solution) to each reaction. Ferric ammonium sulfate stock solution, freshly prepared, was added so that the final concentration of Fe^{2+} was equimolar with the $U2AF^{65}$ protein present in the reaction. The reactions were allowed to equilibrate at room temperature for 30 minutes. The hydroxyl radical reaction was initiated by adding 1 µL of 250 mM sodium ascorbate (freshly prepared or from frozen aliquots) and 1 µL of 20 mM H_2O_2. The reaction was allowed to continue for 2 minutes and then quenched with 3 µL of 3 M sodium acetate pH 5.6 containing 0.5 mg/mL tRNA. The reactions were then brought up in volume with 100 µL of 0.3 M sodium acetate pH 5.6 containing 10 µg/mL of glycogen. The reactions were phenol/EtOH precipitated to isolated the RNA. The RNA pellets were resuspended in denaturing loading dye (8 M urea in 1 x TBE, 1 mg/mL xylene

cyanole), denatured at 65°C for 5 min. then run on a denaturing polyacrylamide gel (Sequagel, National Diagnostics). The gel type and run time depended on the RNA substrate used: 15% for LF1/LF2, 6% for PIP85.b; 2-3 h. The gel was then transferred to blotting paper and dried on a BioRad slab gel drier for at least 1 h. at 80°C. The dried gels were exposed overnight on Phosphor screens which were then scanned using the phosphorimager (Molecular Dynamics Storm 860, ImageQuant 5.0).

14. Cross-linking of Benzophenone-Derivatized U2AF^{65} Proteins to U2AF^{35} in a Minimal System

(a) Photoactivated Cross-linking of Benzophenone-Derivatized U2AF^{65} Proteins

The conditions for 100% bound protein from the native gel mobility shift assay were employed here. RNA (LF1), not labeled with ^{32}P, was used as a substrate for U2AF^{65} and U2AF^{35} binding. The total reaction volume was 25 µL. In reactions containing one type of U2AF^{65} protein and U2AF^{35}, 12 µL of each protein stock solution was added, with 1 µL of 25 nM RNA stock solution added (final concentration 1 nM). In controls containing only one protein, the remaining volume of 12 µL was made up with MOPS buffer (50 mM MOPS pH 7.6, 100 mM KCl, 1 mM DTT, 10% glycerol). Before starting, the Rayonet photoreactor chamber, fitted with 350 nm UV lamps, was cooled down to 4°C in a cold room for at least 2 h. before beginning the reaction. The reactions were allowed to equilibrate at room temperature for 30 min. and then were placed in the wells of a microtitre plate (Nunclon) on ice. The microtitre plate containing all of the reactions and the ice container supporting it were placed inside the Rayonet photoreactor, in the cold room, and exposed to UV radiation for 1 h.

(b) Western Blot to Detect U2AF^{65}

To each set of reactions, SDS loading buffer was added and the samples heated to >95°C for 5 min, then separated by SDS-PAGE, one set of reactions on one gel (16% 200:1 SDS-polyacrylamide gel; 50 min. at 200 V). One set was stained with Coomassie Blue dye to

determine the positions of the proteins on the gel. Another two sets, after separation on SDS-polyacrylamide gels, were transferred on to nitrocellulose membrane using the Bio-Rad Western Blot transfer apparatus (1 h, 100 V). Two sets of reactions were done, one to be detected with mouse anti-U2AF^{65} serum, the other with mouse anti-U2AF^{35} serum. The membranes were blocked for 1 h. at room temperature with 5% milk in TBS buffer. The membranes were then washed with TBS buffer and 7-8 mL of TBS containing (i) mouse anti-U2AF^{65} and (ii) mouse anti-U2AF^{35}, at 150 x dilution of each of the serums was added to each of the membranes. This primary membrane wash was carried out overnight at 4°C. The membranes were then washed thoroughly with TBS buffer and blocked with HRP-linked rabbit anti-mouse for 1 h. at room temperature. The membranes were washed with TBS buffer then subjected to luminol/H_2O_2 treatment to initiate chemiluminescence (all reagents for chemiluminescent reaction from New England Biolabs). The membranes were exposed on to Kodak Bio-Max film immediately afterwards.

SECTION C: RESULTS AND DISCUSSION

U2AF^{65} is an essential splicing factor required in early spliceosome complex formation and the recruitment of U2 snRNP to the spliceosome. U2AF^{65} shares basic domain similarities to the SR proteins, a large family of proteins involved in RNA splicing. However, the ordering of the domains in U2AF^{65} is opposite to the SR protein domain order. The RS domain of U2AF^{65} is required for RNA splicing to occur (Zamore *et al.*, 1992), but the mechanism for this activity has not been well defined. Green & coworkers have speculated that the RS domain of U2AF^{65} may be involved in stabilizing the interaction of U2 snRNP to the branchpoint region of the pre-mRNA (Valcárcel *et al.*, 1996). The attachment of small molecule reporter groups to the RS domain of U2AF^{65} would help to understand its role in U2AF function within the spliceosome, in terms of its interaction with either the pre-mRNA or other splicing factors.

The thioester-cysteine peptide ligation method (Dawson *et al.*, 1994) was modified to become a generally useful strategy for attachment of any functional group to any protein with cysteine at its N-terminus. In the thioester reagent strategy, a thioester derivative of a reporter molecule is ligated to a protein with a cysteine at its N-terminus. This strategy is ideal for the study of U2AF^{65}'s RS domain, which begins near the N-terminus. A set of thioester reagents that allow the study of protein-RNA and protein-protein interations were made and a series of U2AF^{65} mutants beginning with cysteine were generated.

1. Synthesis of Thioester Derivatives of Chemical Reporter Groups

Thioester reagents containing chemical reporter groups were synthesized. 3-Mercaptopropionic acid was used to create the thioester derivatives, but it is possible to use any other thiol compound, keeping in mind the following constraints. One consideration was the effect of the thioester portion on the rate of the ligation reaction, since it is the leaving group in the reaction with cysteine. Kent & coworkers used thiophenol as an additive in the thioester-cysteine peptide ligation reaction (Dawson *et al.*, 1997). Thiophenol exchanges with the starting thioester to form the more reactive phenyl thioester *in situ*, thus

increasing the yield of the ligation reaction and decreasing the time required for the reaction to go to completion. The 3-mercaptopropionic acid group is not a particularly good leaving group but in subsequent derivatization of the U2AF^{65} mutants, it did not adversely affect reactivity of the thioester with the proteins. Another consideration was the effect of the terminal group, a carboxylic acid in this case, on the solubility of the thioester reagent in aqueous buffer or a water-miscible solvent. The carboxylic acid group was thought to be helpful in increasing solubility in aqueous buffers. EDTA-3MPA was completely soluble in water. Benzophenone-3MPA was not soluble in water or any aqueous solution but was soluble in ethanol. Biotin-3MPA was not soluble in water but soluble in an aqueous buffer containing 10% DMSO.

The synthesis of EDTA-3MPA (**1**) (Scheme 1) was as described by Erlanson *et al.* (1996). Since there are two sites for nucleophilic attack within the EDTA dianhydride molecule (**4**), both the monothioester and the *bis*-thioester are produced in this reaction. The monothioester was purified from the *bis*-thioester by HPLC. Although the final yield from the reaction and subsequent purification steps was extremely low, it was necessary to obtain the purest form of EDTA-3MPA since the *bis*-thioester form, if present, would act as a protein cross-linker by joining two U2AF^{65} cysteine mutants through their N-termini.

Benzophenone-3MPA (**2**) was synthesized from the starting compound, 4-benzoylbenzoic acid (**5**) in three steps (Scheme 2). This was accomplished by activation of the acid group as a *p*-nitrophenyl ester (7), via the acid chloride intermediate (6). A more gentle reaction which would eliminate the use of thionyl chloride would be the use of *bis*-(*p*-nitrophenyl)carbonate. One of the products from reaction with this reagent is CO_2 gas, which is evolved during the course of the reaction, and the *p*-nitrophenyl ester is obtained in a single step.

Synthesis of biotin-3MPA (**3**) was carried out in a single step (Scheme 3). The *p*-nitrophenyl ester form of *d*-biotin (**8**) was reacted with 3-mercaptopropionic acid; and the product biotin-3MPA was easily purified from the reaction mixture by forming the quaternary ammonium salt (**3a**) and precipitation from ethyl acetate. The biotin-3MPA

reagent was created to detect the presence of the free N-terminal cysteine in proteins which have been engineered to have this residue. Since it is a thioester, it should react as predicted to covalently attach biotin to a protein with cysteine at its N-terminus. The biotin-3MPA reagent may also be used to detect previous reaction of the N-terminal cysteine to other thioester reagents since a protein that has already reacted with a thioester reagent cannot react a second time with biotin-3MPA. This means that a prior reaction with a thioester may be detected as a loss of signal as compared to the biotin-labelled protein in an assay to detect the presence of biotin-labelled protein.

2. Generation of a Series of U2AF^{65} N-terminal Cysteine Mutants

A series of U2AF^{65} mutants were generated, featuring a Factor Xa protease site, followed with cysteine and the U2AF^{65} sequence. N-terminal cysteine residues were introduced into the U2AF^{65} sequence by creating PCR primers that contained the coding sequence for a Factor Xa protease site and the codon for cysteine, followed by the start of the U2AF^{65} sequence. PCR with this primer and a downstream primer bracketing the sequence of the U2AF^{65} gene on the given template, a plasmid containing the U2AF^{65} gene (U2AF^{65}-pGEX), generated a new DNA fragment containing the U2AF^{65} gene preceded by the protease site and cysteine residue. Once this new gene was inserted into a new plasmid vector and transfected into a host *E. coli* strain, the expressed protein may be cleaved with Factor Xa protease to expose cysteine at the N-terminus. The cloning strategy thus described also allowed flexibility in the form of U2AF^{65} that is eventually expressed. N-Terminal truncation mutants (Δ1-14, Δ1-24, Δ1-64, Δ1-94) were created in order to map various portions of the RS domain with respect to the pre-mRNA or splicing factor involved.

The cloning of N-terminal cysteine U2AF^{65} (full-length) and the N-terminal truncation mutants was straightfoward and no major problems were encountered during the creation of the *E. coli* cell strains carrying the engineered U2AF^{65} vectors. After creation of the cell lines carrying the engineered plasmid vector, the cell lines were tested for their ability to express the correct protein (*i.e.* of the correct molecular weight) before continuing to the next stage, the large scale expression and purification.

The first step to obtaining large amounts of the U2AF^{65} constructs was to grow large cell cultures from the cell stocks. This was a straightforward process and no major problems were encountered except in the case of Δ1-24. The U2AF^{65} mutant Δ1-24 did not express in a high enough amount for a significant amount of protein to be retrieved (*i.e.* at least 1 mg/mL), even with the use of a denaturing preparation to try and maximize the yield of purified protein. A high concentration of protein in large quantities was required on the outset to be of utility in future experiments. For this reason, Δ1-24 was abandoned from further experiments.

After expressing the various protein constructs in the large cell cultures, the next step was to purify the proteins. Because the pET expression vector includes the his-tag, it was initially thought that the easiest method to purify the U2AF^{65} proteins was to use affinity chromatography with Ni^{2+}-NTA resin, following the protocols suggested by the manufacturer, Qiagen. For Δ1-64 and Δ1-94, a denaturing preparation using Ni^{2+}-NTA as a support for refolding the protein was employed. This method was fairly successful in isolating these two proteins in high yield and concentration.

A denaturing preparation was thought to be the most generally useful method in obtaining a high yield of protein. It was found that a denaturing preparation was useful in cases where the level of expression was not extremely high. A drawback of the denaturing preparation was precipitation of large quantities of protein during the refolding of the protein by stepwise dialysis to remove denaturant. Loss of protein through precipitation was partially reduced in the case of Δ1-64 and Δ1-94, both of which were refolded while bound to Ni^{2+}-NTA resin. Another drawback of the denaturing preparation was the possibility of incorrect folding of the protein, as which occurred with Δ1-14 when refolded on SP-sepharose resin. When the level of protein expression was extremely high, as for full length U2AF^{65}, a native preparation with Ni^{2+}-NTA resin could be used. A native preparation avoids the problem of protein precipitation and is a convenient method for obtaining protein very quickly. Although the overall yield of the Ni^{2+}-NTA resin native preparation is lower than the comparable denaturing preparation, this does not present a major problem when the level of expression of the cell culture is extremely high. A Ni^{2+}-NTA native preparation could not be employed for Δ1-14 since it did not bind to Ni^{2+}-NTA resin under any circumstances.

For unknown reasons, Δ1-14 did not bind to Ni^{2+}-NTA resin in any significant amount, even under denaturing conditions. Since the protein is completely unfolded under these conditions, the binding affinity (and therefore the final yield) of the protein should be increased, but in this case it did not have any effect. In order to purify Δ1-14, it was necessary to bind it to SP-sepharose, a negatively-charged ion exchange resin. It was found that if Δ1-14 was bound to SP-sepharose while denatured, the yield was very high but the protein did not fold properly shown by from test proteolysis reactions with Factor Xa. Δ1-14, when refolded on SP-sepharose and then eluted, did not cleave with Factor Xa. However, if the protein was partially refolded (*i.e.* at 1 M urea) while free in solution, and then bound to SP-sepharose, the yield was extremely high and the correct folding conformation appeared to be achieved, since it could be cleaved with Factor Xa.

For both full-length $U2AF^{65}$ and Δ1-14, it was found that a large number of nonspecifically cleaved products were obtained if the proteolysis reaction was allowed to proceed unchecked. A test cleavage reaction with Δ1-14 (Fig. 11A) shows the appearance of lower molecular weight products below the band of the main cleavage product in the SDS polyacrylamide gel after about 30-60 min. from the start of the reaction. This is likely due to the tendency of Factor Xa to nonspecifically cleave after arginine residues, particularly within the context of a basic region of the protein such as the RS domain of $U2AF^{65}$ (New England Biolabs). The Factor Xa site directly precedes the RS domain of $U2AF^{65}$, which contains a high concentration of arginine residues. Non-specifically cleaved products were not visible in the case of Δ1-64 and Δ1-94, both of which do not have the RS domain. The non-specific cleavage of full length $U2AF^{65}$ and Δ1-14 cleavage by Factor Xa appears in a time dependent fashion, suggesting that the first product is the correctly cleaved protein with the N-terminal cysteine. The first product of Factor Xa cleavage is then subject to further cleavage at the arginine residues contained within the RS domain, resulting in a series of lower molecular weight species.

