Structural Model of Lymphocyte Receptor NKR-P1C Revealed by Mass Spectrometry and Molecular Modeling

ARTICLE *in* ANALYTICAL CHEMISTRY · DECEMBER 2012

Impact Factor: 5.64 · DOI: 10.1021/ac302860m · Source: PubMed

analytical chemistry

Article

pubs.acs.org/ac

Structural Model of Lymphocyte Receptor NKR-P1C Revealed by Mass Spectrometry and Molecular Modeling

Daniel Rozbesky,[†,‡] Zofie Sovova,[§,∥] Julien Marcoux,[⊥] Petr Man,[†,‡] Rudiger Ettrich,[§,∥] Carol V. Robinson,[⊥] and Petr Novak*[,†,‡]

[†]Department of Biochemistry, Charles University, Prague, Czech Republic

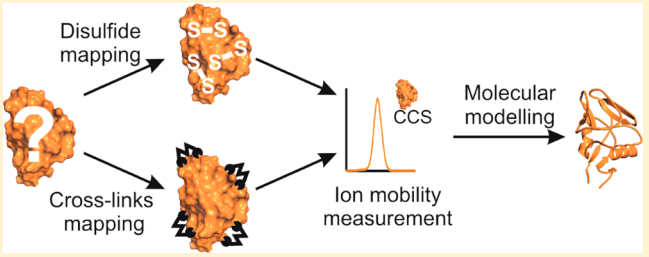
[‡]Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague, Czech Republic

[§]Institute of Nanobiology and Structural Biology, Global Change Research Center, Academy of Sciences of the Czech Republic, Nove Hrady, Czech Republic

[∥]Faculty of Sciences, University of South Bohemia, Nove Hrady, Czech Republic

[⊥]Department of Chemistry, University of Oxford, Oxford, U.K.

ABSTRACT: NKR-P1C is an activating immune receptor expressed on the surface of mouse natural killer cells. It has been widely used as a marker for NK cell identification in different mice strains. Recently we solved a crystal structure of the C-type lectin-like domain of a homologous protein, NKR-P1A, using X-ray crystallography and also described the strategy for rapid characterization of the protein conformation in solution. This procedure utilized chemical cross-linking, hydrogen/deuterium exchange, and molecular modeling. It was found that the solution structure differs from the crystal structure in the conformation of the loop region. The loop, detached from the protein compact core in the crystal structure, is closely attached to the core of the protein in solution. Here we present and interpret the solution structure of the C-type lectin-like domain of NKR-P1C using chemical cross-linking and molecular modeling. The validation of the model and conformation of the loop region in NKR-P1C were addressed using ion-mobility mass spectrometry.



The C-type lectin-like receptor NKR-P1C is an activating immune receptor expressed on the surface of natural killer (NK) cells, key players in antiviral and antitumor immunity.[1] NKR-P1C, originally known as NK1.1 antigen, is a common marker for the identification of NK cells in certain mice strains.[2,3] It also plays a crucial role in the stimulation of NK cell cytotoxicity of target cells and appears to be involved in natural killing of some tumor targets.[4,5] Antibody cross-linking of NKR-P1C receptor leads to Ca^{2+} flux, phosphatidylinositol turnover, kinase activity, and cytokine production. It was also reported that its intracellular tail associates with the tyrosine kinase, p56 (lck).[6] From a structural point of view, the NKR-P1C is a type II transmembrane protein, and in the literature it is described as a disulfide-linked homodimer with a fold typical for the C-type lectin-like family.[7,8] Although the NKR-P1C was discovered more than 20 years ago, its structure remains unknown. Interestingly, NKR-P1 family includes both activating and inhibitory receptors and the C-type lectin-like domain of NKR-P1 family members share considerable homology; the structural origins of ligand-binding specificity are thus of high interest.[9]

Structure of the extracellular part, namely, the C-type lectin-like domain, of a homologous protein, NKR-P1A, was solved recently using X-ray crystallography. However, the fold of the protein found in the crystal differed from the fold of known homologous proteins (e.g., CD69, NKG2D) in the loop region which was detached from the protein compact core and provided significant interaction surface with the other molecules of NKR-P1A in the crystal.[10] In order to validate the structure in solution and at near physiological conditions, we developed a protocol combining chemical cross-linking, deletion of the loop region, hydrogen/deuterium exchange coupled to mass spectrometry, and homology modeling. Surprisingly we found that the loop region was collapsed and closely interacting with the rest of the protein.[11] Sequence and phylogenetic analysis suggests that an extended loop region evolved in concert with target specificity, adapting the chemical properties of the residues to specific target ligands while embedding them in a structurally stable framework, suggesting that the long loop plays a key role in ligand specificity, and that a conformational equilibrium might be involved in ligand binding.[9]

In this study we use a set of methods, ranging from chemical cross-linking with sequence threading using mass spectrometry, restraint-based molecular modeling and steered molecular dynamics, to ion-mobility mass spectrometry, not only to

Received: October 2, 2012
Accepted: December 18, 2012
Published: December 18, 2012

   dx.doi.org/10.1021/ac302860m | *Anal. Chem.* 2013, 85, 1597−1604

describe the overall fold of NKR-P1C but also to elucidate the solution structure of the large extended loop described above.

Traditional methods of protein structure determination rely on high-resolution techniques: X-ray crystallography and NMR spectroscopy. However, with the technological advances and increasing demands to study protein structure of dynamic complexes under physiological conditions,[12] complementary low-resolution methods are becoming more widely used. The main advantages of low resolution techniques are the low sample requirement, which applies to proteins difficult to express/isolate, and the ability to work at near-physiological conditions with dynamic, heterogeneous proteins at low concentrations. Moreover, there is virtually no limitation in terms of protein size, flexibility, and time requirements for both data acquisition and data processing are relatively low. Low-resolution approaches usually apply several methodological concepts in concert, and their main role is in characterization of proteins or protein systems not amenable by NMR or X-ray.[13] An excellent example of this approach is the description of the interaction between the 14-3-3 protein and the regulator of G-protein signaling (with an unstructured N-terminal half) by combination of small-angle X-ray scattering, spectroscopic techniques, hydrogen/deuterium exchange coupled to mass spectrometry and molecular modeling.[14] Other examples include description of quaternary structure of eukaryotic RNAPII complex or assembly of proteasome.[15,16]

