Supplementary Information

Kojak: Efficient analysis of chemically cross-linked protein complexes

Michael R. Hoopmann[1], Alex Zelter[2], Richard S. Johnson[3], Michael Riffle[2], Michael J. MacCoss[3], Trisha N. Davis[2], and Robert L. Moritz[1*]*

[1]Institue for Systems Biology, Seattle, Washington 98109, USA
[2]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA
[3]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA

* Corresponding Author: robert.moritz@systemsbiology.org

Table of Contents

1. Supplementary Information: Additional analysis of the CRY2-FBXL3-SKP1 complex structure

The 81,256 MS/MS spectra were obtained by shotgun MS as described in the main text and analyzed with Kojak to identify peptide spectrum matches (PSMs). PSMs broadly fall into four categories: 1) single peptides, with or without modifications, including half-reacted cross-linker, 2) loop-link peptides in which the cross-linker binds two sites on the same peptide, 3) intraprotein cross-links where both peptides belong to the same protein, and 4) interprotein cross-links between peptides from two different proteins. PSM validation was performed with Percolator after dividing the PSMs into the four categories. Details of the cross-linked PSMs were provided in the main text.

Several comparisons between Kojak, Protein Prospector, and pLink were performed in the main text. The protein sequence database contained 292 total sequences. In addition to CRY2, FBXL3, and SKP1, several *Drosophila melanogaster* proteins were included. The cryptochrome proteins were grown in *Spodoptera frugiperda* (Sf9) insect cells[1]. Because of this fact, a small percentage of contaminating insect proteins would be expected. Prior to running Kojak, the spectra were searched Comet[2] (a standard mass spectrometry proteomic data and protein sequence database search tool) using a custom database with CRY2, FBXL3, and SKP1 and *D. melanogaster* sequences as surrogates for the unknown *S. frugiperda* protein sequences. A 1% FDR was applied the results and any *D. melanogaster* peptides identified were assumed to be homologous in *S. frugiperda*. The *D. melanogaster* protein sequences containing these peptides were then included in the cross-linked sequence database used with Kojak, Protein Prospector, and pLink. Additional lab contaminants, including trypsin, keratins, and several *Saccharomyces cerevisiae* proteins used in the same laboratory were also included in the protein sequence database. A total of 146 target proteins were included in the sequence database. Those sequences were reversed to create an equal number of decoy sequences to produce the final 292 sequences in the database.

Comparative analysis between Kojak and Protein Prospector were performed after validation using Percolator. Cross-linked PSMs were divided into intraprotein and interprotein cross-links, as described in the main text. Percolator was used to validate the PSMs, as it uses a support vector machine (SVM) classifier model, similar to that used in other published analyses[3]. For the Protein Prospector results, following guidelines in the same publication, a limited number of parameters were selected for training the SVM. PPScoreDiff was used, as it was indicated as being the best parameter. As percent TIC matched is not output in the Protein Prospector results, the Protein Prospector score was used as the other parameter. Only a small number of PSMs (585 total) were reported for intraprotein cross-links, and the SVM model failed due to the low percentage (3.1%) of decoys contained in this set. Instead, a 1% FDR was approximated by computing the number of decoy PSMs at a given Protein Prospector score cutoff. A sufficient number of interprotein cross-linked PSMs were available to generate FDR values using the SVM. For Kojak, additional parameters were added to the analysis, as described elsewhere in the Supplementary Information. Both intraprotein and interprotein FDRs were computed separately, and thus an approximate 1% FDR for each set does not represent the FDR of the combined set, which is still expected to be similar.

Additional comparisons of cross-linked PSMs were made for Kojak, Protein Prospector, and pLink. All three algorithms reported cross-links to proteins other than CRY2-FBXL3-SKP1. For interprotein cross-

links, it can be expected that one protein belongs to CRY2-FBXL3-SKP1, with the other belongs to a random background protein included in the database search. Such instances suggest that one peptide in the cross-link is indeed correct and providing sufficient score to be validated when linked to a random peptide[3-4]. Protein Prospector reported the fewest such cross-links, with two, both of which contained a CRY2-FBXL3-SKP1 peptide. Kojak returned three such interactions, of which two contained a CRY2-FBXL3-SKP1 peptide, and pLink reported 23, of which only seven contained a CRY2-FBXL3-SKP1 peptide. It is worth noting that pLink reports results at a fixed 5% FDR, which may contribute to the larger number of unexpected cross-links between background proteins.

Finally, algorithm performance was compared in terms of computation time. Where possible, a single thread on a single core was used for each algorithm. Computation time on a Windows XP desktop was 8.4 hours using Kojak to analyze the CRY2-FBXL3-SKP1 spectra, compared to 34.2 hours when using pLink. Protein Prospector could only be run remotely from its web interface, and required 47.4 hours of computation time. However, it is likely that the analysis was distributed over a cluster from the Protein Prospector server and does not reflect the computation time with a single core. The time savings with Kojak is highly advantageous considering the intensive computing resources required to perform larger studies with whole proteomes.

2. Supplementary Information: Performance of Kojak analytical processes

Spectral processing

The Kojak algorithm was used to process high-resolution precursor and MS/MS spectra prior to database searching to improve algorithm speed and accuracy. Using the true monoisotopic precursor mass instead of the selected ion mass when searching a spectrum narrows the list of target peptide sequences that must be searched to obtain the correct PSM[5] and improves algorithm speed. The monoisotopic peak might not have been identified at the time of data acquisition for some or all spectra. Factors such as low precursor intensity and interfering signals often prevent monoisotopic peak identification. Precursor ion peaks of low abundance at the time of selection are averaged across neighboring scan events in which the ion was observed. The resulting composite spectrum improves isotope peak visibility over noise, and improves isotopic envelope shape (Supplementary Figure 1). The monoisotopic precursor mass is then predicted from the composite using a model-based precursor fitting algorithm[6]. The 81,256 CRY2-FBXL3-SKP1 MS/MS spectra were used to assess the performance improvement using precursor prediction. For these spectra, the monoisotopic peak was identified by the instrument software for only 21,942 (27%) MS/MS spectra. Using Kojak, 39,925 (49%) additional MS/MS spectra were assigned a monoisotopic mass using composite spectrum analysis. The remaining 19,389 (24%) MS/MS spectra were given a series of potential monoisotopic mass predictions equal to approximately 0, -1, and -2 Da from the selected ion mass. Despite the slight increase in search space for the fraction of spectra for which the monoisotopic mass could not be determined, over all the spectra the search space was reduced by 40%, nearly doubling the speed at which PSMs can be identified. The computation time to perform precursor spectral processing is dependent on the number of spectra in the file, and typically takes only one to two minutes for data sets similar to those processed here.

Additional spectral processing was performed on the high-resolution CRY2-FBXL3-SKP1 MS/MS spectra as described in the methods. Spectral processing of MS/MS spectra was performed to collapse isotope distributions to the precursor mass, represent the entire isotope envelope as a single peak intensity, and reduce the number of peaks in the spectra to be analyzed (Supplementary Figure 2). The spectral processing can be applied to any spectrum, but is best performed only on high-resolution spectra where fragment ion isotope envelopes are fully resolved, and thus spectral processing can be optionally disabled by the user. Without spectral processing, 13,591 PSMs were identified from the CRY2-FBXL3-SKP1 spectra at 1% FDR from 19.6 million total spectral peaks. Using spectral processing reduced the number of peaks to 15.0 million peaks (23.3% reduction) and the number of PSMs increased to 13,808. The computation time to perform MS/MS spectral processing is typically less than 1 second per 1,000 spectra.

Though currently integrated into the basic features of the Kojak algorithm, these spectral processing functions have broad applicability for processing spectra prior to analysis with many other algorithms. The open source nature of Kojak allows for these spectral processing functions to be integrated into other tools. Furthermore, these functions can be bypassed from the Kojak parameters for spectra that have been preprocessed using independent tools.