The nonspecific cleavage of full-length $U2AF^{65}$ and Δ1-14 raised a major obstacle in the course of purification. The Factor Xa proteolysis had to be halted at about 50-60% completion, to prevent complete degradation of the correctly cleaved protein product. FPLC

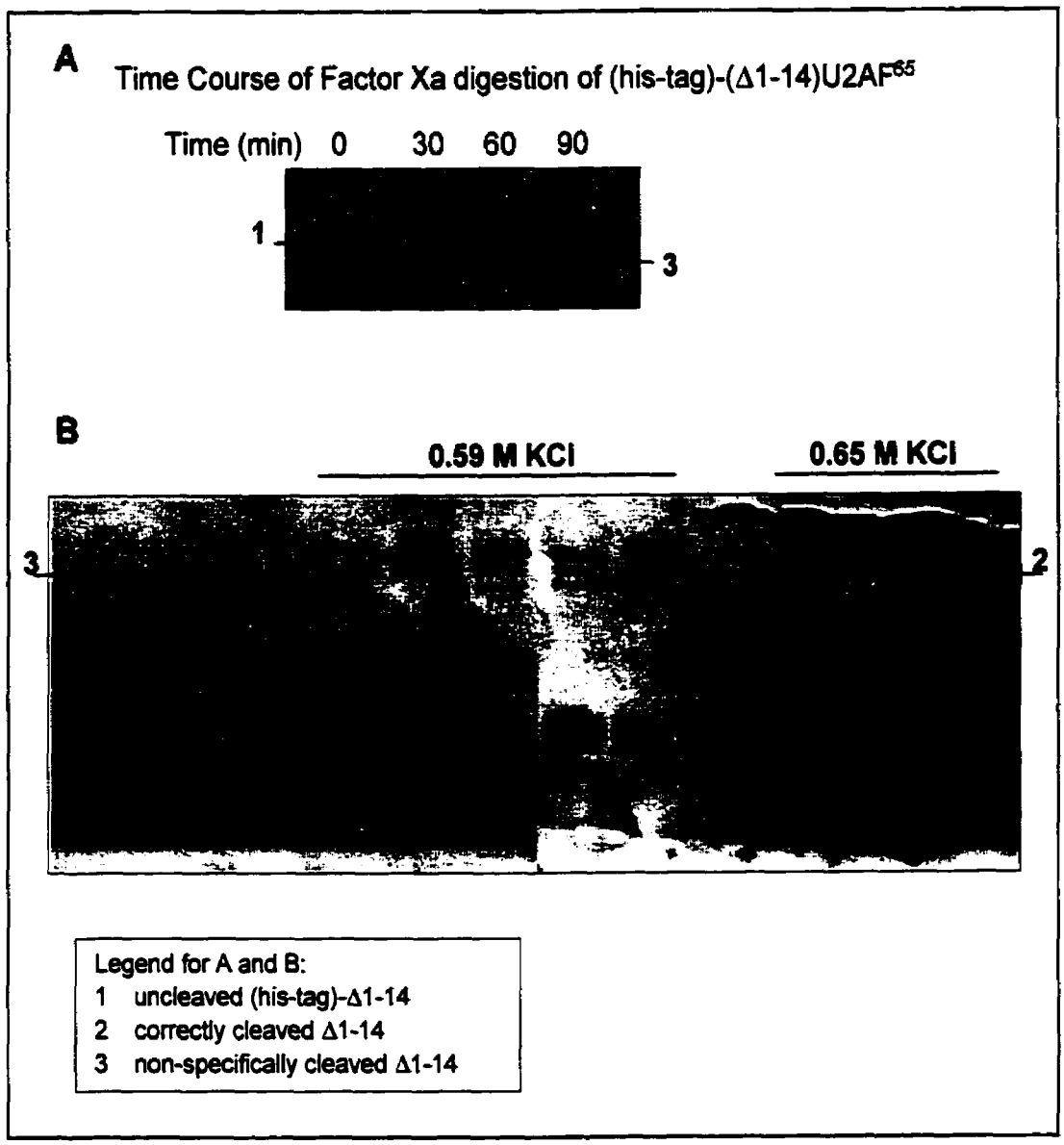


Figure 11: (A) Purification of (EDTA)-(Δ1-14)U2AF^{65} on SP-sepharose, immediately following cleavage with Factor Xa and reaction with EDTA-3MPA. The purification followed a step gradient of KCl, where most of the uncleaved protein (2) eluted at 0.59 M KCl and the cleaved and derivatized protein (3) eluted at 0.65 M KCl. (B) The non-specific cleavage of (Δ1-14)U2AF^{65} by Factor Xa. Urea was added to partially inhibit Factor Xa activity, to allow appreciable amounts of the correctly cleaved product to accumulate.

purification was employed to purify the correctly cleaved protein from non-specific degradation products (Δ1-14 shown in Fig. 11B). Both of these processes used large amounts of protein and as a result, the starting amount of protein had to be very high, >> 1 mg/mL.

3. Derivatization of U2AF^{65} N-Terminal Cysteine Mutants with the Thioester Reagents

After cleavage of the various U2AF^{65} proteins with Factor Xa, the proteins were derivatized with the thioester reagents, EDTA-3MPA, biotin-3MPA, or benzophenone-3MPA in the presence of excess dithiothreitol (DTT). The excess DTT was used to keep all cysteine residues reduced and to convert any thioesters formed from the reaction between internal cysteines and the thioester reagent, back to the starting cysteine residues. The derivatization reaction was straightforward; and the yield was ascertained by the biotin-3MPA assay, described below. In the case of derivatization with benzophenone-3MPA, a fine white precipitate formed during the reaction (most likely excess benzophenone-3MPA, which is not soluble in aqueous solution); the precipitate did not affect the yield of the reaction and did not appear to affect the integrity of the protein.

Since the full-length U2AF^{65} and Δ1-14 produced non-specifically cleaved products when cleaved by Factor Xa, these derivatized proteins were further purified to remove the non-specific cleavage products and the uncleaved protein still present. Both the non-specifically cleaved proteins and the uncleaved proteins still contain the RNA binding domains and could thus interfere in future experiments. To minimize the amount of degradation that occurred, the protein samples were treated with Factor Xa, then reacted with the thioester reagents immediately following the protease reaction. The protein was then purified with the use of an SP-sepharose column and an FPLC pump to maintain an accurate salt gradient. Fig. 11B shows the purification of Δ1-14 using a KCl gradient from 0.1 M to 1.0 M. It was found that the his-tagged protein eluted earlier than the cleaved proteins, so a step gradient was employed. This gradient contained two steps, holding at 0.59 M KCl to elute his-tagged protein, and holding at 0.69 M to elute the cleaved proteins. Most of the non-specific cleavage products eluted from the column well before the his-tagged proteins. Again, this suggests the degradation of the RS domain in these proteins, since they are less basic and therefore have less affinity for the SP-sepharose. The use of the FPLC required a large injection of protein for successful recovery of protein to occur, typically 3-5

mg total protein per sample injection (based on 1 mg/mL concentration; 3-5 mL sample volume) on to the column. This was possibly due protein absorption on to the tubing of the injection loops and column. The recovery of purified cleaved and derivatized protein was approximately 0.2 mg/mL (total volume ~1 mL), which was not very high. The concentration of purified and derivatized Δ1-14 posed further problems in experiments that will be described in the next chapter.

4. An Assay with Biotin-3MPA to Determine the Presence of N-terminal Cysteines in the U2AF^{65} Mutants

The biotin-3MPA reagent was employed in a Western blot immunoassay to determine the presence and reactivity of the N-terminal cysteine in the group of U2AF^{65} mutants. The assay, as carried out on full-length U2AF^{65}, is shown in Fig. 12. Anti-biotin mouse antibody was chosen as the method to identify biotin-labelled protein. The use of a secondary antibody (rabbit anti-mouse conjugated to horseradish peroxidase) and chemiluminescent detection amplified the signal so that very small amounts of protein may be used. This is a useful aspect in conserving protein stocks for further investigation. Two controls were used in the immunoblot assay: the first control is the protein with no N-terminal cysteine. This protein was his-U2AF^{65}, with a 5 kDa his-tag peptide attached to the N-terminus (after treating this protein with Factor Xa protease to remove the his-tag peptide the protein sequence begins with cysteine). The second control was the protein reacted with biotin-3MPA alone. In Fig. 12B, the first lane of the blot contains the first control, "no cysteine". There are several bands apparent, all of which appear at lower molecular weight than the actual his-U2AF^{65}, which may be due to either nonspecific bonding of the antibodies or residual biotin-3MPA adsorbed to the protein. The sample of full-length U2AF^{65} after reaction with biotin-3MPA, shows an extremely strong signal (lane 2, Fig. 12B). In comparison, the next two lanes (lanes 3 and 4) which show full-length U2AF^{65} reacted first with EDTA-3MPA or benzophenone-3MPA and then chased with biotin-3MPA, show no signal. This strongly supports the conclusion that the thioester reaction in each case was successful (*i.e.* no N-terminal cysteine is present to react with biotin-3MPA)

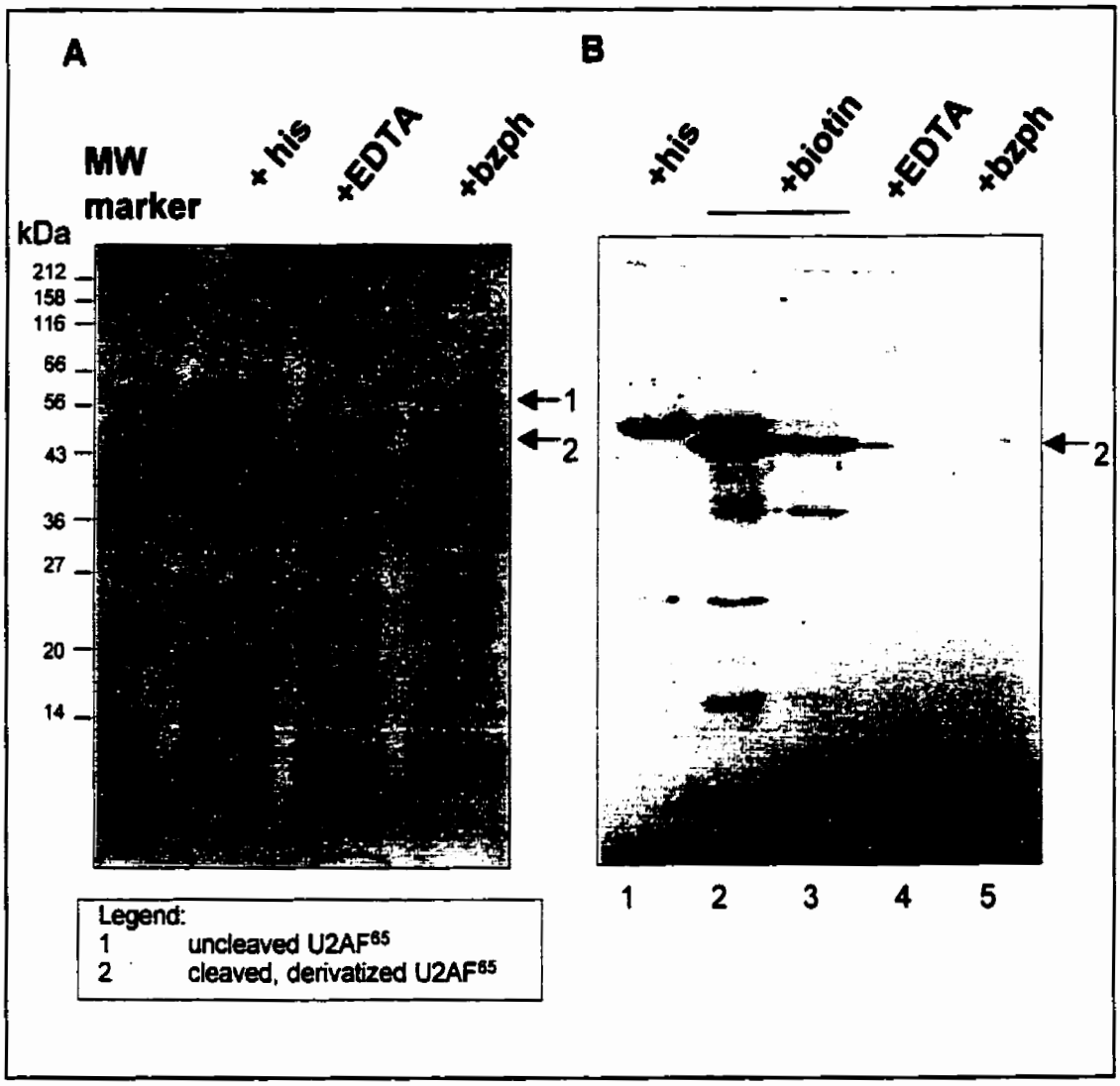


Figure 12: Immunoblot assay for N-terminal cysteines in full-length U2AF^{65} mutant, employing biotin-3MPA. (A) SDS-polyacrylamide gel featuring his-tagged U2AF^{65}, and U2AF^{65} after derivatization with two thioester reagents: (EDTA)-U2AF^{65} and (benzophenone)-U2AF^{65}. (B) Chemiluminescent immunoblot of U2AF^{65} derivatives, using mouse anti-biotin to detect biotin-labelled protein. In lane 1, uncleaved U2AF^{65} with the his-tag still attached was reacted with biotin-3MPA.. U2AF^{65} was reacted with biotin-3MPA to label it with biotin, and shows a very strong signal (lanes 2 and 3). In lanes 4 and 5, U2AF^{65} was first reacted with EDTA-3MPA (4) or benzophenone-3MPA (5), then biotin-3MPA.

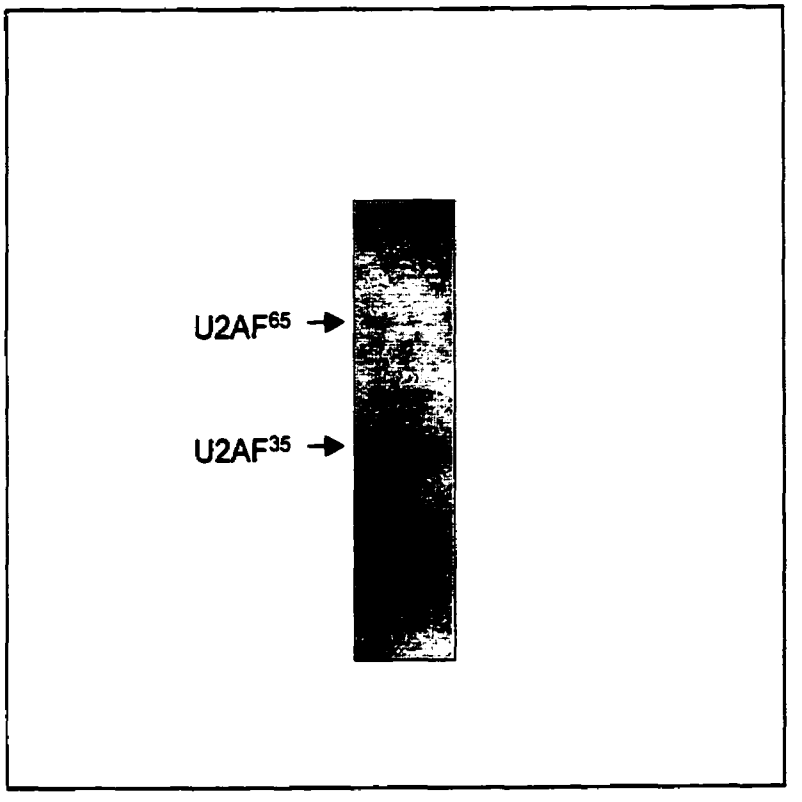


Figure 13: SDS-polyacrylamide gel analysis of $U2AF^{35}$, after a denaturing purification from a transfected *E. coli* stock coexpressing $U2AF^{65}$ and his-tagged $U2AF^{35}$. Ni^{2+}-NTA resin was used to purify $U2AF^{35}$ from the cell lysate.

and that the reaction as carried out, proceeds to completion. The thioester reagents are active and may thus be used to derivatize proteins with a cysteine at the N-terminus.