The concept of chemical cross-linking in combination with sequence threading was introduced more than a decade ago to determine the fold of bovine basic fibroblast growth factor 2 (FGF-2).[17] Since then, the implementation of protein cross-linking has expanded and various problems were addressed using this technique, protein conformational changes,[18] details of protein structure,[11] a mechanism of small proteins,[19] a topology of protein−peptide complexes,[20] or an architecture of multiprotein complexes.[21,22] In chemical cross-linking, side chains of amino acid residues are targeted by bifunctional reagents in which the two reactive groups are interconnected by a spacer arm of variable length. If two residues are within a certain distance from each other, they may be covalently connected by a cross-linking reagent. In this approach, mass spectrometry is used for identification of cross-linked peptides and also for exact localization of the cross-link.[23,24] Identification of cross-linked peptides may provide two types of information. One is simple assignment of protein regions/residues involved in the interaction. This is the main aim of large scale studies mapping the entire cellular interactome.[25] A second way to use information from chemical cross-linking is to derive distance constraints in order to generate low resolution structure. Here, the number of cross-links (distances) dictates the resolution of the final structure. High-resolution structure requires $3N$ distances where N is the number of amino acid residues within a protein. However, low-resolution structure or fold identification can be done with less information, and $N/10$ distances were described as a minimal data set.[17] Therefore, a variety of cross-linkers must be employed to overcome possible problems such as low reactivity[26] and to generate enough constraints.[18,27] Taken from this point of view, disulfide bridges stabilizing the protein structure[28,29] represent a native cross-link and their identification may facilitate protein homology modeling and quality control in in vitro recombinant protein production.[30,31] Tools for disulfide pattern analysis have advanced significantly in the past 2 decades. Mass spectrometric-based strategies of disulfide arrangement are based on either partial reduction of a protein in conjunction with cysteine alkylation[32] or hydrolysis of intact proteins and subsequent analysis of disulfide-linked peptides[33,34] or top down strategies.[35]

In cases where at least one member of a "fold-family" is known, homology modeling can be used to analyze the three-dimensional structure of proteins of unknown structure. The approach chosen in this study is a back and forth between experiment and theory and represents a cycle of structure generation and verification, in which distance constraints derived from disulfides and chemical cross-linking of NKR-P1C are not only directly implemented in the structure modeling but also used in the next cycle for the experimental validation of the proposed structure and the improvement of the modeling accuracy. However, because of differences between the crystal structure and in-solution structure of NKR-P1A mentioned above, we put the strong focus on the loop position also in NKR-P1C. H/D exchange and deletion of the loop was used previously to assess the loop conformation in NKR-P1A. The results were consistent with the output from the chemical cross-linking but instead of deleting a large portion of the protein we used here native electrospray coupled to ion mobility mass spectrometry (IM-MS) to differentiate between the two possible loop conformations.

This technique can be seen as gas-phase chromatography, in which the ionized molecules are transmitted in a drift cell where they are submitted to a weak electric field and a relatively high pressure of neutral gas (typically helium or nitrogen) to separate them on the basis of their collisional cross section (CCS). Therefore, even the species with the same molecular weight but different conformation can be effectively separated. Since its first applications to small proteins in the early 90s,[36] IM-MS has progressed significantly, enabling nowadays the study of systems of increasing complexity. Since the work of Ruotolo et al.[37] confirming that quaternary structures could be maintained in the gas phase, IM-MS has provided additional information about large protein complexes, especially when combined with other structural techniques such as EM,[38] NMR,[39] or cross-linking.[16] This hybrid approach was shown to be invaluable for model filtering[40] or for the analysis of conformational isoforms.[41,42]

EXPERIMENTAL SECTION

Protein Expression and Purification. The extracellular portion of the C57BL/6 mouse NKR-P1C (residues 89−223) was produced in *E. coli* into inclusion bodies. Protein was refolded in vitro and purified using ion-exchange chromatography as described previously.[43] The purified protein was dialyzed against buffer containing 10 mM HEPES (pH 7.0), 150 mM NaCl, and 1 mM NaN_3, and protein was concentrated to 10 mg/mL using Amicon Ultra centrifugal filter unit (Millipore). The monoisotopic mass of intact protein was measured on a 9.4 T Apex Ultra Qe Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (Bruker Daltonics), confirming the identity of the protein.

Disulfide Bonds Determination. The disulfide bonds in NKR-P1C were determined according to the previously published protocol.[34] Briefly, the protein was separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and digested by trypsin or Asp-N protease under nonreducing conditions in the presence of 200 μM cystamine. The peptide mixtures were desalted on a peptide MacroTrap (Michrom Bioresources).

dx.doi.org/10.1021/ac302860m | *Anal. Chem.* 2013, 85, 1597−1604

Chemical Cross-Linking. The NKR-P1C protein was cross-linked at a concentration of 0.2 mg/mL with the homobifunctional cross-linkers disuccinimydyl suberate (DSS) and disuccinimydyl glutarate (DSG). For the cross-linking reaction, NKR-P1C was dialyzed into 50 mM triethylamine/bicarbonate buffer (pH 7.5) with 50 mM NaCl, and cross-linkers DSS and DSG were used as 1:1 (mol/mol) mixtures of nondeuterated and four-times deuterated compounds (d_0/d_4). Freshly prepared stock solutions of cross-linkers (10 mg/mL in DMSO) were added in 10× molar excess to NKR-P1C, and reaction mixtures were incubated for 1 h at room temperature.

Electrophoretic Separation and In Gel Digestion. Following cross-linking, NKR-P1C was separated on a NuPAGE 4−12% Bis-Tris gel using MES running buffer and the bands corresponding to monomeric cross-linked protein were excised. The disulfide bonds were reduced with 50 mM TCEP for 10 min at 90 °C, and free cysteines were alkylated with 50 mM iodoacetamide for 90 min at room temperature in the dark. Trypsin and Asp-N digestion proceeded overnight at 37 °C with an enzyme/protein ratio of 1:20 (w/w). The resulting peptide mixtures were desalted on a peptide MacroTrap column (Michrom Bioresources).

