Supporting Information

Optimizing the Photocontrol of bZIP Coiled Coils with Azobenzene Crosslinkers: Role of the Crosslinking Site


cbic_201500191_sm_misellaneous_information.pdf
EMSA analysis

CREB bZIP CRE binding assay:

Using Igor Pro software, three sets of data were averaged and fit to the Hill equation to determine the apparent $K_d$, i.e., the concentration of protein required for 50% binding to the CRE site:

$$\% \text{CRE bound} = \frac{\text{base} - \text{(max - base)}}{1 + \left(\frac{K_d}{[\text{CRE}]_{\text{app}}}\right)^n}$$

(Eq. 1)

Where $[\text{CRE}]_{\text{is}}$ taken to be the total concentration of CREB bZIP added. Data are shown in Figure S1.

**Figure S1.** EMSA analysis of CREB bZIP binding to CRE site DNA (Cy5-labeled oligo) performed as described in the Experimental section. The apparent $K_d$ is 10 ± 3 nM.

CREB bZIP CRE binding inhibition assay:

The extent of CREB bZIP CRE binding inhibition was quantified by analysis with Image Lab® software (Bio-Rad). Data were fit to a modified Hill equation:
\[
\% \text{CRE bound} = \text{base} - \frac{(\text{max} - \text{base})}{\left[1 + (K_{d_{DNA}}/[\text{CRE}])^n\right]} \quad \text{(Eq. 2)}
\]

Where \(K_{d_{DNA}}\) is determined in the CREB-bZIP/CRE titration described above and [CREB] is given by Eq. 3:

\[
[\text{CREB}]_{\text{free}} = [\text{CREB}]_{\text{total}} - \frac{(K_i + [\text{CREB}]_{\text{total}} + [I]_{\text{total}}) - \sqrt{(K_i + [\text{CREB}]_{\text{total}} + [I]_{\text{total}})^2 - 4[\text{CREB}]_{\text{total}}[I]_{\text{total}}}}{2}
\]

Where \([\text{CREB}]_{\text{total}}\) and \([I]_{\text{total}}\) are the total concentrations of CREB bZIP (fixed) and the dominant negative inhibitor (variable) and \(K_i\) is the dissociation constant for the CREB/inhibitor complex.

(a)
(b)

Dark adapted

365 nm irradiated

[-Ve +Ve 4 6 10 20 40 60 80 100 200 400 600 800 1K 2K 4K
[A2-CREB-x-1] nM]
Figure S2. (a) Representative EMSA gel showing A2-CREB inhibition of CREB bZIP binding to CRE site DNA (Cy5-labeled oligo) performed as described in the Experimental section. Fraction bound is plotted in Fig. 3(a). (b). Representative EMSA gel showing A2-CREB-x-1 inhibition of CREB bZIP binding to CRE site DNA (Cy5-labeled oligo) performed as described in the Experimental section. Fraction bound is plotted in Fig. 3(b). (c). Representative EMSA gel showing A2-CREB-x-4 inhibition of CREB bZIP binding to CRE site DNA (Cy5-labeled oligo) performed as described in the Experimental section. Fraction bound is plotted in Fig. 3(c).
Figure S3. Representative EMSA gel showing A4-CREB inhibition of CREB bZIP binding to CRE site DNA (Cy5-labeled oligo) performed as described in the Experimental section. Fraction bound is plotted in Fig. 5(a).
Figure S4. Representative EMSA gel showing A4-CREB-x-4 inhibition of CREB bZIP binding to CRE site DNA (Cy5-labeled oligo) performed as described in the Experimental section. Fraction bound is plotted in Fig. 5(b).
Mass spectrometry analysis

The azobenzene cross-linked peptides were analyzed using high-performance liquid chromatography coupled with electrospray ionization tandem mass spectrometry (HPLC-MS/MS). Sequencing was carried out using collision-induced dissociation (CID) and all mass spectra were acquired in positive ion mode using a 6538 UHD model quadrupole time-of-flight mass spectrometer equipped with a 1200 Infinity Series HPLC (Agilent Technologies, Santa Clara, CA). Mass spectra were acquired in the 2 GHz extended dynamic range mode and mass calibrations were applied externally without the use of internal reference re-calibration or lock-mass correction.