5. Expression and Isolation of U2AF^{35}

The purification of U2AF^{35} was carried out so that it might be used in conjunction with the derivatized U2AF^{65} mutants. As discussed in the introduction, U2AF^{35} plays an important role in 3' splice site recognition and indirectly influences alternative splicing of introns, as dictated by the length of the polypyrimidine tract. U2AF^{35} was purified to be used in the hydroxyl-radical footprinting experiments with EDTA-derivatized U2AF^{65}, to determine whether it affected the positioning of U2AF^{65} with respect to the pre-mRNA. A major obstacle in the purification of U2AF^{35} was the fact that it was co-expressed with U2AF^{65}. Due to the tight association between these two proteins under normal conditions, a native preparation could not be used. A denaturing preparation was necessary and Ni^{2+}-NTA resin employed to make use of the his-tag present on U2AF^{35}. During the course of the purification, it was found that U2AF^{35} was prone to degradation, particularly during dialysis to remove denaturant. To avoid this problem, U2AF^{35} was refolded on the Ni^{2+}-NTA resin and then eluted with an imidazole-containing buffer. Despite these precautions, a small amount of U2AF^{65} remained in the eluted protein fractions, as seen in the SDS-PAGE analysis (Fig. 13). The concentration of this extraneous U2AF^{65} was quite low and not a concern in future experiments.

6. Activity of the U2AF^{65} N-terminal Cysteine Mutants

U2AF^{65} mutants beginning with cysteine were engineered, purified and then tested for the presence and reactivity of the N-terminal cysteine to the thioester reagents described in Section C.1. Proof of the activity of these mutant U2AF^{65} proteins was required. A number of factors could interfere in protein activity: improper folding of the protein could occur or post-translational modifications that occur in eukaryotic systems that do not occur in the prokaryotic system in which these proteins had been expressed, could prevent activity. In the case of Δ1-14 and Δ1-64 mutants, they were purified by a denaturing preparation. After the purification process was completed it was not possible to say whether or not these proteins had folded properly during renaturation. Since U2AF^{65} had been successfully cloned

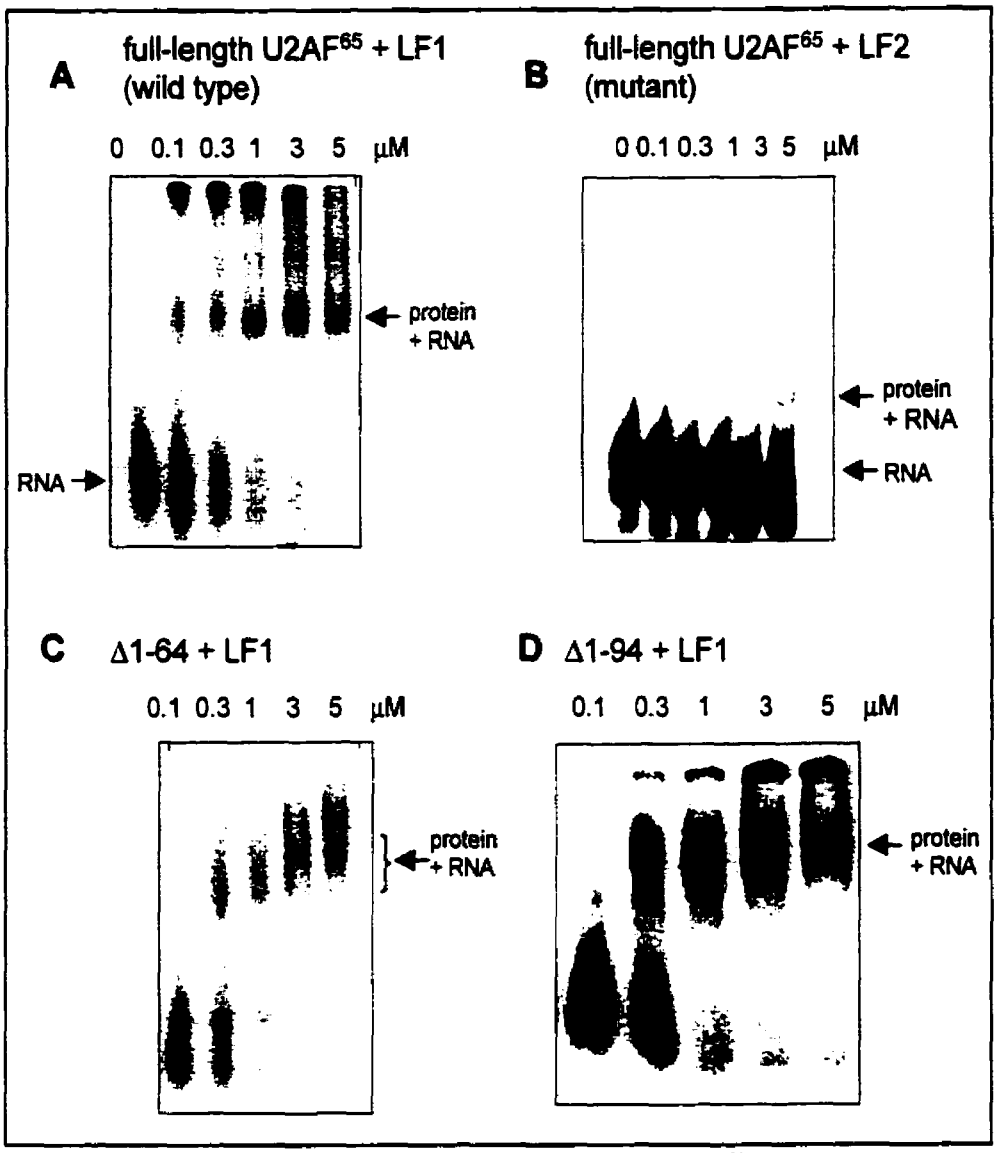


Figure 14: Native gel mobility shift assay with the U2AF^{65} mutants and short ^{32}P-labelled RNAs representing the branchpoint region, polypyrimidine tract and part of the 3' exon were used as the binding substrate. LF1 is the wild type pre-mRNA based on PIP85.b, LF2 is pre-mRNA substrate with the polypyrimidine tract replaced with purines. RNA concentration was 1 nM throughout. (A) Full-length U2AF^{65} binds to LF1, with all of the protein bound at ~1 µM. (B) Full-length U2AF^{65} does not bind to LF2 in any appreciable amount. A very small amount of protein appears to bind at the very highest concentration (far right lane). (C) Δ1-64 and LF1. (D) Δ1-94 and LF1. Both of the ΔRS-U2AF^{65} mutants, Δ1-64 and Δ1-94 bind to LF1 at similar concentrations to full-length U2AF^{65}.

and expressed in *E. coli* on previous occasions (MacMillan *et al.*, 1997), it was believed that eukaryotic post-translational modifications were not required in the case of U2AF^{65}. Two conditions for U2AF^{65} activity were set: (1) the ability to bind to an RNA oligonucleotide containing the consensus sequences of the intron (*i.e.* the branchpoint region, the polypyrimidine tract, and the 3' splice site), with a binding affinity matching the wild type U2AF and (2) in the case of full-length U2AF^{65} and Δ1-14 mutants, the ability to reconstitute RNA splicing in a nuclear extract depleted of U2AF.

To prove that the U2AF^{65} mutants were able to bind to a pre-mRNA substrate containing the intron consensus sequences, a native gel mobility shift assay was employed. The U2AF^{65} proteins were bound to a short RNA oligonucleotide (5' end-labelled with ^{32}P) containing part of the intron, including the branch-point region and the polypyrimidine tract, the 3' splice site, and 15 bases of the 3' exon following. It was expected that all of the U2AF^{65} proteins should be able to bind to the RNA since the N-terminal truncations do not affect the RNA binding domains. Rosbash & coworkers reported a binding affinity (K_d) for U2AF of 1 μM, given an RNA concentration of 0.1 nM (Berglund *et al.*, 1998). In this assay, the gel assay conditions almost identical to the Rosbash assay, except that a lower concentration of tRNA (0.25 mg/mL) was employed in the binding buffer. It was found in preliminary tests that a concentration of 0.5 mg/mL tRNA, as indicated by Rosbash & coworkers, inhibited the entry of the protein-RNA complex into the gel. It was observed that approximately 50% of the full-length U2AF^{65} protein was bound at a concentration of ~1 μM, in agreement with Rosbash's findings (Fig. 14A). Δ1-64 and Δ1-94 also bound to LF1 (Fig. 14C,D) with approximately the same binding affinity as full-length U2AF^{65}. This indicated that the removal of the RS domain did not affect the activity of the RNA binding domains; they appear to behave independently from the RS domain. As a control to distinguish non-specific binding, the binding of full-length U2AF^{65} with LF2, the mutant RNA, was tested. In LF2, the polypyrimidine tract has been replaced by a random purine sequence, so it was expected that U2AF^{65} would not bind to it. As can be seen in Fig. 14B, the U2AF^{65} protein does not bind to LF2.

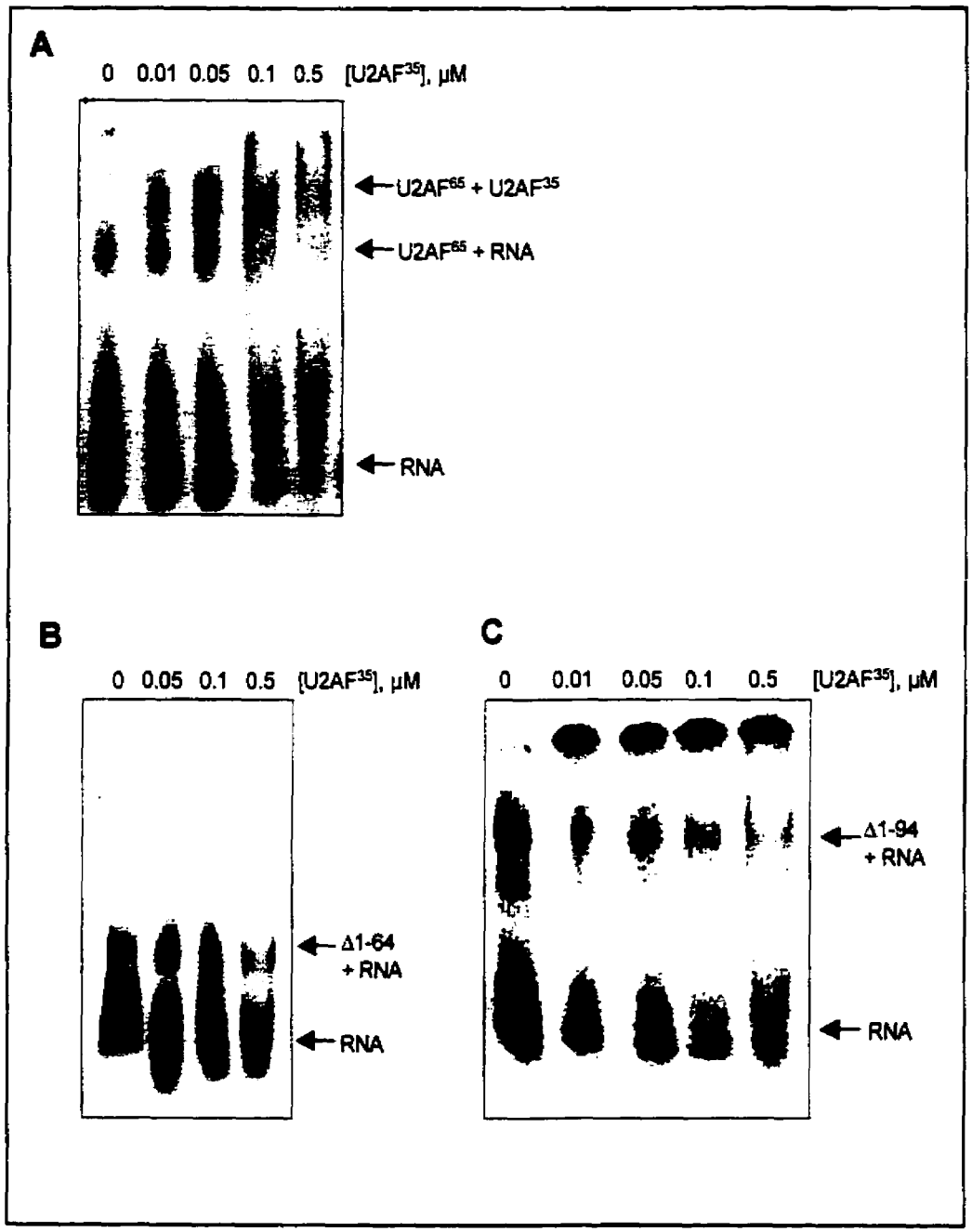


Figure 15: Formation of the U2AF heterodimer, using the $U2AF^{65}$ mutants and wild type $U2AF^{35}$, on the short pre-mRNA substrate LF1. Complexes were observed by native gel mobility shift. (A) Full-length $U2AF^{65}$ concentration was held at ~0.5 μM while $U2AF^{35}$ concentration was increased. (B) Δ1-64 (~0.5 μM in each lane) + $U2AF^{35}$. (C)Δ1-94 (~0.5 μM in each lane) + $U2AF^{35}$.

In the U2AF^{65}-U2AF^{35}-RNA binding assay, U2AF^{65} concentration was held constant at ~1 μM and U2AF^{35} concentration was increased. The U2AF^{35} used in these experiments was expressed and purified from a transfected *E. coli* cell stock, a gift from Donald Rio (UC Berkeley, California). There were concerns that U2AF^{35} expressed from *E. coli* was not active (Sabine Guth, EMBL Heidelberg; personal communication). If the U2AF^{35} was not active, then it would likely not bind to U2AF^{65}. If it was active, then U2AF^{35} was expected to bind to the full-length U2AF^{65}, and possibly Δ1-64 but not Δ1-94, since amino acids 64-182 of U2AF^{65} are believed to be used for interaction with U2AF^{35} (Zhang *et al.*, 1992). This was found to be the case (Fig. 15), with U2AF^{35} binding to the full-length but not Δ1-64 and Δ1-94. This suggested that U2AF^{35} was active and that the binding surfaces of both proteins were of the correct conformation. Δ1-64 did not appear to bind to U2AF^{35} at the given concentrations, which was an unexpected finding. It suggested that the removal of the RS domain affects the region of U2AF^{65} that binds to U2AF^{35}. In summary, the native gel shift assay showed that the U2AF^{65} mutant proteins were capable of binding to polypyrimidine tract within an RNA oligonucleotide, with similar affinity as previously reported.

