

# Structural Analysis of Multiprotein Complexes by Cross-linking, Mass Spectrometry, and Database Searching\*

Alessio Maiolica<sup>‡§</sup>, Davide Cittaro<sup>‡§</sup>, Dario Borsotti<sup>‡§¶</sup>, Lau Sennels<sup>¶</sup>, Claudio Ciferri<sup>||\*\*</sup>, Cataldo Tarricone<sup>||</sup>, Andrea Musacchio<sup>||</sup>, and Juri Rappsilber<sup>¶††</sup>

**Most protein complexes are inaccessible to high resolution structural analysis. We report the results of a combined approach of cross-linking, mass spectrometry, and bioinformatics to two human complexes containing large coiled-coil segments, the NDEL1 homodimer and the NDC80 heterotetramer. An important limitation of the cross-linking approach, so far, was the identification of cross-linked peptides from fragmentation spectra. Our novel approach overcomes the data analysis bottleneck of cross-linking and mass spectrometry. We constructed a purpose-built database to match spectra with cross-linked peptides, define a score that expresses the quality of our identification, and estimate false positive rates. We show that our analysis sheds light on critical structural parameters such as the directionality of the homodimeric coiled coil of NDEL1, the register of the heterodimeric coiled coils of the NDC80 complex, and the organization of a tetramerization region in the NDC80 complex. Our approach is especially useful to address complexes that are difficult in addressing by standard structural methods. *Molecular & Cellular Proteomics* 6:2200–2211, 2007.**

Mass spectrometry-based proteomics is a powerful tool for the analysis of multiprotein complexes (1). Thousands of complexes have been isolated, and their protein compositions have been determined (2–4). Although many complexes will feed into large scale crystallization trials, only a few are likely to reveal their structure. Many protein complexes are heterogeneous, insoluble at the concentrations needed for crystallization, or yield crystals lacking the quality needed for structure determination. When structures are obtained they often comprise only parts of the proteins because difficult areas have been removed to increase solubility or crystallization properties. Cross-linking in conjunction with mass spectrometry is a very promising tool to yield structural information on

proteins and protein complexes that is difficult to address using standard structural methods (5). Just as mass spectra can reveal the identity of the protein components of a complex, if the complex or protein has been cross-linked mass spectra can be used to identify direct proximity of proteins in a complex (6) and aid fold recognition of proteins (7). Although this has been shown in proof of principle (6, 7), general application has yet to be achieved.

Success of mass spectrometry in identifying proteins is largely due to the apparent simplicity and to the automation of protein identification using mass spectrometric data. Three features are central to the automation. (a) Based on an observed peptide mass a list of candidate peptides can be extracted from protein databases. (b) The candidate peptides can be evaluated by assessing their match to the fragmentation spectrum, resulting in a single number, the score. (c) The rate of false identifications can be estimated by computing the likelihood of a random hit. Unfortunately this straightforward automation procedure could so far not be applied to cross-linked peptides.

In the absence of automatic tools similar to those used for normal peptide identification, cross-linking cannot be used routinely for structural analysis of multiprotein complexes. Indeed work based on identifying cross-linked peptides has so far been limited to complexes composed of not more than two different proteins (5). Standard database search tools cannot create a list of candidate cross-linked peptides based on the observed mass. A number of dedicated programs consider pairs of peptides contributing together to the observed mass (7–17). The candidates then need to be validated on the basis of their match to fragmentation spectra. Currently this requires screening the spectra either completely manually or through software assistance (13, 15, 16, 18). No scoring system or algorithm has yet been developed to replace human intervention. Ideally such a scoring algorithm would objectively sort the false from the true matches and add a measure of confidence to the results.

Here we present an algorithm that automatically finds and validates cross-linked peptides using fragmentation spectra, thereby overcoming the key limitation in the analysis of protein cross-links. Proteins are cross-linked using a 1:1 mixture of stable isotope-labeled and non-labeled cross-linker to re-

From the <sup>‡</sup>FIRC Institute of Molecular Oncology Foundation, Via Adamello 16, 20139 Milan, Italy, <sup>¶</sup>Wellcome Trust Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, United Kingdom, and <sup>||</sup>Department of Experimental Oncology, European Institute of Oncology, Via Ripamonti 435, 20141 Milan, Italy

Received, June 12, 2007, and in revised form, October 3, 2007

Published, MCP Papers in Press, October 5, 2007, DOI 10.1074/mcp.M700274-MCP200

duce false positive rates of the process. Proteins are digested using trypsin, and peptides are analyzed by LC-MS/MS prior to data analysis with our algorithm (see Fig. 1). The algorithm was applied to data acquired from two human coiled-coil complexes, the NDEL1-(17–174) homodimer (38 kDa) and the NDC80 heterotetramer (176 kDa). We used a standard database search tool, Mascot (19), and a purpose-built cross-link database (XDB)<sup>1</sup> that contains cross-linked peptides represented as single linear peptides. Our algorithm assigns a score and describes the confidence of each match through comparison with negative controls.

#### EXPERIMENTAL PROCEDURES

**Purification of Protein Complexes**—The NDC80 complex was purified according to our published procedure (20).

NDEL1-(17–174) was purified as described earlier (21) with the following changes. Polymerase chain reaction fragments of human NDEL1 corresponding to amino acids 17–174 of the full-length protein were subcloned in the pGEX6P-1 expression vector (GE Healthcare) and expressed in *Escherichia coli* strain BL21(DE3). The protein was purified by glutathione affinity chromatography. The GST tag was removed using Prescission protease (GE Healthcare), and the resulting sample was further purified by size exclusion chromatography using a Superdex 200 column equilibrated with 10 mM Hepes, pH 7.5, 100 mM NaCl. Treatment of the GST fusion protein with the Prescission protease leaves a 5-residue extension at the N terminus of NDEL1-(17–174), numbered –5 to –1.

**Cross-linking**—Human NDEL1-(17–174) (29  $\mu$ g of protein equivalent to 775 pmol) and human NDC80 complex (15  $\mu$ g of protein equivalent to 86 pmol) were mixed with a 100 $\times$  excess of isotope-labeled cross-linker bis(sulfosuccinimidyl)glutarate (BS<sup>2</sup>G) (Pierce) in a final volume of 150  $\mu$ l of 10 mM Hepes, pH 7.5, 100 mM NaCl at room temperature. The cross-linker, a 1:1 mixture of light BS<sup>2</sup>G-*d*<sub>0</sub> and heavy BS<sup>2</sup>G-*d*<sub>4</sub>, was freshly prepared as a 10 nmol/ $\mu$ l solution in DMSO. The reaction was stopped after 30 min by adding 5  $\mu$ l of 1 M ammonium bicarbonate. Sample buffer was added for separation by SDS-PAGE.

**Digestion**—The samples were electrophoresed through Novex NuPAGE 1-mm 4–12% Tris-glycine gels (Invitrogen) in MOPS buffer (Invitrogen), fixed in 50% methanol, 5% acetic acid, and stained with the colloidal blue kit (Invitrogen). Bands were excised and processed following a standard trypsin digestion procedure (22): reduction in 100 mM DTT for 30 min at room temperature, alkylation with 55 mM iodoacetamide for 30 min at room temperature in the dark, and digestion with 12.5 ng/ $\mu$ l trypsin (proteomics grade, Sigma) overnight at 37 °C. The supernatant was loaded onto StageTips (23), and peptides were eluted in 20  $\mu$ l of 80% acetonitrile, 0.1% trifluoroacetic. The acetonitrile was allowed to evaporate off (Concentrator 5301, Eppendorf AG, Hamburg, Germany), and the volume of each eluate was adjusted to 5  $\mu$ l with 1% trifluoroacetic acid of which 2.5  $\mu$ l, *i.e.* half, were injected for LC-MS/MS analysis.