Two separate strategies for the mass spectrometry analyses were used: (1) top-down MS/MS and (2) bottom-up MS/MS. The top-down approach involves fragmentation of the intact cross-linked peptide followed by sequence analysis of all fragment ions simultaneously. This method requires only small sample quantities (e.g. pmol) and can readily provide extensive sequence information for small- to medium-sized proteins. However, as the size of the protein increases (e.g. > 5-7 kDa) and given the mass resolving power of the TOF instrumentation used here, there is a practical upper limit for the molecular weight of a protein which can be fully sequenced by top-down methods. To complement the top-down analysis and to extend the ability to characterize higher molecular weight targets, a second analysis using bottom-up MS/MS sequencing was incorporated. The bottom-up approach involves first proteolysis of the target with trypsin followed by sequencing of the tryptic peptides via HPLC-MS/MS.

The analysis of MS/MS data obtained from cross-linked peptides requires more involved methods than standard sequencing algorithms to accommodate the modifications to the peptide primary sequence. Therefore, a method was developed using the Protein Prospector on-line resource (prospector.ucsf.edu/) to enable specific sequence analysis of the cross-linked peptides. To illustrate these methods, the following sections show two parallel examples describing MS/MS sequencing of the cross-linked peptide A2-CREB-
x-5. Figure S5 shows the sequence and cross-link location of (a) the intact A2-CREB-x-5 peptide and (b) the tryptic peptide containing the azobenzene cross-link following digestion with trypsin.

![Diagram](image)

**Figure S5.** Sequence and cross-link location of (a) intact A2-CREB-x-5 peptide and (b) a tryptic peptide following digestion. The regions highlighted in blue (including the azobenzene cross-linker) are identified as the custom amino acid (u) for use with the Protein Prospector *MS-Product* tool and have the molecular formulae indicated.

**Top-down MS/MS analysis of intact peptides**

Intact peptides were mass analyzed following online de-salting using a Tricorn 5/50 column packed with Sephadex G-25 size-exclusion media (GE Healthcare Life Sciences, Baie d’Urfe, QC). The mobile phase was 1:1 (v/v) 0.1% aqueous formic acid: methanol flowing at a rate of 100 µL min\(^{-1}\). MS/MS analysis was accomplished using the *TargetedMS2* acquisition mode. The highest abundance precursor charge states were selected for analysis and MS/MS spectra were recorded at sequential collision energies (e.g. 35 V, 40 V, 45 V) to maximize fragment ion sequence coverage.
Peptide sequencing was accomplished using the Protein Prospector *MS-Product* data analysis tool and Figure S6 shows a screenshot of the input dialogue with the input sequence for A2-CREB-x-5. N- and C-terminal sequencing of the cross-linked protein was facilitated by inserting a user-defined amino acid (u, see Figure S5(a)) in place of the cross-linked portion of the sequence. The MS/MS mass lists from each charge state were combined and searched against the b- and y-series fragment ions computed from the protein sequence using a mass tolerance of 10 ppm.

![Figure S6](image)

*Figure S6.* A screenshot showing the *MS-Product* data input dialogue from Protein Prospector (prospector.ucsf.edu/). The example shows the sequence input for A2-CREB-x-5 and the user specified amino acid u=C₉₅H₁₆₅N₁₄O₁₂S₂ shown in Figure S2(a).

Figure S7 shows the results from the top-down sequencing of A2-CREB-x-5. MS/MS data were extracted using the MassHunter software package to obtain reconstructed, collision energy-averaged mass spectra for each targeted precursor charge state. As can be seen from Figure S7, top-down sequencing even of
moderately-sized protein such as A2-CREB-x-5 (~5 kDa) results in complicated MS/MS spectra; therefore, all of the MS/MS spectra were manually checked in conjunction with the MS-Product results to confirm all of the matching sequence ions. The results shown in Figure S7 demonstrate that nearly complete sequence coverage for A2-CREB-x-5 is achieved, with the exception of the cross-linked region of the protein. The cross-link appears to effectively suppress fragmentation between Cys residues and no clear patterns or corresponding sequences could be identified which could be attributed to cleavage of the azobenzene cross-link and subsequent formation of sequence ions.