The next condition for activity was the ability to restore splicing in a mammalian nuclear extract that had been previously depleted of U2AF (ΔU2AF NE). The mutant U2AF^{65} protein, both before and after derivatization with a thioester reagent, were added to the depleted nuclear extract and the various RNA products from splicing were tracked with the use of a ^{32}P-labeled pre-mRNA substrate. Activity was compared against three controls: (1) "wild-type" nuclear extract (*i.e.* untreated and containing the full complement of splicing factors), (2) ΔU2AF nuclear extract, and (3) ΔU2AF nuclear extract with wild type U2AF^{65} added (a reconstitution control). The formation of the lariat intermediate (intron + 3' exon) and the lariat product (intron), after a 60 min. incubation, are apparent as two bands near the top of the gel, when the substrate is body-labelled with ^{32}P-UTP (Fig. 16A). When the pre-mRNA substrate is 5' end-labelled with ^{32}P, the formation of the free 5' exon, and the spliced-together exons (Fig. 16B) may be observed. The full-length and Δ1-14 proteins were active in this assay, as can been seen by the formation of the splicing products. The level of activity is considerably less than the wild type nuclear extract, by comparison of the intensity

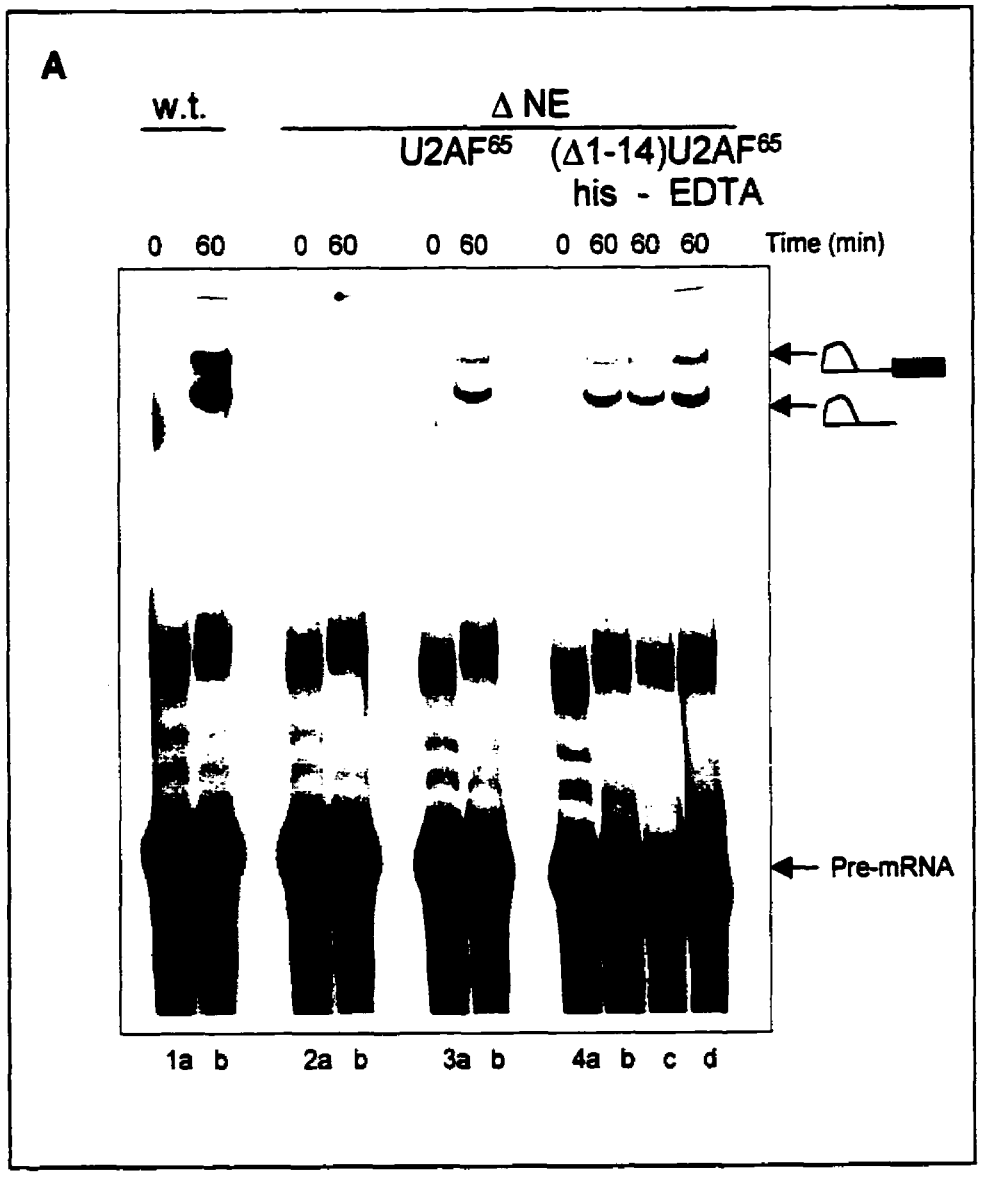


Figure 16: Reconstitution of RNA splicing in U2AF-depleted nuclear extract (ΔNE) by the U2AF^{65} cysteine mutants. The splicing products, the lariat intermediate and the lariat product are indicated with the arrows. The whole NE showed the likely maximum yield of RNA splicing and ΔNE showed the lack of splicing in the absence of U2AF. (A) Splicing of ^{32}P body-labelled PIP85.b pre-mRNA substrate, by Δ1-14. The ability of Δ1-14, to reconstitute splicing was contrasted with whole NE (w.t.) (lanes 1a,b); ΔNE (2a,b); ΔNE + w.t. U2AF^{65} (3a,b). Splicing reconstitution was done with Δ1-14 with the his-tag still attached (4b), after removal of the his-tag by Factor Xa (4c), and after derivatization with EDTA-3MPA (4d).

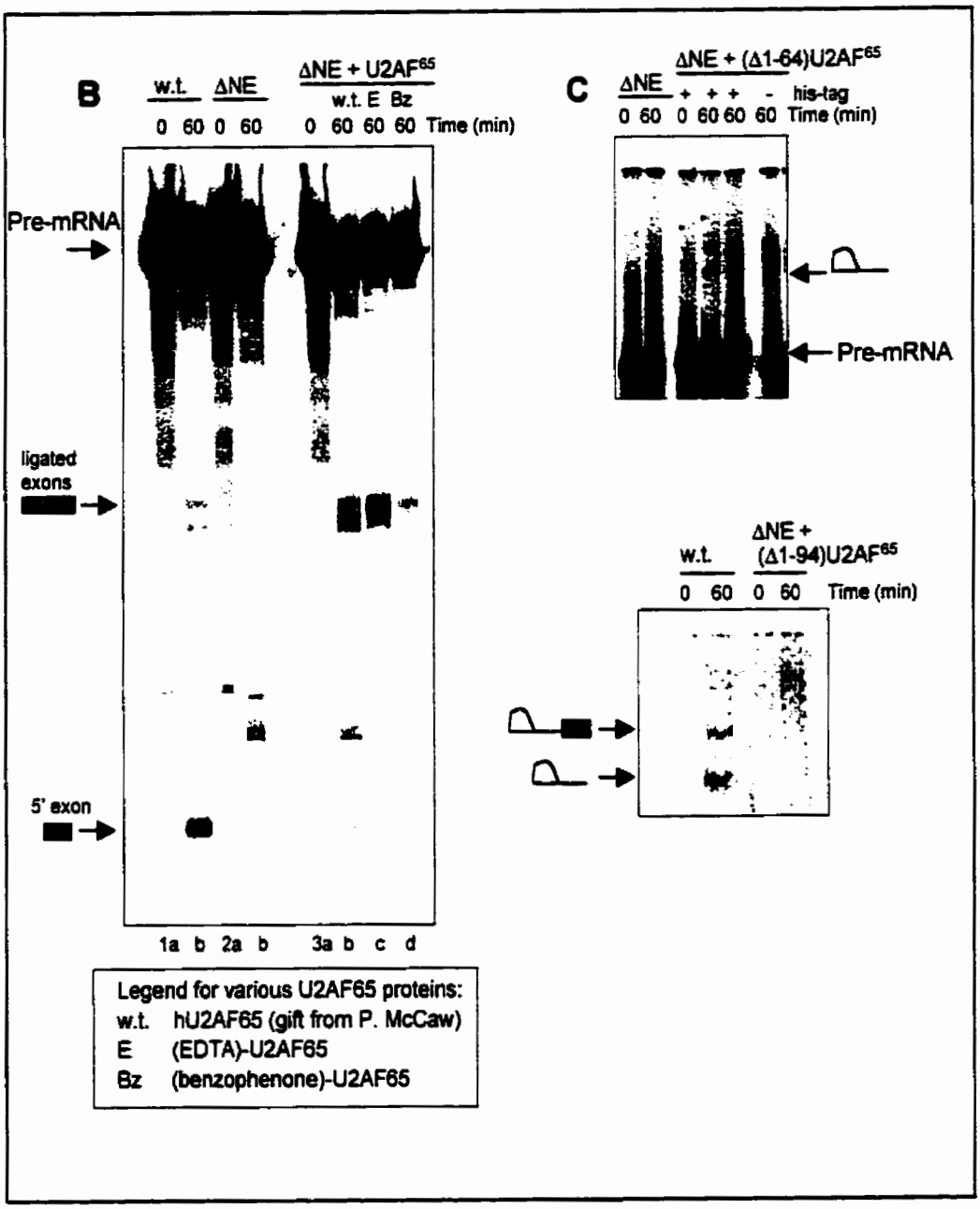


Figure 16 (continued): (B) Reconstitution of splicing of ^{32}P 5' end-labeled PIP85.b substrate by full-length U2AF^{65} (3a,b) and derivatives (3c,d). The splicing products, the spliced-together exons and the 5'-exon (intermediate) are as indicated with the arrows. (C) Splicing reconstitution of ^{32}P body-labeled PIP85.b with the ΔRS-U2AF^{65} mutants, Δ1-64 and Δ1-94. A small amount of lariat product was produced in the reconstitution with Δ1-64. The splicing reconstitution with Δ1-94 shows no splicing products.

of the bands which represent the splicing products. It was also noted that the ability of Δ1-14 to reconstitute splicing diminished with each treatment step, after removal of the his-tag peptide and after derivatization with EDTA-3MPA. This is likely attributable to the nonspecific cleavage of the protein upon Factor Xa treatment, which drastically reduces the amount of active protein present. In the splicing reconstitution with Δ1-64, this protein unexpectedly showed a very low but still detectable level of splicing activity (Fig. 16C). This was not expected since Green & coworkers had earlier reported ΔRS-U2AF^{65} (with the same N-terminal truncation) was unable to reconstitute splicing to any observable level (Valcárcel *et al.*, 1996). In the case of Δ1-94, no splicing activity was observed (Fig. 16C). Considering that the difference between these two proteins is the stretch of amino acids 65-93, which are believed to interact with U2AF^{35} (Zhang *et al.*, 1992), it raises the interesting possibility that this difference may affect the ability to reconstitute splicing. As mentioned before, the procedure to deplete U2AF from the nuclear extract does not entirely remove U2AF^{35}, so that it is likely present in minute but still significant quantities (Wu & Maniatis, 1993). U2AF^{35} binds the 3' splice site and is usually required in introns where the length between the polypyrimidine tract and the 3' splice site is short. This is the case with the substrate used for this assay, the PIP85.b pre-mRNA. The Δ1-64 protein may be able to bind U2AF^{35} weakly (not detectable in the gel shift assay), and is able to reconstitute splicing to a very low level. The Δ1-94 protein is unable to bind U2AF^{35} and is not able to reconstitute splicing. It is possible, considering these observations, that the ability to bind U2AF^{35} affected the splicing activity observed.

7. Hydroxyl Radical Footprinting of EDTA-Derivatized U2AF^{65} Mutants

Hydroxyl radical footprinting of proteins on DNA using Fenton's reagent (Fe^{2+}/H_2O_2) has been well-documented, in particular by Tullius & coworkers (Dixon *et al.*, 1991). The hydroxyl radical footprinting experiment here involved the binding of an EDTA-derivatized U2AF protein to a substrate ^{32}P 5' end-labelled RNA, initiation of the hydroxyl formation reaction by addition of H_2O_2, and quenching of the reaction after a set amount of time. There are several factors which affect the strength of the observed signal. The first

consideration was the concentrations of the reactants, Fe^{2+}, H_2O_2 and sodium ascorbate to be used. The concentration of H_2O_2 was kept at 0.03% v/v to prevent protein degradation, as recommended by Tullius & coworkers. Ascorbate concentration was held at 1 mM, in large excess to Fe^{2+}, to ensure adequate turnover of the iron during the course of the reaction. The Fe^{2+} ion concentration was dictated by the concentration of protein used in the reaction since it had to be added in equimolar amounts. Excess free Fe^{2+} in solution would cause nonspecific cleavage of the RNA. The protein concentration had to be high enough (at least 30-50% bound) to ensure a cleavage pattern would be observable. However, the concentration of protein allowable was constrained by the 10% glycerol content of the storage buffer: glycerol at a concentration greater than 0.5% would severely inhibit the hydroxyl radical cleavage reaction (Dixon *et al.*, 1991). The protein stock solution had to be diluted at least 20 times in the reaction volume to bring the overall glycerol concentration to 0.5%. These two factors required a very high starting concentration of protein. This raised problems with the experiments involving Δ1-14, since the concentration after dilution (~0.1 μM) was well below the stated 1 μM K_d for $U2AF^{65}$. Indeed, experiments with Δ1-14 failed to show any recognizable cleavage pattern. Experiments concentrated primarily on full-length $U2AF^{65}$ and Δ1-64, both of which had a final concentration of ~0.5 μM after dilution.