LC−MS Analysis. After desalting, the samples (disulfide bond assignment or chemical cross-linking) were loaded onto a reverse phased column MAGIC C18 column (0.2 mm × 150 mm, Michrom Bioresources) and separated on a capillary HPLC system (Agilent Technologies) at a flow rate of 4 μL/min under the following gradient conditions: 1−10% B in 1 min, 10−40% B in 29 min, 40−95% B in 5 min, where solvent A was 0.2% formic acid, 2.5% acetonitrile, and 2.5% isopropanol in water and solvent B was 0.16% formic acid in 90% acetonitrile and 5% isopropanol. The column was connected directly to an Apex-ULTRA Qe FTICR mass spectrometer (Bruker Daltonics) equipped with a 9.4 T superconducting magnet using an electrospray ion source. The instrument was calibrated externally using arginine clusters resulting in mass accuracy below 2 ppm.

Data Analysis. Data acquisition and data processing were performed using ApexControl 3.0.0 and DataAnalysis 4.0 (Bruker Daltonics), respectively. The disulfide bonds and cross-links were identified using Links software.[17] The Links algorithm was set to consider the possible single oxidation of methionine, and different cysteine modifications were set for disulfide mapping (disulfide bond formation) and for cross-linking (carbamidomethylation). The enzyme specificities were set to K-/R- for trypsin and -D/-E for Asp-N and the missed cleavages were set to 3. Modification or cross-link of amino groups (protein N-terminus and lysine side chains) by DSS or DSG was considered for cross-linking experiments. The mass error threshold was kept below 3 ppm, and all assigned fragments were verified manually.

Homology Modeling and Molecular Dynamics. Homology modeling followed by a short steepest descent minimization was performed using the MODELER 9v7 package.[44] The recently published crystal structure of mouse NKR-P1A[10] (pdb-code 3M9Z) was used as a template for the core of the protein, and the published structure of human CLEC5A[45] (pdb-code 2YHF) was used for the extended loop region. This structure was chosen as it has the highest homology with our sequence (35%) in the extended loop region. In total, 10 models were generated and the best one was chosen based on the distribution of the torsional angles of amino acids in favored regions in the Ramachandran diagram and on visual inspection of the model (both in Swiss PDBViewer[46] and the MODELER objective function). The three-dimensional model is based on all non-hydrogen atoms. All three disulfide bridges were already created during the restraint-based modeling procedure and refinement was achieved through algorithmic analysis and minimization with the Amber03 force field56 in YASARA[47] after hydrogen atoms were added. Confrontation of the best "blind" model with cross-linking data showed that all distances determined experimentally are in agreement with the model. Constraints for the loop region coming from the experimentally determined distances were solved by nonequilibrium molecular dynamics simulation in vacuo using YASARA with the Amber03 force field applying an external force on the loop atoms in the direction of coordinates within the experimentally determined range. A nonbonded interaction cutoff of 7.8 Å was applied using a multiple time step of 1 fs for intramolecular and 2 fs for intermolecular forces and the simulation was then run at 298 K and constant pressure. The external forces were interactively applied to the loop unless the cross-linking distances were fulfilled. During the nonequilibrium molecular dynamics, the rest of the system was fixed to prevent distortion of the protein core. Additionally, torsion angles in K212 were manually adjusted to reorient the side chain in the opposite direction. Once all experimental constraints were fulfilled, the whole structure was minimized by 2000 steps of simulated annealing minimization.

Native Mass Spectrometry and Ion Mobility Measurement. The protein was buffer exchanged in 200 mM ammonium acetate pH 7.4 using a micro BioSpin-6 device (Biorad) and diluted in the same buffer to reach 13 μM concentration. The sample was then loaded into gold-coated nanoflow capillaries prepared in house.[48] IM-MS data was acquired on a quadrupole ion-mobility time-of-flight mass spectrometer (Synapt HDMS, Waters) modified as described previously such as the collisional cross section (CCS) can be determined directly by plotting the drift time versus the reciprocal drift voltage[40,49] as long as the temperature and pressure in the IM cell are determined accurately. The capillary, cone, and trap collision energy were set at 1.6 kV, 20 V, and 10 V, respectively, in order to minimize unfolding in the gas phase. The backing pressure was set at 6−8 mbar, and the pressure in the drift cell was set at 2.17 Torr using a flow of helium of 50 mL/min. The temperature outside the drift cell was monitored with a thermocouple (Omega Engineering). The drift voltage was increased from 50 to 200 V in 5−30 V steps. The drift time for each charge state and each drift voltage were obtained using MassLynx and DriftScope (Waters). Calibration was performed using a 10 mg/mL solution of cesium iodide. Theoretical CCS for the NKR-P1C and its homologue NKR-P1A (pdb code 3M9Z) were calculated using the projection approximation (PA) method implemented in MOBCAL and scaled, as described previously.[50]

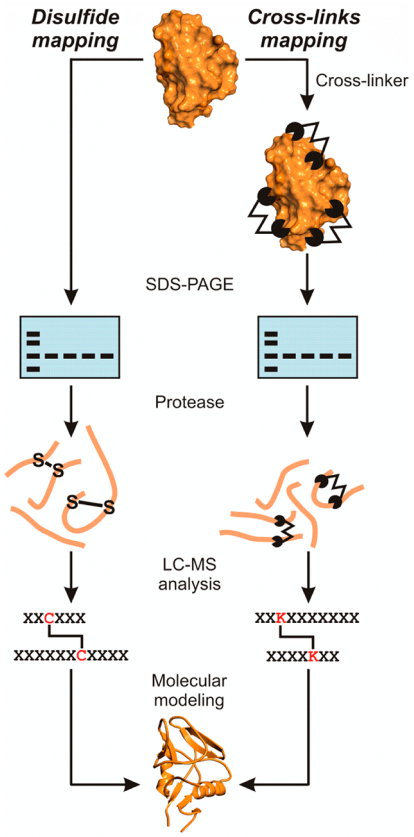
RESULTS AND DISCUSSION

Disulfide Mapping of Mouse NKR-P1C. The C-type lectin-like domain of recombinant mouse NKR-P1C (NKR-P1C) contains 135 amino acid residues including six cysteine residues. In order to determine the cysteine status, we measured the mass of the intact protein by electrospray (ESI)-FTICR MS. Three disulfide bonds seemed to be present in the NKR-P1C as the monoisotopic mass of the intact protein $[M + H]^+$ of 15 394.4421 Da was in agreement with the predicted monoisotopic mass calculated with the assumption of

dx.doi.org/10.1021/ac302860m | Anal. Chem. 2013, 85, 1597−1604

all paired cysteine residues. The experimental setup for the disulfide mapping is depicted in Scheme 1. Digestion of the