Figure S7. Top-down MS/MS analysis of intact A2-CREB-x-5 using the MS-Product online tool from Protein Prospector. Sequence coverage is indicated by the symbols for the b- (⊥) and y-series ions (⊥). The MS/MS spectrum shows all fragment ions in the 1+ charge state and the matching peaks are highlighted in red and are labeled with the appropriate b- or y-series sequence match. The mass errors of all the identified sequence ions are also shown.
Bottom-up Sequencing

(i) HPLC-MS/MS Analysis of Tryptic Digests

To 50 μL of a solution of 50 μM dominant negative peptide in 10 mM sodium phosphate buffer pH 7.0 was added 5 μL of 500 mM ammonium bicarbonate solution (pH 8) followed by 5 μL of Trypsin Gold® (1 μg/μL in 50 mM acetic acid). The pH of the mixture was adjusted to pH 8 using the ammonium bicarbonate solution and the mixture was incubated in the dark for overnight at 37°C. The following day the samples were analyzed using tandem mass spectrometry.

Tryptic digests were analyzed by data-dependent MS/MS (AutoMS2) following separation by reverse phase chromatography. The column was a 50x2 mm Jupiter Proteo C12 stationary phase (Phenomenex, Torrance, CA) and the mobile phase components (A and B) were 0.1% aqueous formic acid and acetonitrile, respectively, flowing at a rate of 250 μL min⁻¹. Peptides were separated over a 15 minute gradient from 5 to 40% B followed by a 3 minute ramp to 100% B. The AutoMS2 parameters were adjusted to target 3 precursor ions for MS/MS (3 Hz) per MS cycle (4 Hz). The highest intensity precursor ions were targeted first and subsequently prioritized by charge state (2+, 3+, 1+). Active exclusion was also implemented to target lower-intensity co-eluting precursor ions.

(ii) Data analysis

HPLC-MS/MS data were analyzed using the MassHunter Qualitative Analysis software package version B.06.00 SP1 with BioConfirm B.06.00 (Agilent Technologies, Santa Clara, CA). The total ion chromatogram (TIC) was analyzed using the Find by Molecular Feature algorithm to extract molecular weights from compounds in the MS analysis. Protein sequences and cross-link modifications were defined and digested in silico using MassHunter. Tryptic peptides were identified using intact peptide masses (MS) and sequence coverage was confirmed using peptide fragmentation (MS/MS). The mass error tolerance for sequence matching was 5 ppm.
Figure S8 shows the results of the bottom-up peptide sequencing following tryptic digestion of A2-CREB-x-5. The upper panel (a) shows a legend defining the sequence information displayed including the tryptic peptides T1, T2, etc. (green bars), cross-linked tryptic peptide T4+T6 and the b- and y-series fragment ions. Panel (b) shows the TIC and with the peaks from the tryptic peptides identified. In addition to each of the peptides T1, T2, T3, T5 and T4+T6 there is also a partially digested peptide identified as T3-T6. It was also observed that the protein was successfully digested within the cross-linked portion of the sequence at two Lys residues.

The MassHunter software package was used to sequence the non-crosslinked peptides shown in Figure S8(a); however, the software does not support MS/MS sequencing of the cross-linked peptides. Therefore, MS-Product was also applied to the tryptic cross-linked tryptic peptide as was done for the top-down analysis. The user-defined amino acid (u, see Figure S5(b)) was inserted in place of the cross-linked portion of the sequence the b- and y-series fragment ions were matched using a mass tolerance 5 ppm. The MS/MS spectrum following fragmentation of the cross-linked tryptic peptide T4+T6 is shown in Figure S8(c). In this case, additional sequence information is obtained where fragmentation has occurred along the cross-linked region. In particular, y-series sequence ions adjacent to the C-terminal Cys (C-ELK) were observed from the digested peptide which were not observed in the top-down analysis.
Figure S8. Results from the bottom-up sequencing of A2-CREB-x-5, following digestion with trypsin and HPLC-MS/MS analysis. (a) Sequence coverage map showing the identified tryptic peptides and the b- and y-series fragment ions. (b) The total ion chromatogram (TIC) from the HPLC-MS and the identified peaks. (c) MS/MS spectrum of the cross-linked tryptic peptide from xCZRS and the sequence ions identified using MS-Product.