A hydroxyl radical cleavage experiment to determine the location of the $U2AF^{65}$ RS domain with respect to the pre-mRNA was carried out with the full-length mutant, (EDTA)-$U2AF^{65}$. An initial attempt to footprint (EDTA)-$U2AF^{65}$ on a full-length pre-mRNA (PIP85.b, containing the 5' and 3' exons), failed to show any cleavage whatsoever. To isolate the cleavage signal, the short RNA oligonucleotide LF1 was employed for the experiment. It was reasoned that $U2AF^{65}$ would interact with LF1 in the same manner as with PIP85.b, since LF1 contains the branchpoint region, the polypyrimidine tract and the 3' splice site (including part of the 3' exon). Also, since this was a minimal experiment only involving $U2AF^{65}$, there would be no concerns as to the missing elements upstream of the LF1 sequence. (EDTA)-$U2AF^{65}$ produced sharp and definite cleavage bands at the three bases immediately upstream of the branchpoint adenosine (bases 5-9, lane 4, Fig. 17A). The

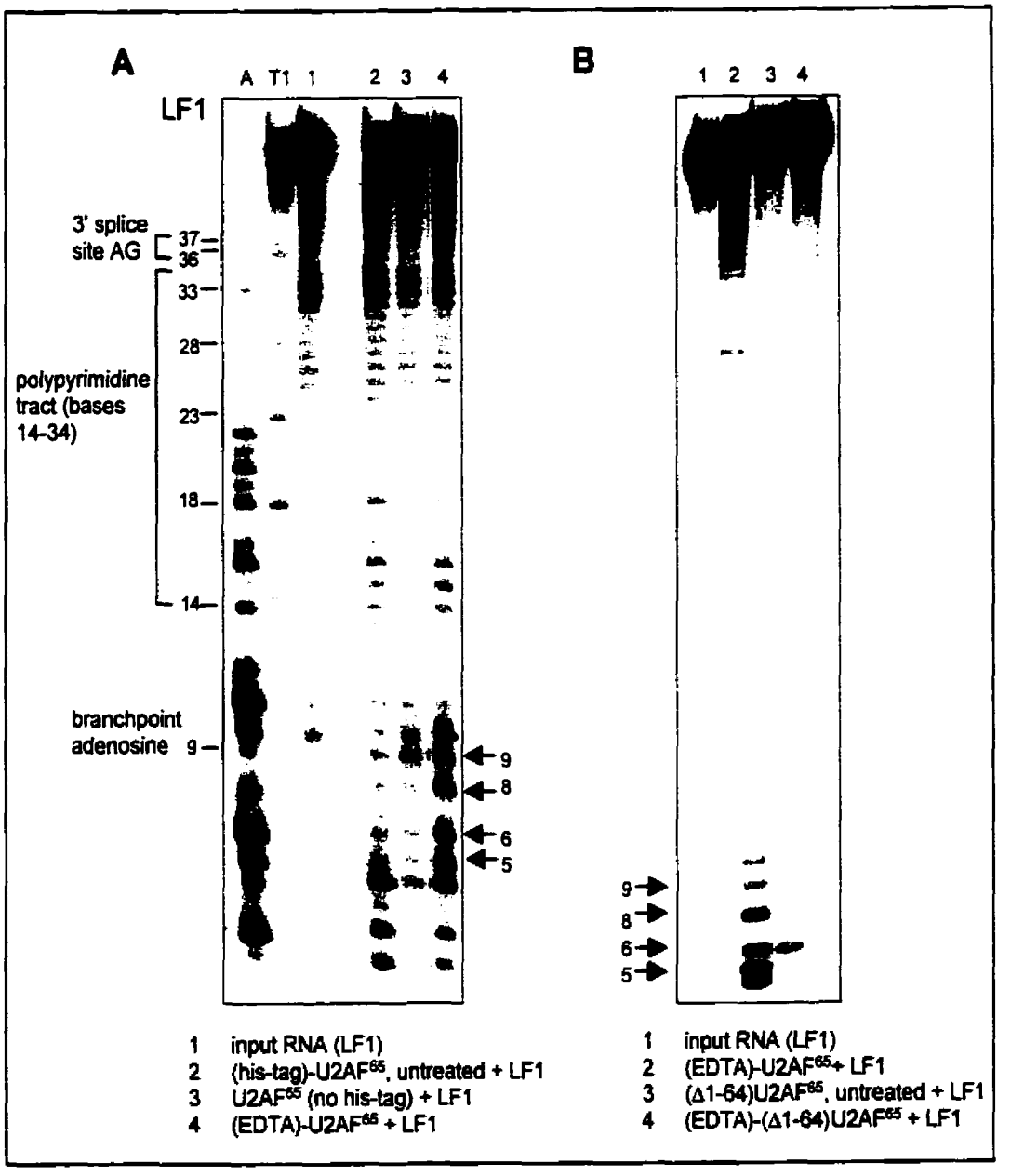


Figure 17: Site-directed hydroxyl radical footprinting of EDTA-derivatized U2AF^{65} mutants on a short RNA (LF1). LF1 represents the intron (branchpoint region, polypyrimidine tract) and the first 15 bases of the 3' exon of PIP85.b. A digestion of LF2 by RNase A (A) and RNase T1 (T1) produces a ladder by which the position within the sequence is identified. (A) EDTA-derivatized full-length U2AF^{65} causes cleavages at bases immediately upstream of the branchpoint adenosine. (B) (EDTA)-(Δ1-64)U2AF^{65} does not cause any observable cleavages on the RNA (lanes 3, 4), compared to (EDTA)-U2AF^{65} (bands at positions 4-9, lane 2).

intensity of these bands is well above the background level observed in the controls (lanes 1-3, Fig. 17A), which consist of (1) RNA with various combinations of the chemical reagents (Fe^{2+}, H_2O_2, ascorbate), (2) uncleaved his-tagged protein + RNA, and (3) cleaved (no his-tag) underivatized protein + RNA). There are faint bands that appear in the control containing underivatized $U2AF^{65}$ (lane 2 of Fig. 17A) at positions 5-8 and strong bands at positions 1-4. This could be due to the binding of $U2AF^{65}$ to the PPT causing exposed portions of the RNA substrate to be susceptible to random cleavage by free Fe^{2+}/H_2O_2 in solution (*i.e.* protection footprinting). However, it was thought that the cleavage pattern in the control was due to the possible contamination of the $U2AF^{65}$ protein sample with metal ions. The control in lane 2 is full-length $U2AF^{65}$ with the his-tag peptide at its N-terminus; the his-tag peptide may have chelated Ni^{2+} from the Ni^{2+}-NTA resin that was used to purify the protein. Nickel is a group VIII transition metal like iron and can exist in +2 and +3 oxidation states; it should therefore also oxidize in the presence of H_2O_2 to form hydroxyl radicals, thus acting as a nuclease. After treatment with Factor Xa to remove the his-tag, these cleavage bands are no longer apparent and all bands in the region of positions 1-9 appear relatively faint (lane 3, Fig. 17A). This implies that the his-tag peptide in uncleaved $U2AF^{65}$ is responsible for the cleavages at positions 1-4 and that protection footprinting of the protein is not being observed.

The cleavage bands at positions 5-9 in the reaction with (EDTA)-$U2AF^{65}$ are most likely due to the presence of the Fe^{2+}-EDTA nuclease at the N-terminus of the protein. The appearance of these bands strongly suggested the RS domain of $U2AF^{65}$ is in proximity to the branchpoint region. This is in agreement with the cross-linking data reported by Green & coworkers (Valcárcel *et al.*, 1996). As a further control, the same experiment was carried out with LF2 RNA as the substrate. LF2 represents the same sequence within PIP85.b as LF1 except for the polypyrimidine tract, which has been replaced with purines. With LF2 as the substrate, cleavage occurred every base along the length of the RNA. Also, in comparison to LF1, the cleavage bands were very intense. It is possible that the protein, since it is unable to bind to LF2, is free in solution and thus releases hydroxyl radicals which cause random cleavage of any nearby RNA.

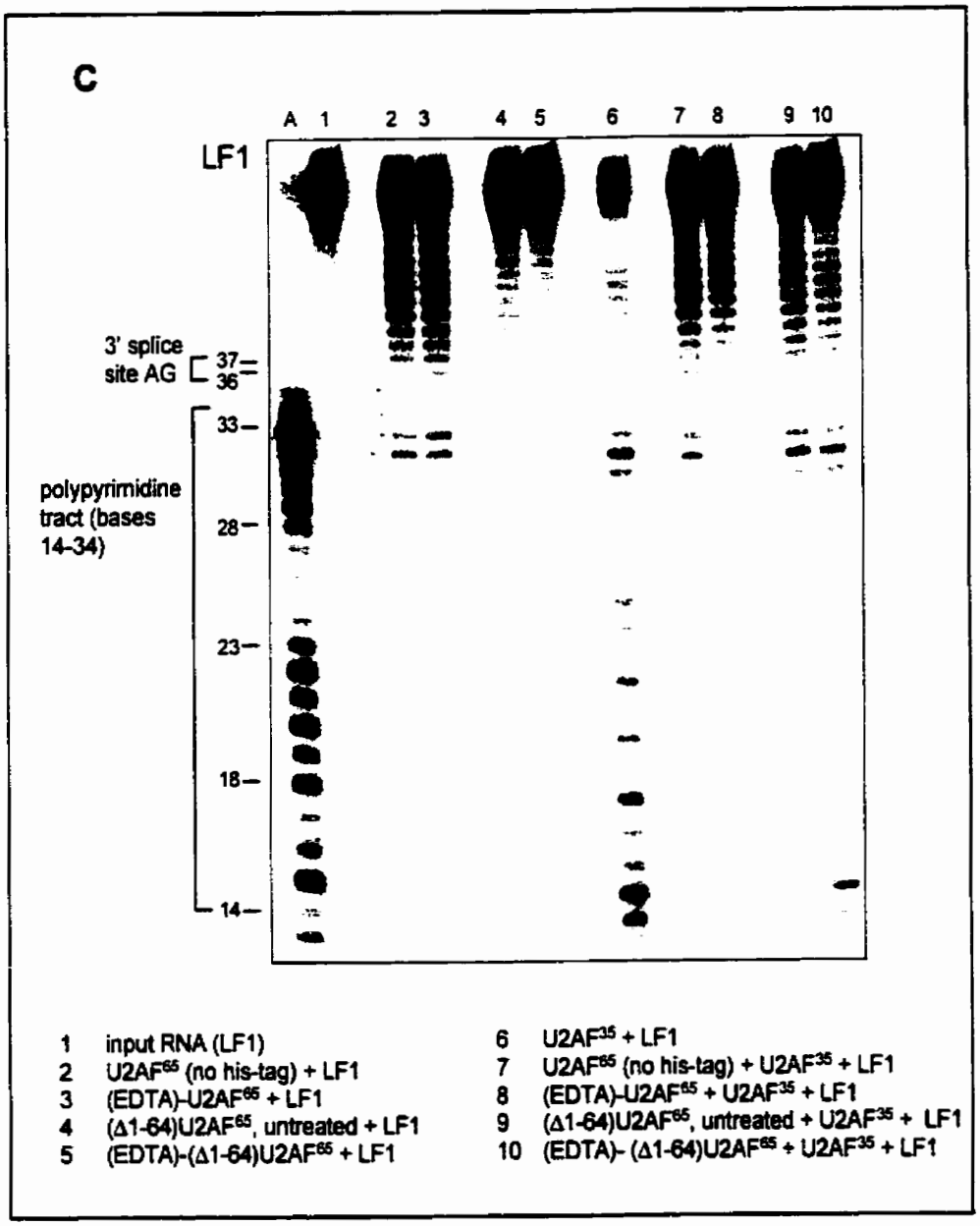


Figure 17 (continued): (C) Hydroxyl radical footprinting of EDTA-derivatized full-length U2AF^{65} and Δ1-64 with U2AF^{35}, on 5' end-labelled LF1 pre-mRNA substrate. The RNA fragments were run further into the gel to identify any changes after the addition of U2AF^{35} that may have occurred near the 3' splice site.

The next hydroxyl radical footprinting experiment that was attempted was with Δ1-64 on LF1 (Fig. 17B). It was hoped that (EDTA)-(Δ1-64)U2AF^{65} would produce a cleavage pattern indicating the location of both the end of the RS domain and the beginning of the U2AF^{35} site of interaction (which interacts with amino acids 64-182 of U2AF^{65}; Zhang *et al.*, 1992). Given that the N-terminus of the RS domain was likely located at the branchpoint, it was possible that the N-terminus of Δ1-64 would footprint to a location several bases downstream from the branchpoint. However, since the N-terminus of Δ1-64 marks to beginning of the U2AF^{35} interaction site, it was also possible that it would be located somewhere near the 3' splice site since U2AF^{35} recognizes the 3' splice site. This issue could not be resolved since (EDTA)-(Δ1-64)U2AF^{65} did not produce any significant cleavages that were observable above background levels. Since the concentration of the protein was sufficient for RNA binding to occur, it was concluded that the N-terminal residue in Δ1-64 is at a position that is not in proximity to the RNA. However, this conclusion is not believed to be applicable to the wild type protein; *i.e.* the location of the N-terminal residue of Δ1-64 does not necessarily reflect the true location of amino acid 64 in the wild type protein, with respect to the pre-mRNA. This is a limitation to the thioester reagent/N-terminal cysteine strategy. The requirement of an N-terminal cysteine necessarily involves the use of N-terminal truncations in order to study locations deeper within the protein sequence. These truncation mutants may have a disordered and unpredictable structure at the N-terminus that is different from the wild-type, unless it is certain that the protein in question is composed of highly ordered, modular domains. In this respect, the scanning cysteine mutagenesis method is more useful for studying internal protein structure since it may be used for all types of proteins (Akabas *et al.*, 1992).

The lack of cleavage bands appearing at positions 1-9 in the hydroxyl radical cleavage experiment with Δ1-64, both underivatized and derivatized with EDTA (lanes 3 & 4, Fig. 17B) again suggest that protection footprinting does not occur and that the cleavage pattern observed for (EDTA)-U2AF^{65} is specific to the presence of EDTA at its N-terminus (lane 2, Fig. 17B).

Another hydroxyl radical footprinting experiment was the introduction of U2AF^{35} to the U2AF^{65}/LF1 system (Fig. 17C). Since U2AF^{35} binds to the 3' splice site, it was thought that Fe^{2+}-EDTA nuclease at the N-terminus of U2AF^{65} would not only cause cleavage at the branchpoint region but perhaps also at the 3' splice site due to the binding activity of each subunit of the U2AF heterodimer. In order to assess the effect of U2AF^{35}, it was added to (EDTA)-(full-length)U2AF^{65} and to (EDTA)-(Δ1-64)U2AF^{65}, to form the U2AF heterodimer before initiating the hydroxyl radical reaction. Reaction mixtures were run further into the gel to observe any changes that occurred at sites near the 3' splice site. No significant cleavages in the 3' splice site region were observed above background levels. It was difficult to observe any change since the intensity of the (degradation) bands in the region of the 3' splice site were quite high. However, it did not seem like any difference was apparent between the control containing no U2AF^{35} and the heterodimer reaction. Since U2AF^{35} interacts with amino acids 64-182 of U2AF^{65}, it is possible that the U2AF^{35}-RNA complex is too far removed from the Fe^{2+}-EDTA nuclease located at U2AF^{65}'s N-terminus for any cleavage at the 3' splice site to occur.