Scheme 1. Schematic Workflow of Mass Spectrometric Disulfide and Cross-Link Mapping[a]



[a]In the analysis of disulfide pattern, the protein is separated using SDS-PAGE containing cystamine to prevent disulfide bonds from scrambling. Protein is subsequently in-gel digested, and peptides are analyzed by capillary liquid chromatography coupled to high-resolution mass spectrometry. Accurate masses of intact peptides and disulfide linked peptides are subjected to database searching to identify disulfide-linked peptides. In chemical cross-linking, the protein is cross-linked in solution with a mixture of light and heavy cross-linkers and cross-linked protein is separated using SDS-PAGE. Cross-links are identified in the same fashion as disulfide-linked peptides. Finally, disulfide pattern and distance constraints derived from cross-links are used in homology modeling.

protein by Asp-N or trypsin proteases generated peptides containing disulfide-linkages that were then subjected to reverse phase capillary HPLC followed by ESI-FTICR MS. After Asp-N digestion, an unambiguous disulfide pattern of NKR-P1C was determined. The first disulfide bond between Cys94 and Cys105 was assigned based on the dipeptide with the monoisotopic mass $[M + H]^+$ of 2951.2893 (Figure 1a). The other cystine dipeptide with the monoisotopic mass $[M + H]^+$ of 3468.7075 (Figure 1b) indicated the presence of a bond between Cys122 and Cys 210. Finally, the monoisotopic mass $[M + H]^+$ of 2910.4065 (Figure 1c) matches the dipeptide with a link between Cys189 and Cys202. Trypsin digestion provided verification of the assigned pattern by identification of a cystine dipeptide with the monoisotopic mass $[M + H]^+$ of 2840.3059 (disulfide bond between Cys189 and Cys202) and tripeptide containing the four remaining cysteines, all involved in disulfide

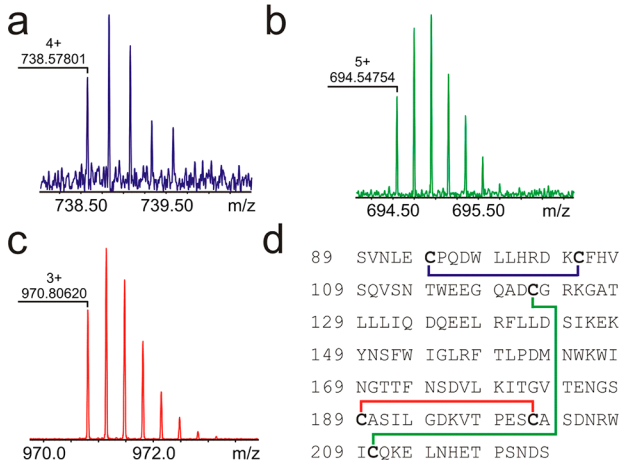


Figure 1. Disulfide mapping of mouse NKR-P1C. (a) Representative mass spectra of cystine peptide Cys94/Cys105, (b) cystine peptide Cys122/Cys210, and (c) cystine peptide Cys189/Cys202 from Asp-N digest. (d) Protein sequence of the C-type lectin-like domain of NKR-P1C with schematic drawing of disulfide arrangement.

bonds (monoisotopic mass $[M + H]^+$ 4859.1958). All experimentally determined disulfide linkages in NKR-P1C are summarized in Table 1 and in Figure 1d.

Table 1. Summary of Disulfide Bond Analysis by LC−MS[a]

Cys-Cys cross-link	disulfide-linked peptides(s)	theoretical $[M + H]^+$	experimental $[M + H]^+$	error (ppm)
Cys94/ Cys105[a]	89-96/103-120	2951.2934	2951.2893	1.5
Cys122/ Cys210[a]	121-133/205-221	3468.7111	3468.7075	1.2
Cys189/ Cys202[a]	176-194/195-204	2910.4071	2910.4065	0.3
Cys94/ Cys105[t]	89-124/208-212	4859.1973	4859.1958	0.3
Cys122/ Cys210[t]	89-124/208-212	4859.1973	4859.1958	0.3
Cys189/ Cys202[t]	180-196/197-207	2840.3037	2840.3059	0.6

[a]Identified cross-linked peptides and corresponding disulfide bonds are shown in the first two columns. Upper indices [a] and [t] in the first column indicate the protease (a, AspN; t, trypsin) by which the protein was digested. Theoretical and experimental masses and mass errors are shown in the third and fourth columns, respectively.

Chemical Cross-Linking and Mass Spectrometry. The presence of nine lysine residues in NKR-P1C provides a good basis for structural studies using chemical cross-linking of primary amines. The schematic workflow of the cross-linking and cross-links identification is depicted in Scheme 1. The NKR-P1C was incubated with 10× molar excess of DSS or DSG. Both cross-linkers were applied as a 1:1 (mol/mol) mixture of the nondeuterated and deuterated (d_0/d_4) derivate to then facilitate the identification of cross-linked peptides. The concentration of cross-linking reagents was optimized to give the highest relative yield of protein containing a single intramolecular cross-link, which minimized the likelihood of distorting the tertiary structure. The reactions were not quenched but the cross-linkers were allowed to hydrolyze in order to reduce the complexity of the reaction mixture. Under these conditions, no aggregates were observed on SDS-PAGE and position of the cross-linked NKR-P1C corresponded to the

dx.doi.org/10.1021/ac302860m | Anal. Chem. 2013, 85, 1597−1604

control represented by the unmodified protein (Figure 2a). Following the in-gel reduction and alkylation of cysteines, the