**Nano-LC-MS/MS and Data Analysis**—The proteins, after digestion with trypsin, were analyzed by LC-MS/MS using an HPLC system (1100 binary nanopump, Agilent, Palo Alto, CA) coupled on line to an ion trap FTICR hybrid mass spectrometer (LTQ-FT, ThermoElectron, Bremen, Germany). C<sub>18</sub> material (ReproSil-Pur C18-AQ 3  $\mu$ m, Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) was packed into a spray emitter (75- $\mu$ m inner diameter, 8- $\mu$ m opening, 70-mm length; New Objectives) using an air pressure pump (Proxeon Biosystems,

Odense, Denmark) to prepare an analytical column with a self-assembled particle frit (24). Mobile phase A consisted of water, 5% acetonitrile, and 0.5% acetic acid, and mobile phase B consisted of acetonitrile and 0.5% acetic acid. The samples were loaded from an Agilent 1100 autosampler onto the column at a 700 nl/min flow rate. The gradient had a flow rate of 300 nl/min, and the percentage of buffer B varied linearly from 0 to 20% in the first 77 min and then from 20 to 80% in a further 15 min. We used a SIM method for mass acquisition (25) with one low resolution FT-MS scan (fill target, 1,000,000 ions; resolution, 25,000; maximum fill time, 2 s; mass range, *m/z* 300–1575). The three most intense signals (dynamic exclusion for 180 s) were selected for SIM (fill target, 500,000 ions; maximum fill time, 50 ms; window, *m/z* 22) in the FTICR cell and MS<sup>2</sup>/MS<sup>3</sup> in the ion trap (normal scan; wideband activation; fill target, 10,000 ions; maximum fill time, 100 ms). Each cycle lasted  $\sim$ 3 s.

In principle, precursor selection for MS/MS could be directed onto doublet signals focusing the fragmentation on candidate cross-linked peptides. However, the low data quality of the usually weak signals of cross-linked peptides in a full FT-MS spectrum makes an FT-SIM scan necessary for reliable observation of both signals of a doublet. Directed selection of precursors would require the acquisition of the fragmentation spectrum to follow the SIM scan. However, the MS/MS spectrum can be recorded in the ion trap part of our LTQ-FT in parallel to the SIM scan being recorded in the FT cell. Recording both spectra in parallel, regardless of the multiplicity of the precursor, is more time-economic than recording first the SIM and then an MS/MS scan for peptides with doublet signals. The doublet information is instead used for post-acquisition data filtering. High mass accuracy FT-MS/MS would have a different time economy. Because the MS/MS spectrum has to be recorded after the MS and SIM spectra, doublet directed sequencing is highly advisable.

Peaks were picked from the raw data files using DTA Supercharge (version 0.94, made available by SourceForge, Inc.) with the following settings: precursor mass deviation, *m/z* 0.08; smart picking for MS/MS activated; maximum search level, 8. The four lists, one for each band in the gel, contained in total the fragment information of 14,871 precursors. A peak list was then created for light precursors, and a corresponding peak list was created for heavy precursors. The apparent occurrence of isotopic doublets, which indicates the presence of the cross-linker, was used to enrich the dataset for spectra of cross-linked peptides. For the selection of doublets, all SIM scans were extracted from the raw data files by a custom program written in the “.NET”-integrated programming language C# using the XDA-api (Xcalibur Development kit, Thermo Inc.). We then extracted, for all precursors in the complete list, the *m/z*, charge state, and scan number. The appropriate scan was located in the SIM file, and it was determined whether the precursor had a partner signal with intensity 0.4–2.5 $\times$  at plus or minus  $4.025 \pm 0.01$  Da. The partner intensity threshold was imposed to take into account that peptides containing the non-deuterated and deuterated cross-linker show shifted elution profiles due to the difference in isotope composition of the two species. This shift can result in partner peak intensity ratios different from 1:1, depending on the timing of the SIM acquisition with respect to the elution profiles of the peptides. The threshold values were determined by inspecting SIM scans of peptides that we identified as being modified with the hydrolyzed cross-linker. If there was a matching signal above the precursor *m/z*, the precursor was taken as a candidate peptide containing the light form of the cross-linker, and the MS/MS peak list of the precursor was added to the peak list of light precursors. Equivalently if a matching signal was found below the precursor *m/z*, the precursor was taken as a candidate peptide containing the heavy form of the cross-linker, and the MS/MS peak list of the precursor was added to the peak list of heavy precursors. This process resulted in a total of 1452 queries, *i.e.* 10% of the acquired data.

<sup>1</sup> The abbreviations used are: XDB, cross-link database; BS<sup>2</sup>G, bis(sulfosuccinimidyl)glutarate; SIM, selected ion monitoring.

**Peptide Identification via Mascot Database Search**—The complete peak list for each band from the gel was searched against Swiss-Prot ([www.expasy.org/sprot](http://www.expasy.org/sprot)), to which the exact sequences of the recombinant proteins under investigation were added, using Mascot (version 2.0) with the following parameters: monoisotopic mass values; peptide tolerance, 0.08 Da; MS/MS tolerance, 0.5 Da; instrument, ESI-TRAP; fully tryptic specificity; cysteine carbamidomethylation as fixed modification; oxidation on methionine and hydrolyzed cross-linker on protein N terminus, lysine, serine, and tyrosine as variable modifications; two missed cleavage sites allowed. The results of this first search were used in three ways. First, the peptides matching the protein complex members were used to estimate the mass accuracy of the analysis. We took as the mass accuracy the mass deviation that included 97% of the identified peptides (588 peptides). This value was ~4.5 ppm for all four sets of data. For comparison, the average deviation was 1.3 ppm. Second, we could see the extent of side reaction of the amine-specific cross-linker with serine and tyrosine. We detected only a very small number of serine-containing peptides being modified, and there was no indication of tyrosine modification. Therefore, we did not further consider these modifications in our analysis. Third, the identified proteins were selected for the construction of the XDB. Although we worked with a purified complex, in addition to the four expected proteins several other proteins were present in the sample as judged from the gel; presumably they were contaminants from the expression system. One approach would be to identify all proteins present and to include them in the XDB. This would, however, unnecessarily inflate XDB as we are only interested in those proteins actually found in the respective fraction we analyze. By searching Swiss-Prot for each analysis we ensure that we consider exactly those proteins that can be detected in the respective gel band. To also identify cross-linked peptides using a standard database search tool like Mascot required a special database and a separate search. The selected proteins were digested *in silico* allowing for up to two missed cleavages. The obtained peptides were filtered, to contain either an internal Lys or the protein N terminus, and joined up in all possible pairwise combinations. It is essential to have both linear permutations of a peptide pair, *i.e.* AB and BA, to allow the complete matching of fragments (see also Fig. 1 and “Results”). Creating one entry per peptide pair has the disadvantage of resulting in many short entries and occupying more memory than combining the peptides in linear succession. If a protein P gives a peptide set [a, b, c] and a protein Q gives a peptide set [A, B, C], then XDB would contain a single protein in which the peptides of the proteins P and Q were concatenated in a single sequence as caabacbbccCAABACB-BCCcAaAbAcBaBbBcCaCbC. The search program will create from this sequence all possible pairs in both permutations: aa, ab, ba, bb, bc, cb, cc, ca, ac, etc. This way of constructing a cross-link database is a very condensed way of writing all possible pairwise combinations, in our example leading to 30 letters instead of 54, *i.e.* resulting in almost 50% compaction. Note that the peptides concatenated in XDB contain missed cleavage sites. The search will also create chimeric peptides containing parts of the two original peptides. These are known false positives. The reversed cross-link database was obtained by inverting the entire cross-link database, *i.e.* writing the sequence from C terminus to N terminus. The cross-link database was searched with the peak lists of light and heavy precursors using Mascot with the parameters: monoisotopic mass values; peptide tolerance, 0.08 Da; MS/MS tolerance, 0.5 Da; instrument, ESI-TRAP; fully tryptic specificity; cysteine carbamidomethylation as fixed modification; light or heavy cross-linker hydrolyzed and oxidation on methionine as variable modifications; five missed cleavage sites allowed. For the second control using a wrong mass for the cross-linker, 3 Da were added to the correct masses of the heavy and light cross-linker in the modification file of Mascot.