Database Searching Efficiency

Peptide identification was performed using database searching algorithms contained in the Kojak software. In common with traditional database search algorithms, theoretical fragment ion masses were predicted from peptide sequences parsed out of a protein sequence database. A major limitation to cross-link search algorithms is the computational search space that expands quadratically with protein sequence database size to consider all possible cross-links, as defined in equation 1:

$$x = \frac{n(n+1)}{2}$$

where *x* is the number of cross-linked peptide combinations, and *n* is the number of candidate peptides. This equation represents searching all combinations of peptides against a spectrum, which rapidly becomes prohibitive as database size increases because of the approximately n^2 peptide combinations that must be searched.

Although methods exist that reduce the search space through efficient ordering and traversal,[7-8] these solutions still scale quadratically with database size and thus may remain prohibitive for large databases despite the improvements in efficiency. An alternative strategy that scales linearly with database size searches only single peptides, but with a modification mass equal to difference from the precursor mass.[9-11] Kojak uses this modified peptide approach in a two-pass strategy, as described in the methods. To illustrate the differences in performance, a small set of 991 spectra were alternatively searched against databases increasing in size using the two-pass method and the broader approach that finds all combinations of peptides prior to searching (ordered approach). In these comparisons, the ordered approach was traversed using the published method[8], and were filtered by precursor monoisotopic mass. The same software functions were used for data reading, processing, and spectral scoring in both approaches, and all tests were performed on a single computational thread. With the ordered approach, searching a small sequence database of only 20 proteins was much faster than the two-pass method at 6 vs. 31 seconds (Supplementary Table 4). However, the ordered approach became significantly slower when increasing the database size to 200 proteins. The two-pass method was seven-fold faster at this database size. Indeed, the ordered approach showed quadratic increases in computation time with database size while the two-pass method remained linear with increasing database size (Supplementary Figure 3). Though many cross-linking analyses restrict database size to only those proteins involved in the protein complex being studied, analyses of larger complexes or cellular lysates[12] are becoming more routine. In these cases, the two-pass method provides a computationally efficient approach.

3. Supplementary Information: Kojak/Percolator input parameter descriptions

Parameters for non-linked and loop-linked PSMs:
Score: PSM score reported by Kojak
dScore: difference in score between the PSM and the next best match for that spectrum
Charge: charge state of the precursor ion
Mass: mass (in Daltons) of the precursor ion
PPM: mass difference between observed precursor ion mass and PSM, in parts per million
Len: length of the peptide sequence

Parameters for cross-linked PSMs:
Score: PSM score reported by Kojak
dScore: difference in score between the PSM and the next best match for that spectrum
NormRank: summed rank of the two peptides in the top scoring list
PPScoreDiff: score of the worse matching peptide in the pair
Charge: charge state of the precursor ion
Mass: mass (in Daltons) of the precursor ion
PPM: mass difference between observed precursor ion mass and PSM, in parts per million
LenShort: length of the short peptide sequence
LenLong: length of the long peptide sequence
LenSum: sum of the two peptide lengths

4. Supplementary Information: Step-by-step Kojak operation instructions

1. Create a configuration file in any text editor and save as ASCII text. An example is provided at the top of this document. Provide the required input/output files: a FASTA database, MS data file, output file name, and Percolator input file name. Use absolute paths if not operating in the current working directory.
2. Open a command prompt and navigate to the Kojak application folder, if it is not in your system path.
3. On the command line, type "Kojak config.txt" where config.txt can be replaced with the file name used for the configuration file in step 1.
4. After Kojak analysis completes, filter intraprotein or interprotein cross-links using a spreadsheet style application (e.g. Microsoft Excel). Export the filtered list as tab-delimited text.
5. Compute FDR using Percolator version 2.07 (https://github.com/percolator/percolator) by typing "Percolator perc_input.txt > perc_output.txt" where perc_input.txt is your filtered Kojak results from step 4.
6. Open perc_output.txt in any text editor and extract PSMs at the desired FDR cutoff (q-value column).

5. Supplementary Information: Algorithm parameters used

Kojak:

Parameters for CRY2-FBXL3-SKP1 analysis:

```
percolator_version      =       2.07

# Parameters used to described the data being input to Kojak
enrichment              =       0       #Values between 0 and 1 to describe 18O APE. For example, 0.25 equals 25 APE.
instrument              =       0       #Values are: 0=Orbitrap, 1=FTICR, such as Thermo LTQ-FT
MS1_centroid            =       0       #0=no, 1=yes
MS2_centroid            =       0       #0=no, 1=yes
MS1_resolution          =       50000   #Resolution at 400 m/z
MS2_resolution          =       25000   #Resolution at 400 m/z

# Cross-link and mono-link masses allowed. May have more than one of each parameter.
#1 = amine reactive, 2=carboxyl, 3=sulfhydryl
cross_link_mass         =       1       1       138.0680742
mono_link_mass          =       1       155.0946
mono_link_mass          =       1       156.0786

# Fixed modifications.
fixed_modification      =       C       57.02146

# Differential modifications. @ = N-terminus
max_mods_per_peptide    =       2
modification            =       K       14.015894
modification            =       K       28.031788
modification            =       K       42.047682
modification            =       M       15.9949
modification            =       @       42.01055

enzyme      =           [KR]|{P}

# Parameters used in Kojak analysis
threads                 =       1
decoy_filter            =       reverse #identifier for all decoys in the database. Default value is "random" (without quotes)
fragment_bin_offset     =       0
fragment_bin_size       =       0.03
max_miscleavages        =       4       #number of missed trypsin cleavages allowed
max_peptide_mass        =       8000.0  #largest allowed peptide mass in Daltons
min_peptide_mass        =       500.0   #lowest allowed peptide mass in Daltons
max_spectrum_peaks      =       0       #top N peaks to use during analysis. 0 uses all peaks.
ppm_tolerance_pre       =       15.0    #mass tolerance on precursor when searching
prefer_precursor_pred   =       0       #prefer precursor mono mass predicted by instrument software.
search_dimers           =       0       #0=no, 1=yes
relaxed_analysis        =       1       #0=no, 1=yes
top_count               =       250     #number of top scoring single peptides to combine in relaxed analysis
spectrum_processing     =       1       #0=no, 1=yes
use_comet_xcorr         =       0       #0=no, 1=yes
```

Parameters for Cop9 signalosome were the same except for:

```
MS2_centroid          =       1
MS1_resolution        =       120000
cross_link            =       1       1       138.0680742
cross_link            =       1       1       150.1434042
mono_link             =       1       155.0946
mono_link             =       1       156.0786
mono_link             =       1       167.16993
mono_link             =       1       168.15393
modification          =   K    14.015894
modification          =   K    28.031788
modification          =   K    42.047682
modification          =   M    15.9949
threads               =       2
fragment_bin_offset   =       0.4
fragment_bin_size     =       1.0005
ppm_tolerance_pre     =       10.0
spectrum_processing=          0
```