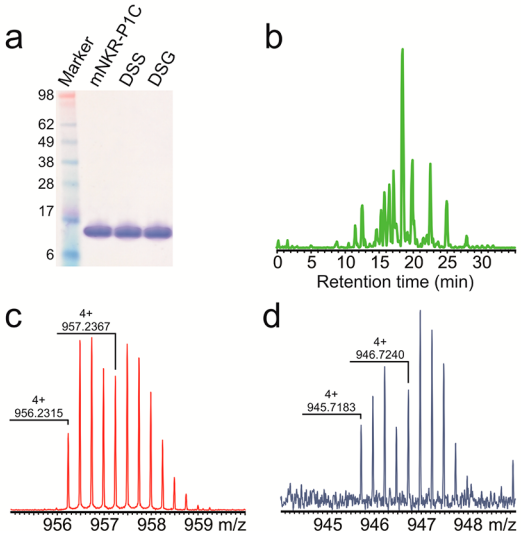


Figure 2. Chemical cross-linking and mass spectrometric cross-links identification. (a) Coomassie stained SDS PAGE gel of unmodified NKR-P1C which was used as a sample control, and the NKR-P1C cross-linked with DSS-d_0/d_4 and DSG-d_0/d_4. (b) Representative total ion chromatogram of the reverse-phase capillary HPLC and FTICR MS of trypsin-digested NKR-P1C cross-linked with DSS. (c, d) Representative mass spectra of peptides 125-139/208-223 cross-linked with DSS-d_0/d_4 or DSG-d_0/d_4. Because the protein was cross-linked using a 1:1 mixture of light and heavy cross-linker, the peptides containing the cross-linker are recognized by their distinct doublet isotope pattern with a mass difference of 4.025 Da in the deconvoluted mass spectra.

NKR-P1C was digested by trypsin or Asp-N. Generated peptides were separated by capillary HPLC connected to ESI-FTICR MS (Figure 2b). Although the peptides with deuterated cross-linkers showed small retention time shift compared to their nondeuterated counterparts, the difference did not complicate the data analysis. The cross-links were identified using Links software,[17] which matches a peak list obtained from the LC–MS analysis with theoretical masses of all possible products. The use of heavy and light cross-linkers resulted in characteristic isotopic pattern of peptides bearing the cross-linker (mass difference of 4.025 Da in the deconvoluted mass spectra, Figure 2c,d). This specific signature, together with high mass accuracy (below 2 ppm) provided unambiguous assignment of the cross-links. All unique cross-linked residues derived from the identified cross-links are listed in Table 2. Out of 10 cross-linked residues identified, 7 were formed with DSS and 3 with DSG. Similar cross-linking pattern was observed using DSS and DSG, with the exception of Lys146-Lys148, Lys148-Lys196, and Lys104-Lys125, which were not detected with the shorter cross-linker (DSG). All experiments including disulfide bond assignment as well as chemical cross-linking were done in triplicate and all presented cross-links were identified in all three repetitions. There were no cross-links found in just one or two experiments confirming the overall structural homogeneity of the protein.

Molecular Modeling. The three-dimensional model of NKR-P1C shown in Figure 3a has the typical C-type lectin-like fold and three disulfide bridges. Within the whole cross-linking network (Figure 3b) only Lys166 and Lys179 are part of the

Table 2. Table of Identified Cross-Links[a]

cross-linking reagent	cross-linked peptide(s)	Lys-Lys cross-link	theoretical d_0-[M + H]$^+$	experimental d_0-[M + H]$^+$	error (ppm)
DSS	103-120/121-133[a]	K104-K125	3703.7847	3703.7915	1.8
DSS	125-139/208-223[t]	K125-K212	3821.9015	3821.9043	0.6
DSS	140-157[t]	K146-K148	2367.2747	2367.2663	2.7
DSS	147-157/180-207[t]	K148-K196	4488.1446	4488.1382	1.5
DSS	162-175/195-204[a]	K166-K196	2944.3142	2944.3142	0.3
DSS	176-194/195-204[a]	K179-K196	3164.5338	3164.5356	0.5
DSS	186-200/201-213[a]	K196-K212	3338.5059	3338.5059	0.9
DSG	125-139/208-223[t]	K125-K212	3779.8545	3779.8511	1.0
DSG	162-175/195-199[a]	K166-K196	2368.0914	2368.0859	2.5
DSG	176-194/195-204[a]	K179-K196	3122.4868	3122.4951	2.5

[a]First column indicates the cross-linking reagent used, second column shows the peptide/dipeptide based on which the cross-link (third column) was identified. Type of protease used for digestion of the cross-linked protein is shown as the upper indices in the second column: a, AspN; t, trypsin. Theoretical and experimental masses of nondeuterated cross-link and mass errors are shown in the fourth, fifth, and sixth columns, respectively.

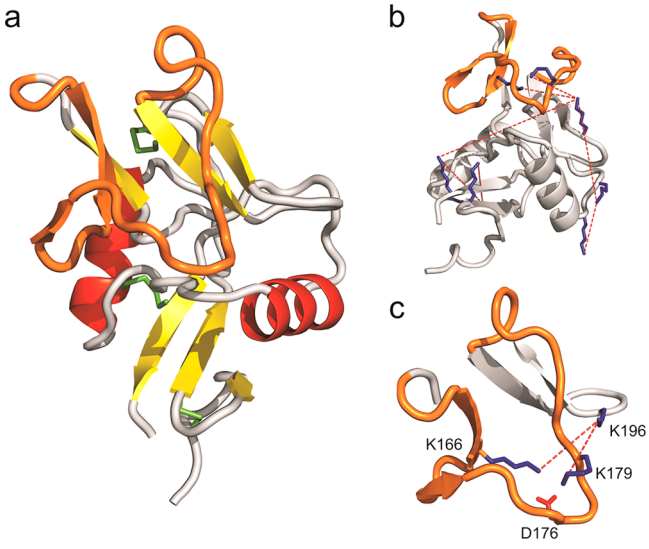


Figure 3. (a) Structural model of the mouse NKR-P1C with the loop region colored in orange and disulfide bonds colored in green. (b) Model of NKR-P1C with lysine residues involved in chemical cross-linking. Cross-linked residues are highlighted using dashed line. (c) Detail view on the loop region with K166 and K179 interacting with the negatively charged aspartic acid D176. These two residues were shown to be cross-linked to K196.