**Identification of Cross-links by Analyzing the Database Search Results**—The database search retrieves the peptides that match the observed mass. Mascot does not return all candidates but only those considered non-random based on an initial matching of fragments.<sup>2</sup> Mascot already uses some fragment information to select higher value candidates than obtained on the basis of the measured peptide mass alone. The cross-linked peptides can be found as miscleaved peptides in the output of the database search. The score used for expressing the quality of match between a spectrum and a cross-linked peptide is presented under “Results.” When calculating our score, we considered all b- and y-ions of the cross-linked peptide. Other ions such as those resulting from loss of water or ammonia, internal fragments, and multiply charged fragments are observable (26) but not currently included in the algorithm. Considering all possible fragments results in a large number of mass values and lowers the selectivity of the score at our current, low mass accuracy ( $\pm 0.5$  Da). For the Mascot-independent matching of precursor masses with predicted cross-linked peptides we wrote a Perl script that computes all predicted cross-linked peptides matching to precursors within 4.5 ppm deviation based on input protein sequences and number of missed cleavages allowed (two) and the amino acid required for the linkage (lysine or protein N terminus). Note that we focus on those products composed of two linked peptides and containing a single cross-linker molecule. Including other cross-link products is possible but increases the search space and consequently the background of the data analysis. Currently data of peptides containing more than one cross-linker do not contribute to the background because, not being identified as a doublet of 4-Da spacing, they are filtered out.

## RESULTS AND DISCUSSION

### Algorithm

**Spotting Candidate Signals of Cross-linked Peptides**—We used a cross-linker targeting amino groups, BS<sup>2</sup>G, in a 1:1 mixture of its unlabeled light and labeled heavy form (the latter containing four deuterium atoms) (27). The use of a light/heavy mixture results in doublet mass signals for those cases in which the cross-linker was incorporated between two peptides or alternatively on a single peptide. We began our analysis by selecting from the entire LC-MS/MS dataset only those fragmentation spectra of peptides with doublet signals, thus focusing our analysis on fragmentation products of cross-linked peptides (Fig. 1a). It should be noted that the doublet information serves solely for the reduction of data to focus the analysis onto likely cross-linked peptides. In this way, the false positive rate of the database search is minimized. For small datasets such as obtained for a single protein or a small complex this will not be necessary, and any cross-linker can be used.

**Assigning Candidate Peptide Pairs**—Two observations allowed the construction of a special database for the identification of cross-linked peptides. First, a cross-linked peptide has the same mass of a peptide obtained by fusing the two linked peptides via a normal peptide bond and adding the hydrolyzed cross-linker as a modification (Fig. 1b). Second, from its ends up to the linkage site the “linearized” virtual peptide generates the same fragments of the cross-linked

<sup>2</sup> J. Cottrell, personal communication.

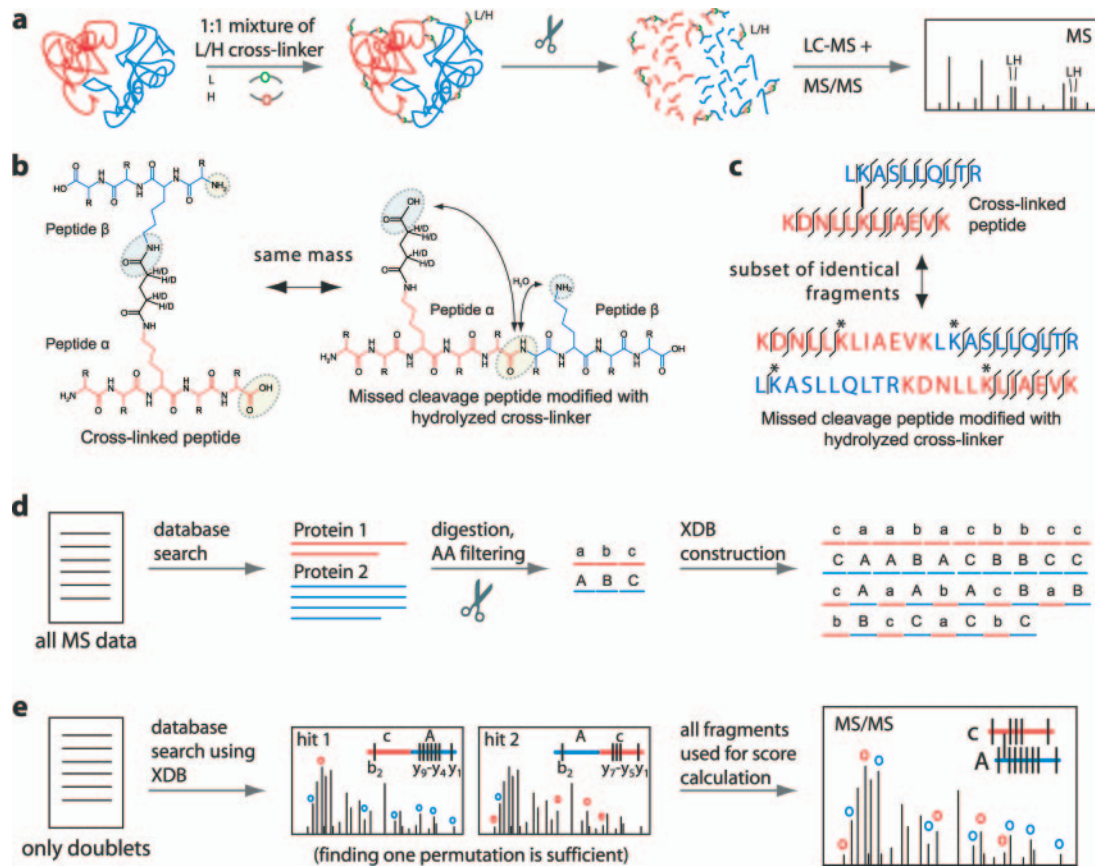


Fig. 1. *a*, a protein complex is cross-linked using a 1:1 mixture of the light (*L*; unlabeled) and heavy (*H*; stable isotope-labeled) versions of a cross-linking agent. The complex is then digested, and peptides are analyzed by LC-MS/MS. Peptides containing the cross-linker are recognized as doublets in the MS spectrum. *b*, a cross-linked peptide can be linearized without changing the mass into a missed cleavage peptide carrying a hydrolyzed cross-linker as modification for the purpose of using standard database search algorithms. *c*, the single bond fragments of a cross-linked peptide coincide with those of the two permutations of the linearized peptide between peptide termini and linked residues (asterisks). Each of the two linearized peptides accounts for a subset of the possible single bond fragments of the cross-linked peptides, and together they account for the complete set. *d*, the acquired fragmentation data are used in standard database searching to identify the cross-linked proteins. The sequences of these proteins are used to construct an XDB in which all peptides that are considered candidate partners in cross-linking are combined linearly with each other in permutations  $\alpha\beta$  and  $\beta\alpha$ . *e*, the fragmentation data are used to search XDB. A candidate cross-link between peptide  $\alpha$  and  $\beta$  is found as a missed cleavage  $\alpha\beta$  (hit 1) and/or  $\beta\alpha$  (hit 2). The candidates are then scored as cross-linked peptides.

peptide (Fig. 1c). The two possible permutations of the linearized peptide ( $\alpha\beta$  and  $\beta\alpha$ ) cover the entire set of possible single bond fragments of the cross-linked peptide, which are usually the most intense signals. If a linearized version of all possible cross-linked peptides were built, a standard database search tool should make it possible to find cross-linked peptides using fragmentation data very much like any ordinary peptide. Thus, in our XDB, every peptide in a target protein is combined with every other peptide in a linear sequence and in both permutations (Fig. 1d). We took into account cross-links within a protein and between proteins. A standard database search algorithm can now find matches to the observed mass of cross-linked peptides simply by allowing for missed cleavages of the enzyme used for digestion and considering the hydrolyzed cross-linker as a variable modification. Mascot (19) can thus be used in its normal function as a database

search tool to find peptides matching the experimental data. The analysis of the non-linked peptides gives a clear indication of the mass accuracy of the measurement in MS and MS/MS and which proteins to include in XDB.

For small complexes, the candidate list of cross-linked peptides returned from the database search is short enough for manual validation. However, the number of candidates increases dramatically with the size of the complex. This increase of complexity follows the third power of the number ( $n$ ) of tryptic peptides, assuming the complexity of the dataset increases linearly with  $n$  and the database size increases with  $n^2$ . Only a score, ideally probabilistic, can free the investigator from having to validate every match manually. A score expressing the quality of match between a spectrum and a linear peptide, such as the Mascot score, does not fulfill this function. The database search program computes the score

based on the fragment matches to the linearized version of the cross-linked peptide (Fig. 1e). Both permutations of the linearized version together account for all single bond cleavages of the cross-linked peptide. However, as only one permutation at a time is considered by the database search program, only a subset of observed fragments is matched, whereas the other subset is not and thus lowers the score.