Parameters for 26S proteasome were the same except for:

```
MS2_centroid          =       1
MS1_resolution        =       60000
cross_link            =       2       2       138.09055
cross_link            =       2       2       146.14076
cross_link            =       1       2       -18.010595
mono_link             =       2       156.10111
mono_link             =       2       164.15132
modification          =   K    14.015894
modification          =   K    28.031788
modification          =   K    42.047682
modification          =   M    15.9949
threads               =       2
fragment_bin_offset   =       0.4
fragment_bin_size     =       1.0005
prefer_precursor_pred         =       1
spectrum_processing=          0
```

pLink parameters

[pLink]
sample.num=1
sample1.spectra.instrument=HCD
sample1.spectra.format=mgf
sample1.spectra.path=C:\pLink\data4
sample1.spectra.title=test1
enzyme.name=Trypsin
database.path=C:\pLink\data4\RJAZ10-comet-plus-short.fasta
database.name=DB_small
max_miss_site=4
mod.fixed.total=1
mod.fixed.1=Carbamidomethyl_C
mod.variable.total=5
mod.variable.1=_Oxidation_M_15.9949
mod.variable.2=_Methyl_K_14.015894
mod.variable.3=_N-Acetyl_Protein_42.01055
mod.variable.4=_di-Methylation_K_28.031301
mod.variable.5=_tri-Methyl_K_42.047682
linker.total=1
linker.name1=BS3
noninterexport=true
peptide_tol_total=1
peptide_tol1=15
peptide_tol_type1=ppm
peptide_tol_base1=0.000000
peptide_tol_base_type1=ppm
peptide_tol2=15
peptide_tol_type2=ppm
peptide_tol_base2=1.007825035
peptide_tol_base_type2=Da
peptide_tol3=15
peptide_tol_type3=ppm
peptide_tol_base3=2.01565007
peptide_tol_base_type3=Da
peptide_tol4=15
peptide_tol_type4=ppm
peptide_tol_base4=3.023475105
peptide_tol_base_type4=Da
peptide_tol5=15
peptide_tol_type5=ppm
peptide_tol_base5=4.03130014
peptide_tol_base_type5=Da
filter_peptide_tol_base=0,1.007825035,2.01565007,3.023475105,4.03130014
filter_peptide_tol_lb=-10,-10,-10,-10,-10
filter_peptide_tol_ub=10,10,10,10,10
filter_peptide_tol_type=ppm
evalue_max=10000
show_peptide_tol_type=ppm
bin.path=C:\pLink\pLink_release\
output.path=C:\pLink\data4\

Protein Prospector parameters: version: **5.12.2**

aa_modified_1: **K**
aa_modified_2: **K**
aa_modified_3: **K**
aa_modified_4: **Protein N-term**
aa_modified_5: **Protein N-term**
aa_modified_6: **Protein N-term**
allow_non_specific: **at 0 termini**
const_mod: **Carbamidomethyl (C)**
data_source: **List of Files**
database: **User Protein**
dna_frame_translation: **3**
enzyme: **Trypsin**
expect_calc_method: **Linear Tail Fit**
expect_coeff_file: **AllThree.exp.1**
fragment_masses_tolerance: **10**
fragment_masses_tolerance_units: **ppm**
full_pi_range: **1**
high_pi: **10.0**
input_filename: **lastres**
input_program_name: **msfit**
instrument_name: **ESI-Q-high-res**
link_aa: **C->C**
link_search_type: **DSS**
low_pi: **3.0**
max_hits: **9999999**
max_saved_tag_hits: **1000**
missed_cleavages: **4**
mod_c_term_type: **Peptide**
mod_comp_ion: **K**
mod_defect: **0.00048**
mod_end_nominal: **4038**
mod_max_z: **4**
mod_n_term: **1**
mod_n_term_type: **Protein**
mod_range_type: **Da**
mod_start_nominal: **438**
mod_uncleaved: **1**
msms_full_mw_range: **1**
msms_max_modifications: **2**
msms_max_peaks: **80**
msms_max_reported_hits: **5**
msms_mod_AA: **Carbamyl (N-term)**
msms_mod_AA: **Dimethyl (Uncleaved K)**
msms_mod_AA: **Met-loss (Protein N-term M)**
msms_mod_AA: **Methyl (K)**
msms_mod_AA: **Oxidation (M)**
msms_mod_AA: **TriMethyl (Uncleaved K)**
msms_mod_AA: **Xlink:DSS1 (Protein N-term)**
msms_mod_AA: **Xlink:DSS1 (Uncleaved K)**
msms_mod_AA: **Xlink:DSS2 (Protein N-term)**

msms_mod_AA: **Xlink:DSS2 (Uncleaved K)**
msms_parent_mass_systematic_error: **0**
msms_parent_mass_tolerance: **15**
msms_parent_mass_tolerance_units: **ppm**
msms_pk_filter: **Max MSMS Pks**
msms_precursor_charge_range: **2 3 4 5**
msms_prot_high_mass: **125000**
msms_prot_low_mass: **1000**
parent_mass_convert: **monoisotopic**
report_title: **BatchTag**
script_filename: **script**
search_key: **zeOT9B1t7sjQ1nT3**
search_name: **batchtag**
species: **All**
use_instrument_ion_types: **1**

6. Supplementary Tables

Supplementary Table 1: CRY2-FBXL3-SKP1 interprotein cross-links, 1% FDR

Peptide	Protein #1	Position #1	Protein #2	Position #2
TFNIKNDFTEEEEAQVR(5)--NSSEEGTAEKSK(10)	Skp1dd	128	Fbxl3	22
NSSEEGTAEKSK(10)--KVIQWCTHHK(1)	Fbxl3	22	Skp1dd	51
LLKMSSCPHVSPAGILCVADQCHGLR(3)--MKQIYQQLSR(2)	Fbxl3	206	mCRY2	503
ATHPELIKQIIK[155.09]R(8)--KENQWCEEK(1)	Fbxl3	102	Skp1dd	141
SLSSLKIDDTPVDDPSLK(6)--MKQIYQQLSR(2)	Fbxl3	180	mCRY2	503
MAKEAGVEVVTENSHTLYDLDR(3)--NSSEEGTAEKSK(10)	mCRY2	136	Fbxl3	22
ATHPELIK[155.09]QIIKR(12)--KENQWCEEK(1)	Fbxl3	106	Skp1dd	141
ATHPELIKQIIK[156.08]R(8)--KENQWCEEK(1)	Fbxl3	102	Skp1dd	141
NSSEEGTAEKSK(10)--TVANMIKGK(7)	Fbxl3	22	Skp1dd	114
ATHPELIKQIIK(8)--KENQWCEEK(1)	Fbxl3	102	Skp1dd	141
VLVANNSDTLKLLK(11)--MKQIYQQLSR(2)	Fbxl3	203	mCRY2	503
TMLEDLGMDPVPLPNVNAAILKK(22)--IIELNGQKPPLTYK(8)	Skp1dd	50	mCRY2	163
IDDTPVDDPSLKVLVANNSDTLK(12)--MKQIYQQLSR(2)	Fbxl3	192	mCRY2	503
IDDTPVDDPSLKVLVANNSDTLK(12)--IIELNGQKPPLTYK(8)	Fbxl3	192	mCRY2	163
NSSEEGTAEKSK(10)--GKTPEEIRK(2)	Fbxl3	22	Skp1dd	116
ATHPELIKQIIK(8)--ENQWCEEK(8)	Fbxl3	102	Skp1dd	149
NSSEEGTAEKSK(10)--GKTPEEIR(2)	Fbxl3	22	Skp1dd	116
ATHPELIKQIIK(8)--KENQWCEEK(9)	Fbxl3	102	Skp1dd	149
NSSEEGTAEKSK(10)--KENQWCEEK(1)	Fbxl3	22	Skp1dd	141
TFNIKNDFTEEEEAQVR(5)--NSSEEGTAEKSK[156.08]K(10)	Skp1dd	128	Fbxl3	22
CFEFELNQPATSYLKATHPELIK(15)--ENQWCEEK(8)	Fbxl3	94	Skp1dd	149
MKQIYQQLSR(2)--KENQWCEEK(1)	mCRY2	503	Skp1dd	141
HSNHLQYVSFKVDSSK(11)--KENQWCEEK(9)	Fbxl3	118	Skp1dd	149
CFEFELNQPATSYLKATHPELIK(15)--KENQWCEEK(9)	Fbxl3	94	Skp1dd	149
NSSEEGTAEKSK(10)--MKQIYQQLSR(2)	Fbxl3	22	mCRY2	503
HSNHLQYVSFKVDSSK(11)--KENQWCEEK(1)	Fbxl3	118	Skp1dd	141
NSSEEGTAEKSK(10)--KTFNIK(1)	Fbxl3	22	Skp1dd	123
CFEFELNQPATSYLKATHPELIK(15)--K[156.08]ENQWCEEK(9)	Fbxl3	94	Skp1dd	149
HSNHLQYVSFKVDSSK(11)--TVANMIKGK(7)	Fbxl3	118	Skp1dd	114
YSLEQIHWEVSKHLGR(12)--IIELNGQKPPLTYKR(14)	Fbxl3	416	mCRY2	169

