# thermoscientific

# In-Depth Characterization of Intact Protein Standards Using HRAM Top Down Mass Spectrometry with Multiple MSMS Strategies

Helene Cardasis<sup>1</sup>, Romain Huguet<sup>1</sup>, Chris Mullen<sup>1</sup>, Stephane Houel<sup>1</sup>, Luca Forneli<sup>2</sup>, Rosa Viner<sup>1</sup>, Viktorija Vitkovske<sup>1</sup>, Shanhua Lin<sup>1</sup>, Seema Sharma<sup>1</sup>, Hongqian Yang<sup>3</sup>, Vlad Zabrouskov<sup>1</sup>, Neil Kelleher<sup>2</sup> <sup>1</sup> Thermo Fisher Scientific, San Jose, CA USA; <sup>2</sup> Northwestern University, Chicago, IL USA; <sup>3</sup> Thermo Fisher Scientific, Stockholm, Sweden

## **ABSTRACT**

Purpose: Demonstrate unique characteristics and effectiveness of various dissociation mechanisms for intact protein identification and characterization.

Methods: Collection and analysis of high resolution CID, HCD, ETD, and UVPD data on various proteins at various energies or reaction times.

**Results:** Each fragmentation mechanism generates unique data that, together, maximizes sequence coverage for improved protein identification and proteoform characterization. Considerations for optimizing each dissociation mechanism with respect to proteins representing a MW range from 9kDa to 50kDa are presented.

## INTRODUCTION

Complete and accurate characterization of intact proteins by mass spectrometry is both possible and increasingly popular today thanks to the latest technological developments made in LC and MS hardware, instrument control software, and data processing software. Here we demonstrate the dissociative behavior of four proteins from the recently released Pierce<sup>™</sup> Intact Protein Standard Mix representing a MW range of 9kDa to 50kDa, with four different modes of ion dissociation (CID, HCD, ETD, and UVPD) available on the Orbitrap<sup>™</sup> Fusion<sup>™</sup> Lumos<sup>™</sup>. For each dissociation mode, we test three different normalized collision energies or reaction/ irradiation times. We aim to illustrate attributes of each of these modes on intact proteins, and ultimately inform method development for top down proteomics applications. While we focused here was on single mode techniques to highlight the specific uniqueness of each mode of dissociation, mixed mode dissociation techniques (ex. EThcD) are also available and can be highly beneficial for both identification and structural characterization.

Ion trap CID employs m/z selective slow heating to produce b- and y- type product ions via many low energy-imparting collisions with He atoms, resulting in minimal secondary dissociation of product ions. This is advantageous, unless post translational modification (PTM) loss is the primary fragmentation pathway. HCD also produces b- and y- type ions through "fast heating" induced relatively fewer, but higher energy-imparting collisions with N<sub>2</sub> gas molecules in a non-m/z selective manner. This makes subsequent over-fragmentation of product ions a risk, but also overcomes the limitation presented by primary loss of labile PTMs. By contrast, ETD generates c- and z- type ions through the abstraction of electrons from a donor reagent anion. To accommodate the resulting radical site, the cation almost instantaneously undergoes rearrangement leading to bond cleavage without internal energy transfer. As such, PTMs are preserved by this mode of dissociation. Intact charge reduced dissociation products from lower charge state precursors can at times dominate spectra, however mild activation of these species through techniques such as EThcD can overcome this limitation. Finally, UVPD, the most recently introduced mode of dissociation on the Orbitrap<sup>TM</sup> Fusion<sup>TM</sup> Lumos<sup>TM</sup> is initiated by irradiation of the precursor ions with photons from a 213nm UV laser, proceeds though multiple dissociation pathways. This results in formation of a-, b-, c-, x-, y-, and z- type fragment ions, many only observed with this mode of dissociation.

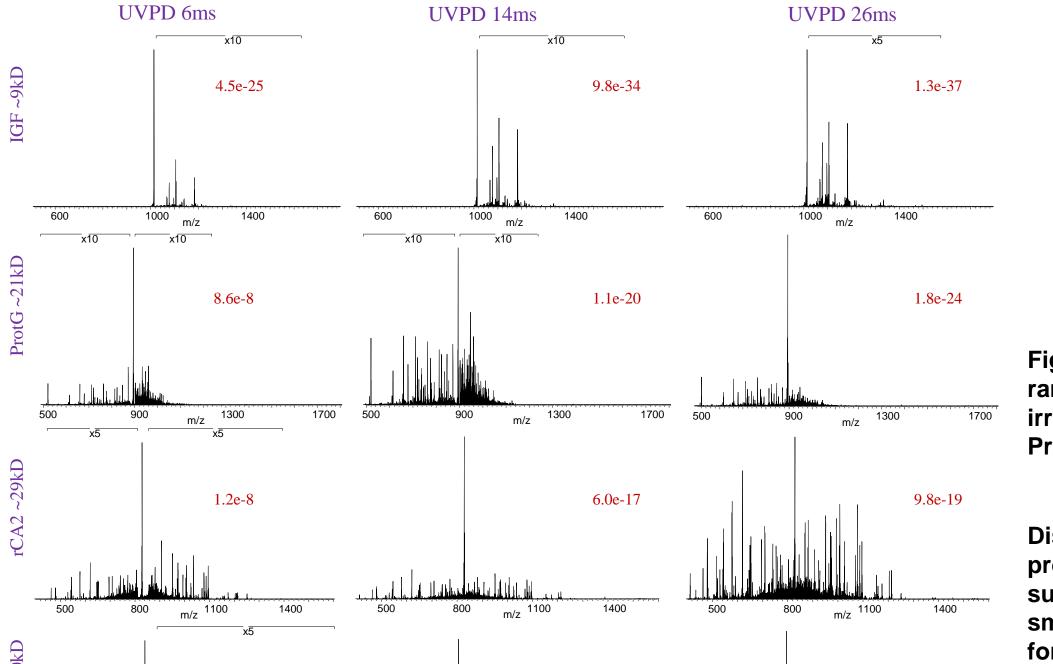
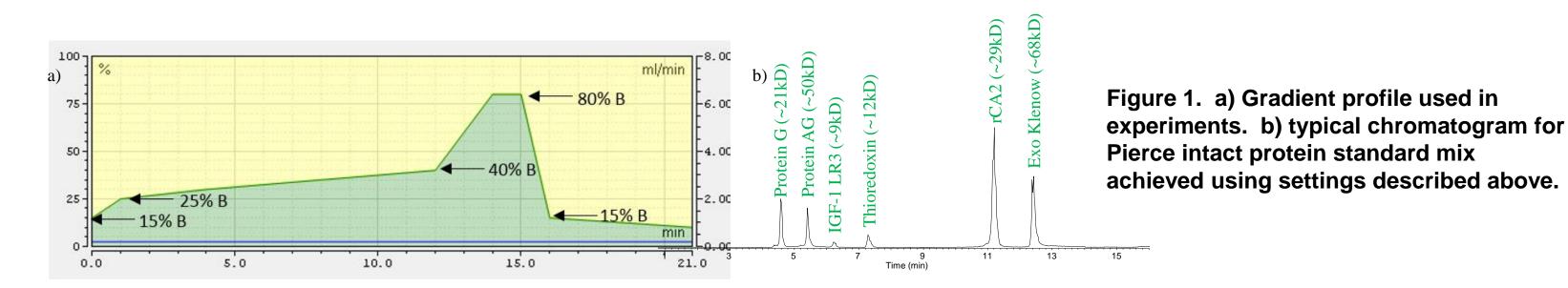


Figure 5: UVPD analysis of 4 different proteins ranging in MW from 9kD to 50kD, at 3 different irradiation times. Inset numbers in red are ProSightPC P-scores.

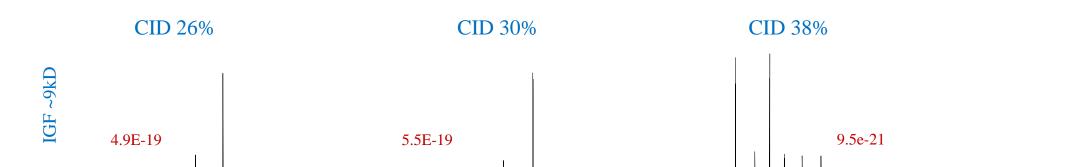
Dissociation here happens at a speed proportional to the MW of the precursor, and as such we see rich spectra produced for the smaller proteins, but a high unresolved baseline for ProteinAG, indicating overfragmentation.