loop region, whereas the other lysine residues are located in the globular domain. The experimental constraints matched conformations in which the loop is packed onto the protein surface (cross-links between Lys196 from the core region and Lys166 or Lys179 from the loop region) in a similar way as in the dengue virus receptor CLEC5A.[45] Furthermore, these distances excluded the conformation in which the extended

dx.doi.org/10.1021/ac302860m | Anal. Chem. 2013, 85, 1597–1604

loop points away from the globular core as seen in the crystal structure of NKR-P1A.[10] However, the cross-link between loop lysine residues Lys166 and Lys179 was not observed although these two residues are close to each other in the model. Since the absence of cross-links can point on some structural details,[51] we propose the following explanation. The absence of cross-link between Lys166 and Lys179 on one side and existing cross-links between Lys166 and Lys196 or Lys179 and Lys196 on the other side may be explained by a hydrogen bond network between positively charged side chains of Lys166 or Lys179 and negatively charged aspartic acid Asp176 (Figure 3c). Hydrogen bonding between Lys166 and Asp176 will leave Lys179 reactive toward Lys196 and vice versa, and bonding between Lys179 and Asp176 will allow formation of cross-link between Lys166 and Lys196.[26] In addition, such conformation would disfavor cross-linking between Lys166 and Lys176. Indeed, as no cross-links between these two residues were ever detected, this interaction should be stable in time and in solution. Furthermore, Lys166 is part of a β-sheet that plays a major role in the C-type lectin-like fold and thus constrains its fluctuation. Finally, all experimentally determined distances are correctly represented in the structural model and are consistent with the ability of DSS and DSG to cross-link primary amines on residues. The extended loop hereby seems to be packed onto the protein surface partially stabilizing the second β-sheet, which is probably the most populated conformation in solution.

Native Mass Spectrometry and Ion Mobility Measurement. In order to verify the fold of NKR-P1C, native mass spectrometry with ion mobility measurements were performed. Under native conditions, the protein sprays as a monomer (Figure 4a) with main charge states ranging from 6+ to 8+. The measurement of the drift time under increasing acceleration voltages for these charge states (Figure 4b) enabled us to calculate a collisional cross section (CCS) for each one of these charge states. The values for the charge states 6+, 7+, and 8+ were 1444 ± 10 Å^2, 1495 ± 16 Å^2, and 1491 ± 16 Å^2, respectively. The monomodal and symmetrical distributions of the arrival time distributions (ATD) (Figure 4c) strongly suggest that the protein is present in a unique, very homogeneous conformation. The value obtained for the lowest charge state (1444 Å^2), which is usually the least activated in the gas phase,[52] was very close to the theoretical CCS calculated for our model (1411 Å^2, Figure 4c). As a comparison, the more extended X-ray structure obtained for the homologue NKR-P1A (pdb-code 3M9Z) shows a much higher theoretical CCS of 1618 Å^2. The difference between these two CCS (~15%) is well above the usual experimental error (±6%).[40]

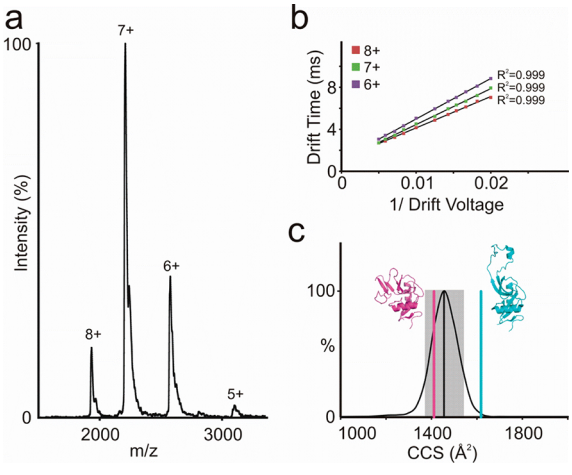


Figure 4. Ion mobility-mass spectrometry: (a) mass spectra of NKR-P1C under native conditions (200 mM ammonium acetate pH 7.4). (b) Plot of the drift times obtained for three different charge states and increasing drift voltages. (c) Arrival time distribution (ATD) obtained for the 6+ charge state (the black line represents the average ATD and the shaded area the ±6% experimental error). The pink and blue lines correspond to the theoretical CCS calculated for the model of NKR-P1C and its homologue NKR-P1A (pdb-code 3M9Z), respectively.

In this study, we combined molecular modeling together with chemical cross-linking and ion-mobility mass spectrometry to describe the structure of the extracellular part of an NK cell receptor NKR-P1C. Despite its high sequence similarity with NKR-P1A, all attempts to solve the high-resolution structure of NKR-P1C using X-ray crystallography failed. The crystals either did not grow or were of poor quality. Therefore we decided to use a combination of low-resolution MS-based methods with homology modeling to obtain the structure of this important lymphocyte receptor.

First, we obtained distance constraints using disulfide mapping and chemical cross-linking of NKR-P1C using amine-reactive cross-linkers DSS and DSG. In disulfide mapping, we first measured the mass of the intact protein by high-resolution ESI-FTICR MS. The experimental mass fitted very well to the theoretical mass with all six cysteine residues in oxidized form, confirming that the protein contains three disulfide bridges. Detailed LC−MS analysis of the disulfide-linked peptides generated from the Asp-N and tryptic digest revealed the unambiguous disulfide pattern. The identified disulfide arrangement in NKR-P1C is in agreement with the highly conserved disulfide pattern of C-type lectin-like domain[53] and with putative disulfide linkages reported previously[43] (Figure 1 and Table 1). Chemical cross-linking using nondeuterated and deuterated cross-linking agents[54] provided seven distances for DSS and four nonredundant constraints for DSG. Three DSS cross-links were cross-confirmed from the DSG experiment, and for the four remaining DSS cross-links the equivalents of DSG cross-links were not identified. This observation is consistent with the fact that cross-linkers with longer spacer arms are more likely to form cross-links compared to shorter ones.[55] However, it is unrealistic to combine data from cross-linkers with different spacer arms in order to determine lower boundaries for space distances in a quantitative way.[56]