Supplementary Table 2: CRY2-FBXL3-SKP1 intraprotein cross-links, 1% FDR

Peptide	Protein	Position #1	Position #2
AAAAVVAATVPAQSMGADGASSVHWFR(1)--LFKEWGVTR(3)	mCRY2	2	107
KPAVAVSSQQMESCR(1)--LWDLYKK(6)	mCRY2	183	292
AAAAVVAATVPAQSMGADGASSVHWFR(1)--KLNSR(1)	mCRY2	2	86
AAAAVVAATVPAQSMGADGASSVHWFR(1)--KPAVAVSSQQMESCR(1)	mCRY2	2	183
LQSSDGEIFEVDVEIAKQSVTIK(17)--MPSIK(1)	Skp1dd	22	1
AAAAVVAATVPAQSMGADGASSVHWFR(1)--DAAIMKMAK(6)	mCRY2	2	133
MAKEAGVEVVTENSHTLYDLDR(3)--LTFEYDSEPFGKER(12)	mCRY2	136	125
AAKCIIGVDYPRPIVNHAETSR(3)--LKGFPSR(2)	mCRY2	477	455
MELPK[155.09]KPAVAVSSQQMESCR(6)--LWDLYKK(6)	mCRY2	183	292
MAKEAGVEVVTENSHTLYDLDR(3)--LFKEWGVTR(3)	mCRY2	136	107
LLKMSSCPHVSPAGILCVADQCHGLR(3)--SLSSLKIDDTPVDDPSLK(6)	Fbxl3	206	180
YIYEPWNAPESVQKAAK(14)--YLPKLK(4)	mCRY2	474	453
TFNIKNDFTEEEEAQVR(5)--KENQWCEEK(9)	Skp1dd	128	149
TFNIKNDFTEEEEAQVR(5)--GKTPEEIR(2)	Skp1dd	128	116
TFNIKNDFTEEEEAQVRK(5)--GKTPEEIR(2)	Skp1dd	128	116
MELPK[156.08]KPAVAVSSQQMESCR(6)--LWDLYKK(6)	mCRY2	183	292

TFNIKNDFTEEEEAQVR(5)--GKTPEEIRK(2)	Skp1dd	128	116
SLSSLKIDDTPVDDPSLK(6)--HSNHLQYVSFKVDSSK(11)	Fbxl3	180	118
YIYEPWNAPESVQK[156.08]AAKCIIGVDYPRPIVNHAETSR(17)--LKGFPSR(2)	mCRY2	477	455
KPAVAVSSQQMESCR(1)--LFKEWGVTR(3)	mCRY2	183	107
KPAVAVSSQQMESCR(1)--KVK[156.08]R(1)	mCRY2	183	293
YIYEPWNAPESVQK[155.09]AAKCIIGVDYPRPIVNHAETSR(17)--LKGFPSR(2)	mCRY2	477	455
KPAVAVSSQQMESCR(1)--LKGFPSR(2)	mCRY2	183	455
KTFNIK[156.08]NDFTEEEEAQVR(1)--GKTPEEIR(2)	Skp1dd	123	116
SLSSLKIDDTPVDDPSLK(6)--VLVANNSDTLKLLK(11)	Fbxl3	180	203
CFEFELNQPATSYLKATHPELIK(15)--QIIKR(4)	Fbxl3	94	106
KPAVAVSSQQMESCR(1)--LWDLYKK[155.09]VK(6)	mCRY2	183	292
MAKEAGVEVVTENSHTLYDLDR(3)--YIYEPWNAPESVQKAAK(14)	mCRY2	136	474
AAKCIIGVDYPRPIVNHAETSR(3)--YLPKLK(4)	mCRY2	477	453
YIYEPWNAPESVQKAAK(14)--LKGFPSR(2)	mCRY2	474	455
LLKMSSCPHVSPAGILCVADQCHGLR(3)--IDDTPVDDPSLKVLVANNSDTLK(12)	Fbxl3	206	192
YIYEPWNAPESVQKAAK(14)--LFKEWGVTR(3)	mCRY2	474	107
HSNHLQYVSFKVDSSK(11)--QIIKR(4)	Fbxl3	118	106
NSSEEGTAEKSK(10)--QIIKR(4)	Fbxl3	22	106
KENQWCEEK(9)--KTFNIK(1)	Skp1dd	149	123
AAKCIIGVDYPRPIVNHAETSR(3)--YLPKLK[155.09]GFPSR(4)	mCRY2	477	453
KENQWCEEK(9)--TVANMIKGK(7)	Skp1dd	149	114
KAWVANYERPR(1)--KVK[156.08]R(1)	mCRY2	246	293
KPAVAVSSQQMESCR(1)--LWDLYKK[156.08]VK(6)	mCRY2	183	292
YIYEPWNAPESVQKAAK[156.08]CIIGVDYPRPIVNHAETSR(14)--LKGFPSR(2)	mCRY2	474	455
DAAIMKMAK[155.09]EAGVEVVTENSHTLYDLDR(6)--LFKEWGVTR(3)	mCRY2	133	107
TFNIKNDFTEEEEAQVR(5)--KENQWCEEK(1)	Skp1dd	128	141
ENQWCEEK(8)--KTFNIK(1)	Skp1dd	149	123
AAKCIIGVDYPRPIVNHAETSR(3)--YLPK[155.09]LKGFPSR(6)	mCRY2	477	455
KPAVAVSSQQMESCR(1)--KLNSR(1)	mCRY2	183	86
LTFEYDSEPFGKER(12)--DAAIMKMAK(6)	mCRY2	125	133
KENQWCEEK(9)--GKTPEEIR(2)	Skp1dd	149	116
YIYEPWNAPESVQKAAK(14)--LDKHLER(3)	mCRY2	474	241
KENQWCEEK(1)--GKTPEEIR(2)	Skp1dd	141	116
KAWVANYERPR(1)--KVKR(3)	mCRY2	246	295
KAWVANYERPR(1)--KVK[155.09]R(1)	mCRY2	246	293
KPAVAVSSQQMESCR(1)--K[156.08]VKR(3)	mCRY2	183	295
KAWVANYERPR(1)--LDKHLER(3)	mCRY2	246	241
QSVTIKTMLEDLGMDPVPLPNVNAAILK(6)--MPSIK(1)	Skp1dd	28	1
LLKMSSCPHVSPAGILCVADQCHGLR(3)--HSNHLQYVSFKVDSSK(11)	Fbxl3	206	118
IIELNGQKPPLTYK(8)--MKQIYQQLSR(2)	mCRY2	163	503
KTFNIK[155.09]NDFTEEEEAQVR(1)--GKTPEEIR(2)	Skp1dd	123	116
TVANMIKGK(7)--ENQWCEEK(8)	Skp1dd	114	149
IIELNGQKPPLTYKR(14)--KLNSR(1)	mCRY2	169	86
YIYEPWNAPESVQKAAK(14)--MKQIYQQLSR(2)	mCRY2	474	503
GKTPEEIR(2)--KTFNIK(1)	Skp1dd	116	123
SLSSLKIDDTPVDDPSLK(6)--ATHPELIKQIIK(8)	Fbxl3	180	102
HSNHLQYVSFKVDSSK(11)--VLVANNSDTLKLLK(11)	Fbxl3	118	203
KTFNIKNDFTEEEEAQVR(6)--GKTPEEIR(2)	Skp1dd	128	116
LFKEWGVTR(3)--DAAIMKMAK(6)	mCRY2	107	133
MELPKK[155.09]PAVAVSSQQMESCR(5)--LWDLYKK(6)	mCRY2	182	292
ENQWCEEK(8)--GKTPEEIR(2)	Skp1dd	149	116
YIYEPWNAPESVQKAAK[155.09]CIIGVDYPRPIVNHAETSR(14)--LKGFPSR(2)	mCRY2	474	455
AAKCIIGVDYPRPIVNHAETSR(3)--YLPKLK[156.08]GFPSR(4)	mCRY2	477	453
IDDTPVDDPSLKVLVANNSDTLK(12)--QIIKR(4)	Fbxl3	192	106
SLSSLKIDDTPVDDPSLK(6)--QIIKR(4)	Fbxl3	180	106
KENQWCEEK(1)--TVANMIKGK(7)	Skp1dd	141	114
MKQIYQQLSR(2)--LKGFPSR(2)	mCRY2	503	455
IIELNGQKPPLTYKR(14)--LDKHLER(3)	mCRY2	169	241