## **MATERIALS AND METHODS**

Pierce intact protein standard mix (A33526) was purchased from Fisher Scientific and each vial was reconstituted in 100ul HPLC grade water prior to use. Proteins were separated over a 20minute gradient (**Figure 1a**) at 200ul/min using a Dionex Ultimate 3000 UHPLC system fitted with a 2.1 mm MabPac<sup>TM</sup> RP LC column. Solvent A was 0.1% formic acid in LCMS grade water (Fisher Scientific LS118-1) and solvent B was 0.1% formic acid in LCMS grade acetonitrile (Fisher Scientific LS120-1). Full scan MS data was collected at 15k resolution in the Orbitrap, with alternating targeted MS2 scans at either 60k (CID, HCD) or 120k resolution (ETD, UVPD). A single charge state of each protein near the center of the charge envelope was selected at random for isolation and fragmentation. As such, precursor charge state selection within a protein is not considered here, though it can be a major variable affecting extent of dissociation. In all cases, precursors were isolated by the quadrupole using a 3Da window. For ETD, anion target value was reduced to 5e4 to reduce reaction kinetics in an attempt to avoid over-fragmentation of large highly charged precursors. Data was collected in a targeted fashion, and MS2 were manually averaged, then deconvoluted using Xtract in QualBrowser. Xtracted raw files were submitted to ProSightPC 4.1 for fragment ion assignment. The Pierce Intact Protein Standard Mix database (.pscw) was downloaded directly from the Proteinacious database warehouse (http://proteinaceous.net/database-warehouse/).



## RESULTS



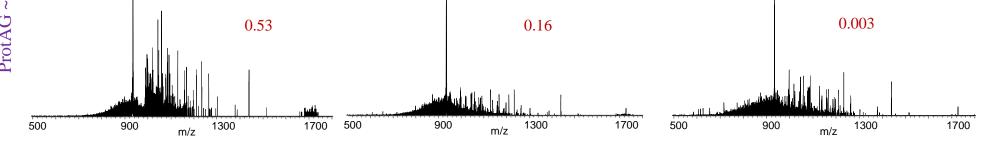


Figure 6: Sequence coverage maps for each of 4 proteins analyzed by each of the 4 modes of dissociation. Each map represents the results from the spectra to the left with the best P-score.

#### IGF ~9kDa

```
CID NMIFIPAMPLISISILIFVINGPRTL GAELVD<sup>25</sup>

<sup>26</sup> ALQFV GDRGFYFNKPTGYGSSSR<sup>50</sup>

<sup>31</sup> APQTGIVDE FRSDLRRLEMY A<sup>75</sup>

<sup>76</sup> [PLK[PAKSAC

NMFP]A]MPLS[S[LFVN]GPRTL GAELVD<sup>25</sup>

<sup>26</sup> ALQFV GDRGFYFNKPTGYGSSSR<sup>50</sup>

<sup>31</sup> APQTGIVDE FRSDLRRLEMY A<sup>75</sup>

<sup>32</sup> ALQFV GDRGFYFNKPTGYGSSSR<sup>50</sup>

<sup>31</sup> APQTGIVDE FRSDLRRLEMY A<sup>75</sup>

<sup>32</sup> ALQFV GDRGFYFNKPTGYGSSSR<sup>50</sup>

<sup>31</sup> APQTGIVDE FRSDLRRLEMY A<sup>75</sup>

<sup>31</sup> APQTGIVDE FRSDLRRLEMY A<sup>75</sup>
```

#### Protein G ~ 21kDa

CIE	N M D P Y P L P K T D T Y K L I L N G K T L K G E T <sup>25</sup> <sup>26</sup> T T E A V D A A T A E K V F K Q Y A N D N G V D G <sup>50</sup> <sup>31</sup> E W T Y D D A T K T F T V T E K P E V I D A S E L <sup>75</sup> <sup>76</sup> T P A V T T Y K L V I N G K T L K G E T T T K A V <sup>100</sup> <sup>101</sup> D A E T A E K A F K Q Y A N D N G V D G V W T Y D <sup>125</sup> <sup>126</sup> D A T K T F T V T E M V T E V P G D A P T E P E K <sup>150</sup> <sup>151</sup> P E A S I L V L V P L L T P A T P A K D D A K K A E T A G C	N M D P Y P]L P K T]D T]Y]K]L I]L]N]G]K]T]L]K]G]E]T <sup>25</sup> <sup>26</sup> T]T]E]A V]D A A T]A E]K]V F K]Q Y]A N]D]N G V]D G <sup>50</sup> <sup>51</sup> E W T]Y]D]D A T K T F T V T E K P]E V]I D A]S E L <sup>75</sup> <sup>76</sup> T P A V T T Y K]L V I]N]G K T L K]G E T T]T K A V <sup>100</sup> <sup>101</sup> ]D A E T A E K A F]K]Q Y A N D N G V D G V W T Y[D <sup>125</sup> <sup>126</sup> [D A T K T F T V T[E M V T]E V P G D]A P T]E P E[K <sup>150</sup> <sup>151</sup> P E A S[I P L]V P L]T P A T P[I A K]D[D]A[K]K]D[D <sup>175</sup> <sup>176</sup> [T]K]K]E[D]A K]K P E A]K K D D]A[K]K]K A E T A G C
HCD	N       M D P Y P L P K T D T Y K L I L N G K T L K G E T       25         26       T T E A V D A A T A E K V F K Q Y A N D N G V D G       50         51       E W T Y D D A T K T F T V T E K P E V I D A S E L       75         76       T P A V T T Y K L V I N G K T L K G E T T T K A V       100         101       D A E T A E K A F K Q Y A N D N G V D G V W T Y D       125         126       D A T K T F T V T E M V T E [V P G D A [P T E [P E K 150	N M D P Y P L P K T D T Y K L I L N G K T L K G E T <sup>25</sup> <sup>26</sup> T T E A V D A A T A E K V F K Q Y A N D N G V D G <sup>50</sup> <sup>51</sup> E W T Y D D A T K T F T V T E K P E V I D A S E L <sup>75</sup> <sup>76</sup> T P A V T T Y K L V I N G K T L K G E T T T K A V <sup>100</sup> <sup>101</sup> D A E T A E K A F K Q Y A N D N G V D G V W T Y D <sup>125</sup> <sup>126</sup> D A T K T F T V T E M V T E V P G D A P T E P E K <sup>150</sup>

151 PLEASI PLL V PLL T PLATPIAKDDAKKDD

76 ТККЕДАКК<mark>ІР</mark>ЕАККДДАККАЕТА<u></u> С

Ѕ Н Н W G Y G K H N G P E H W H K D F P I A N G E

R]Q]S P V]D I]D]T]K A V V Q]D P A]L K P L]A L]V Y

GIELA TISIR RMVNNGH SENVIEIVIDDISODK

<sup>76</sup> A V<mark>I K D G P L T G T Y R L V Q F H</mark>F H W G S S D

<sup>01</sup> D Q G S E H T V D R K K Y A A E L H L V H W N T K 26 Y G D F G T A A O O P D G L A V V G V F L K V G D

<sup>51</sup> A N P A L Q K V L D A L D S I K T K P P S T D F P <sup>1</sup> <sup>76</sup> N F D P G S L L P N V L N Y W T Y P G S L T T P P <sup>2</sup>

<sup>01</sup> L L E S V T W I V L K E P I S V S S Q Q M L K F R <sup>2</sup>

<sup>226</sup> T L N F N A E G E P E L L M L A N W R P A Q P L K

SHHWGYGKHNÌGÌPEHWHKDFÌPÌIÌANGE

R QÌSÌPÌV D I D T K AÌV V QÌDÌP A L KÌPÌL A L V Y

<sup>51</sup> G E A T S R R M V N N G H S F N V E Y D D S O D K

<sup>76</sup> A V L K D G P L T G T Y R L V Q F H F H W G S S D

**DQGSELHTVDRKKYAAELHLVHWNTK** 

<sup>26</sup> YGDFGTAAQQ]PDGLAVVGVFLKVGD

AN PALQKVLDALDSIKTKGKSTDFP

<sup>76</sup> [Ν F DĮΡ<mark>]</mark>G S L LĮΡ N V L N Y W T YĮΡĮG S LĮT TĮΡ<mark>]</mark>Ρ 2

<sup>201</sup>LLESVTWIVLKELPLISVSSQQMLKFR<sup>2</sup>

<sup>26</sup> T L N F N A E G E P E L L M L A N W R P A Q P L K

N R Q V R G F P K (

**NRQVRGFPK** 

#### 176 [T K K E D[A K K[P E A K K D D A K K A E T A G C **rCA2 ~ 29kDa**

<sup>51</sup> P E A S ILP LLVLPLLTLPLALTLP I A K D DLA K K D D <sup>175</sup>

```
      N
      S H H W G Y G K H N G P E HÌW H KÌD FÌPÌ I ÀINÌG E
      23

      26
      R QÌSÌP VÌDÌ I DÌT K AÌVÌVÌQ DÌP A L K P L AÌL V Y
      50

      51
      G E A T S R R M V N N G H S F N V E Y D D S Q D K
      75

      76
      A V L K D G P L T G T Y R L V Q F H F H W G S S D 100

      101
      D Q G S E H T V D R K K Y A A E L H L V H W N T K
      125

      126
      Y G D F G T AÌA Q Q P D G L A V V G V F L K V G D 150

      151
      A N P A L Q K V L D A L D S I K T K G K S T D F P
      173

      176
      N F D P G S L L P N V L N Y W T Y P G S L TĮTĮP [P
      200

      201
      L L E S VĮTĮW I V L K EĮP I S V S S Q Q M L K F R
      225

      226
      T L N F N A E LG EĮP E LLL ML LA LNĮW R P A Q P L K
      250

      251
      N R Q V R G F P K C
      251
```