**Scoring**—We developed a scoring algorithm that expresses how well an identified cross-linked peptide agrees with the experimental fragmentation spectrum, based on an algorithm recently used for MS<sup>3</sup> scoring (25). This score does not provide an absolute answer, but together with the estimation of false positives described below it can be used to conclude whether an identification is correct or not. The scoring algorithm uses a probabilistic approach that considers the fragments of a cross-linked peptide. The scoring includes only matches with the most intense ions of the spectrum in a given *m/z* window. Ion matching is performed with the same tolerance used in the database search. To describe the chance of matching the sequence of a cross-linked peptide to a fragmentation spectrum we use the binomial distribution

$$S = \binom{n}{k} p^k (1-p)^{n-k} \quad (\text{Eq. 1})$$

where *k* is the number of matched ion masses, *n* is the number of calculated fragments in the mass range under consideration, and *p* is the probability of a random match for a fragment mass. For convenience, *S* is reported as a score similar to the Mascot score.

$$S' = -10 \log_{10} S \quad (\text{Eq. 2})$$

The probability of a random match (*p*) for fragment masses is given by the number of considered peaks (*N*) in an *m/z* window of width *W* and the mass accuracy ( $\Delta m$ ) used in the database search as Equation 3.

$$p = N2\Delta m/W \quad (\text{Eq. 3})$$

We found empirically that selection of the four most intense peaks in an *m/z* window of width 100 Da represents the best parameters for scoring our data. For ion trap fragmentation data (mass error,  $\pm 0.5$  Da) the probability of a random match is 0.04. The same values have been used in the algorithm for RS<sup>3</sup> scoring (25) and another algorithm used for spectrum matching (28).

**False Positive Estimation**—Database searches conducted under conditions that yield only false results and no true cross-links give an estimate of how many wrong identifications we expect at a given score cutoff. In conventional large scale protein identification experiments, false positive rates are determined by searching against a control database containing reversed or otherwise falsified sequences (29). The operator can then determine for any score cutoff the rate of random, false matches in the database search by looking at

how many matches were found using the same score cutoff in the control database search. We adapted the same concept for our XDB creating and searching a reversed XDB. As a second negative control we conduct a database search against XDB but using a false mass for the cross-linker. This is a more strict control because in any false match one of the two peptides may actually be correct, and only the second one may be wrong.

#### *Coiled Coil Analysis: the NDEL1 Homodimer*

Kinetochores are complex proteinaceous scaffolds that represent the site of attachment of chromosomes to the mitotic spindle (30). Several coiled-coil proteins play essential roles at the kinetochore. These include NDEL1 and the members of the NDC80 complex. High resolution structural analysis of large coiled-coil complexes such as the four-protein NDC80 complex and the NDEL1 homodimer suffers from the general difficulty of crystallizing coiled-coil domain-containing complexes. In particular, the elongated shape of these domains and the difficulties in determining the register and overall organization of heterologous and/or antiparallel coiled coils restrain the potential of protein engineering for designing stable constructs for crystallization.

As a proof of principle to establish our approach, we tried to detect sites of cross-link on the coiled-coil domain of NDEL1 (residues 17–174, indicated as NDEL1-(17–174)) (Fig. 2a). NDEL1 is a regulator of cytoplasmic dynein that acts by forming a complex with LIS1 (31). NDEL1 localizes to the centrosome where it is implicated in centrosomal separation and centrosomal maturation and for mitotic entry (32). NDEL1 also localizes to kinetochores during mitosis and is believed to regulate dynein function there during the process of microtubule-kinetochore attachment (33). NDEL1 binds LIS1, the product of a gene that is mutated in type I lissencephaly (31). A previous structural and biochemical analysis of LIS1 and its interaction with NDEL1 suggested that NDEL1 might need to form an antiparallel coiled coil to bind to LIS1, but this was not corroborated by structural analysis (21).

To study whether NDEL1 forms parallel or antiparallel dimers, we carried out a cross-linking analysis on a recombinant form of NDEL1-(17–174). We efficiently cross-linked NDEL1-(17–174) to form a dimer (Fig. 2b). Analysis of the cross-linked protein upon tryptic digestion by LC-MS/MS and Mascot searches using XDB gave us three matches, all of which passed manual inspection (see Fig. 2c and for additional annotated spectra Supplemental Fig. 1). Cross-links I and II involved overlapping sequences and could therefore be unambiguously identified as sites of interprotomer cross-linking (Fig. 3a). It is clear from Fig. 3b that these two sites are only compatible with a parallel arrangement of the  $\alpha$ -helices of the NDEL1 coiled-coil region. Cross-link III involved different tryptic peptides and therefore could not be identified unambiguously as an intra- or interprotomer cross-link.



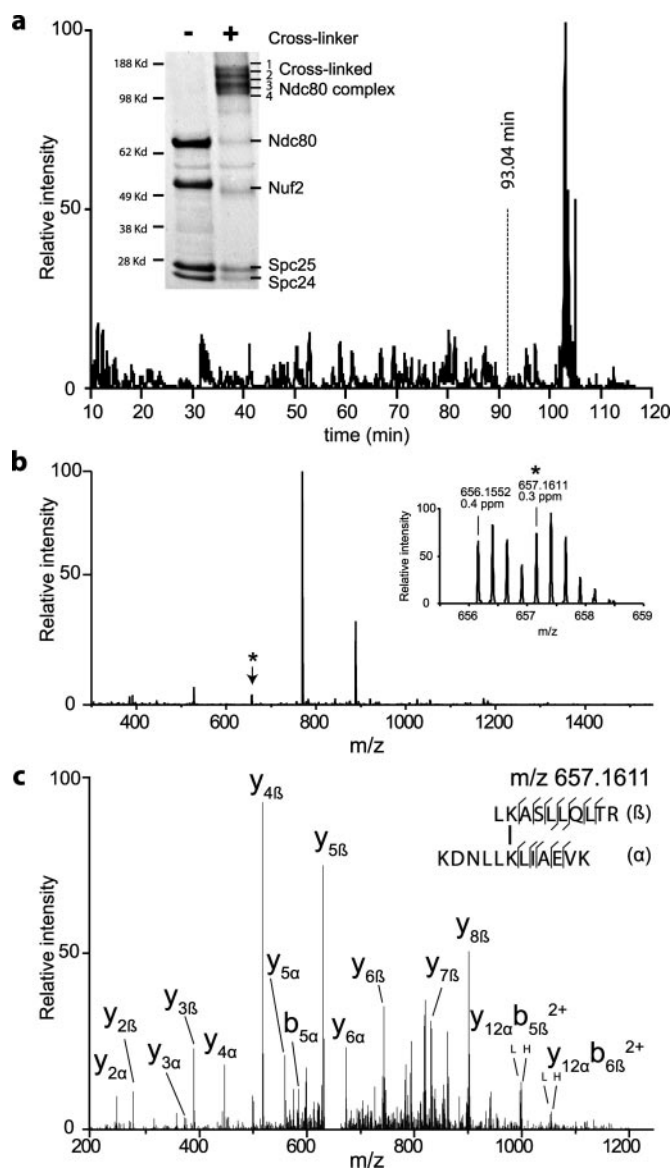


FIG. 4. **Analysis of the NDC80 complex.** *a*, base peak chromatogram of band 2 with SDS-PAGE gel of the control and cross-linked complex as *inset*. The excised and analyzed bands are labeled 1–4. *b*, full FT-MS spectrum recorded at 93.04 min. The peak at  $m/z$  657.1611 denoted with an asterisk was selected for FT-SIM (shown as *inset*, revealing the signal as doublet) and ion trap MS/MS (*c*). *c*, fragmentation spectrum of  $m/z$  657.1611 and the matched cross-linked peptide. Only the most intense fragments are annotated in the spectrum.

large database, as a result of its size, that virtually any precursor mass produces a match even at high mass accuracy (4.5 ppm). Mascot, as a routine tool for matching fragmentation spectra with peptide sequences, condensed the list of matches to 125 by using XDB, yielding a 10-fold reduction of the data. Nevertheless the number of matches was still very large for manual validation. Next the matches were sorted using our scoring algorithm. This gave a possibility to start the manual validation with the best quality data. However, manual

validation is not free of error. Its success rate is unknown and furthermore depends on subjective criteria. Given controls, the score can be used to assign a degree of confidence to candidate cross-linked peptides. This is how scores are used in protein identification, and this is how we planned to use our scoring algorithm in protein structure determination.