K[156.08]ENQWCEEK(9)--KTFNIK(1)	Skp1dd	149	123
GKTPEEIRK(2)--GKTPEEIRK(2)	Skp1dd	116	116
KAWVANYERPR(1)--K[156.08]VKR(3)	mCRY2	246	295
K[156.08]TFNIKNDFTEEEEAQVR(6)--GKTPEEIR(2)	Skp1dd	128	116
SLSSLKIDDTPVDDPSLK(6)--RHSNHLQYVSFKVDSSK(12)	Fbxl3	180	118
KENQWCEEK(1)--KTFNIK(1)	Skp1dd	141	123
KAWVANYERPR(1)--LWDLYKK(6)	mCRY2	246	292
KENQWCEEK(1)--GKTPEEIRK(2)	Skp1dd	141	116
KAWVANYERPR(1)--LKGFPSR(2)	mCRY2	246	455
KAWVANYERPR(1)--K[155.09]VKR(3)	mCRY2	246	295
TVANMIKGK(7)--KTFNIK(1)	Skp1dd	114	123
KENQWCEEK(9)--KENQWCEEK(9)	Skp1dd	149	149
YIYEPWNAPESVQKAAK(14)--DAAIMKMAK(6)	mCRY2	474	133
KENQWCEEK(9)--GKTPEEIRK(2)	Skp1dd	149	116
LDKHLER(3)--KLNSR(1)	mCRY2	241	86
KAWVANYERPR(1)--LWDLYK[156.08]KVK(7)	mCRY2	246	293
MELPKK[156.08]PAVAVSSQQMESCR(5)--MELPKKPAVAVSSQQMESCR(6)	mCRY2	182	183
MELPKK[156.08]PAVAVSSQQMESCR(5)--MELPKKPAVAVSSQQMESCR(5)	mCRY2	182	182
GKTPEEIRKTFNIK[156.08]NDFTEEEEAQVR(9)--KENQWCEEK(1)	Skp1dd	123	141
KAWVANYERPR(1)--LWDLYK[155.09]KVK(7)	mCRY2	246	293
LFKEWGVTR(3)--LDKHLER(3)	mCRY2	107	241
IDDTPVDDPSLKVLVANNSDTLK(12)--SLSSLKIDDTPVDDPSLK(6)	Fbxl3	192	180
KTFNIK(1)--KTFNIK(1)	Skp1dd	123	123
LFKEWGVTR(3)--LKGFPSR(2)	mCRY2	107	455
IIELNGQKPPLTYKR(14)--KAWVANYERPR(1)	mCRY2	169	246
YIYEPWNAPESVQKAAK(14)--YLPKLK[156.08]GFPSR(4)	mCRY2	474	453

Supplementary Table 3: CRY2-FBXL3-SKP1 cross-linked site evaluation compared between the different algorithms and the published PDB structure.

Protein #1	Position #1	Protein #2	Position #2	Cα-Cα Distance (Å)[a]	Kojak	Protein Prospector	pLink
Fbxl3	1	Fbxl3	22	n/a	no	yes	no
Fbxl3	102	Fbxl3	180	23.83	yes	no	no
Fbxl3	102	Skp1dd	141	14.29	yes	yes	yes
Fbxl3	102	Skp1dd	149	n/a	yes	no	no
Fbxl3	106	Fbxl3	118	20.68	yes	yes	no
Fbxl3	106	Fbxl3	180	24.81	yes	yes	no
Fbxl3	106	Fbxl3	192	29.84	yes	yes	no
Fbxl3	106	Fbxl3	203	29.05	no	yes	no
Fbxl3	106	Skp1dd	141	17.20	yes	yes	yes
Fbxl3	118	Fbxl3	180	11.40	yes	yes	yes
Fbxl3	118	Fbxl3	203	21.47	yes	no	no
Fbxl3	118	Fbxl3	206	15.17	yes	no	no
Fbxl3	118	Skp1dd	114	25.73	yes	yes	yes
Fbxl3	118	Skp1dd	141	25.21	yes	no	no
Fbxl3	118	Skp1dd	149	n/a	yes	no	no
Fbxl3	180	Fbxl3	192	16.40	yes	yes	no
Fbxl3	180	Fbxl3	203	11.01	yes	yes	yes
Fbxl3	180	Fbxl3	206	4.55	yes	yes	yes
Fbxl3	180	mCRY2	503	14.81	yes	yes	yes
Fbxl3	192	Fbxl3	206	15.98	yes	yes	yes
Fbxl3	192	mCRY2	163	56.67	yes	yes	yes
Fbxl3	192	mCRY2	503	30.33	yes	no	no
Fbxl3	203	mCRY2	503	21.39	yes	yes	yes
Fbxl3	206	mCRY2	503	14.56	yes	yes	yes
Fbxl3	22	Fbxl3	106	n/a	yes	yes	yes
Fbxl3	22	Fbxl3	25	n/a	no	yes	no

Fbxl3	22	mCRY2	133	n/a	no	yes	yes
Fbxl3	22	mCRY2	136	n/a	yes	no	no
Fbxl3	22	mCRY2	503	n/a	yes	no	yes
Fbxl3	22	Skp1dd	114	n/a	yes	yes	yes
Fbxl3	22	Skp1dd	116	n/a	yes	yes	yes
Fbxl3	22	Skp1dd	123	n/a	yes	yes	yes
Fbxl3	22	Skp1dd	128	n/a	yes	yes	yes
Fbxl3	22	Skp1dd	141	n/a	yes	yes	yes
Fbxl3	22	Skp1dd	50	n/a	no	no	yes
Fbxl3	22	Skp1dd	51	n/a	yes	yes	yes
Fbxl3	24	Skp1dd	116	n/a	no	no	yes
Fbxl3	24	Skp1dd	123	n/a	no	yes	no
Fbxl3	24	Skp1dd	128	n/a	no	yes	no
Fbxl3	25	Fbxl3	118	n/a	no	yes	no
Fbxl3	25	Fbxl3	203	n/a	no	yes	no
Fbxl3	25	Skp1dd	123	n/a	no	yes	no
Fbxl3	25	Skp1dd	141	n/a	no	yes	no
Fbxl3	323	mCRY2	455	31.94	no	no	yes
Fbxl3	323	Skp1dd	141	62.95	no	no	yes
Fbxl3	416	mCRY2	169	17.94	yes	no	no
Fbxl3	94	Fbxl3	106	22.20	yes	no	no
Fbxl3	94	Skp1dd	149	n/a	yes	no	no
mCRY2	107	mCRY2	133	16.19	yes	yes	yes
mCRY2	107	mCRY2	136	13.76	yes	yes	yes
mCRY2	107	mCRY2	183	36.86	yes	no	no
mCRY2	107	mCRY2	241	45.67	yes	yes	yes
mCRY2	107	mCRY2	246	54.95	no	no	yes
mCRY2	107	mCRY2	455	61.99	yes	yes	yes
mCRY2	107	mCRY2	474	71.28	yes	no	no
mCRY2	125	mCRY2	133	12.42	yes	yes	no
mCRY2	125	mCRY2	136	16.43	yes	yes	yes
mCRY2	133	mCRY2	169	38.68	no	yes	no
mCRY2	133	mCRY2	241	43.73	no	yes	yes
mCRY2	133	mCRY2	455	59.55	no	yes	no
mCRY2	133	mCRY2	474	68.42	yes	no	no
mCRY2	136	mCRY2	474	71.99	yes	no	no
mCRY2	163	mCRY2	503	28.07	yes	no	no
mCRY2	169	mCRY2	241	30.33	yes	no	no
mCRY2	169	mCRY2	246	29.79	yes	no	no
mCRY2	182	mCRY2	182	n/a	yes	no	no
mCRY2	182	mCRY2	183	3.83	yes	no	no
mCRY2	182	mCRY2	246	27.29	no	yes	no
mCRY2	182	mCRY2	292	17.86	yes	no	no
mCRY2	183	mCRY2	246	24.72	no	yes	yes
mCRY2	183	mCRY2	292	15.86	yes	yes	yes
mCRY2	183	mCRY2	293	18.12	yes	yes	yes
mCRY2	183	mCRY2	295	21.28	yes	yes	yes
mCRY2	183	mCRY2	455	49.06	yes	yes	no
mCRY2	2	mCRY2	107	n/a	yes	yes	yes
mCRY2	2	mCRY2	133	n/a	yes	no	yes
mCRY2	2	mCRY2	183	n/a	yes	no	no
mCRY2	2	mCRY2	86	n/a	yes	yes	yes
mCRY2	241	mCRY2	246	10.31	yes	yes	yes
mCRY2	241	mCRY2	293	16.43	no	yes	no
mCRY2	241	mCRY2	295	19.32	no	yes	no
mCRY2	241	mCRY2	455	32.46	no	yes	yes
mCRY2	241	mCRY2	474	44.47	yes	yes	no
mCRY2	246	mCRY2	292	13.93	yes	yes	yes