Restraint-based computational modeling was then used to generate a model that represents the experimentally determined constraints with a minimum of violations. Molecular dynamics was used to refine the model and to describe the most populated protein conformers in solution. In addition to the positional constraints obtained from the disulfide mapping and from cross-linking experiments, the model needs to preserve the overall C-type lectin-like fold in these simulations, as the protein core is strongly conserved, and the template and modeled structure share a sequence identity of 88%. However, as crystal structures are rigid contrary to protein dissolved in solution, we allowed the side chains in the core to be more flexible and adapt to the given experimentally determined constraints. Specific attention was paid to the extended loop region (P161−G187) proposed to be involved in protein−ligand interactions and ligand specificity. The only crystal structure published to date for the entire NKR-P1 family,

dx.doi.org/10.1021/ac302860m | Anal. Chem. 2013, 85, 1597−1604

mouse NKR-P1A, shows this extended loop pointing away from the protein core, in a conformation in which the loop would be fully exposed to the solvent. Such a conformation could be clearly excluded from the cross-links of the protein in solution. This was further supported by IM-MS measurements corresponding to the compact form of the molecule based on the experimentally derived collisional cross section. Therefore, in the most populated conformation in solution, NKR-P1C most likely adopts the conformation similar to the solution structure of NKR-P1A. Our model enables us to describe this conformation on an atomic scale, showing its packing onto the protein surface while partially stabilizing the second β-sheet. This demonstrates not only the principal high flexibility of this region, with conformations of the extended loop ranging from full solvent exposure to a close attachment to the protein core. As this loop is proposed to play a key role in ligand specificity and varies in primary sequence in the NKR-P1 family members, we might consider its concrete dynamical behavior as being not a collective property of the whole family but rather a specific property of NKR-P1C. This loop might thus contribute to ligand-binding specificity and consequently to the differentiation between activating and inhibitory receptors.

In conclusion, this is the first report describing combined use of chemical cross-linking, homology modeling, and ion mobility mass spectrometry for the purpose of structure elucidation of a protein with no known high-resolution structure. We show that the structure of the C-type lectin-like domain of NKR-P1C receptor is in solution similar to the homologous protein NKR-P1A and that its loop region is attached to the protein core. This also nicely demonstrates how ion mobility mass spectrometry can be used in validation of low-resolution structures in cases where different possible conformations differ in their cross sections.

AUTHOR INFORMATION

Corresponding Author

*E-mail: pnovak@biomed.cas.cz. Phone: +420 2 4106 2156.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by Charles University Grant Agency (Grant 403211/2011), Charles University (Project UNCE_204025/2012), the Grant Agency of the Czech Republic (Grants P207/10/1040, P303/09/0477, and P207/10/1934), the Ministry of Education, Youth and Sports of the Czech Republic (Grant MSM6007665808), University of South Bohemia (Grant GAJU 170/2010/P), and the Academy of Sciences of the Czech Republic (Grant AVOZ60870520).

REFERENCES

(1) Vivier, E.; Raulet, D. H.; Moretta, A.; Caligiuri, M. A.; Zitvogel, L.; Lanier, L. L.; Yokoyama, W. M.; Ugolini, S. *Science* **2011**, *331*, 44−49.
(2) Glimcher, L.; Shen, F. W.; Cantor, H. *J. Exp. Med.* **1977**, *145*, 1−9.
(3) Koo, G. C.; Peppard, J. R. *Hybridoma* **1984**, *3*, 301−303.
(4) Karlhofer, F. M.; Yokoyama, W. M. *J. Immunol.* **1991**, *146*, 3662−3673.
(5) Reichlin, A.; Yokoyama, W. M. *Immunol. Cell Biol.* **1998**, *76*, 143−152.
(6) Ljutic, B.; Carlyle, J. R.; Filipp, D.; Nakagawa, R.; Julius, M.; Zuniga-Pflucker, J. C. *J. Immunol.* **2005**, *174*, 4789−4796.
(7) Ryan, J. C.; Turck, J.; Niemi, E. C.; Yokoyama, W. M.; Seaman, W. E. *J. Immunol.* **1992**, *149*, 1631−1635.
(8) Giorda, R.; Trucco, M. *J. Immunol.* **1991**, *147*, 1701−1708.
(9) Sovova, Z.; Kopecky, V., Jr.; Pazderka, T.; Hofbauerova, K.; Rozbesky, D.; Vanek, O.; Bezouska, K.; Ettrich, R. *J. Mol. Model.* **2011**, *17*, 1353−1370.
(10) Kolenko, P.; Rozbesky, D.; Vanek, O.; Kopecky, V., Jr.; Hofbauerova, K.; Novak, P.; Pompach, P.; Hasek, J.; Skalova, T.; Bezouska, K.; Dohnalek, J. *J. Struct. Biol.* **2011**, *175*, 434−441.
(11) Rozbesky, D.; Man, P.; Kavan, D.; Chmelik, J.; Cerny, J.; Bezouska, K.; Novak, P. *Anal. Chem.* **2012**, *84*, 867−870.
(12) Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R.; Berkowitz, S. A.; Houde, D. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 323−337.
(13) Serpa, J. J.; Parker, C. E.; Petrotchenko, E. V.; Han, J.; Pan, J.; Borchers, C. H. *Eur. J. Mass Spectrom. (Chichester, Eng.)* **2012**, *18*, 251−267.
(14) Rezabkova, L.; Man, P.; Novak, P.; Herman, P.; Vecer, J.; Obsilova, V.; Obsil, T. *J. Biol. Chem.* **2011**, *286*, 43527−43536.
(15) Lasker, K.; Phillips, J. L.; Russel, D.; Velazquez-Muriel, J.; Schneidman-Duhovny, D.; Tjioe, E.; Webb, B.; Schlessinger, A.; Sali, A. *Mol. Cell. Proteomics* **2010**, *9*, 1689−1702.
(16) Stengel, F.; Aebersold, R.; Robinson, C. V. *Mol. Cell. Proteomics* **2012**, *11*, R111 014027.
(17) Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5802−5806.
(18) Jacobsen, R. B.; Sale, K. L.; Ayson, M. J.; Novak, P.; Hong, J.; Lane, P.; Wood, N. L.; Kruppa, G. H.; Young, M. M.; Schoeniger, J. S. *Protein Sci.* **2006**, *15*, 1303−1317.
(19) Ahrman, E.; Lambert, W.; Aquilina, J. A.; Robinson, C. V.; Emanuelsson, C. S. *Protein Sci.* **2007**, *16*, 1464−1478.
(20) Schulz, D. M.; Ihling, C.; Clore, G. M.; Sinz, A. *Biochemistry* **2004**, *43*, 4703−4715.
(21) Maiolica, A.; Cittaro, D.; Borsotti, D.; Sennels, L.; Ciferri, C.; Tarricone, C.; Musacchio, A.; Rappsilber, J. *Mol. Cell. Proteomics* **2007**, *6*, 2200−2211.
(22) Chen, Z. A.; Jawhari, A.; Fischer, L.; Buchen, C.; Tahir, S.; Kamenski, T.; Rasmussen, M.; Lariviere, L.; Bukowski-Wills, J. C.; Nilges, M.; Cramer, P.; Rappsilber, J. *EMBO J.* **2010**, *29*, 717−726.
(23) Rappsilber, J. *J. Struct. Biol.* **2011**, *173*, 530−540.
(24) Novak, P.; Haskins, W. E.; Ayson, M. J.; Jacobsen, R. B.; Schoeniger, J. S.; Leavell, M. D.; Young, M. M.; Kruppa, G. H. *Anal. Chem.* **2005**, *77*, 5101−5106.
(25) Bruce, J. E. *Proteomics* **2012**, *12*, 1565−1575.
(26) Novak, P.; Kruppa, G. H.; Young, M. M.; Schoeniger, J. *J. Mass Spectrom.* **2004**, *39*, 322−328.
(27) Novak, P.; Kruppa, G. H. *Eur. J. Mass Spectrom. (Chichester, Eng.)* **2008**, *14*, 355−365.
(28) Matsumura, M.; Signor, G.; Matthews, B. W. *Nature* **1989**, *342*, 291−293.
(29) Wedemeyer, W. J.; Welker, E.; Narayan, M.; Scheraga, H. A. *Biochemistry* **2000**, *39*, 4207−4216.
(30) Pitt, J. J.; Da Silva, E.; Gorman, J. J. *J. Biol. Chem.* **2000**, *275*, 6469−6478.
(31) Jensen, P. K.; Harrata, A. K.; Lee, C. S. *Anal. Chem.* **1998**, *70*, 2044−2049.
(32) Smith, D. L.; Zhou, Z. R. *Methods Enzymol* **1990**, *193*, 374−389.
(33) Wallis, T. P.; Pitt, J. J.; Gorman, J. J. *Protein Sci.* **2001**, *10*, 2251−2271.
(34) Pompach, P.; Man, P.; Kavan, D.; Hofbauerova, K.; Kumar, V.; Bezouska, K.; Havlicek, V.; Novak, P. *J. Mass Spectrom.* **2009**, *44*, 1571−1578.
(35) McLafferty, F. W.; Fridriksson, E. K.; Horn, D. M.; Lewis, M. A.; Zubarev, R. A. *Science* **1999**, *284*, 1289−1290.