Using a score of 15 at peptide mass error 4.5 ppm (Fig. 5) left us with 69 matches having 90–100% confidence, leading to the identification of 26 cross-linked peptide pairs with unique sequence containing 25 different linkage sites (Table 1). We designed two negative controls to determine the false positive rate at this score and peptide mass accuracy cutoff: (a) assuming a false mass for the cross-linker and (b) searching against the inverted version of XDB. At 4.5 ppm, the accuracy of our measurement for peptide masses, 2436 false positives were obtained taking both controls together. Remarkably the number of false positives was reduced to 55 using fragment ion information in Mascot to search XDB. Assuming a similar number of false positives for our experiment, however, these controls predict that up to half of our 125 initial matches might be false positives. The score is a quality measure that ranks the candidates according to their match to the fragment ions (Fig. 5a). Through the control searches we could estimate the confidence ( $C$ ) associated with our identifications for any score cutoff as

$$C = 1 - (N_c/N_r) \quad (\text{Eq. 4})$$

where  $N_c$  is the number of hits above the cutoff in the control search and  $N_r$  is the number of hits above the cutoff in the real search. Conducting this calculation within score-windows of five units and applying a cut-off at 90–100% confidence resulted in our high confidence list of 69 matches. Fig. 5c illustrates the high specificity achieved by our scoring algorithm despite the relatively low accuracy ( $\pm 0.5$  Da) of our fragment data. Other types of instruments can yield fragment data of 10–100 $\times$  higher accuracy than those obtained here and will result in even higher specificity. Note in Fig. 5a that 20 cross-links are found with a score larger than 45 but no false positives regardless of the peptide mass accuracy. The algorithm can therefore also be used for data obtained with instruments, such as a stand alone ion trap, that provide less accurate data for peptide masses than those we used for our analyses. Note also in Fig. 5a the virtual absence of any match with a score below 5. This is the effect of the random match filter of Mascot. Our score would also have been able to cope with these random matches. For a complete list of the 69 high confidence matches see Supplemental Table 1, and for additional annotated spectra see Supplemental Fig. 2.

A detailed view into the 69 high confidence matches reveals an important insight into the current limitations of the analysis and points toward a possible solution. In eight instances, both forms of the cross-link, light and heavy, were identified within the same LC-MS/MS run (Supplemental Table 1). The observation that only one of the pair was identified in all other cases

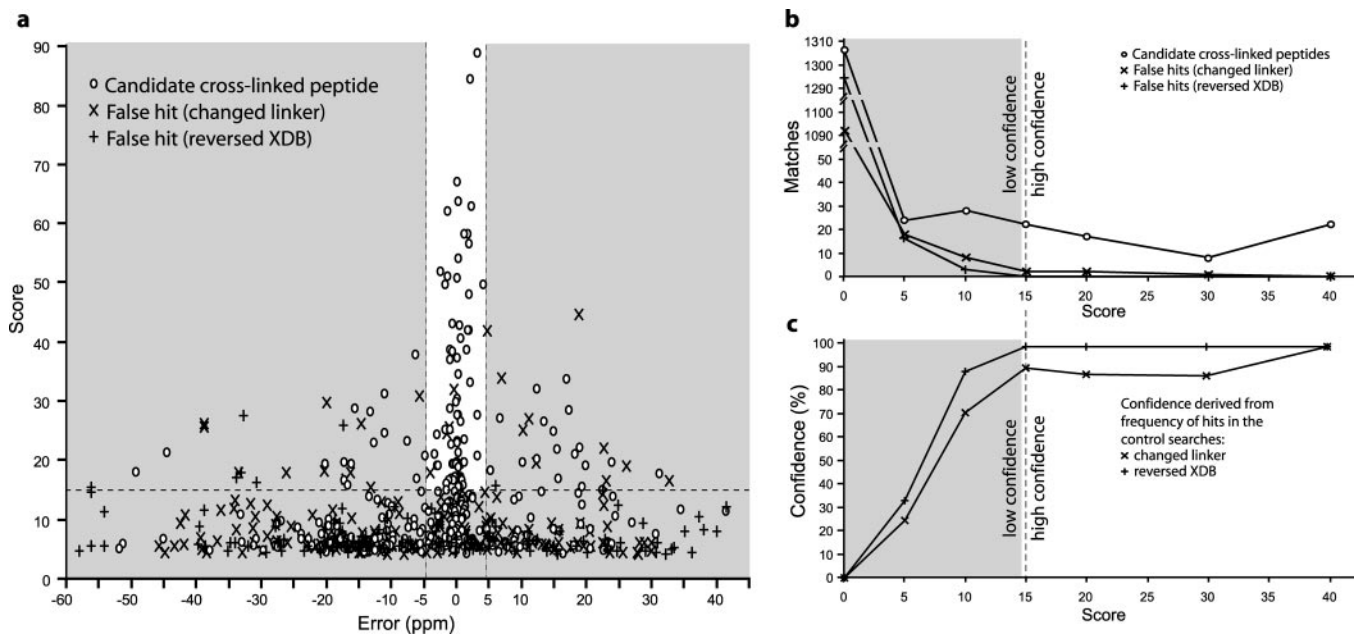


FIG. 5. a, each match of the combined data of four analyses is plotted with its score and its mass deviation between the observed mass and the mass of the matched cross-linked peptide (○). The same data are searched as negative control either assuming a 3-Da heavier cross-linker mass for both forms of the cross-linker, light and heavy (×), or against a reversed sequence database (+). The lines at 4.5 ppm indicate the mass accuracy of the measurement. The line at score 15 indicates 95% confidence as estimated by the hits of the two negative controls. The lack of matches with scores below 5 is a result of the prefiltering of random matches. b, the number of hits are plotted over the score using the data of a at 4.5 ppm accuracy. Each data point sums the counts up to the next point, i.e. counts at score 20 give the count of hits having score 20–30. The number of matches with scores below 5 was computed separately as described under “Experimental Procedures.” c, confidence of the cross-linked peptides plotted over score. The confidence  $C^*$  is calculated in each score range as  $C^* = (1 - (N_c^*/N_r^*)) \times 100$  where  $N_c^*$  is the number of hits in the control search and  $N_r^*$  is the number of hits in the real search falling into the respective score range. The shaded region designates the region of confidence below 90–100%.

indicates that the analysis was not exhaustive. The same line of reasoning based on SILAC (stable isotope labeling by amino acids in cell culture) labeled peptide pairs demonstrated recently the limited depth of analysis for complex peptide mixtures (38). As a result of the incompleteness of our LC-MS/MS analyses we are possibly missing out on a significant proportion of cross-links. A solution would be to focus the data acquisition on potential cross-links. Using the doublet detection as a selection criterion for MS/MS is unfortunately not possible with the current version of our instrument software. However, we noticed that nearly all identified cross-links (64 of 69) were observed with  $z > 2$ . In contrast, about 50% of all selected precursors had  $z = 2$ . In agreement with this we observed that restricting the selection of precursors for MS/MS on those with  $z > 2$  allows reducing the background and focusing better on cross-linked peptides.<sup>4</sup> Importantly the high charge states of cross-linked peptides can be utilized for the enrichment of these species using strong cation exchange.<sup>4</sup>

The 25 linkage points we obtained provide a detailed picture of the NDC80 complex and illustrate the potential of using cross-linking and mass spectrometry in structure elucidation (Fig. 6). So far, the full-length human and yeast NDC80 com-

plexes have failed to provide diffraction quality crystals. Scanning force microscopy and electron microscopy studies on the reconstituted human NDC80 complex (20) or on its yeast homolog (35) have indicated, however, that the complex is ~57 nm long. The architecture of the NDC80 complex is characterized by the existence of two dimeric subcomplexes consisting of the SPC24 and SPC25 subunits and of the NUF2 and NDC80 subunits, respectively (20, 35). As already explained above, the two globular regions have been proposed to contain the globular, N-terminal heads of NDC80-NUF2 and the globular, C-terminal heads of SPC24-SPC25. The central shaft has been proposed to contain the coiled-coil regions in the N-terminal portions of SPC24 and SPC25 and in the C-terminal portions of NUF2 and NDC80 with a tetramerization domain containing the C-terminal tails of NDC80-NUF2 and the N-terminal heads of SPC24-SPC25.