Finally, the hydroxyl footprinting experiment was attempted in the context of the whole spliceosome. This was initially done in 20% ΔU2AF NE, with EDTA-(full-length)U2AF^{65} added. The concentration of the nuclear extract was at the minimum possible to support splicing (Steven Chaulk, unpublished data). 20% Nuclear extract contains an overall glycerol concentration of 4% v/v, which is considerably above the 0.5% limit set by Tullius (Dixon *et al.*, 1991). No cleavage was observed in all cases, including all of the controls, meaning that the glycerol concentration was too high to allow any cleavage from occurring. In order to carry out this experiment, it would be necessary to exchange the buffer of the nuclear extract for a glycerol free buffer or use a method to isolate spliceosome complexes. Spliceosome complexes may be isolated by immunoprecipitation from solution using antibodies directed against one or more of the splicing factors. This approach has been employed by Sharp & coworkers to study the kinetics of protein associations to the pre-mRNA at different time points in splicing (MacMillan *et al.*, 1994).

In summary, the hydroxyl radical footprinting experiments with U2AF^{65} derivatized with Fe^{2+}-EDTA nuclease provided evidence to support Green & coworkers' finding, that the U2AF^{65} RS domain is in close proximity to the branchpoint region. From the observed cleavage pattern on the RNA substrate, the N-terminus of the full-length U2AF^{65} appeared to be immediately adjacent to the branchpoint adenosine. These experiments showed the utility of the EDTA-3MPA reagent as a chemical probe in protein-RNA interactions. The experiments also highlighted the shortcomings of the Fe^{2+}-EDTA nuclease system. It is extremely sensitive to glycerol concentration, as was noted by Tullius & coworkers (Dixon *et al.*, 1991). The inhibition of hydroxyl radical cleavage of RNA due to increased glycerol concentration was evident in the experiment involving the nuclear extract. The glycerol content proved to be problematic, since glycerol is an essential component in the protein storage buffers; it is necessary to protect the stability of the proteins during freezing and thawing. The sensitivity of Fe^{2+}-EDTA to glycerol content sets a limit on the volume of protein that may be added to the reaction volume. This requires the starting concentration of protein to be high enough so that after dilution, a significant proportion of protein will still be bound to the RNA. The Fe^{2+}-(EDTA-3MPA) reagent is therefore not suitable for use in cases where the protein to be studied cannot be generated in sufficiently high concentration, at least 2 μM in the case of U2AF^{65}. To avoid the problem of glycerol concentration, another thioester derivative of a small molecule nuclease could be use instead of EDTA-3MPA. Like Fe^{2+}-EDTA, (Cu^{2+})-1,10-phenanthroline is a nuclease reagent that may be activated by H_2O_2 to produce hydroxyl radicals (Sigman *et al.*, 1991). Bruice & coworkers reported that the nuclease activity of (Cu^{2+})-1,10-phenanthroline is very hardy under most conditions that would inhibit other nucleases (*ibid.*). In particular, it is not inhibited by high glycerol concentration. A thioester derivative of 1,10-phenanthroline could therefore be used to footprint Δ1-14 and Δ1-94 on the pre-mRNA, both of which could not be used in the experiments with EDTA-3MPA due to the lower concentrations of the protein stocks available.

8. Protein Cross-linking of Benzophenone-Derivatized U2AF^{65} Mutants to U2AF^{35}

Benzophenone deriviatives of full-length U2AF^{65} and Δ1-64 were tested for their ability to cross-link to U2AF^{35}. This was to determine the activity of the benzophenone reagent in a system where the interaction of the two proteins has already been characterized to some extent. The method of detection of the cross-linked products was a chemiluminescent Western blot using anti-U2AF^{65} and anti-U2AF^{35} antibodies from mouse sera (a kind gift from Tom Maniatis, Harvard University). From initial tests, it was apparent that the antibody solutions were of extremely low titre. This required a large amount of protein to be used in the assay. Also, the anti-U2AF^{35} antibody was rather non-specific, binding to most of the proteins present in the assay. LF1 RNA is added as a substrate for the binding of the U2AF heterodimer. The conditions of the U2AF^{65}/U2AF^{35}/RNA native gel mobility shift assay were followed, to ensure that all of the proteins were in the heterodimer-RNA complex. The reactions were first allowed to equilibrate at room temperature and then transferred to sample wells embedded in ice, to slow the tumbling of the proteins in solution. The cross-linking reaction was initiated by exposure to 350 nm ultraviolet radiation, delivered in a Rayonet photochemical reactor. The Rayonet reactor tended to heat up significantly during the course of the reaction. In order to maintain the low temperature, the reactor was placed in a cold room and the reactions were kept on ice inside the reactor.

As mentioned earlier, amino acids 64-182 of U2AF^{65} are believed to be involved in interaction with U2AF^{35} (Zhang *et al.*, 1992). The identification of these amino acids was done by engineering a series of deletion mutants of U2AF^{65} and testing their ability to bind U2AF^{35}. In this assay, the benzophenone reagent was present at amino acid 1 in the full-length U2AF^{65} and at amino acid 65 in Δ1-64. The radius of benzophenone's activity is approximately 3 Å, centred on the ketone oxygen (Dormán & Prestwich, 1994). The radius of reactivity is also dependent on the length and flexibility of the linker between the benzophenone moiety and the protein. The linker between benzophenone and the first amino acid of the protein is an amide bond, which is short and inflexible. This would likely prevent it from reacting with amino acid residues that were not in the immediate vicinity.

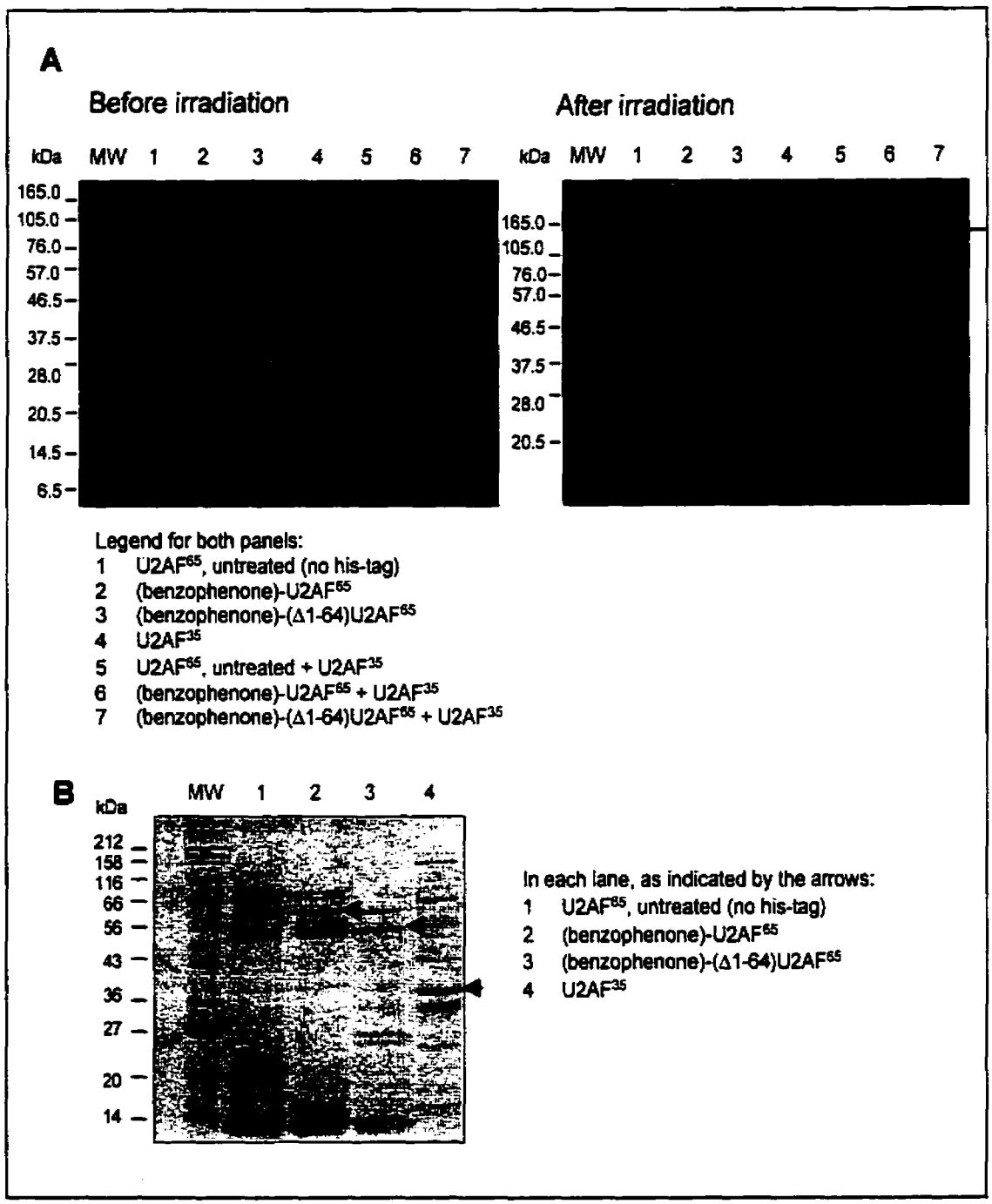


Figure 18: Western blot of benzophenone-derivatized full-length U2AF^{65}, (Δ1-64)U2AF^{65}, with U2AF^{35} after 350 nm irradiation (1 h.). (A) Detection with mouse anti-U2AF^{65} antibody. The appearance of the new band after irradiation in the reaction containing (benzophenone)-(Δ1-64)U2AF^{65} + U2AF^{35} is as indicated with the arrow. It has an approximate molecular weight slightly greater than 105 kDa. (B) SDS-polyacrylamide gel showing the same sample set 1-4 as in (A), before irradiation.

Considering the previous biochemical data and the distance constraint on the reactivity of the benzophenone moiety, it was thought that perhaps (benzophenone)-(full-length)U2AF^{65} would not cross-link to U2AF^{35}. Although Δ1-64 did not appear to bind U2AF^{35} in the gel shift assay, it was thought possible that it could still bind at low levels given the circumstantial evidence in the splicing reconstitution assay. (Benzophenone)-(Δ1-64)U2AF^{65} was thought to be more likely to cross-link to U2AF^{35}, since the truncation would place the benzophenone group at the border of the site of interaction between the two subunits. Even if the binding is weak between Δ1-64 and U2AF^{35}, the possibility for cross-linking remains, given that benzophenone may be excited repeatedly during exposure to the U.V. radiation. The samples were therefore irradiated for an hour, which was felt to be a sufficient amount of time for cross-linking to occur. Δ1-94 could not be tested in this assay because the concentration of the stock solution was not high enough, due to a lower level of expression compared to Δ1-64. In the Western blot, a new band with a mobility slight greater than ~105 kDa was observed in the photo-irradiated sample containing benzophenone-(Δ1-64) and U2AF^{35} (lane 7, After Irradiation, Fig. 18A). A band running at ~106 kDa is not apparent in the other samples containing other forms of the U2AF heterodimer (lane 5: underivatized full-length U2AF^{65} + U2AF^{35}; lane 6: (benzophenone)-U2AF^{65} + U2AF^{35}). The amount of cross-linked product was quite low, as indicated by the very low intensity of the band in the auto-exposure of the Western blot, despite using the highest amount of protein possible. The amount of cross-linked product was too low to be detected by Coomassie Blue staining of the SDS-polyacrylamide gel; a distinct band which appears to represent a new cross-linked product was not distinguishable from background bands in a silver-stained SDS-polyacrylamide gel (gel not shown).

A cross-link between benzophenone-(Δ1-64)U2AF^{65} and U2AF^{35} appears to form, in agreement with the prediction made earlier that the N-terminus of (Δ1-64)U2AF^{65} is closer to the site of interaction with U2AF^{35} than the N-terminus of full-length U2AF^{65}. The Western blot assay to detect cross-linked products in this minimal system was simple and is likely most useful for systems involving a few different proteins which must be present in sufficiently high concentrations for the yield of cross-linked product to be observable in the

blot. To improve the yield of the cross-linked product and thus observe a stronger signal in the Western blot, higher concentrations of the proteins (the $U2AF^{65}$ mutants and $U2AF^{35}$) should be obtained.

The Western blot assay to detect cross-links between the benzophenone derivatized $U2AF^{65}$ and other proteins may be applied to other circumstances. The interaction of poly(A) polymerase (PAP) with the N-terminus of $U2AF^{65}$ should be detectable as a cross-linked product in the Western blot assay using the benzophenone-derivatized U2AF proteins. The Δ1-14 protein would most likely cross-link to PAP, since the benzophenone moiety would be very near the proposed site of interaction with PAP.

Another future experiment with the benzophenone-derivatized $U2AF^{65}$ proteins would be an anti-U2AF immunoprecipitation of cross-linked spliceosome complexes from U2AF-depleted nuclear extract with added benzophenone-$U2AF^{65}$. This experiment would help identify other splicing factors that the RS domain of $U2AF^{65}$ may come in contact with at different time points during the splicing process. Therefore, a factor in this experiment would be the amount of time allowed between equilibration of the nuclear extract and the photo-initiated cross-linking. The assembly of the spliceosome components occurs in a dynamic fashion, with snRNPs and other protein factors entering and leaving the spliceosome at different times during the course of splicing (MacMillan *et al.*, 1994). U2AF enters and leaves the spliceosome at the very earliest stages of spliceosome formation. Another factor that could affect the cross-links made is the presence of ATP since some of the spliceosome complexes formed are ATP-dependent (Moore *et al.*, 1993; MacMillan *et al.*, 1994). The cross-links obtained using benzophenone-$U2AF^{65}$ proteins could be compared against cross-links obtained using a bi-functional protein cross-linker that may be added exogenously to a protein complex. Examples of bi-functional cross-linking reagents are the *bis*-N-hydroxysuccinimidyl esters (Wong, 1993). Protein cross-linking using an exogenous protein cross-linker is less specific than the cross-links observed with a protein with a site-specifically attached benzophenone moiety; the use of such a reagent may help to identify all other proteins coming into contact with $U2AF^{65}$.

9. Future Experiments

Synthesis of a thioester derivative of a small nuclease reagent that is hardy under most experimental conditions, such as (Cu^{2+})-1,10-phenanthroline, would by highly useful as a replacement for EDTA-3MPA. The activity of Fe^{2+}-EDTA is inhibited by high concentrations of common buffers such as PIPES, and the presence of glycerol in the reaction mixture (Dixon *et al.*, 1991). Using a thioester derivative nuclease reagent such as (Cu^{2+})-1,10-phenanthroline would allow one to use higher concentrations of $U2AF^{65}$ and associated proteins, since there would be no restriction on the level of glycerol present. This would allow for a higher level of protein-RNA complex to be present and a stronger cleavage "signal" would be detected.