(36) Smith, R. D.; Loo, J. A.; Loo, R. R. O.; Busman, M.; Udseth, H. R. *Mass Spectrom. Rev.* **1991**, *10*, 359−451.
(37) Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. *Science* **2005**, *310*, 1658−1661.
(38) Lane, L. A.; Fernandez-Tornero, C.; Zhou, M.; Morgner, N.; Ptchelkine, D.; Steuerwald, U.; Politis, A.; Lindner, D.; Gvozdenovic, J.; Gavin, A. C.; Muller, C. W.; Robinson, C. V. *Structure* **2011**, *19*, 90−100.
(39) Baldwin, A. J.; Lioe, H.; Hilton, G. R.; Baker, L. A.; Rubinstein, J. L.; Kay, L. E.; Benesch, J. L. *Structure* **2011**, *19*, 1855−1863.
(40) Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V. *J. Am. Chem. Soc.* **2012**, *134*, 3429−3438.
(41) Hilton, G. R.; Thalassinos, K.; Grabenauer, M.; Sanghera, N.; Slade, S. E.; Wyttenbach, T.; Robinson, P. J.; Pinheiro, T. J.; Bowers, M. T.; Scrivens, J. H. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 845−854.
(42) Bagal, D.; Valliere-Douglass, J. F.; Balland, A.; Schnier, P. D. *Anal. Chem.* **2010**, *82*, 6751−6755.
(43) Rozbesky, D.; Kavan, D.; Chmelik, J.; Novak, P.; Vanek, O.; Bezouska, K. *Protein Expression Purif.* **2011**, *77*, 178−184.
(44) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779−815.
(45) Watson, A. A.; Lebedev, A. A.; Hall, B. A.; Fenton-May, A. E.; Vagin, A. A.; Dejnirattisai, W.; Felce, J.; Mongkolsapaya, J.; Palma, A. S.; Liu, Y.; Feizi, T.; Screaton, G. R.; Murshudov, G. N.; O'Callaghan, C. A. *J. Biol. Chem.* **2011**, *286*, 24208−24218.
(46) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714−2723.
(47) Krieger, E.; Koraimann, G.; Vriend, G. *Proteins* **2002**, *47*, 393−402.
(48) Hernandez, H.; Robinson, C. V. *Nat. Protoc.* **2007**, *2*, 715−726.
(49) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. *Anal. Chem.* **2010**, *82*, 9557−9565.
(50) Benesch, J. L. P.; Ruotolo, B. T. *Curr. Opin. Struct. Biol.* **2011**, *21*, 641−649.
(51) Novak, P.; Young, M. M.; Schoeniger, J. S.; Kruppa, G. H. *Eur. J. Mass Spectrom. (Chichester, Eng.)* **2003**, *9*, 623−631.
(52) Clemmer, D. E.; Jarrold, M. F. *J. Mass Spectrom.* **1997**, *32*, 577−592.
(53) Zelensky, A. N.; Gready, J. E. *FEBS J* **2005**, *272*, 6179−6217.
(54) Muller, D. R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O. *Anal. Chem.* **2001**, *73*, 1927−1934.
(55) Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R. *Mol. Cell. Proteomics* **2010**, *9*, 1634−1649.
(56) Back, J. W.; de Jong, L.; Muijsers, A. O.; de Koster, C. G. *J. Mol. Biol.* **2003**, *331*, 303−313.