The list of cross-link sites we identified is consistent with this prediction (Table I and Fig. 6). A number of intramolecular cross-links bridge seven (IX, X, XI, XVI, and XXVI) or 10 residues (VII and VIII). This corresponds to two or three turns of a helix, bringing the linked residues again onto the same side of the helix, and supports a helix as a secondary structure element in the region of these cross-links. The lysine residues of cross-link XIV are in positions g and e of the predicted coiled coil supporting the prediction locally as well. In agreement

<sup>4</sup> J. Rappsilber, unpublished results.



TABLE I  
Summary of cross-linked peptides found in the NDC80 complex

Cross-link	Peptide $\alpha$	Peptide $\beta$	Protein		Residue $\beta$	Alternative residue $\alpha$	Alternative residue $\beta$	No. of observations	$m/z^a$	Error <sup>a</sup> $z^a$ Score <sup>a</sup>	Modifications <sup>a</sup>	Cross-link <sup>a</sup>		
			1	2									$\alpha$	$\beta$
I	SQDVNKGGLYTPQTK	<u>Nter</u> -GPLGSMETLSF-qPR	NDC80	NUF2	26	-5 <sup>b</sup>		1	803.1502	0.2	4	51	Oxidation (Met)	L
II	<u>Nter</u> -GPLGSMETLSFPR	NKILTGADGK	NUF2	NUF2	-5 <sup>b</sup>	21		1	840.4338	2.3	3	63	Oxidation (Met)	L
III	<u>Nter</u> -GPLGSMETLSFPR	AKR	NUF2	NUF2	-5 <sup>b</sup>	115		1	930.9874	2.2	2	33		L
IV	ILTGADGKNLTK	KVSLFGK	NUF2	NDC80	29	53		1	526.8058	0.0	4	17		L
V	LSINKPTSER	KVSLFGK	NDC80	NDC80	47	53		1	505.2877	0.4	4	28		L
VI	IFKDLGYPFALSK	VSLFGKR	NDC80	NDC80	156	59		5	600.8386	0.0	4	30		L
VII	EKEPNRLESRL	KLK	NDC80	NDC80	288	298	NDC80 Lys-421; NUF2 Lys-338	4	371.6145	-0.5	5	43		L
VIII	EKEPNR	KLK	NDC80	NDC80	288	298	NDC80 Lys-421; NUF2 Lys-338	2	419.2397	3.3	3	28		L
IX	KSNISEK	TKR	NUF2	NUF2	213	221		1	435.5773	0.3	3	19		L
X	IVDSPEKLK	EKMK	NUF2	NUF2	250	257		5	553.64	1.3	3	24		L
XI	MKDTVQK	LKNAR	NUF2	NUF2	259	266		1	388.2223	0.3	4	20		H
XII	LQNIIDNQKYSVADIER	LKNYK	NDC80	NUF2	360	252		1	696.6275	0.3	4	37		H
XIII <sup>c</sup>	KEKLATAQFK	LKNAR	NUF2	NUF2	351/349	266		4	930.536	-2.9	2	24		L
XIV	AQVVVPLKELLNETEEE-INK	KIQDLSNDR	NDC80	NUF2	462	299		3	709.3712	1.3	5	58		L
XV	TLKEEVQK	KLK	NDC80	NDC80	504	421	NDC80 Lys-298; NUF2 Lys-338	4	365.2199	0.1	4	20		L
XVI	TEENSFKR	KEK	NUF2	NUF2	342	349		1	505.2646	-0.1	3	17		H
XVII	TLKEEVQK	LMIVKK	NDC80	NUF2	504	348		2	455.0173	1.3	4	20	Oxidation (Met)	L
XVIII	TLKEEVQKLDLDLYQQK	KEK	NDC80	NUF2	504/509	349		3	621.0902	0.2	4	27		H
XIX	VTTINQEIQIK	<u>Nter</u> -GPLGSMAAFR	NUF2	SPC24	400	-5 <sup>b</sup>		6	630.8513	1.6	4	39		H
XX	<u>Nter</u> -GPLGSMAAFR	KVGNLQK	SPC24	NDC80	-5 <sup>b</sup>	577		2	682.6912	2.1	3	42	Oxidation (Met)	L
XXI	LKSQEIFLNK	<u>Nter</u> -GPLGSMAAFR	NUF2	SPC24	418	-5 <sup>b</sup>		12	812.113	3.4	3	89		L
XXII	<u>Nter</u> -VEDELALFDK	EKLK	SPC25	NUF2	1	416		2	598.9944	0.7	3	26		H
XXIII	EILTMEKEVAQSLNNAK	LSVKLK	SPC24	SPC25	60	50		1	680.6424	0.5	4	43	Oxidation (Met)	H
XXIV	LKEEER	KATLIK	SPC25	NDC80	52	633		2	393.7307	0.4	4	34		L
XXV	KDNLLKJAEVK	LKASLLQLTR	SPC25	SPC24	84	98		2	657.1611	0.3	4	64		H
XXVI	ETIISTANKANAER	LKR	SPC25	SPC25	122	129		1	480.7668	0.3	4	21		H

<sup>a</sup> Data given for the observation with the best match. The data for all observations can be found in Supplemental Table 1. The annotated spectrum of the best observation can be found for all cross-links in Supplemental Fig. 2.

<sup>b</sup> Numbering refers to a linker region (sequence GPLGS, numbered -5 to -1) that precedes the natural N terminus of NUF2 or SPC24 (numbered 1) and that is retained at the N termini of these proteins after proteolytic cleavage from the GST tag used for purification.

<sup>c</sup> The data equally support the N-terminal lysine of the  $\alpha$ -peptide being the N-terminal residue in the  $\beta$ -peptide: EKLATAQFK and KLKNAR.

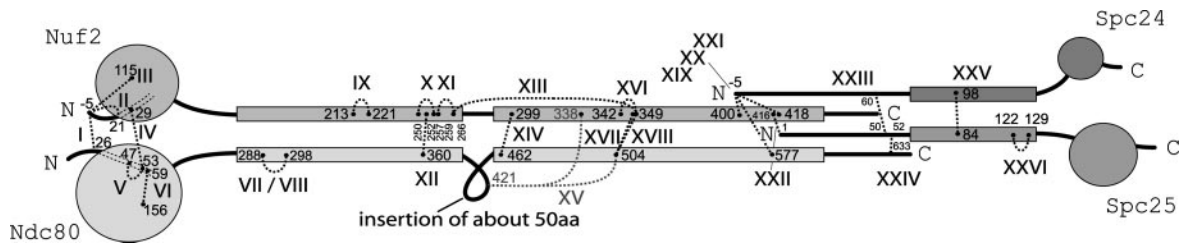


FIG. 6. **Model of the NDC80 complex fitting best to the observed cross-links.** Amino acid residue numbers are *Arabic numbers*, and cross-links are *Roman numbers*. For a full list of residue numbers for all cross-links see Table I. aa, amino acids.

with the previous analyses (20, 35), we found several linkages connecting the SPC24 and SPC25 subunits (cross-links XXIII and XXV) and the NUF2 and NDC80 subunits (cross-links I, IV, XII, XIV, XVII, and XVIII). These linkages are consistent with a parallel orientation of the predicted coiled coils in both subcomplexes. As far as the SPC24-SPC25 subcomplex is concerned, cross-links XXIII and XXV approximately define the register of the two chains in the coiled coil as they reveal a 10- to 14-residue offset between SPC24 and SPC25 (residues 60<sup>SPC24</sup>/50<sup>SPC25</sup> and 98<sup>SPC24</sup>/84<sup>SPC25</sup>).

The analysis of cross-links in the NDC80-NUF2 subcomplex provides a useful framework to understand the register of the two chains in the coiled-coil regions. Cross-links XII, XIV, XVII, and XVIII all map to the predicted coiled-coil region of this subcomplex. Sites XIV, XVII, and XVIII represent a consistent set with separations of ~45 residues on NDC80 and ~50 residues on NUF2. Indeed we show below that the register of the coiled coil defined by these cross-links remains unaltered until the C-terminal end of the NDC80-NUF2 subcomplex. Conversely the distance between site XII (residues 360<sup>NDC80</sup>/252<sup>NUF2</sup>) and site XIV (residues 462<sup>NDC80</sup>/299<sup>NUF2</sup>) is different for the NDC80 and NUF2 chains (102 residues on NDC80 and 47 residues on NUF2). The discontinuity on the NDC80 chain correlates with a drop in the coiled coil prediction between residues 420 and 460 of NDC80 (data not shown), suggesting the presence of a non-coiled-coil loop region extending from the NDC80 chain around this region. This interpretation is reinforced by the presence of two long range intramolecular links bridging roughly equivalent positions in NUF2 and NDC80 (XIII and XV).