N SHHWGYGKHNGPEHWHKDF)P]IANGE 24 HCDD<sup>26</sup> RQJS)P)V)DJIDTKA)VVQDPALKPLALVY 56 51 GEATSRRMVNNGHSF)N)VE)YDDSQDK 77 76 AVLKDGPLTGTYRLVQFHFHWGSSD 100 101 DQGSEHTVDRKKYAAELHLVHWNTK 122 126 YGDFGT]AAQQPDG]LAVVGVFLKVGD 150 151 ANPALQKVLDALDSIKTKGKSTDFP 172 176 NFDPGSLLPNVLNYWTYPGSL[T[T[P]P 200 201 LL[E[SV[T[WIVL[KE[PI]SVSSQQML[KFR 22] 226 TLNFNALE[G]E[PELL[L[M]LLA[N]WRPAQPLK 250 251 NRQVRGFPKC

## Discussion

#### Protein AG ~50kDa

26 NG F I Q S L K D D P S Q S A N V L G E A Q K L N 50 51 D S Q A P K A D A Q Q N N F N K D Q Q S A F Y E I 75 76 L NMP N L N E A Q R NG F I Q S L K D D P S Q S 100 101 T N V L G E A K K L N E S Q A P K A D N N F N K E 125 126 Q Q N A F Y E I L NMP N L N E E Q R NG F I Q S 150 26 N G F I Q S L K D D P S Q S A N V L G E A Q K L N S 51 D S Q A P K A D A Q Q N N F N K D Q Q S A F Y E I 51 D S Q A P K A D A Q Q N N F N K D Q Q S A F Y E I 51 D S Q A P K A D A Q Q N N F N K D Q S A F Y E I L N M P N L N E A Q R NG F I Q S L K D D P S Q S 100 101 T N V L G E A K K L N E S Q A P K A D N N F N K E 125 126 Q Q N A F Y E I L N M P N L N E E Q R NG F I Q S 150 126 Q Q N A F Y E I L N M P N L N E E Q R NG F I Q S 150 126 Q Q N A F Y E I L N M P N L N E E Q R NG F I Q S 150 126 Q Q N A F Y E I L N M P N L N E E Q R NG F I Q S 150 126 Q Q N A F Y E I L N M P N L N E E Q R NG F I Q S 150 126 Q Q N A F Y E I L N M P N L N E E Q R NG F I Q S 150 126 Q Q N A F Y E I L N M P N L N E E Q R NG F I Q S 150 127 D Z M Z M Z M Z M Z M Z M Z M Z M Z M Z	מי	$CID \xrightarrow{N} a Q H D E a Q Q N A F Y Q Y L N M P N L N A D Q R^{25}$	N AQHDEAQQNAFYQVLNMPNLNADQR 25 FTD
VPD       Intrinsition of the second stress of the second strese of the second stress of the second stress of the se	D	CID <sup>26</sup> NGFIQSLKDDPSQSANVLGEAQKLN <sup>50</sup>	
VPD       Introl GEAKKLINESQAPKADNNEKKEI       Introl GEAKKLINESQAPKADNNEKKEI         UVPD       Introl GEAKKLINESQAPKADNNEKKEI       Introl GEAKKLINESQAPKADNNEKKEI         Introl GEAKKLINESQAPKADNNEKKEI       Introl GEAKKLINESQAPKADNNEKKEI       Introl GEAKKLINESQAPKADNNEKKEI         Introl GEAKKLINESQAPKADNKERKEI       Introl GEAKKLINESQAPKADNKERKEI       Introl GEAKKLINESQAPKADNKERKEI         Introl GEAKKLINESQAPKADNKERKEI       Introl GEAKKLINESQAPKADANKERKEI       Introl GEAKKLINESQAPKADANKERKEI         Introl GEAKKLINESQAPKADANKERKEI       Introl GEAKKLINESQAPKADANKERKEI       Introl GEAKKLINESQAPKADANKERKEI         Introl GEAKKLINESQAPKADANKEI       Introl GEAKKLINESQAPKADANKEI       Introl GEAKKLINESQAPKADANKEI         Introl GEAKKLINESQAPKADANKEI       Introl GEAKKLINESQAPKADANKEI       Introl GEAKKLINESQAPKADANKEI         Introl GEAKKLINESQAPKADANKEI       Introl GEAKKLINESQAPKADANKEI       Introl GEAKKLINESQAPKADANKEI       Introl GEAKKLINESQAPKADANKEI         INTROL GEAKKLINESQAPKADINERKEI       Introl GEAKKLINESQAPKADINERKEI       Introl GEAKKLINESQAPKADINERKEI       Introl GEAKKLIN		51 D S Q A P K A D A Q Q N N F N K D Q Q S A F Y E I 75	51 D S Q A P K A D A Q Q N N F N K D Q Q S A F Y E I 75
VPD       In QQNAFYET LINMPNI.NEEQRNGFIQS       In QQNAFYET LINMPNI.NEEQRNGFIQS         In Koddpsqsanlisseakkinesqqapkin       In QQNAFYET LINMPNI.NEEQRNGFIQS       In QQNAFYET LINMPNI.NEEQRNGFIQS         ADNKFNKEQQNAFYET LINNINE       ADNKFNKEQQNAFYET LINNINE       In DAQAPKADNKFNKEQQNAFYET LINNINE       In DAQAPKADNKFNKEQQNAFYET LINNINE         In Labakkindaqapkeednnskpite       In GRNSRGSVDASELTPAVTTYKLVIN       In DAQAPKADNKFNKEQQNAFYET       In Labakkindaqapkeednnskpite         In GRNSRGSVDASELTPAVTTYKLVIN       In Koft I CS LKDDPSVSK       In Koft I CS LKDDPSVSK       In Koft I CS LKDDPSVSK         In Koft I CS LKDDPS QSANLLAEEXK       In Koft I CS LKDDPSVSK       In Koft I CS LKDDPSVSK       In Koft I CS LKDDPSVSK         In CD       In Gentrike QNAFYET I LINNINE       In Koft I CS LKDDPSVSK       In Koft I CS LKDDPSVSK       In Koft I CS LKDDPSVSK         In CD       In Gentrike QNAFYET I LINNINE       In CS LKDDPSVSK       In CS LKDDPSVSK       In CS LKDDPSVSK       In CS LKDDPSVSK         IN CO I S LKDDPS S SANVLGE AGKLIN       In CS LKDDPSVSK		76 L N M P N L N E A Q R N G F I Q S L K D D P S Q S 100	76 L N M P N L N E A Q R N G F I Q S L K D D P S Q S 100
VPD       III LKDDPSQSANLLSEAKKINESQAPKII       III LKDDPSQSANLLSEAKKINESQAPKII         III KDDPSQSANLLSEAKKINESQAPKII       III KDDPSQSANLLSEAKKINESQAPKII       III KDDQAPKADNKFNKEQQNAFYEILHLPNINEE         III KDDQAPKADNKFNKEQQNAFYEILHLPNINEE       III KDDQAPKADNKFNKEQQNAFYEILHLPNINEE       III KDDQAPKADNKFNKEQQNAFYEILIIII         III KDDQAPKADNKFNKEQQNAFYEILHLPNINEE       III KDDQAPKADNKFNKEQQNAFYEILIIIII       III KDDQAPKADNKFNKEQQNAFYEILIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII		101 T N V L G E A K K L N E S Q A P K A D N N F N K E 125	101 T N V L G E A K K L N E S Q A P K A D N N F N K E 125
<ul> <li>TOP NORTHER QUARTYSTILL LPNLNE E</li> <li>ADNKFNKE QQNAFYSTIL HLPNLNE E</li> <li>QNNGFIQSLKDDPSQSANLLAEAKK</li> <li>QNNGFIQSLKDDPSQSANLLAEAKK</li> <li>UNDAQAPKADNKFNKEQQNAFYSTIL</li> <li>HLPNLTEEQRNGFIQSLKDDPSQSANLLAEAKK</li> <li>HLPNLTEEQRNGFIQSLKDDPSQSANLAEAKK</li> <li>HLPNLTEEQRNGFIQSLKDDPSQSANLAAEAKK</li> <li>GRNSRGSVDASELTPAVTTYKLVIN</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> </ul>		126 Q Q N A F Y E I L N M P N L N E E Q R N G F I Q S 150	126 Q Q N A F Y E I L N M P N L N E E Q R N G F I Q S 150
<ul> <li>TOP NORTHER QUARTYSTILL LPNLNE E</li> <li>ADNKFNKE QQNAFYSTIL HLPNLNE E</li> <li>QNNGFIQSLKDDPSQSANLLAEAKK</li> <li>QNNGFIQSLKDDPSQSANLLAEAKK</li> <li>UNDAQAPKADNKFNKEQQNAFYSTIL</li> <li>HLPNLTEEQRNGFIQSLKDDPSQSANLLAEAKK</li> <li>HLPNLTEEQRNGFIQSLKDDPSQSANLAEAKK</li> <li>HLPNLTEEQRNGFIQSLKDDPSQSANLAAEAKK</li> <li>GRNSRGSVDASELTPAVTTYKLVIN</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> <li>GRNSRGSVDASELTPAVTTYKLVINGKTLK</li> </ul>	VPD	151 L K D D P S Q S A N L L S E A K K L N E S Q A P K 175	151 L K D D P S Q S A N L L S E A K K L N E S Q A P K 175