mCRY2	246	mCRY2	293	11.71	yes	yes	yes
mCRY2	246	mCRY2	295	12.53	yes	yes	no
mCRY2	246	mCRY2	455	34.44	yes	yes	yes
mCRY2	246	mCRY2	503	46.16	no	yes	yes
mCRY2	29	mCRY2	107	25.14	no	yes	no
mCRY2	29	mCRY2	246	31.93	no	yes	no
mCRY2	29	mCRY2	503	40.93	no	yes	no
mCRY2	292	mCRY2	295	5.89	no	yes	no
mCRY2	453	mCRY2	474	12.06	yes	yes	yes
mCRY2	453	mCRY2	477	8.14	yes	yes	yes
mCRY2	455	mCRY2	474	12.11	yes	yes	yes
mCRY2	455	mCRY2	477	9.08	yes	yes	yes
mCRY2	455	mCRY2	503	40.36	yes	yes	yes
mCRY2	474	mCRY2	503	43.28	yes	no	no
mCRY2	86	mCRY2	107	28.14	no	yes	yes
mCRY2	86	mCRY2	133	34.35	no	yes	no
mCRY2	86	mCRY2	169	44.95	yes	yes	yes
mCRY2	86	mCRY2	183	25.34	yes	no	no
mCRY2	86	mCRY2	241	24.23	yes	yes	yes
Skp1dd	1	Skp1dd	22	n/a	yes	yes	yes
Skp1dd	1	Skp1dd	28	n/a	yes	no	no
Skp1dd	114	Skp1dd	123	14.19	yes	yes	yes
Skp1dd	114	Skp1dd	141	19.84	yes	yes	yes
Skp1dd	114	Skp1dd	149	n/a	yes	no	no
Skp1dd	116	Skp1dd	116	n/a	yes	no	no
Skp1dd	116	Skp1dd	123	10.14	yes	yes	yes
Skp1dd	116	Skp1dd	128	15.92	yes	yes	yes
Skp1dd	116	Skp1dd	141	17.27	yes	yes	yes
Skp1dd	116	Skp1dd	149	n/a	yes	no	no
Skp1dd	123	Skp1dd	123	n/a	yes	yes	no
Skp1dd	123	Skp1dd	141	21.65	yes	yes	yes
Skp1dd	123	Skp1dd	149	n/a	yes	no	no
Skp1dd	128	Skp1dd	141	20.56	yes	yes	yes
Skp1dd	128	Skp1dd	149	n/a	yes	no	no
Skp1dd	141	mCRY2	107	54.74	no	yes	yes
Skp1dd	141	mCRY2	241	68.08	no	yes	yes
Skp1dd	141	mCRY2	503	36.84	yes	yes	yes
Skp1dd	141	Skp1dd	141	n/a	no	no	yes
Skp1dd	149	Skp1dd	149	n/a	yes	no	no
Skp1dd	50	mCRY2	163	40.28	yes	no	no

[a] Distances were computed from PDB 4I6J. n/a refers to residues not included in the crystal structure.

Supplementary Table 4: Comparison of database search speeds for 991 MS/MS spectra using different approaches to spectral searching.

Method[a]	Protein Sequences	Peptide Sequences[b]	Peptide Spectrum Comparisons[c]	Time (min)
Full	20	2537	4.17E+05	0.1
Full	200	62057	2.92E+08	72.63
Full	2000	299192	4.20E+09	1098.8
Full	4000	547170	1.37E+10	3633.75
Two-pass	20	2537	3.21E+06	0.52
Two-pass	200	62057	7.28E+07	10.45
Two-pass	2000	299192	2.67E+08	39.72
Two-pass	4000	547170	4.79E+08	72.47

[a] Full refers to the exhaustive search of all peptide combinations within the precursor mass tolerance. Two-pass refers to the efficient search for single peptides with a variable modification mass.

[b] Only tryptic peptides with an internal lysine residue are counted.

[c] For the full search mode, this number represents the number of times two peptide sequences paired were compared to spectra. For the relaxed search mode, this is the number of times single peptide sequences were compared to spectra.

Supplementary Table 5: Kojak parameter descriptions

Parameter	Value	Description
cross_link	<number> <number> <number>	The first two numbers are the codes for the linker chemistry on either end (e.g. amine reactive, carboxyl reactive, etc.) The third number is the mass modification of the cross-linker.
database	<string>	A FASTA protein sequence database to parse for peptides. Should include decoy sequences.
decoy_filter	<string>	The identifier for all decoy sequences in the protein sequence database
enrichment	<number>	A value between 0 and 1 describing the 18O atom percent excess. For example, 0.25 is 25 APE. A value of 0 turns off enrichment.
enzyme	<string>	Set of characters that describe the enzyme digestion rules.
fixed_modification	<character> <number>	The amino acid character and modification mass for static modifications.
fragment_bin_offset	<number>	Offset to cross-correlation array
fragment_bin_size	<number>	Size (in Daltons) of the bins in the cross-correlation array.
instrument	<number>	0=Orbitrap, 1=FTICR
max_miscleavages	<string>	The maximum number of missed trypsin cleavages allowed.
max_mods_per_peptide	<number>	Maximum number of differential modifications allowed on a peptide
max_peptide_mass	<number>	The largest peptide mass allowed
max_spectrum_peaks	<number>	The top N MS/MS peaks to use during analysis. Setting to 0 uses all peaks.
min_peptide_mass	<number>	The smallest peptide mass allowed
modification	<character> <number>	The amino acid character and modification mass for differential modifications. @=N-terminal modification
mono_link	<number> <number>	A combination of linker code and additional mass of a half-reacted cross-linker.
MS_data_file	<string>	The mass spectrometry data file. Acceptable formats include .mzXML and .mzML
MS1_centroid	<number>	Indicates if MS data are already centroided. 0=no, 1=yes
MS1_resolution	<number>	Estimated MS resolution at 400 m/z
MS2_centroid	<number>	Indicates if MS/MS data are already centroided. 0=no, 1=yes
MS2_resolution	<number>	Estimated MS/MS resolution at 400 m/z
output_file	<string>	A file name to hold the Kojak output.
percolator_file	<string>	A file name to hold the Kojak output in a format that can be directly imported to Percolator.
percolator_version	<number>	The version number of Percolator. This outputs the results to the correct format.
ppm_tolerance_pre	<number>	Mass tolerance on precursor mass when matching peptides.
prefer_precursor_pred	<number>	Use pre-computed accurate precursor m/z value, if present in data file. 0=no, 1=yes.
relaxed_analysis	<number>	Uses faster relaxed analysis mode instead of exhaustive search mode. 0=off, 1=on
search_dimers	<number>	Tests for non-covalent dimerizations. These dimerizations are not the result of cross-linking but do occur very rarely during analysis, in which two peptides are isolated as a single molecule with a combined mass. 0=no, 1=yes
spectrum_processing	<number>	Collapses MS/MS isotope peaks to the monoisotopic m/z. 0=off, 1=on
threads	<number>	Number of threads to use on multi-core systems
top_count	<number>	Top N single peptides to combinatorially mine during relaxed analysis mode
use_comet_xcorr	<number>	Use Comet's Xcorr scoring metric instead of default Kojak scoring metric. 0=no, 1=yes

