

De Novo Mass Spectrometry Sequencing and Characterization of Species-Specific Peptides from Nucleoside Diphosphate Kinase B for the Classification of Commercial Fish Species Belonging to the Family Merlucciidae

Mónica Carrera,^{*,†} Benito Cañas,[‡] Carmen Piñeiro,[†] Jesús Vázquez,[§] and José Manuel Gallardo[†]

Departamento Tecnología de Alimentos, Instituto de Investigaciones Marinas (IIM-CSIC),
 Vigo, Pontevedra, Spain, Departamento Química Analítica, Facultad de Química,
 Universidad Complutense de Madrid, Spain, and Laboratorio Química de Proteínas y Proteómica,
 Centro de Biología Molecular Severo Ochoa, Madrid, Spain

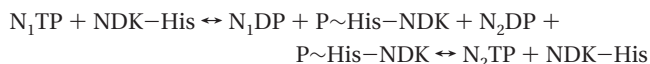
Received April 5, 2007

The characterization by *de novo* peptide sequencing of the different protein nucleoside diphosphate kinase B (NDK B) from all the commercial hakes and grenadiers belonging to the family Merlucciidae is reported. A classical proteomics approach, consisting of two-dimensional gel electrophoresis, tryptic *in-gel* digestion of the excised spots, MALDI-TOF MS, LC-MS/MS, and nanoESI-MS/MS analyses, was followed for the purification and characterization of the different isoforms of the NDK B. Fragmentation spectra were used for *de novo* peptide sequence. A high degree of homology was found between the sequences of all the species studied and the NDK B sequence from *Gillichthys mirabilis*, which is accessible in the protein databases. Particular attention was paid to the differential characterization of species-specific peptides that could be used for fish authentication purposes. These findings allowed us to propose a rapid and effective classification method, based in the detection of these biomarker peptides using the selective ion reaction monitoring (SIRM) scan mode in mass spectrometry.

Keywords: proteomics • nucleoside diphosphate kinase • fish • *de novo* sequencing • mass spectrometry • selective ion reaction monitoring • NDK • SIRM

Introduction

The family of nucleoside diphosphate kinases (NDK) (EC 2.7.4.6) comprises a specific group of ubiquitous housekeeping phosphotransferases. They are implicated in maintaining the pool of intracellular nucleoside triphosphates (NTPs) required for biosynthesis, catalyzing the non-substrate-specific conversion of nucleosides diphosphate (NDP) to nucleoside triphosphate (NTP).¹ Particularly, NDK catalyzes the reversible transfer of a terminal γ -phosphoryl group (P) from NTP to NDP involving a conserved histidine residue (His) as intermediate, by means of a ping-pong mechanism:



Such enzymes have been highly conserved throughout evolution (>40% identity between prokaryotes and eukaryotes),

having been isolated from a variety of organisms and cell types, including prokaryotes, lower and higher eukaryotes, mouse, rat, and human.^{2–4} They are ubiquitously distributed, particularly in the nucleus, in the cytoplasm associated to the microtubules, on the cell surface, in the mitochondria, and in the extracellular fluid.⁵

NDKs have been implicated in the biological regulation of growth, development, apoptosis, and differentiation. However, other relevant functions have also been reported, as the suppression of metastasis in tumors,⁶ the activity of histidine-dependent protein kinase,⁷ and the 3'–5' exonuclease activity.⁸ Moreover, there is also evidence about that NDK proteins promote tumor formation and play various roles in normal development and cellular proliferation.^{4,9}

In higher animals, NDKs are encoded by several genes, called *nm23* (*nonmetastatic* 23) genes, on the basis of their reported action as tissue specific metastasis inhibitors.⁶ Eight *nm23* genes have been identified in the human genome;⁶ two of them, known as *nm23-H1* and *nm23-H2*, are the most studied. These genes code for polypeptide chains, called NDK A and NDK B, which present a high degree of similarity (88%). They have about 150 residues and a molecular weight between 16 and 20 kDa. However, their isoelectric points are different, because of the substitution of several acidic amino acids in NDK A by basic

* To whom correspondence should be addressed. Mónica Carrera, Instituto de Investigaciones Marinas (IIM-CSIC), Eduardo Cabello 6, E-36208 Vigo, Pontevedra, Spain; Phone, +34 986 231930; Fax, +34 986 292762; E-mail, mcarrera@iim.csic.es.

† Instituto de Investigaciones Marinas.

‡ Universidad Complutense de Madrid.