ETD Indequarkad nk k nk eqq na fyteil in in landqap kad nk f nk eqq na fyteil in land nk f nk eqq na fyteil in in landqap kad nk f nk eqq na fyteil in in landqap kad nk f nk eqq na fyteil in in landqap kad nk f nk eqq na fyteil in in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in in nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land kk in landqap kad nk f nk eqq na fyteil in land k		176 A D N K F N K E Q Q N A F Y E I L H L P N L N E E 200	176 A D N K F N K E Q Q N A F Y E I L H L P N L N E E 200
UVPD UVPD UVPD Introductor of the second		201 Q R N G F I Q S L K D D P S Q S A N L L A E A K K 225	201 Q R N G F I Q S L K D D P S Q S A N L L A E A K K 225
ETD International set of the set o		226 L N D A Q A P K A D N K F N K E Q Q N A F Y E I L 250	226 L N D A Q A P K A D N K F N K E Q Q N A F Y E I L 250
ETD and set treav daata keverkey and set treav daat kerkey kever <td></td> <td>251 H L P N L T E E Q R N G F I Q S L K D D P S V S K 275</td> <td>251 H L P N L T E E Q R N G F I Q S L K D D P S V S K 275</td>		251 H L P N L T E E Q R N G F I Q S L K D D P S V S K 275	251 H L P N L T E E Q R N G F I Q S L K D D P S V S K 275
ETD "     G KT L K G ET TT TE AV DA AT A E KV F K QY 30 A N D NG V D G EWT Y D D AT KT FT V T E KV F X QY E Y I D AS E L T P AV TT Y K L V I NG KT L K 4 G ET TT TK AV D A E T A E KA F K QY AN D NG 4 G ET TT TK AV D A E T A E KA F K QY AN D NG 4 G ET TT TK AV D A E T A E KA F K QY AN D NG 4 E S T AC W A Q H D E A Q Q N A FY Q Y Q I NM P N L NA D Q R 3 N A Q H D E A Q Q N A FY Q Y Q I N M P N L NA D Q R 3 N A Q H D E A Q Q N A FY Q Y Q I N M P N L NA D Q R 3 N A Q H D E A Q Q N A FY Q Y Q I N M P N L NA D Q R 3 N A Q H D E A Q Q N A FY Q Y Q I N M P N L NA D Q R 3 N A Q H D E A Q Q N A FY Q Y Q I N M P N L NA D Q R 3 N A Q H D E A Q Q N A FY Q Y Q I N M P N L NA D Q R 3 N A Q H D E A Q Q N A FY Q Y Q I N M P N L NA D Q R 3 N A Q H D E A Q Q N A FY Q I N M P N L NA D Q R 3 N A Q H D E A Q Q N A FY Q I L N M P N L NA D Q R 3 N A Q H D E A Q Q N A FY Q I L N M P N L NA D Q R 3 N Y L G E A KK L N E Q A P K A D A Q A P K A D A N F N K E Q A A FY E I L N M P N L NE E Q R N F I Q S L K D D P S Q S A N L L S E A KK L N E S Q A P K 4 N A Q H D E A Q A P K A D A Q A P K A D A Q A P K A D A N F N K E Q A A FY E I L N M P N L NE E Q R N F I Q S L K D D P S Q S A N L L S E A KK L N E S Q A P K 3 L K D D P S Q S A N L L S E A KK L N E S Q A P K 3 L K D D P S Q S A N L L S E A KK L N E S Q A P K 3 L K D D P S Q S A N L L S E A KK L N E S Q A P K 3 L K D D P S Q S A N L L S E A KK L N E S Q A P K 3 L K D D P S Q S A N L L S E A KK L N E S Q A P K 4 L K D D P S Q S A N L L A E A KK 2 L K D D P S Q S A N L L A E A KK 2 L K D D P S Q S A N L L A E A KK 2 L K D D P S Q S A N L L A E A KK 2 L K D D P S Q S A N L L A E A KK 2 L K D D P S Q S A N L L A E A KK 2 L K D D P S Q S A N L L A E A KK 2 L K D D P S Q S A N L L A E A KK 2 L K D D P S Q S A N L L A E A KK 2 L K D D P S Q S A N L L A E A KK 2 L K D D Y S Q S A N L L A E A KK 2 L K D D Y S Q A A Y Y D A		276 E I L A E A K K L N D A Q A P K E E D N N K P I E 300	276 E I L A E A K K L N D A Q A P K E E D N N K P I E 300
ETD       IIII AND NG V D G EWT Y D D AT KT FT VT EKP 737 SE EVI D AS ELT PA VT TY KL VI NG KT L K 40 G ET TT KA V D A ET A E KA F K Q Y AN D NG KT L K 40 G ET TT KA V D A ET A E KA F K Q Y AN D NG KT L K 40 G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K 40 G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K 40 G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K 40 G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K 40 G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K G ET TT KA V D A E TA E KA F K Q Y AN D NG KT L K G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET TT KA V D A E TA E KA F K Q Y AN D NG G G ET T T KA V D A E TA E KA F K Q Y AN D NG G G E T T KA V D A E TA E KA F K Q Y AN D NG G G E T T KA V D A E TA E KA F K Q Y AN D NG G G E T T KA V D A E TA E KA F K Q Y AN D NG G G E T T KA V D A E TA E KA F K Q Y		301 G R N S R G S V D A S E L T P A V T T Y K L V I N 325	301 G R N S R G S V D A S E L T P A V T T Y K L V I N 325
ETD Solution of the second sec		326 G K T L K G E T T T E A V D A A T A E K V F K Q Y 350	326 G K T L K G E T T T E A V D A A T A E K V F K Q Y 350
ETD       #*** GETTTKAAVDAETAEKAFKQYANDNG #***       #**** GETTTKAVDAETAEKAFKQYANDNG #***         #************************************		351 A N D N G V D G E W T Y D D A T K T F T V T E K P 375	351 ANDNGVDGEWTYDDATKTFTVTEKP 375
<ul> <li>W D G VW TY D D A T KT FT VT E MV[T]E[V[P] L</li> <li>W D G VW TY D D A T KT FT VT E MV[T]E[V[P] L</li> <li>W D G VW TY D D A T KT FT VT E MV[T]E[V[P] L</li> <li>W D G VW TY D D A T KT FT VT E MV[T]E[V[P] L</li> <li>W D G VW TY D D A T KT FT VT E MV[T]E[V[P] L</li> <li>W D G VW TY D D A T KT FT VT E MV[T]E[V[P] L</li> <li>W D G VW TY D D A T KT FT VT E MV[T]E[V[P] L</li> <li>W D G VW TY D D A T KT FT VT E MV[T]E[V[P] L</li> <li>W D G VW TY D D A T KT FT VT E MV[T]E[V[P] L</li> <li>W D G VW TY D D A T KT FT VT E MV[T]E[V[P] L</li> <li>W D G VW TY D D A T KT FT VT E MV[T]E[V[P] L</li> <li>W D G VW TY D D A T KT FT V[T E MV[T]E[V[P] L</li> </ul>		376 E V I D A S E L T P A V T T Y K L V I N G K T L K 400	376 E V I D A S E L T P A V T T Y K L V I N G K T L K 400