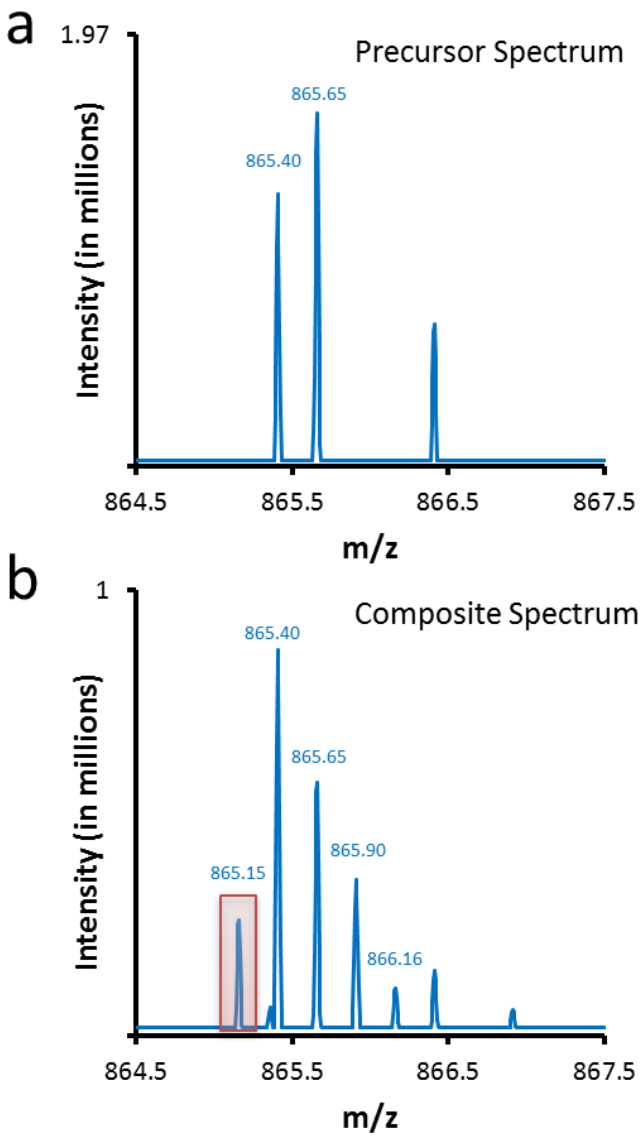
Supplementary Table 6: Kojak PSMs for Cop9 and 26S at 5% FDR

Interprotein XL			
Dataset	PSMs	Non-redundant peptide sequences	Unique XL sites[1]
Cop9 Signalosome	223	45	37
26S proteasome	130	43	41

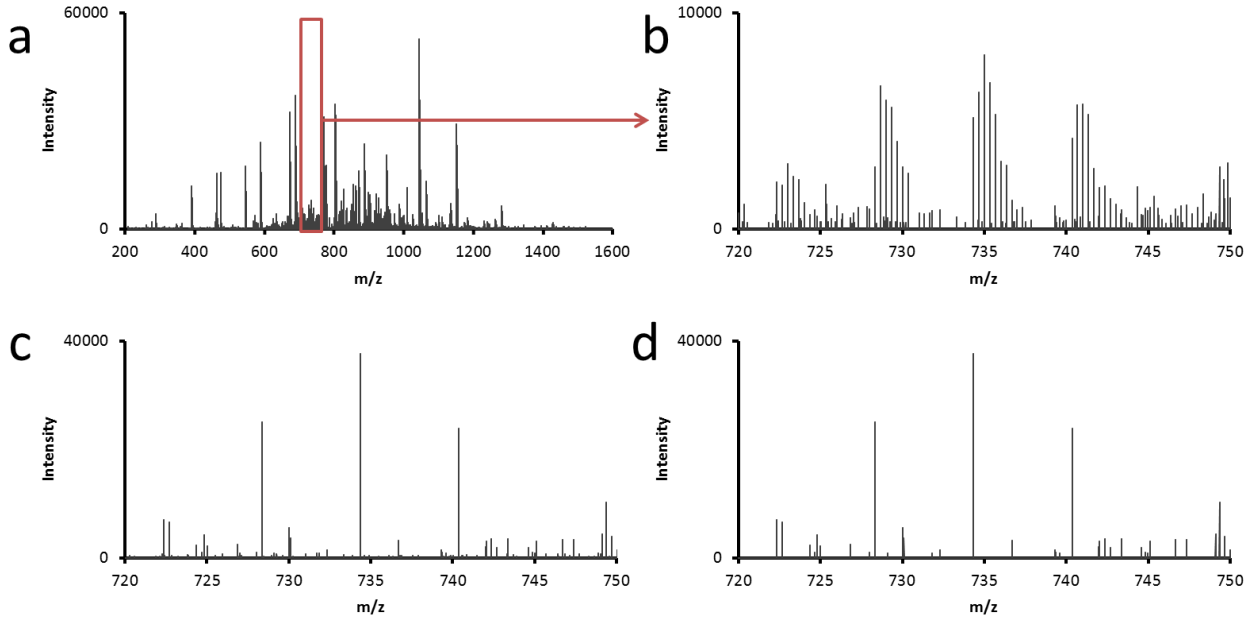
Intraprotein XL			
Dataset	PSMs	Non-redundant peptide sequences	Unique XL sites
Cop9 Signalosome	544	103	57
26S proteasome	485	216	198

[1]Represents the novel linkage between two residues by absolute protein position, thus removing redundant associations derived from elongated peptide sequences that result from enzymatic missed cleavages.