As mentioned earlier, a hydroxyl radical cleavage experiment should be attempted with the $U2AF^{65}$ mutants, derivatized with a nuclease thioester reagent (either EDTA-3MPA or a 1,10-phenanthroline thioester), in the context of the spliceosome, in nuclear extract. Anti-U2AF antibody may be used to precipitate complexes containing $U2AF^{65}$ from the reaction mixture. Hydroxyl radical cleavage could be initiated at various time points to study the movement of $U2AF^{65}$ through early spliceosome complexes. Reed & coworkers have postulated the ATP-dependent formation of an "E* complex" (following E complex, just prior to A complex formation), in which $U2AF^{65}$ binding to the polypyrimidine tract is destabilized by phosphorylation by an unknown kinase activity (Champion-Arnaud *et al.*, 1995). If the phosphorylation of $U2AF^{65}$ is responsible for binding destabilization, the addition of ATP to a reaction mixture, containing E complexes, would therefore trigger this change in binding affinity, and a corresponding weakening of cleavage signal should be observable after ATP addition. Furthermore, the $U2AF^{65}$ RS domain is the most likely site of phosphorylation. The activity of some SR proteins is regulated by phosphorylation of serine residues found in their RS domains by SR protein kinases (Manley & Tacke, 1996). These SR protein kinases may also phosphorylate U2AF (Gui *et al.*, 1994a; Gui *et al.*, 1994b). Therefore, the RS-deletion $U2AF^{65}$ mutants Δ1-64 and Δ1-94 should not be susceptible to phosphorylation-dependent destabilization of RNA binding. This lack of binding

destabilization should be apparent in a hydroxyl radical cleavage experiment involving nuclease-derivatized Δ1-64 and Δ1-94 in nuclear extract, after ATP addition.

A series of minimal experiments to study E complex formation could also be carried out, using nuclease-derivatized U2AF^{65} mutants plus 1-3 proteins which U2AF^{65} is known to interact with, such as SF1/mBBP and SAP 155, to observe the effect on the positioning of its RS domain with respect to the pre-mRNA.

In studying the protein-protein interactions of the U2AF^{65} RS domain, it may be more useful to synthesize a thioester derivative of benzophenone with a longer, more flexible linker between the thioester group and the benzophenone group. It was felt that benzophenone-3MPA was too specific a reagent since only a low level of cross-linking was obtainable between (benzophenone)-(Δ1-64)U2AF^{65} and U2AF^{35}. A longer, flexible tether between the protein and benzophenone would detect a wider range of protein contacts.

As discussed earlier, the benzophenone-derivatized U2AF^{65} mutants should be subjected to cross-linking in U2AF-depleted nuclear extract to identify other splicing factors the RS domain may be coming in contact with, particularly in the presence of ATP. In the formation of the postulated ATP-dependent E* complex, the U2AF^{65} RS domain should come in contact with a kinase which has yet to be identified. The ATP dependence of U2AF^{65} interactions with other proteins within the E complex may be studied by comparing cross-linked products obtained before and after the addition of ATP. If a kinase phosphorylates U2AF^{65} in the E complex, it should no interact with the RS deletion mutants Δ1-64 and Δ1-94. If full-length U2AF^{65} is found to cross-link to a protein only after addition of ATP, and the same does not occur for Δ1-64 and Δ1-94, then this protein may possibly be the unknown kinase postulated by Reed & coworkers (Champion-Arnaud *et al.*, 1995).

The N-terminal cysteine truncation mutants of U2AF^{65} may be chemically ligated to peptides with C-terminal thioester groups, which have been made by solid phase synthesis (*cf.* Fig. 5A, Section A). The purpose of this is to embed a reporter group within the context of the whole U2AF^{65} protein, instead of being confined to placing reporter groups at the N-

termini of the truncation mutants. The U2AF^{65} N-terminal truncations go up to amino acid 94; solid phase peptide synthesis allows the synthesis of peptides of up to 100 amino acids long (Merrifield, 1965). Dawson & coworkers have developed a solid phase resin support that allows the synthesis of a peptide which will have a thioester at its C-terminus upon cleavage from the support (Hackeng *et al.*, 1999); such peptides may then be used in a thioester-cysteine ligation to a protein with an N-terminal cysteine. Solid phase synthesis would allow the insertion of an unnatural amino acid, which contains a reporter group as its residue, into the peptide sequence. This reporter group may be an activatable nuclease reagent such as (Cu^{2+})-1,10-phenanthroline or a protein cross-linking reagent such as benzophenone. This peptide ligation method would allow one to insert a reporter group at any point within the RS domain or within the loop between the RS domain and the first RRM of U2AF^{65}. The ligated U2AF^{65} protein should function like the wild type U2AF^{65}, given the many examples in the literature of functional proteins produced by chemical ligation (reviewed in Dawson & Kent, 2000). As discussed earlier in Section C.7, it is possible that the N-terminal truncation mutants have a loosely defined structure due to the disruption (by truncation) of what would normally be a compactly folded domain. Therefore, there may be an advantage to inserting the reporter group within the protein, without disturbing the domain structure. The series of U2AF^{65} N-terminal cysteine mutants that was created here would allow such experiments to be done.

A final set of experiments that may be carried out on U2AF^{65} in the future involve the study of the individual RNA binding domains (RRMs). All three RRMs are required for high affinity binding of U2AF^{65} to the polypyrimidine tract. The thioester reagent strategy may be used here as well. The three binding domains may be cloned separately or in sets (*e.g.* RRM1-RRM2, RRM2-RRM3), with the Factor Xa cleavage site followed by cysteine engineered into the N-termini of each of these peptides. This would allow the attachment of a nuclease thioester reagent to the N-terminus of the RRM peptides. The attachment of a nuclease reagent could give information about the placement of these peptides on the PPT and how the binding of each RRM affects the other RRMs.

The solution NMR structures of U2AF^{65}'s RRM1 and RRM2 were elucidated, giving an idea of the RNA binding surfaces of each module (Ito *et al.*, 1999). The structure of U2AF^{65}'s RRM3 has yet to be determined. RRM3 in particular deserves attention. RRM3 is not as necessary for RNA binding as RRM1 or RRM2 (loss of RRM3 from U2AF^{65} results in a 10^1-10^2 diminution of binding affinity; Zamore *et al.*, 1992). However, U2AF^{65}'s RRM3 has been found to interact with SF1/mBBP (Berglund *et al.*, 1998; Rain *et al.*, 1998) and SAP 155, a protein associated with U2 snRNP (Gozani *et al.*, 1998). In the E complex, SF1/mBBP binds to the branchpoint region while U2AF binds to the polypyrimidine tract; the two proteins bind cooperatively to the pre-mRNA (Berglund *et al.*, 1998). Reed & coworkers postulated that SAP 155 displaces SF1/mBBP from the branchpoint sequence and from its interaction with U2AF^{65} during the formation of A complex, thus acting as the vehicle for U2 snRNP recruitment to the branchpoint region of the pre-mRNA (Gozani *et al.*, 1998). The cloning of U2AF^{65}'s RRM3 as suggested above may be useful in NMR or X-ray crystallography studies, to identify the RNA binding surfaces and the sites of protein interaction within this domain.

Conclusion

The purpose of this research was to employ compounds that would react with and modify proteins, thus allowing the experimenter to probe protein structure and function. In Part I and II, a specific family of compounds was chosen, the thiol reagents, and ion channels as the family of proteins to be investigated. The choice of thiol reagents was based on the reactivity of cysteine's thiol side chain as a nucleophile and the advent of the MTS reagents, which offers selective reactivity with thiol residues. Ion channels are complex proteins and little is known about their structure and how they function, despite their essential function within nerve function. The development of new tools to study them is required; indirect methods such as the use of thiol reagents to modify channels and observation of subsequent behaviour have led to a large portion of the current information on ion channel structure and function.

In Part I, the design, synthesis and testing of a novel MTS reagent was described. MTSAC was created as a reagent to be used in scanning cysteine mutagenesis experiments, to identify residues within or near the pore of an ion channel. MTSAC incorporates three chemical groups: (1) an MTS group which confers site-specific reaction to cysteine residues, (2) a charged amino group which will block ion flow and (3) a carbamate bond between the MTS group and the amino group. The isomerization of the carbamate bond within MTSAC changes the position of the charged amino group at the terminus of MTSAC; the charged amino group transiently blocks ion flow through the channel. If MTSAC is positioned near the pore of the channel, its isomerization is detectable as small steps in current across the open channel. The time between each current step corresponds to the timescale for the carbamate bond isomerization. MTSAC was tested in a model ion channel, gramicidin. MTSAC would be useful in cases where the ion channel being investigated opens and closes at a rate slower than the isomerization of MTSAC. If the ion channel to be surdied meets this criterion and current steps are observed (similar to those seen in gramicidin'-S-MTSAC), it would be a definitive indication that MTSAC was indeed covalently linked to the ion channel, and in a position near the pore region of the channel. Annother important aspect of MTSAC was that its effect on ion flow was correlated with its distance from the pore of gramicidin. If this effect holds true in tests on other ion channels, then MTSAC would be useful as a qualitative indicator of distance from the channel's pore region.

In Part II, the design and synthesis of a series of thiol-reactive analogs of lidocaine was described. Lidocaine is a local anesthetic drug that reversibly blocks sodium ion channels. It is an important drug in the treatment of cardiac arrhythmia. It is also commonly used in electrphysiological studies of sodium ion channels to study the structure of the pore region and the gating mechanisms of the channel. A thiol-reactive version of lidocaine and other local anesthetics would be very useful reagents in the application of scanning cysteine mutagenesis experiments on the sodium ion channel. This is because the local anesthetic binding site in the sodium channel is found within the pore region of the channel below the selectivity filter. Scanning cysteine mutagenesis of residues thought to line the pore would allow these thiol-reactive local anesthetics to be anchored to these positions, and the proximity of these residues to the local anesthetic binding site may be determined by observing sodium channel block. From previous electrophysiological studies, two major functional domains of the sodium channel, the selectivity filter and the inactivation gate, appear to be near the local anesthetic binding site (Balser, 1999). Backx & coworkers created a series of MTS derivatives of benzocaine, with an alkyl ether linking chain (0-6 atoms long) between the MTS group and the benzocaine group (Li *et al.*, 1999). Benzocaine is uncharged at physiological pH and may only enter the channel through the postulated hydrophobic pathway (Hille, 1992). Also, it is not a long-acting local anesthetic; it has lower affinity for the local anesthetic binding site compared to another drug such as lidocaine. It may be useful to create thiol-reactive analogs of other local anesthetics which have different properties from benzocaine. In studies to determine the proximity of residues thought to be within the selectivity filter or the inactivation gate to the local anesthetic binding site, it may be useful to create an MTS derivative of a local anesthetic that is capable of entering the channel via the hydrophilic pathway. In this instance, lidocaine was chosen to be the starting point in the creation of a new series of thiol-reactive local anesthetic reagents to study ion channel structure. In contrast to benzocaine, it is a longer acting anesthetic, and it may be charged at physiological pH and therefore enter the channel via the postulated hydrophilic pathway (Hille, 1992).

A series of halogenated versions of lidocaine was synthesized and the conversion of one of these compounds to the MTS version was accomplished. These lidocaine analogs may be used in future experiments involving scanning cysteine mutagenesis on sodium channel, to identify residues near the local anesthetic binding site. In contrast to lidocaine itself, the

thiol-reactive lidocaine analogs should cause long-acting block of the channels. The "propyl-MTS" derivative of lidocaine, DMPA-N(Et)(PrMTS), has the advantage of selective reaction to cysteine residues and it is expected that this compound should only cause irreversible block in the cardiac sodium channel which contains a cysteine residue near the pore region and not the skeletal sodium channel, which has no free cysteine residue near the pore region. The other halogenated versions of lidocaine may not be as selective in which residue they will react with, so irreversible block could also be observed with skeletal sodium channels. In future experiments, it is hoped that new MTS derivatives of lidocaine may be synthesized, so that the effect of varying chain length and linker characteristics (*e.g.* using unsaturated bonds to decrease chain flexibility) between the MTS group and the lidocaine portion of the molecule on channel blockage may be explored.

In Part III, the essential RNA splicing factor U2AF was studied. Chemical probes were attached to the large subunit of U2AF, $U2AF^{65}$. Kent's strategy of linking a peptide ending in a thioester to another peptide with an N-terminal cysteine was modified to become a generally useful method of attachment of a thioester derivative of a reporter group to a protein with an N-terminal cysteine. From previous biochemical studies, the RS domain of $U2AF^{65}$ was shown to play an extremely important role in the activity of $U2AF^{65}$. Since the RS domain occurs near the N-terminus of $U2AF^{65}$, it was an ideal candidate for the use of the thioester reagent strategy. A series of N-terminal cysteine mutants of $U2AF^{65}$ were engineered, cloned and purified from transfected *E. coli* cultures. Along with EDTA-3MPA, a previously synthesized thioester derivative of a nuclease reagent (Erlanson *et al.*, 1996), benzophenone-3MPA and biotin-3MPA were also synthesized. Biotin-3MPA was shown to be very useful in identifying the presence of the N-terminal cysteine in the $U2AF^{65}$ mutants, and it was also used for detecting a prior reaction of the N-terminal cysteine to another thioester reagent. The attachment of EDTA-3MPA was to allow the site-specific attachment of the Fe^{2+}-EDTA nuclease reagent to the N-terminus of the $U2AF^{65}$ mutants. Tethering a nuclease reagent to the N-terminus of the $U2AF^{65}$ RS domain would thus allow one to monitor its position relative to the pre-mRNA. The EDTA-derivatized full-length $U2AF^{65}$ mutant generated a site-specific cleavage pattern on the substrate pre-mRNA that suggested the RS domain was in close proximity to the branchpoint adenosine, in agreement with an earlier finding by Green & coworkers (Valcárcel *et al.*, 1996). The EDTA-derivatized ΔRS deletion mutant, Δ1-64, did not appear to generate any site-specific cleavages on the

substrate pre-mRNA, suggesting that this portion of the U2AF^{65} protein may be to far from the RNA to have an effect. Fe^{2+}-EDTA was found to be easily inhibited by the reaction conditions required for the hydroxyl radical footprinting experiments. The synthesis of a thioester derivative of a more hardy nuclease reagent such as (Cu^{2+})-1,10-phenanthroline would likely aid in further footprinting studies of U2AF^{65} on pre-mRNA. The benzophenone derivatives of full-length and Δ1-64 U2AF^{65} were used in a cross-linking study with U2AF^{35}. Δ1-64 cross-linked at a very low level to U2AF^{35}, while the full-length U2AF^{65} did not appear to cross-link to U2AF^{35}. This suggested that benzophenone-3MPA is a very specific reagent which will only cross-link to another protein in the immediate vicinity. It may be more useful to design and synthesize a benzophenone thioester reagent with a long, flexible tethering group between the benzophenone group and the site of attachment to the protein. Future studies of U2AF^{65} should involve the nuclease or benzophenone derivatized U2AF^{65} mutants to obtain more information about the changes that take place during early spliceosome complex formation (*i.e.* the transition from E to A complex) and the possibility that the phosphorylation of the RS domain affects U2AF binding affinity, thus regulating its activity. Future studies of U2AF^{65} should also focus on the RNA binding domains (RRMs). The same thioester reagent strategy could be applied to cloned peptides representing the RRMs. This strategy would allow the site-specific attachment of a nuclease reagent to study how the binding of each domain to the pre-mRNA affects the other domains.