451 ESTAC HCD N A QH D E A Q Q NA F Y Q Y L NM P N L NA D Q R 25 N A QH D E A Q Q NA F Y Q Y L NM P N L NA D Q R 26 N A QH D E A Q Q NA F Y Q Y L NM P N L NA D Q R 26 N A Q H D E A Q Q NA F Y Q Y L NM P N L NA D Q R 26 N A Q H D E A Q Q NA F Y Q Y L NM P N L NA D Q R 26 S A N G F I Q S L K D D P S Q S A N Y L G E A Q K L N 30 S D S Q A P K A D A Q Q N N F N K D Q Q S A F Y E I 173 S D S Q A P K A D A Q Q N N F N K D Q Q S A F Y E I 173 S D S Q A P K A D A Q Q N N F N K D Q Q S A F Y E I 173 S D S Q A P K A D A Q Q N N F N K D Q Q S A F Y E I 173 S D S Q A P K A D A Q Q N N F N K D Q S A F Y E I 173 S D S Q A P K A D A Q Q N N F N K D Q S A P K A D N N F N K E 253 S D S Q A P K A D A Q Q N N F N K D Q S A P K A D N N F N K E 253 S D S Q A P K A D N K F N K E Q Q N A F Y E I L NM P N L N E A Q R N G F I Q S L K D D P S Q S A N L L S E A K K L N E S Q A P K A D N K F N K E Q Q N A F Y E I L H L P N L N E E Q R N G F I Q S L K D D P S Q S A N L L S E A K K L N E S Q A P K A D N K F N K E Q Q N A F Y E I L H L P N L N E E 270 S A D N K F N K E Q Q N A F Y E I L H L P N L N E E 270 S A D N K F N K E Q Q N A F Y E I L H L P N L N E E 270 S A D N K F N K E Q Q N A F Y E I L H L P N L N E E 270 S A D N K F N K E Q N A F Y E I L H L P N L N E E 270 S A D N K F N K E Q N A F Y E I L H L P N L N E E Q R N G F I Q S L K D D P S V S K 275 S A D N K F N K E Q N A F Y E I L A E A K K L N D A Q A P K A D N K F N K E Q N A F Y E I L A E A K K L N D A Q A P K A D N K F N K E Q N A F Y E I L A E A K K L N D A Q A P K A D N K F N K E Q N A F Y E I L A E A K K L N D A Q A P K A D N K F N K E Q N A F Y E I L A E A K K L N D A Q A P K A D N K F N K E Q N A F Y E I L A E A K K L N D A Q A P K A D N K F N K E Q N A F Y E I L A E A K K L N D A Q A P K A D N K F N K E Q N A F Y E I L A E A K K L N D A Q A P K A D N K F N K E Q N A S E L T P A V T Y K L V I N G K L K Q Y A Y D A T K T K K V Y	ETD	401 G E T T T K A V D A E T A E K A F K Q Y A N D N G 425	401 G E T T T K A V D A E T A E K A F K Q Y A N D N G 425
HCD In a QHD E a QQ NA FIYÎQÎYÎLÎNMÎP NL NA DQ R In a QHD E a QQ NA FIYÎQÎYÎLÎNMÎP NL NA DQ R In a QHD E a QQ NA FIYÎQÎYÎLÎNMÎP NL NA DQ R In a QHD E a QQ NA FIYÎQÎYÎLÎNMÎP NL NA DQ R In a QHD E a QQ NA FIYÎQÎYÎLÎNMÎP NL NA DQ R In a QA PK A D A QQ NN F NK DQ QS AÎFY EÎI In trivil ge a KK L NE SQ A PK A D NN F NK E In trivil ge a KK L NE SQ A PK A D NK F NK E QQ NA F Y E I L H L P N L NE E In trivil ge a KK L ND A QA PK E E D NN K P I E In trivil ge a KK L ND A QA PK E E D NN K P I E In trivil ge a KK L ND A QA PK E E D NN K P I E In trivil ge a KK L ND A QA PK E E D NN K P I E In trivil ge a KK L ND A QA PK E E D NN K P I E In trivil ge a KK L ND A QA PK E E D NN K P I E In trivil ge a KK L ND A QA PK E E D NN K P I E In trivil ge a KK L ND A QA PK E E D NN K P I E In trivil ge a KK L ND A QA PK E E D NN K P I E In trivil ge a KK L ND A QA PK E E D NN K P I E		426 V D G V W T Y D D A T K T F T V T E M V T L V P L 450	426 V D G V W T Y D D A T K T F T V T E M V T E V P L 450
HCD26 NGFIQSLKDDPSQSANVLGEAQKLN26 NGFIQSLKDDPSQSANVLGEAQKLN26 NGFIQSLKDDPSQSANVLGEAQKLN20 UVPD20 SQAPKADAQQNNFNKDQQSAFYEI25 NGFIQSLKDDPSQSANVLGEAQKLN26 NGFIQSLKDDPSQSANVLGEAQKLN27 NGFIQSLKDDPSQSANVLGEAQKLN20 UVPD20 SQAPKADAQQNNFNKDQQSAFYEI26 NGFIQSLKDDPSQSANVLGEAQKLN27 LNMPNLNEAQRNGFIQSLKDDPSQS20 LNMPNLNEAQRNGFIQSLKDDPSQS20 LNMPNLNEEQRNGFIQSLKDDPSQS20 SQAPKADANFNL26 QQNAFYEI21 LNMPNLNEEQRNGFIQS20 LNMPNLNEEQRNGFIQS20 LNMPNLNEEQRNGFIQS20 SQNAFYEI21 LNMPNLNEEQQNAFYEI21 LNMPNLNEEQRNGFIQS21 LNMPNLNEEQRNGFIQS20 LNMPNLNEEQRNGFIQS21 GQNAFYEI21 LNDAPSQSANLLSEAKKLNESQAPK21 QNNFNKEQQNAFYEI21 LNDAPSQSANLLAEAKKLNESQAPK21 QNNFNKEQQNAFYEI21 QRNGFIQSLKDDPSQSANLLAEAKK22 LNDAQAPKADNKFNKEQQNAFYEI20 QNAFYEI20 QNAFYEI20 QNAFYEI22 LNDAQAPKADNKFNKEQQNAFYEI20 QNAFYEI20 QNAFYEI20 QNAFYEI20 QNAFYEI23 QRNGFIQSLKDDPSQSANLLAEAKK20 QNAFYEI20 QNAFYEI20 QNAFYEI20 QNAFYEI24 LNDAQAPKADNKFNKEQQNAFYEI20 QNAFYEI20 QNAFYEI20 QNAFYEI20 QNAFYEI25 HLPNLTEEQRNGFIQSLKDDPSVSK20 QNAFYEI20 QNAFYEI20 QNAFYEI20 QNAFYEI26 GKTLKGETTTEAVDAATAEKVFKQY30 GRNSRGSVDASELTPAVTTYKLVIN30 GRNSRGSVDASELTPAVTTYKLVIN30 GNSRGSVDASELTPAVTTYKLVIN36 GKTLKGETTTEAVDAATAEKVFKQY36 ANDNGVDGEWTYDDATKTFTVTEKP37 EVIDASELTPAVTTYKLVINGKTLK36 GKTLKGETTTEAVDAATAEKVFKQY37 EVIDASELTPAVTTYKLVINGKTLK37 EVIDASELTPAVTTYKLVINGKTLK37 EVIDASELTPAVTTYKLVINGKTLK37 EVIDASELTPAVTTYKLVINGKT		451 <b>E S T A</b> $\subset$	451 ESTAC
UVPD 10 S Q A P K A D A Q Q NN F N K D Q Q S A F Y E I 175 10 S Q A P K A D A Q Q NN F N K D Q Q S A F Y E I 175 10 S Q A P K A D A Q Q NN F N K D Q Q S A F Y E I 175 10 S Q A P K A D A Q Q NN F N K D Q S A F Y E I 175 10 S Q A P K A D A Q Q NN F N K D Q S A F Y E I 175 10 S Q A P K A D A Q Q NN F N K D Q S A F Y E I 175 10 S Q A P K A D N F N K E 125 10 S Q A P K A D N F N K E 125 10 S Q A P K A D N F N K E 125 10 S Q A P K A D N F N K E 125 10 S Q A P K A D N F N K E 125 10 S Q A P K A D N F N K E 125 10 S Q A P K A D N F N K E 125 10 S Q A P K A D N F N K E 125 10 S Q A P K A D N F N K E 125 10 S Q A F Y E I L N P N L N E E Q R N G F I Q S L K D D P S Q S A N L L S E A K K L N E S Q A P K A Z 10 S Q A P K A D N K F N K E Q Q N A F Y E I L H L P N L N E E 200 10 S Q R N G F I Q S L K D D P S Q S A N L L A E A K K 225 10 S Q N A F Y E I L H L P N L N E E Q R N G F I Q S L K D D P S Q S A N L L A E A K K 225 10 S Q N A F Y E I L A E A K K L N D A Q A P K E E D N N K P I E 200 10 S G R N S R G S V D A S E L T P A V T T Y K L V I N 25 10 S G R N S R G S V D A S E L T P A V T T Y K L V I N 25 10 S G K T L K G E T T T E A V D A A T A E K V F K Q Y 305 10 S G R N S R G S V D A S E L T P A V T T Y K L V I N 325 10 S G K T L K G E T T T E A V D A A T A E K V F K Q Y 305 10 S G R N S R G S V D A S E L T P A V T T Y K L V I N S 25 10 S G K T L K G E T T T E A V D A A T A E K V F K Q Y 305 10 S G K T L K G E T T T E A V D A A T A E K V F K Q Y 305 10 S G K T L K G E T T T E A V D A A T A E K V F K Q Y 305 10 S G K T L K G E T T T K A V D A E T A E K A F K Q Y A N D N G 425 10 S G K T L K A V D A E T A E K A F K Q Y A N D N G 425 10 S G K T L K A V D A E T A E K A F K Q Y A N D N G 425 10 S G K T L K A V D A E T A E K A F K Q Y A N D N G 425 10 S G K T L K A V D A E T A E K A F K Q Y A N D N G 425 10 S G K T L K A V D A E T A E K A F K Q Y A N D N G 425 10 S G K T L K A V D A E T A E K A F K Q Y A N D N G 425 10 S G K T L K A V D A E T A E K A F K Q Y A N D N G 425 10 S M D N G V D G		HCD N A Q H D E A Q Q N A F]Y]Q]V]L]N M]P N L N A D Q R 25	