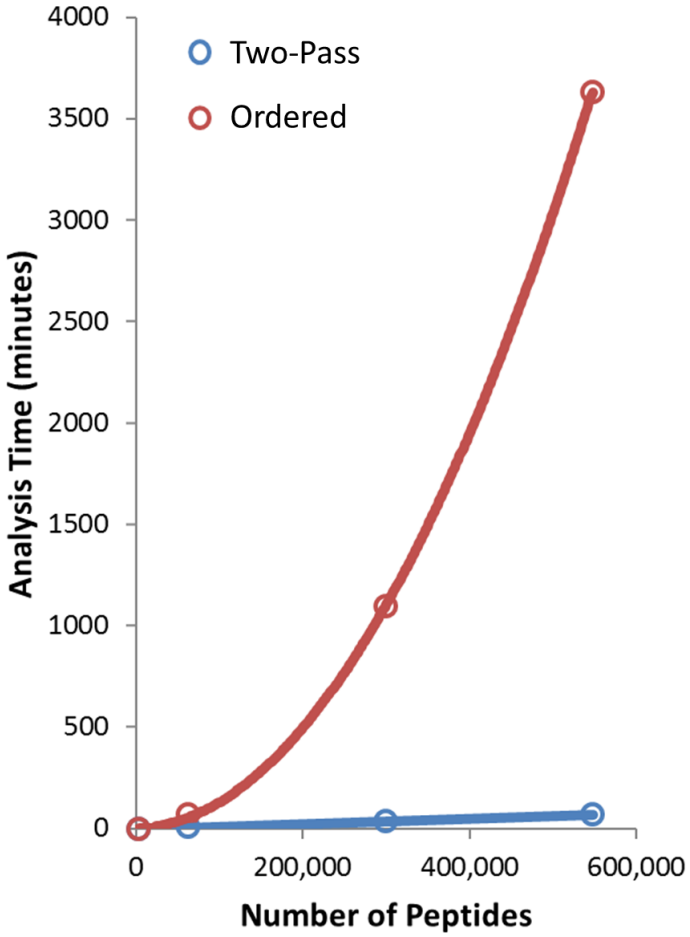
7. Supplementary Figures



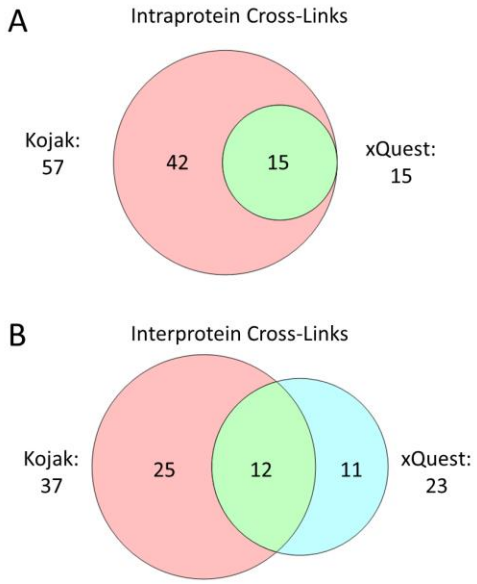
Supplementary Figure 1: Precursor ion spectral processing. (a) The observed precursor ion isotopic cluster is difficult to resolve because of low intensity and missing isotope peaks at the time of selection for MS/MS. (b) A composite spectrum is computed by averaging the ion signals across neighboring scans. The composite spectrum correctly illustrates a larger portion of the precursor isotope envelope. The monoisotopic peak is now resolved (red box) where it could not be identified from the original spectrum.



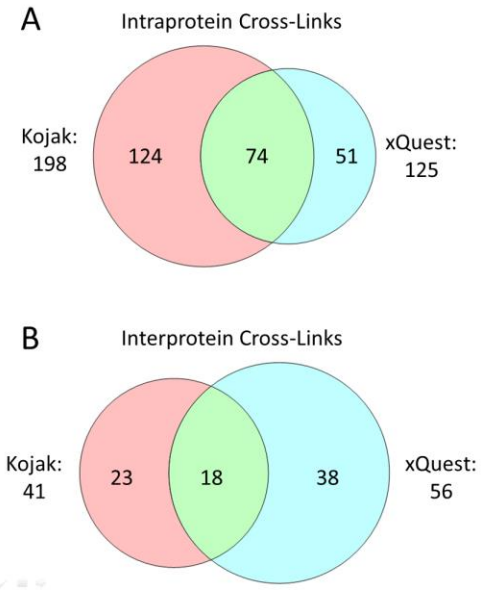
Supplementary Figure 2: MS/MS spectral processing. A raw spectrum is shown in (a) with fine detail for a narrow m/z region (b). (c) Isotopic peak clusters are summed and collapsed to the monoisotopic peak, simultaneously increasing signal strength and reducing the number of peaks in the spectrum. (d) Additional noise reduction is performed by keeping only the top 500 peaks in the spectrum.



Supplementary Figure 3: Computation times for the two search modes. The full search mode time increases quadratically when increasing the size of the search database. The relaxed search mode time remains linear with increasing database size, making possible large proteome database searches.



Supplementary Figure 4: Comparison of Cop9 signalosome intraprotein and interprotein cross-links.



Supplementary Figure 5: Comparison of 26S proteasome intraprotein and interprotein cross-links.

8. Supplementary References

1. Xing, W.; Busino, L.; Hinds, T. R.; Marionni, S. T.; Saifee, N. H.; Bush, M. F.; Pagano, M.; Zheng, N., SCF(FBXL3) ubiquitin ligase targets cryptochromes at their cofactor pocket. *Nature* **2013,** *496* (7443), 64-8.
2. Eng, J. K.; Jahan, T. A.; Hoopmann, M. R., Comet: an open-source MS/MS sequence database search tool. *Proteomics* **2013,** *13* (1), 22-4.
3. Trnka, M. J.; Baker, P. R.; Robinson, P. J.; Burlingame, A. L.; Chalkley, R. J., Matching cross-linked peptide spectra: only as good as the worse identification. *Molecular & cellular proteomics : MCP* **2014,** *13* (2), 420-34.
4. Walzthoeni, T.; Claassen, M.; Leitner, A.; Herzog, F.; Bohn, S.; Forster, F.; Beck, M.; Aebersold, R., False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nature methods* **2012,** *9* (9), 901-3.
5. Hsieh, E. J.; Hoopmann, M. R.; MacLean, B.; MacCoss, M. J., Comparison of database search strategies for high precursor mass accuracy MS/MS data. *Journal of proteome research* **2010,** *9* (2), 1138-43.
6. Hoopmann, M. R.; Finney, G. L.; MacCoss, M. J., High-speed data reduction, feature detection, and MS/MS spectrum quality assessment of shotgun proteomics data sets using high-resolution mass spectrometry. *Analytical chemistry* **2007,** *79* (15), 5620-32.
7. Choi, S.; Jeong, J.; Na, S.; Lee, H. S.; Kim, H. Y.; Lee, K. J.; Paek, E., New algorithm for the identification of intact disulfide linkages based on fragmentation characteristics in tandem mass spectra. *Journal of proteome research* **2010,** *9* (1), 626-35.
8. Petrotchenko, E. V.; Borchers, C. H., Application of a fast sorting algorithm to the assignment of mass spectrometric cross-linking data. *Proteomics* **2014,** *14* (17-18), 1987-9.
9. Singh, P.; Shaffer, S. A.; Scherl, A.; Holman, C.; Pfuetzner, R. A.; Larson Freeman, T. J.; Miller, S. I.; Hernandez, P.; Appel, R. D.; Goodlett, D. R., Characterization of protein cross-links via mass spectrometry and an open-modification search strategy. *Analytical chemistry* **2008,** *80* (22), 8799-806.
10. Yang, B.; Wu, Y. J.; Zhu, M.; Fan, S. B.; Lin, J.; Zhang, K.; Li, S.; Chi, H.; Li, Y. X.; Chen, H. F.; Luo, S. K.; Ding, Y. H.; Wang, L. H.; Hao, Z.; Xiu, L. Y.; Chen, S.; Ye, K.; He, S. M.; Dong, M. Q., Identification of cross-linked peptides from complex samples. *Nature methods* **2012,** *9* (9), 904-6.
11. Zelter, A.; Hoopmann, M. R.; Vernon, R.; Baker, D.; MacCoss, M. J.; Davis, T. N., Isotope signatures allow identification of chemically cross-linked peptides by mass spectrometry: a novel method to determine interresidue distances in protein structures through cross-linking. *Journal of proteome research* **2010,** *9* (7), 3583-9.
12. Chavez, J. D.; Weisbrod, C. R.; Zheng, C.; Eng, J. K.; Bruce, J. E., Protein interactions, post-translational modifications and topologies in human cells. *Molecular & cellular proteomics : MCP* **2013,** *12* (5), 1451-67.