References

Akabas, M. H.; Stauffer, D. A.; Xu, M.; Karlin, A. "Acetylcholine receptor channel structure probed in cysteine-substituted mutants." (1992) *Science* **258**, 307-10.

Abovich, N.; Liao, X. C.; Rosbash, M. "The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition." (1994) *Genes Dev.* **8**, 843-854.

Barrett, J. E.; Lucero, C. M.; Schultz, P. G. "A model for hydride transfer in thymidylate synthase based on unnatural amino acid mutagenesis." (1999) *J. Am. Chem. Soc.* **121**, 7965-7966.

Beck, D. L.; Stump, W. T.; Hall, K. B. "Defining the orientation of the human U1A RBD1 on its UTR by tethered-EDTA(Fe) cleavage." (1998) *RNA* **4**, 331-339.

Berglund, J. A.; Abovich, N.; Rosbash, M. "A cooperative interaction between U2AF^{65} and mBBP/SF1 facilitates branchpoint region recognition." (1998) *Genes & Development* **12**, 858-867.

Blencowe, B. J.; Bowman, J. A. L.; McCracken, S.; Rosonina, E. "SR-related proteins and the processing of messenger RNA precursors." (1999) *Biochem. Cell. Biol.* **77**, 277-291.

Burge, C. B.; Tuschl, T.; Sharp, P. A. "Splicing of Precursors to mRNAs by the Spliceosomes" (1999) in *The RNA World 2nd ed.*, eds. Gesteland, R. F.; Cech, T. R. and Atkins, J. F. (Cold Spring Harbor Laboratory Press, Plainview, New York), pp. 525-560.

Champion-Arnaud, P.; Gozani, O.; Palandjian, L.; Reed, R. "Accumulation of a novel spliceosomal complex on pre-mRNAs containing branch site mutations." (1995) *Mol. Cell. Biol.* **15** (10), 5750-5756.

Conte, M. R.; Grüne, T.; Ghuman, J.; Kelly, G.; Ladas, A.; Matthews, S.; Curry, S. "Structure of tandem RNA recognition motifs from polypryimidine tract binding protein reveals novel features of the RRM fold." (2000) *EMBO J.* **19** (12), 3132-3141.

Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. "Synthesis of proteins by native chemical ligation." (1994) *Science* **266**, 776-779.

Dawson, P. E.; Churchill, M. J.; Ghadiri, M. R.; Kent, S. B. H. "Modulation of reactivity in native chemical ligation through the use of thiol additives." (1997) *J. Am. Chem. Soc.* **119** (9), 4325-4329.

Dawson, P. E.; Kent, S. B. H. "Synthesis of native proteins by chemical ligation." (2000) *Annu. Rev. Biochem.* **69**, 923-960.

Dignam, J. D.; Lebovitz. R. M.; Roeder, R. G. "Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei." (1983) *Nucleic Acids Research* **11** (5), 1475-1489.

Dixon, W. J.; Hayes, J. J.; Levin, J. R.; Weidner, M. F.; Dombroski, B. A.; Tullius, T. D. "Hydroxyl radical footprinting." (1991) *Methods in Enzymology* **208**, 380-413.

Dormán, G. and Prestwich, G. D. "Benzophenone photophores in biochemistry." (1994) *Biochemistry* **33** (19), 5661-5672.

Doyle, D. A.; Morais Cabral, J.; Pfuetzner, R. A.; Kuo, A,; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. "The structure of the potassium channel: Molecular basis of K^+ conduction." (1998) *Science* **280**, 69-77.

Erlanson, D. A.; Chytil, M.; Verdine, G. L. "The leucine zipper domain controls the orientation of AP-1 in the NFAT•AP-1•DNA complex." (1996) *Chemistry & Biology* **3** (12), 981-991.

Ellman, J.; Mendel, D.; Anthony-Cahill, S.; Noren, C. J.; Schultz, P. G. "Biosynthetic method for introducing unnatural amino acids site-specifically into proteins." (1991) *Methods in Enzymology* **202**, 301-336.

Fleckner, J.; Zhang, M.; Valcárcel, J.; Green, M. R. "U2AF^{65} recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction." (1997) *Genes Dev.* **11**, 1864-1872.

Gui, J.-F.; Lane, W.; Fu, X. D. "A serine kinase regulates intracellular localization of splicing factors in the cell cycle." (1994a) *Nature* **369**, 678-683.

Gui, J. F.; Tronchère, H.; Chandler, S. D.; Fu, X. D. "Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors." (1994b) *Proc. Natl. Acad. Sci. USA* **91**, 10824-10828.

Gutte, B.; Merrifield, R. B. "The total synthesis of an enzyme with ribonuclease A activity." (1969) *J. Am. Chem. Soc.* **91**, 501-2.

Gozani, O.; Potashkin, J.; Reed, R. "A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site." (1998) *Mol. Cell. Biol.* **18** (8), 4752-4760.

Grabowski, P. J.; Padgett, R. A.; Sharp, P. A. "Messenger RNA splicing *in vitro*: an excised intervening sequence and a potential intermediate." (1984) *Cell* **37**, 415-427.

Gunderson, S. I.; Vagner, S.; Polycarpou-Schwarz, M.; Mattaj, I. W. "Involvement of the carboxyl terminus of vertebrate poly(A) polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation." (1997) *Genes Dev.* **11**, 761-773.

Ito, T.; Muto, Y.; Green, M. R.; Yokoyama, S. "Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF^{65})." (1999) *EMBO J.* **18** (16), 4523-4534.

Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. "Protein synthesis by native chemical ligation: Expanded scope by using straightforward methodology." (1999) *Proc. Natl. Acad. Sci. USA* **96**, 10068-10073.

Hall, K. B. and Fox, R. O. "Directed cleavage of RNA with protein-tethered EDTA-Fe." (1999) *Methods: A Companion to Methods in Enzymology* **18**, 78-84.

Kanaar, R.; Roche, S. E.; Beall, E. L.; Green, M. R.; Rio, D. R. "The conserved pre-mRNA splicing factor U2AF from *Drosophila*: Requirement for viability." (1993) *Science* **262**, 569-573.

Kramer, A. "The structure and function of proteins involved in mammalian pre-mRNA splicing." (1996) *Annu. Rev. Biochem.* **65**, 367-409.

Lee, C.-G.; Zamore, P. D.; Green, M. R.; Hurwitz, J. "RNA annealing activity is intrinsically associated with U2AF." (1993) *J. Biol. Chem.* **268**, 13472-13478.

Liu, C.-F.; Tam, J. P. "Peptide segment ligation strategy without use of protecting groups." (1994) *Proc. Natl. Acad. Sci. USA* **9**, 6584-6588.

MacMillan, A. M.; Query, C. Q.; Allerson, C. R.; Chen, S.; Verdine, G. L.; Sharp, P. A. "Dynamic association of proteins with the pre-mRNA branch region." (1994) *Genes Dev.* **8**, 3008-3020.

MacMillan, A. M.; McCaw, P. S.; Crispino, J. D.; Sharp, P. A. "SC35-mediated reconstitution of splicing in U2AF-depleted nuclear extract." (1997) *Proc. Natl. Acad. Sci. USA* **94**, 133-136.

Maniatis, T.; Firtsch, E. F.; Sambrook, J. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, 1982.

Manley, J. L. and Tacke, R. "SR proteins and splicing control." (1996) *Genes Dev.* **10**, 1569-1579.

Martell, A. E. "Chelates of Ascorbic Acid: Formation and Catalytic Properties" (1982) in *Ascorbic Acid: Chemistry, Metabolism and Uses*, eds. Seib, P. A. and Tolbert B. M. (A.C.S. Publications, Washington, D.C.).

Merendino, L.; Guth, S.; Bilbao, D.; Martinez, C.; Valcárcel, J. "Inhibition of *msl-2* splicing by Sex-lethal reveals interaction between U2AF^{35} and the 3' splice site AG." (1999) *Nature* **402**, 838-841.

Merrifield, R. B. "Automated synthesis of peptides." (1965) *Science 150*, 178-185.

Misteli, T. and Spector, D. L. "Protein phosphorylation and the nuclear organization of pre-mRNA splicing." (1997) *Trends Cell Biol.* 7, 135-138.

Moore, M. J.; Query, C. C.; Sharp, P. A. "Splicing of precursors to mRNA by the spliceosome" (1993) in *The RNA World*, eds. Gesteland, R. F. and Atkins, J. F. (Cold Spring Harbor Laboratory Press, Plainview, New York), pp.303-357.

Moore, M. J. "Intron recognition comes of AGe." (2000) *Nature Struct. Biol.* 7 (1), 14-16.

Muir, T. W.; Sondhi, D.; Cole, P. A. "Expressed protein ligation: a general method for protein engineering." (1998) *Proc. Natl. Acad. Sci. USA* **95**, 6705-6710.

Nagai, K.; Oubridge, C.; Ito, N.; Avis, J.; Evans, P. "The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA." (1995) *TIBS* **20**, 235-240.

Pogelzelski, W. K. and Tullius T. D. "Oxidative strands scission of nucleic acids: Routes initiated by hydrogen abstraction from the sugar moiety." (1998) *Chemical Reviews* **98**, 1089-1107.

Potashkin, J.; Naik, K.; Wentz-Hunter, K. "U2AF homolog required for splicing in vivo." (1993) *Science* **262**, 573-575.

Ptashne, M. and Gann, A. F. "Activators and targets." (1990) *Nature* **346**, 329-331.

Puglisi, J. D.; Blanchard, S. C.; Green, R. "Approaching translation at atomic resolution." (2000) *Nature Struct. Biol.* 7 (10), 855-861.

Query, C. C.; Moore, M. J.; Sharp, P. A. "Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model." (1994) *Genes Dev.* **8**, 587-597.

Rain, J.-C.; Rafi, Z.; Rhani, Z.; Legrain, P.; Kramer, A. "Conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1." (1998) *RNA* **4**, 551-565.

Reed, R. "The organization of 3' splice-site sequences in mammalian introns." (1989) *Genes Dev.* **3**. 2113-2123.

Rymond, B. C.; Rosbash, M. "Cleavage of 5' splice site and lariate formaion are independent of 3' splice site splicing." (1985) *Nature* **317**, 735-7.

Sigman, D. S.; Kuwabara, M. D.; Chen, C.-H. B.; Bruice, T. W. "Nuclease activity of 1,10-phenanthroline copper in study of protein-DNA interactions." (1991) *Methods in Enzymology* **208**, 414-433.

Singh, R.; Valcárcel, J.; Green, M. R. "Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins." (1995) *Science* **268**, 1173-1176.

Stodjl, D. F.; Bell, J. C. "SR protein kinases: the splice of life." (1999) *Biochem. Cell. Biol.* **77**, 298-298.

Tacke, R.; Chen, Y.; Manley, J. L. "Sequence–specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer." (1997) *Proc. Natl. Acad. Sci. USA.* **94**, 1148-1153.

Tam, J. P.; Yu, Q.; Miao, Z. "Orthogonal ligation strategies for peptide and protein." (1999) *Biopolymers* **51**, 311-332.

Tullius, T. D. and Dombroski, B. A. "Hydroxyl radical "footprinting": High-resolution information about DNA-protein contacts and application to λ repressor." (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5469-5473.

Vagner, S.; Vagner, C. Mattaj, I. W. "The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF^{65} to couple 3'-end processing and splicing." (2000) *Genes Dev.* **14**, 403-413.

Valcárcel, J.; Gaur, R. K.; Singh, R.; Green, M. R. "Interaction of U2AF^{65} SR region with pre-mRNA of branch point and promotion base pairing with U2 snRNA." (1996) *Science* **273**, 1706-1709.

Walling, C. *Free Radicals in Solution.* (1957) John Wiley & Sons, New York.

Wang, J.; Dong, Z.; Bell, L. R. "Sex-lethal interactions with protein and RNA: Roles of glycine-rich and RNA binding domains." (1997) *J. Biol. Chem.* **35**, 22227-22235.

Wang, X. and Padgett, R. A. "Hydroxyl radical "footprinting" of RNA: Application to pre-mRNA splicing complexes." (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7795-7799.

Wieland, T.; Schneider, G. (1953) *Liebigs. Ann. Chem.***583**, 159.

Wong, S. S. *Chemistry of Protein Conjugation and Cross-linking*. Boca Raton, Florida: CRC Press, 1993.

Wu, J. Y. and Maniatis, T. "Specific interactions between proteins implicated in splice site selection and regulated alternative splicing." (1993) *Cell* **75**, 1061-1070.

Wu, S.; Romfo, C. M.; Nilsen, T. W.; Green, M. R. "Functional recognition of the 3' splice site AG by the splicing factor U2AF^{35}." (1999) *Nature* **402**, 832-835.

Xiao, S.-H. and Manley, J. L. "Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing." (1997) *Genes Dev.* **11**, 334-344.

Xiao, S.-H. and Manley, J. L. "Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2." (1998) *EMBO J.* **17**, 6359-6367.

Yeakley, J. M.; Tronchère, H.; Olesen, J.; Dyck, J. A.; Wang, H.-Y.; Fu, X.-D. "Phosphorylation regulates in vivo interaction and molecular targeting of serine-arginine-rich pre-mRNA splicing factors." (1999) *J. Cell Biol.* **145**, 447-455.

Zamore, P. D. and Green, M. R. "Identification, purification and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor." (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9243-9247.

Zamore, P. D. and Green, M. R. "Biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distrubution." (1991) *EMBO J.* **10**, 207-214.

Zamore, P. D.; Patton, J. G.; Green, M. R.. "Cloning and domain structure of the mammalian splicing factor U2AF." (1992) *Nature* **355**, 609-614.

Zhang, M.; Zamore, P. D., Carmo-Fonseca, M.; Lamond, A. I.; Green, M. R. "Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit." (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8969-8773.

Zorio, D. A. R.; Lea, K.; Blumenthal. T. "Cloning of *Caenorhabditis* U2AF^{65}: an alternative spliced RNA containing a novel exon." (1997) *Mol. Cell. Biol.* **17**, 946-953.

Zorbas, H.; Rogge, L.; Meisterernst, M.; Winnaker, E.-L. "Hydroxyl radical footprints reveal novel structural features around the NF I binding site in adenovirus DNA." (1989) *Nucleic Acids Res.* **17**, 7735.

Zorio, D. A. R.; Blumenthal, T. "Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans.*" (1999) *Nature* **402**, 835-838.

Zuo, P. and Maniatis, T. "The splicing factor U2AF^{35} mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing." (1996) *Genes Dev.* **10**, 1356-1368.