10110110210310410		11CD 26 NGFIQSLKDDPSQSANVLGEAQKLN 50	26 NGFIQSLKDD)PSQSANVLGEAQKLN 50 ${ m UVPD}$
101 TNVLGEAKKLNESQAPKADNNFNKE 125101 TNVLGEAKKLNESQAPKADNNFNKE 125102 TNVLGEAKKLNESQAPKADNNFNKE 125103 TNVLGEAKKLNESQAPK 175104 TNVLGEAKKLNESQAPK 175105 QSANLLSEAKKLNESQAPK 175105 QSANLLSEAKKLNESQAPK 175106 QNAFYEILHLPNLNEE 200107 ADNKFNKEQQNAFYEILHLPNLNEE 200107 ADNKFNKEQQNAFYEILHLPNLNEE 200107 ADNKFNKEQQNAFYEILBE QRNGFIQSLKDDPSQSANLLAEAKK 225201 QRNGFIQSLKDDPSQSANLLAEAKK 225201 QRNGFIQSLKDDPSQSANLLAEAKK 225201 QRNGFIQSLKDDPSVSK 225201 QRNGFIQSLKDDPSVSK 225201 HLPNLTEEQRNGFIQSLKDDPSVSK 225201 GRNSRGSVDASELTPAVTTYKLVIN<		51 D S Q A P K A D A Q Q N N F N K D Q Q S A] F Y E] I 75	51 D S Q A P K A D A Q Q N N F N K D Q Q S A F Y E I 75
UVPD126 QQNAFYEIILNMPNLNEEQRNGFIQS120 QQNAFYEILNMPNLNEEQRNGFIQS120 QQNAFYEILNMPNLNEEQRNGFIQS130 QQNAFYEILNMPNLNEEQRNGFIQS151 LKDDPSQSANLLSEAKKLNESQAPK176 ADNKFNKEQQNAFYEILHLPNLNEE151 LKDDPSQSANLLSEAKKLNESQAPK151 LKDDPSQSANLLSEAKKLNESQAPK176 ADNKFNKEQQNAFYEILHLPNLNEE200 QRNGFIQSLKDDPSQSANLLAEAKK220 QRNGFIQSLKDDPSQSANLLAEAKK220 QRNGFIQSLKDDPSQSANLLAEAKK201 QRNGFIQSLKDDPSQSANLLAEAKK220 LNDAQAPKADNKFNKEQQNAFYEIL200 QRNGFIQSLKDDPSVSK221 QRNGFIQSLKDDPSVSK201 QRNGFIQSLKDDPSVSK225 LNDAQAPKADNKFNKEQQNAFYEIL200 QRNGFIQSLKDDPSVSK225 LNDAQAPKADNKFNKEQQNAFYEIL201 GRNSRGSVDASELTPAVTTYKLVIN241 LPNLTEEQRNGFIQSLKDDPSVSK276 EILAEAKKLNDAQAPKEEDNNKPIE200 201 GRNSRGSVDASELTPAVTTYKLVIN202 GKTLKGETTTEAVDAATAEKVFKQY301 GRNSRGSVDASELTPAVTTYKLVIN325 301 GRNSRGSVDASELTPAVTTYKLVIN326 GKTLKGETTTEAVDAATAEKVFKQY302 GKTLKGETTTEAVDAATAEKVFKQY326 GKTLKGETTTEAVDAATAEKVFKQY326 GKTLKGETTTEAVDAATAEKVFKQY327 221 ANDNGVDGEWTYDDATKTFTVTEKP303 GRNSRGSVDASELTPAVTTYKLVINGKTLK301 GRNSRGSVDASELTPAVTTYKLVIN326 301 GRNSRGSVDASELTPAVTTYKLVIN304 GKTLKGETTTEAVDAATAEKVFKQY326 GKTLKGETTTEAVDAATAEKVFKQY326 GKTLKGETTTEAVDAATAEKVFKQY305 GKTLKGETTTKAVDAETAEKAFKQYANDNG401 GETTTKAVDAETAEKAFKQYANDNG401 GETTTKAVDAETAEKAFKQYANDNG402 VDGVWTYDDATKTFTVTEMVTEVPL400 GVWTYDDATKTFTVTEMVTEVPL404 VDGVWTYDDATKTFTVTEMVTEVPL400 VDGVWTYDDATKTFTVTEMVTEVPL		76 <b>] L]N]M P N L N E A Q R N G F I Q S L K D D P S Q S 100</b>	76 L N MÌP N L N E A Q R N G F I Q S L K D DÌP S Q S 100
UVPD151L K D D P S Q S A N L L S E A K K L N E S Q A P K 176151L K D D P S Q S A N L L S E A K K L N E S Q A P K 176176A D N K F N K E Q Q N A F Y E I L H L P N L N E E 200176A D N K F N K E Q Q N A F Y E I L H L P N L N E E 200176A D N K F N K E Q Q N A F Y E I L H L P N L N E E 200201Q R N G F I Q S L K D D P S Q S A N L L A E A K K 225201Q R N G F I Q S L K D D P S Q S A N L L A E A K K 225226L N D A Q A P K A D N K F N K E Q Q N A F Y E I L 250201Q R N G F I Q S L K D D P S V S K 275226L N D A Q A P K A D N K F N K E Q Q N A F Y E I L 250201H L P N L T E E Q R N G F I Q S L K D D P S V S K 275226E I L A E A K K L N D A Q A P K E E D N N K P I E 200201G R N S R G S V D A S E L T P A V T T Y K L V I N 325301G R N S R G S V D A S E L T P A V T T Y K L V I N 325301G R N S R G S V D A S E L T P A V T T Y K L V I N 325326G K T L K G E T T T E A V D A A T A E K V F K Q Y 350326G K T L K G E T T T E A V D A A T A E K V F K Q Y 350326G K T L K G E T T T E A V D A A T A E K V F K Q Y 350326G K T L K G E T T T E A V D A A T A E K V F K Q Y 350326G K T L K G E T T T E A V D A A T A E K V F K Q Y 350326G K T L K G E T T T E A V D A A T A E K V F K Q Y 350326G K T L K G E T T T E A V D A A T A E K V F K Q Y 350326G K T L K G E T T T E A V D A A T A E K V F K Q Y 350327E V I D A S E L T P A V T T Y K L V I N G K T L K 400401G E T T T K A V D A E T A E K A F K Q Y A N D N G 425426V D G V W T Y D D A T K T F T V T E MIV T E [V] L 4504		101 T N V L G E A K K L N E S Q A P K A D N N F N K E 125	101 T N V L G E A K K L N E S Q A]P K A D N N F N K E 125
UVPD176 adnkfnkeqqnafyeilhlpnlnee176 adnkfnkeqqnafyeilhlpnlnee176 adnkfnkeqqnafyeilhlpnlnee201 qRngfiqslkddpsqsanllaeakk220176 adnkfnkeqqnafyeil201 qRngfiqslkddpsqsanllaeakk220201 qRngfiqslkddpsqsanllaeakk220201 QRngfiqslkddpsqsanllaeakk220221 hlpnlteeqRngfiqslkddpsvsk220221 hlpnlteeqRngfiqslkddpsvsk222225 lhgnltegram221 hlpnlteeqRngfiqslkddpsvsk220221 hlpnlteeqRngfiqslkddpsvsk276 eillaeakklundaqapkeednnkpie200201 gRnsrgsvdaseltpavttyklvin225202 gknsrgsvdaseltpavttyklvin226226 gktlkgettteev227203 grup226 gktlkgettteev227221 hlpnltee204 grup205 grup226 gktlkgettteev227205 grup226 gktlkgettteev228 gktlkgettteev228206 grup229229220 grup207 grup229220 grup220 grup208 grup220 grup220 grup220 grup209 grup220 grup220 grup220 grup201 grup220 grup220 grup220 grup202 grup220 grup220 grup220 grup203 grup220 grup220 grup220 grup204 grup220 grup220 grup220 grup205 grup220 grup220 grup220 grup206 grup220 grup220 grup220 grup207 grup220 grup220 grup220 grup208 grup220 grup220 grup220 grup209 grup220 grup220 grup<		126 Q Q N A F Y E]I]L N M]P N L N E E Q R N G F I Q S 150	126 QQNAFYEILNMPNLNEEQRNGFIQS 150