Overall these data suggest that there is an interruption in the coiled-coil region of the NDC80-NUF2 subcomplex, approximately centered on residue 440 of human NDC80, that might represent a site of flexibility in the NDC80 complex rod. Coiled-coil predictions on the NDC80 complex of *Saccharomyces cerevisiae* display a very similar pattern with an ~60-residue interruption of the coiled coil approximately centered on residue 475 (data not shown). Visualization of the yeast NDC80 complex by electron microscopy confirms the speculation that this represents a site of flexibility in the NDC80 rod (35). A majority of the rotary shadowed particles showed evident bending of the NDC80 rod at about a third of its length from the N terminus (35), which is fully consistent with the

position of a non-coiled-coil segment in the central shaft of the NDC80 complex (Fig. 6). Future work will have to address the functional significance of the 40–50-amino acid insertion in the coiled coil of NDC80. Upstream from site XII, the coiled-coil predictions for NDC80 and NUF2 envision a ~110-residue coiled-coil segment for both chains, suggesting that the pairing defined by site XII extends ~110 residues upstream from this site, *i.e.* from the point in which the coiled-coil forms as the NUF2 and NDC80 chains emerge from the N-terminal globular regions.

In summary, our cross-linking studies shed light on the register of the coiled coils within the shaft of the NDC80 complex. In combination with low resolution visualization approaches and prediction methods of common usage, this approach allows probing the structural complexity of a large protein assembly. In this respect, the set of cross-links between different NDC80 subcomplexes (cross-links between NDC80-NUF2 and SPC24-SPC25) are particularly useful in defining the organization of the tetramerization domain. Consistent with the predicted organization of the NDC80 complex (20, 35), these sites involve the C-terminal regions of NDC80-NUF2 and the N-terminal regions of SPC24-SPC25 (XIX–XXIV). Specifically we identified cross-links between residues 400<sup>NUF2</sup> or 418<sup>NUF2</sup> and the N terminus of a five-residue extension of SPC24 that was created by the Prescission protease after cleavage from the GST tag (sites XIX and XXI). We also found a cross-link between 577<sup>NDC80</sup> and the same N-terminal extension of SPC24 (site XX). We therefore suggest that 577<sup>NDC80</sup> faces the midpoint between 400<sup>NUF2</sup> and 418<sup>NUF2</sup> (409<sup>NUF2</sup>). This pairing is fully consistent with the register of the coiled coil established by the set of cross-links XIV, XVII, and XVIII. Specifically if the coiled coil of NDC80 and NUF2 run uninterrupted C-terminal of site XVIII (504<sup>NDC80</sup> and 349<sup>NUF2</sup>), one would expect that ~70 residues downstream from site XVIII, and therefore around residues 574<sup>NDC80</sup> and 419<sup>NUF2</sup>, the NDC80 and NUF2 chains should still be paired. Within the accuracy allowed by our cross-linking experiment, this prediction is in good agreement with our argument (see above) that 577<sup>NDC80</sup> faces 409<sup>NUF2</sup>. Thus, we can conclude that the NDC80-NUF2 coiled coil runs roughly uninterrupted after the predicted loop around 440<sup>NDC80</sup>.

We also found that 416<sup>NUF2</sup> cross-links to the N terminus of SPC25 (site XXII) and that 633<sup>NDC80</sup> cross-links to 52<sup>SPC25</sup>

(site XXIV). Based on arguments developed in the previous paragraph, 416<sup>NUF2</sup> is expected to face ~584<sup>NDC80</sup>. Because 416<sup>NUF2</sup> faces the N terminus of SPC25, also 584<sup>NDC80</sup> is predicted to be close to 1<sup>SPC25</sup>. ~50 residues downstream from this point, the NDC80 and SPC25 chains have advanced at the same pace as revealed by the cross-links between 633<sup>NDC80</sup> and 52<sup>SPC25</sup> (site XXIV). These data also suggest that the C terminus of NDC80 extends beyond the C terminus of NUF2 by some 10–15 residues. On the other hand, the N terminus of SPC24 might advance the N terminus of SPC25 by ~7–10 residues as revealed by their cross-links with NUF2 or NDC80. This is roughly consistent with the offset between these chains described above.

Based on the cross-linking analysis presented here, we carried out extensive additional subcloning to test the expression properties and solubility of different subcomplexes of the NDC80 complex. This work allowed the construction of new expression constructs to generate a stable recombinant version of the tetramerization domain.<sup>5</sup> More importantly, the determination of the register of the coiled coils in the NDC80-NUF2 and SPC24-SPC25 subcomplexes promoted the generation of engineered constructs of the NDC80 complex containing direct fusions of truncated versions of the coiled-coil segments of NDC80 and SPC25 and of NUF2 and SPC24, resulting in the creation of a chimeric dimeric NDC80 complex (data not shown). In this arrangement, residues 80–286 of NDC80 were fused to residues 118–224 of SPC25, and residues 1–169 of NUF2 were fused to residues 122–197 of SPC24. This strategy, which bypassed the requirement for a tetramerization domain, resulted in the production of a stable mini-NDC80 complex (named NDC80<sup>bonsai</sup>) that readily crystallized and whose crystal structure was determined at ~3.0-Å resolution.<sup>6</sup>

The geometry of residues cross-linked in site XXVI, 122<sup>SPC25</sup> and 129<sup>SPC25</sup>, can be observed in the crystal structure of NDC80<sup>bonsai</sup>. The C $\alpha$  atoms of these residues are ~10.5 Å away from each other, a distance that the 7.7-Å BS<sup>2</sup>G cross-linker and the ~6-Å flexible side chain of lysine can easily bridge. Indeed there are several cross-links with an equivalent positioning of linkage sites in the primary sequence in our dataset, such as sites X and XI. The coiled-coil register displayed by the crystal structure of NDC80<sup>bonsai</sup> is consistent with that revealed by the cross-linking analysis described here (data not shown). Thus, our cross-linking studies were instrumental in tailoring appropriate manipulations within the complex architecture of the NDC80 complex, such as the one described above that led to its crystallization or such as those that might be required to design deletion mutants to probe the significance of different segments of a coiled-coil region in a large complex.

<sup>5</sup> C. Ciferri and A. Musacchio, unpublished results.

<sup>6</sup> C. Ciferri, A. Maiolica, J. Rappsilber, and A. Musacchio, manuscript in preparation.

## Conclusion

Cross-linking and mass spectrometry can potentially yield important structural information for multiprotein complexes in the absence of high resolution structural information. Using our novel data analysis algorithm we report the resolution and interpretation of an unprecedented number of cross-links for a 176-kDa, four-protein complex. The density of observed linkages results in a hitherto unobtainable depth of detail and gives clear directions for minimal constructs of the NDC80 complex to obtain a crystal structure. From the simplicity of our algorithm and the ability of mass spectrometry to detect peptides without large bias toward their sequence follows the option of complementing any study of multiprotein complexes with a topological analysis. This significantly increases the focus of follow-up experiments to reveal functional and structural aspects of the complexes. Our link into Mascot as a standard database search tool can be taken up by the providers of this and alternative tools and should ensure professional support and easy access to a wide range of researchers. Our database search strategy does not depend upon the use of isotope labels in any way and is hence equally applicable to work with non-labeled cross-linkers. However, the use of labeled cross-linkers allowed us to reduce 10-fold the size of our NDC80 dataset (~14,000 to ~1400 fragmentation spectra) and was important to reduce the false positive rate. Using more accurate fragment masses than in our study and automation of data analysis now in place, even larger structures than the NDC80 complex can be addressed in the near future.

*Acknowledgments*—We thank Mogjiborahman Salek for assistance; Matthias Mann, Bill Earnshaw, and Jimi-Carlo Bukowski-Wills for valuable comments on the manuscript; and Pierce for early access to the cross-linkers.