201 Q R N G F I Q S L K D D P S Q S A N L L A E A K K 225202 L N D A Q A P K A D N K F N K E Q Q N A F Y E I L 250203 L N D A Q A P K A D N K F N K E Q Q N A F Y E I L 250204 L N D A Q A P K A D N K F N K E Q Q N A F Y E I L 250205 H L P N L T E E Q R N G F I Q S L K D D P S V S K 275206 E I L A E A K K L N D A Q A P K E E D N N K P I E 200201 G R N S R G S V D A S E L T P A V T T Y K L V I N 325202 G K T L K G E T T T E A V D A A T A E K V F K Q Y 350203 G K T L K G E T T T E A V D A A T A E K V F K Q Y 350204 G K T L K G E T T T E A V D A A T A E K V F K Q Y 350205 A N D N G V D G E W T Y D D A T K T F T V T E K P 375206 E V I D A S E L T P A V T T Y K L V I N G K T L K 400401 G E T T T K A V D A E T A E K A F K Q Y A N D N G 425402 V D G V W T Y D D A T K T F T V T E [V] P L 450404 V D G V W T Y D D A T K T F T V T E [V] P L 450		151 L K D D P S Q S A N L L S E A K K L N E S Q A P K 175	151 L K D D P S Q S A N L L S E A K K L N E S Q A P K 175
226LNDAQAPKADNKFNKEQQNAFYEIL250226LNDAQAPKADNKFNKEQQNAFYEIL250251HLPNLTEEQRNGFIQSLKDDPSVSK251HLPNLTEEQRNGFIQSLKDDPSVSK251HLPNLTEEQRNGFIQSLKDDPSVSK251276EILAEAKKLNDAQAPKEEDNNKPIE270EILAEAKKLNDAQAPKEEDNNKPIE270271EILAEAKKLNDAQAPKEEDNNKPIE300301GRNSRGSVDASELTPAVTTYKLVIN325301GRNSRGSVDASELTPAVTTYKLVIN325326GKTLKGETTTEAVDAATAEKVFKQY350326GKTLKGETTTEAVDAATAEKVFKQY350351ANDNGVDGEWTYDDATKTFTVTEKP375351ANDNGVDGEWTYDDATKTFTVTEKP375376EVIDASELTPAVTTYKLVINGKTLK400401GETTTKAVDAETAEKAFKQYANDNG425426VDGVWTYDDATKTFTVTEMVTEVPL450426VDGVWTYDDATKTFTVTEMVTEVPL	UVPD	176 A D N K F N K E Q Q N A F Y E I L H L P N L N E E 200	176 A D N K F N K E Q Q N A F Y E I L H L P N L N E E 200
251 H L P N L T E E Q R N G F I Q S L K D D P S V S K 275251 H L P N L T E E Q R N G F I Q S L K D D P S V S K 275276 E I L A E A K K L N D A Q A P K E E D N N K P I E 300276 E I L A E A K K L N D A Q A P K E E D N N K P I E 300301 G R N S R G S V D A S E L T P A V T T Y K L V I N 325301 G R N S R G S V D A S E L T P A V T T Y K L V I N 325326 G K T L K G E T T T E A V D A A T A E K V F K Q Y 350301 A N D N G V D G E W T Y D D A T K T F T V T E K P 375376 E V I D A S E L T P A V T T Y K L V I N G K T L K 400376 E V I D A S E L T P A V T T Y K L V I N G K T L K 400401 G E T T T K A V D A E T A E K A F K Q Y A N D N G 425426 V D G V W T Y D D A T K T F T V T E MI V T E [V] P L 450		201 Q R N G F I Q S L K D D P S Q S A N L L A E A K K 225	201 Q R N G F I Q S L K D D P S Q S A N L L A E A K K 225
276E I L A E A K K L N D A Q A P K E E D N N K P I E 300301G R N S R G S V D A S E L T P A V T T Y K L V I N 325302G K T L K G E T T T E A V D A A T A E K V F K Q Y 350303301304G K T L K G E T T T E A V D A A T A E K V F K Q Y 350305306306G K T L K G E T T T E A V D A A T A E K V F K Q Y 350307S V D A S E L T P A V T T Y K L V I N G K T L K 400308G K T L K G E T T T E A V D A A T A E K V F K Q Y 350309S V D A S E L T P A V T T Y K L V I N G K T L K 400301G E T T T K A V D A E T A E K A F K Q Y A N D N G 425401G E T T T K A V D A E T A E K A F K Q Y A N D N G 425426V D G V W T Y D D A T K T F T V T E MIV T E [V] P L 450		226 L N D A Q A P K A D N K F N K E Q Q N A F Y E I L 250	226 L N D A Q A P K A D N K F N K E Q Q N A F Y E I L 250
301G R N S R G S V D A S E L T P A V T T Y K L V I N301G R N S R G S V D A S E L T P A V T T Y K L V I N325326G K T L K G E T T T E A V D A A T A E K V F K Q Y350326G K T L K G E T T T E A V D A A T A E K V F K Q Y350321A N D N G V D G E W T Y D D A T K T F T V T E K P375351A N D N G V D G E W T Y D D A T K T F T V T E K P375376E V I D A S E L T P A V T T Y K L V I N G K T L K400401G E T T T K A V D A E T A E K A F K Q Y A N D N G426426V D G V W T Y D D A T K T F T V T E MUY T E [V] P L450426V D G V W T Y D D A T K T F T V T E M V T E V P L450		251 H L P N L T E E Q R N G F I Q S L K D D P S V S K 275	251 H L P N L T E E Q R N G F I Q S L K D D P S V S K 275
326G K T L K G E T T T E A V D A A T A E K V F K Q Y326G K T L K G E T T T E A V D A A T A E K V F K Q Y350351A N D N G V D G E W T Y D D A T K T F T V T E K P375351A N D N G V D G E W T Y D D A T K T F T V T E K P375376E V I D A S E L T P A V T T Y K L V I N G K T L K376E V I D A S E L T P A V T T Y K L V I N G K T L K376E V I D A S E L T P A V T T Y K L V I N G K T L K401401G E T T T K A V D A E T A E K A F K Q Y A N D N G425426V D G V W T Y D D A T K T F T V T E M V T E [V]P L450		276 E I L A E A K K L N D A Q A P K E E D N N K P I E 300	276 E I L A E A K K L N D A Q A P K E E D N N K P I E 300
351 ANDNGVDGEWTYDDATKTFTVTEKP351 ANDNGVDGEWTYDDATKTFTVTEKP376 EVIDASELTPAVTTYKLVINGKTLK351 ANDNGVDGEWTYDDATKTFTVTEKP376 EVIDASELTPAVTTYKLVINGKTLK376 EVIDASELTPAVTTYKLVINGKTLK401 GETTTKAVDAETAEKAFKQYANDNG425426 VDGVWTYDDATKTFTVTEMVTEVPL450		301 G R N S R G S V D A S E L T P A V T T Y K L V I N 325	301 G R N S R G S V D A S E L T P A V T T Y K L V I N 325
376E V I D A S E L T P A V T T Y K L V I N G K T L K 400401G E T T T K A V D A E T A E K A F K Q Y A N D N G 425402V D G V W T Y D D A T K T F T V T E M V T E [V]P L 450403404404G E T T T K A V D A E T A E K A F K Q Y A N D N G 425405V D G V W T Y D D A T K T F T V T E M V T E [V]P L 450		326 G K T L K G E T T T E A V D A A T A E K V F K Q Y 350	326 G K T L K G E T T T E A V D A A T A E K V F K Q Y 350
401 GETTTKAVDAETAEKAFKQYANDNG       425         401 GETTTKAVDAETAEKAFKQYANDNG       425         426 VDGVWTYDDATKTFTVLTEMLVTELVLPL       426 VDGVWTYDDATKTFTVTEMVTEVPL		351 A N D N G V D G E W T Y D D A T K T F T V T E K P 375	351 ANDNGVDGEWTYDDATKTFTVTEKP 375
426 V D G V W T Y D D A T K T F T VLT E MLV T ELVLP L 450 426 V D G V W T Y D D A T K T F T V T E M V T E V P L 450		376 E V I D A S E L T P A V T T Y K L V I N G K T L K 400	376 E V I D A S E L T P A V T T Y K L V I N G K T L K 400
		401 G E T T T K A V D A E T A E K A F K Q Y A N D N G 425	401 G E T T T K A V D A E T A E K A F K Q Y A N D N G 425
451 ESTAC 451 ESTAC		426 V D G V W T Y D D A T K T F T VLT E MLV T ELVLP L 450	426 V D G V W T Y D D A T K T F T V T E M V T E V P L 450
		451 $\mathbf{E}$ $\mathbf{S}$ $\mathbf{T}$ $\mathbf{A}$ $\subset$	451 ESTAC

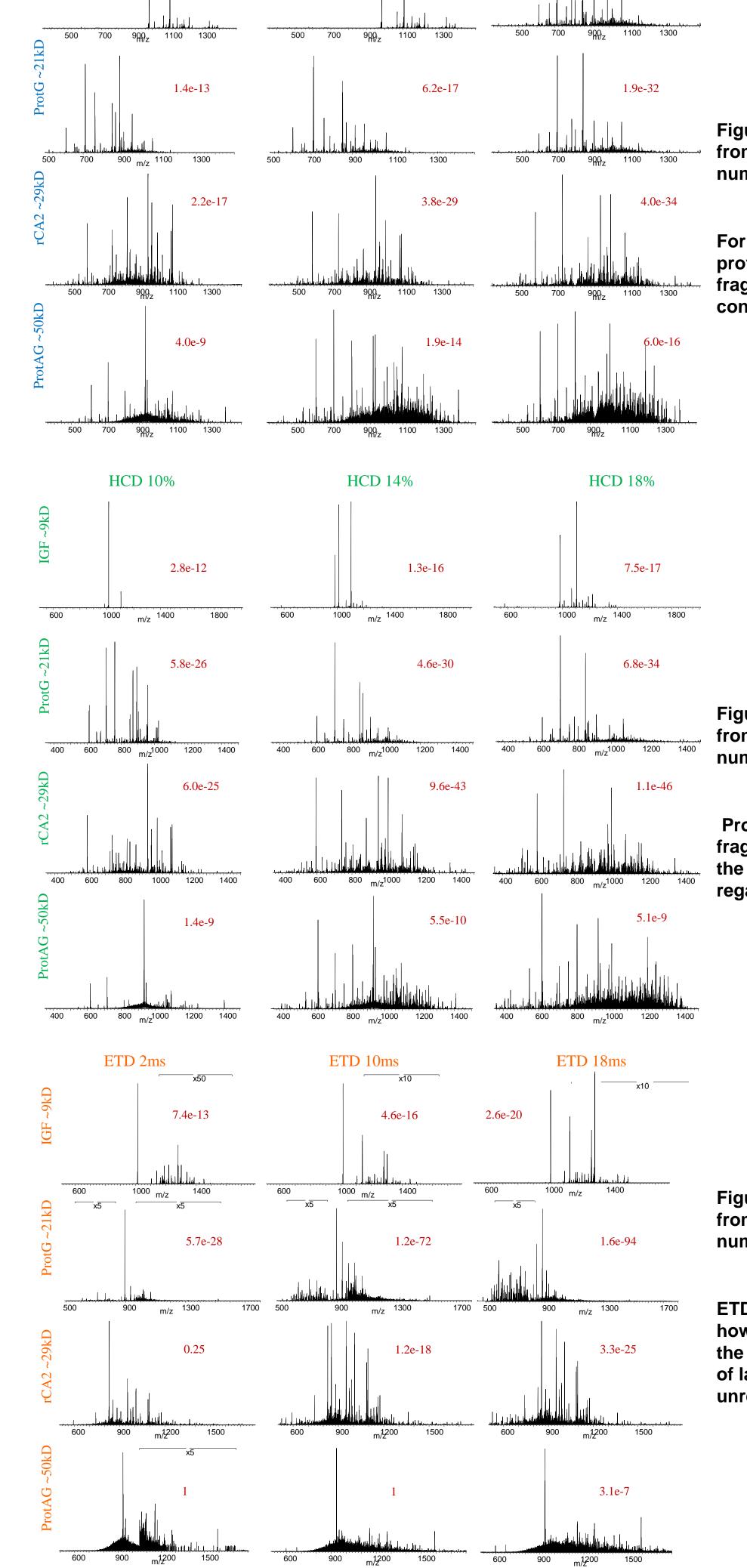


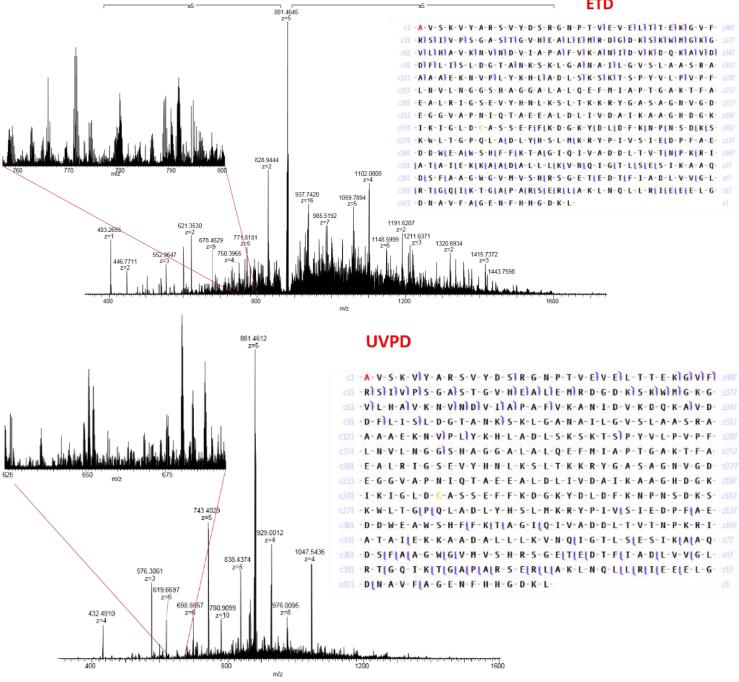
Figure 2: CID analysis of 4 different proteins ranging in MW from 9kD to 50kD, at 3 different collision energies. Inset numbers in red are ProSightPC P-scores.

For top-down analysis if non-modified intact proteins, CID provides the benefit of limited secondary fragmentation, overfragmentation, and formation of internal fragments that is consistent across the mass range.

Figure 3: HCD analysis of 4 different proteins ranging in MW from 9kD to 50kD, at 3 different collision energies. Inset numbers in red are ProSightPC P-scores.

Provided that energies are carefully chosen to avoid overfragmentation, this mode of fragmentation is efficient across the mass range, and provides well resolved fragments regardless of presence of PTMs. The versatility afforded by the Orbitrap<sup>™</sup> Fusion<sup>™</sup> Lumos<sup>™</sup> with respect to the multiple types of available dissociation modes is a clear advantage for top down analyses. Additionally, the Pierce Intact Protein Standard Mix provides an ideal sample for method development (both data acquisition and data analysis) and quality control. In using this sample for method optimization we have highlighted both strengths and weaknesses of our current technology. We are able to obtain extensive sequence coverage for the proteins in the sample up to 30kD on a chromatographic time scale. We do, however, still struggle with MS2 analysis of larger proteins. Multiple challenges contribute to this problem. First, by both ETD and UVPD, larger proteins dissociate much faster than smaller proteins, whether due to their higher charge state, or higher cross section, respectively. In this work, we decreased the anion target value in an attempt to reduce ETD reaction rate (the kinetics of the ETD reaction as we perform it here are first order with respect to anion concentration) and minimize over fragmentation of Protein AG (50kD), though this helped only marginally. Other ion manipulation techniques such as ion parking have been shown to address this problem. Second, larger proteins can of course fragment at more positions, thereby diluting signal among more potential product ions. CID and HCD benefit here from preferential fragmentation at weaker bonds, concentrating signal to fewer possible product ions. Because ETD and UVPD are democratic in their bond cleavage, this is a significant challenge that currently can only be overcome with significant signal averaging. **Figure 7** demonstrates the efficiency of ETD and UVPD on enolase, a 46kD protein, when ~500 transients are averaged.

An added challenge presented by over fragmentation is the production of internal ions. These low abundance, unresolved overlapping product ions create a high baseline that varies across the m/z range. Deconvolution algorithms generally use the averagine model to assign monoisotopic mass, but the large number of overlapping peaks confound such algorithms due to experimental isotopic distributions that deviate too far from theoretical. We continue to work toward addressing these challenges



#### Figure 7: ETD and UVPD of enolase (~46kD); ~500 averaged transients.

Figure 4: ETD analysis of 4 different proteins ranging in MW from 9kD to 50kD, at 3 different reaction times. Inset numbers in red are ProSightPC P-scores.

ETD spectra of the smaller proteins are extremely rich, however because reactions proceed at rates proportional to the square of the precursor charge state, overfragmenation of larger proteins is common and evidenced by the high, unresolvable baseline seen in all ProtAG spectra.

## CONCLUSIONS

■ The multiple modes of dissociation available on the Orbitrap<sup>TM</sup> Fusion<sup>TM</sup> Lumos<sup>TM</sup> present a clear advantage for intact protein identification and characterization, enabling extensive sequence coverage and PTM mapping capabilities.

• The Pierce Intact Protein Standard Mix is an ideal sample for top down method development, optimization, and quality control.

• Many challenges remain in top down analysis, particularly with respect to large proteins. We are actively working to address these.

## REFERENCES

- 1. Syka, J.E.P et al. "Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry." PNAS vol. 101 no. 26, 9528-9533
- Cannon J.R. et al. "Hybridizing Ultraviolet Photodissociation with Electron Transfer Dissociation for Intact Protein Characterization" Anal. Chem. 2014, 86(21), pp 10970-10977
- 3. McLuckey S.A. et al. "Ion parking during ion/ion reactions in eletrodynamic ion traps". Anal. Chem, 2002, 74(2), pp 336-346

## ACKNOWLEDGEMENTS

The authors would like to acknowledge John E.P Syka, David Horn, and Tara Schroeder for helpful discussion.

## TRADEMARKS/LICENSING

© 2016 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries. This information is not intended to encourage use of these products in any manner that might infringe the intellectual property rights of others.

ThermoFisher SCIENTIFIC