\* This work was supported by the European Community through a Marie Curie Excellence Grant and the Italian Association for Cancer Research (AIRC). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

§ These authors contributed equally to this work.

\*\* Present address: Howard Hughes Medical Inst./University of California, 742 Stanley Hall, MS 3220, Berkeley, CA 94720-3220.

‡‡ To whom correspondence should be addressed. Tel.: 44-0131-651-7057; Fax: 44-0131-650-5379; E-mail: [juri.rappsilber@ed.ac.uk](mailto:juri.rappsilber@ed.ac.uk).

## REFERENCES

1. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198–207
2. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciati, C. M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelman, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002) Functional

- organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147
3. Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutlier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreaux, M., Muskaf, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jepsen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W., Figeys, D., and Tyers, M. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183
  4. Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpefeld, B., Edelmann, A., Heurter, M. A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A. M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B., and Superti-Furga, G. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636
  5. Sinz, A. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom. Rev.* **25**, 663–682
  6. Rappsilber, J., Siniosoglou, S., Hurt, E. C., and Mann, M. (2000) A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal. Chem.* **72**, 267–275
  7. Young, M. M., Tang, N., Hempel, J. C., Oshiro, C. M., Taylor, E. W., Kuntz, I. D., Gibson, B. W., and Dollinger, G. (2000) High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5802–5806
  8. Chen, T., Jaffe, J. D., and Church, G. M. (2001) Algorithms for identifying protein cross-links via tandem mass spectrometry. *J. Comput. Biol.* **8**, 571–583
  9. Back, J. W., Sanz, M. A., De Jong, L., De Koning, L. J., Nijtmans, L. G., De Koster, C. G., Grivell, L. A., Van Der Spek, H., and Muijsers, A. O. (2002) A structure for the yeast prohibitin complex: structure prediction and evidence from chemical crosslinking and mass spectrometry. *Protein Sci.* **11**, 2471–2478
  10. Taverner, T., Hall, N. E., O'Hair, R. A., and Simpson, R. J. (2002) Characterization of an antagonist interleukin-6 dimer by stable isotope labeling, cross-linking, and mass spectrometry. *J. Biol. Chem.* **277**, 46487–46492
  11. Collins, C. J., Schilling, B., Young, M., Dollinger, G., and Guy, R. K. (2003) Isotopically labeled crosslinking reagents: resolution of mass degeneracy in the identification of crosslinked peptides. *Bioorg. Med. Chem. Lett.* **13**, 4023–4026
  12. Kruppa, G. H., Schoeniger, J., and Young, M. M. (2003) A top down approach to protein structural studies using chemical cross-linking and Fourier transform mass spectrometry. *Rapid Commun. Mass Spectrom.* **17**, 155–162
  13. Schilling, B., Row, R. H., Gibson, B. W., Guo, X., and Young, M. M. (2003) MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J. Am. Soc. Mass Spectrom.* **14**, 834–850
  14. de Koning, L. J., Kasper, P. T., Back, J. W., Nessen, M. A., Vanrobaeys, F., Van Beeumen, J., Gherardi, E., de Koster, C. G., and de Jong, L. (2006) Computer-assisted mass spectrometric analysis of naturally occurring and artificially introduced cross-links in proteins and protein complexes. *FEBS J.* **273**, 281–291
  15. Seebacher, J., Mallick, P., Zhang, N., Eddes, J. S., Aebersold, R., and Gelb, M. H. (2006) Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing. *J. Proteome Res.* **5**, 2270–2282
  16. Gao, Q., Xue, S., Doneanu, C. E., Shaffer, S. A., Goodlett, D. R., and Nelson, S. D. (2006) Pro-CrossLink. Software tool for protein cross-linking and mass spectrometry. *Anal. Chem.* **78**, 2145–2149
  17. Anderson, G. A., Tolic, N., Tang, X., Zheng, C., and Bruce, J. E. (2007) Informatics strategies for large-scale novel cross-linking analysis. *J. Proteome Res.* **6**, 3412–3421
  18. Petrotchenko, E. V., Olkhovik, V. K., and Borchers, C. H. (2005) Isotopically coded cleavable cross-linker for studying protein-protein interaction and protein complexes. *Mol. Cell. Proteomics* **4**, 1167–1179
  19. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567
  20. Ciferri, C., De Luca, J., Monzani, S., Ferrari, K. J., Ristic, D., Wyman, C., Stark, H., Kilmartin, J., Salmon, E. D., and Musacchio, A. (2005) Architecture of the human Ndc80-Hec1 complex, a critical constituent of the outer kinetochore. *J. Biol. Chem.* **280**, 29088–29095
  21. Tarricone, C., Perrina, F., Monzani, S., Massimiliano, L., Kim, M. H., Dere-wenda, Z. S., Knapp, S., Tsai, L. H., and Musacchio, A. (2004) Coupling PAF signaling to dynein regulation: structure of LIS1 in complex with PAF-acetylhydrolase. *Neuron* **44**, 809–821
  22. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858
  23. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nano-electrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670
  24. Ishihama, Y., Rappsilber, J., Andersen, J. S., and Mann, M. (2002) Micro-columns with self-assembled particle frits for proteomics. *J. Chromatogr. A* **979**, 233–239
  25. Olsen, J. V., and Mann, M. (2004) Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13417–13422
  26. Gaucher, S. P., Hadi, M. Z., and Young, M. M. (2006) Influence of crosslinker identity and position on gas-phase dissociation of Lys-Lys crosslinked peptides. *J. Am. Soc. Mass Spectrom.* **17**, 395–405
  27. Muller, D. R., Schindler, P., Towbin, H., Wirth, U., Voshol, H., Hoving, S., and Steinmetz, M. O. (2001) Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Anal. Chem.* **73**, 1927–1934
  28. Beer, I., Barnea, E., Ziv, T., and Admon, A. (2004) Improving large-scale proteomics by clustering of mass spectrometry data. *Proteomics* **4**, 950–960
  29. Moore, R. E., Young, M. K., and Lee, T. D. (2002) Qscore: an algorithm for evaluating SEQUEST database search results. *J. Am. Soc. Mass Spectrom.* **13**, 378–386
  30. Cleveland, D. W., Mao, Y., and Sullivan, K. F. (2003) Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell* **112**, 407–421
  31. Vallee, R. B., and Tsai, J. W. (2006) The cellular roles of the lissencephaly gene LIS1, and what they tell us about brain development. *Genes Dev.* **20**, 1384–1393
  32. Mori, D., Yano, Y., Toyo-oka, K., Yoshida, N., Yamada, M., Muramatsu, M., Zhang, D., Saya, H., Toyoshima, Y. Y., Kinoshita, K., Wynshaw-Boris, A., and Hirotsune, S. (2007) NDEL1 phosphorylation by Aurora-A kinase is essential for centrosomal maturation, separation, and TACC3 recruitment. *Mol. Cell. Biol.* **27**, 352–367
  33. Liang, Y., Yu, W., Li, Y., Yu, L., Zhang, Q., Wang, F., Yang, Z., Du, J., Huang, Q., Yao, X., and Zhu, X. (2007) Nudel modulates kinetochore association and function of cytoplasmic dynein in M phase. *Mol. Biol. Cell* **18**, 2656–2666
  34. Ciferri, C., Musacchio, A., and Petrovic, A. (2007) The Ndc80 complex: hub of kinetochore activity. *FEBS Lett.* **581**, 2862–2869
  35. Wei, R. R., Sorger, P. K., and Harrison, S. C. (2005) Molecular organization of the Ndc80 complex, an essential kinetochore component. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5363–5367
  36. Wei, R. R., Schnell, J. R., Larsen, N. A., Sorger, P. K., Chou, J. J., and Harrison, S. C. (2006) Structure of a central component of the yeast kinetochore: the Spc24p/Spc25p globular domain. *Structure (Lond.)* **14**, 1003–1009
  37. Wei, R. R., Al-Bassam, J., and Harrison, S. C. (2007) The Ndc80/HEC1 complex is a contact point for kinetochore-microtubule attachment. *Nat. Struct. Mol. Biol.* **14**, 54–59
  38. de Godoy, L. M., Olsen, J. V., de Souza, G. A., Li, G., Mortensen, P., and Mann, M. (2006) Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol.* **7**, R50
  39. Lupas, A., Van Dyke, M., and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164