§ Centro de Biología Molecular Severo Ochoa.

Table 1. List of Species and Subspecies Included in the Study

species or subspecies	common name	origin
<i>M. merluccius</i>	European hake	Spanish coasts
<i>M. capensis</i>	Cape hake	South Africa
<i>M. senegalensis</i>	Senegalense hake	Northwest Africa
<i>M. polli</i>	Benguela hake	Northwest Africa
<i>M. paradoxus</i>	Deep-water Cape hake	South Africa
<i>M. hubbsi</i>	Patagonian hake	South America
<i>M. gayi</i>	Peruvian or Chilean hake	South America
<i>M. australis polylepsis</i>	Austral hake	South America
<i>M. australis australis</i>	Austral hake	New Zealand coasts
<i>M. productus</i>	Pacific hake	North America
<i>M. bilinearis</i>	Silver hake	North America
<i>Ma. novaezelandiae novaezelandiae</i>	Blue grenadier	New Zealand coasts
<i>Ma. novaezelandiae magellanicus</i>	Patagonian grenadier	South America

M. (*Merluccius* genus); *Ma.* (*Macruronus* genus).

amino acids in NDK B. The first NDK amino acid sequences were obtained from genes cloned in *Dictyostelium discoideum*³ and *Myxococcus xanthus*². In general, the eukaryotic enzymes are homohexamers, whereas several of the prokaryotic enzymes are homotetramers.¹⁰ Whether a tetramer or an hexamer, NDKs active sites are structurally identical, and almost all the residues involved in them are fully invariant from bacteria to humans.¹¹

Up to now, only 33 NDK sequences for Teleostei are included in the protein databases (UniProt, April 2007), including only two entries that correspond to the isoform B: one belonging to zebra fish (*Danio rerio*) (accession number: Q9IAD3, UniProt)¹² and the other to a long-jawed mud sucker (*Gillichthys mirabilis*) (Q9DFL9, Uniprot).¹³ The rest of the entries are uncharacterized or uncommon isoforms. Although in fish species NDK activity still is poorly studied, it has been reported as a good marker for: (i) tissue biosynthesis in specimens maintained in metal contaminated environments;^{14–16} (ii) variation of the biochemical composition of tissues with growth in juvenile Atlantic cod;¹⁷ and (iii) digestive capacity, growth, and swimming endurance in Atlantic cod.^{18,19}

In a previous work, due to the differences found in MALDI-TOF MS spectra from tryptic peptides, sarcoplasmic NDK proteins proved to be, in five closely related hake species, a potential molecular marker for fish identification.²⁰ Now, using a combination of high-resolution 2-DE analysis, MALDI-TOF MS, and *de novo* sequencing by ESI-IT MS, we have characterized the NDK B proteins from eleven different closely related species, commercially available, belonging to the Merlucciidae family. Efforts were made in peptide *de novo* sequencing, as genomes from members belonging to Merlucciidae family still await sequencing. In fact, only two Teleostei species genomes have been completely sequenced: *Tetraodon nigroviridis*²¹ and *Takifugu rubripes*²² and works for the complete sequencing of the zebrafish (*Danio rerio*) genome were started in 2001.

In addition, particular attention was focused on the characterization and identification of several species-specific peptides for each of the species included in the Merlucciidae family. These specific peptide sequences can be used, by fisheries as a good tool for seafood product authentication, providing a rapid and effective identification and guaranteeing the quality of the foodstuffs to consumers.²³

Materials and Methods

Fish Material. Ten different hake species, including two different subspecies from *M. australis* and two grenadier

subspecies belonging to the *Macruronus novaezelandiae* species, were employed in this study (Table 1). The specimens were frozen on board at -30°C , except for European hake, with special care in keeping their morphological characters in good shape, and shipped by plane to the laboratory for the analyses. The weight of every specimen studied was in the range of 3–6 kg. At least 10 fish specimens belonging to each different species were subjected to taxonomical study according to their anatomical and morphological features by an expert marine biologist and by genetic identification in the Food Biochemistry laboratory from Marine Research Institute (Vigo, Pontevedra, Spain) and with the fishID Kit based on the amplification of specific mitochondrial DNA sequences (Bionostra SL., Madrid, Spain). Among them, five correctly identified individuals from each of species were considered and selected as the representative species for all the study.

Extraction of Sarcoplasmic Proteins. Sarcoplasmic proteins extraction was carried out by homogenizing 5 g of white muscle from each of the species in 10 mL of 10 mM Tris-HCl buffer, pH 7.2, supplemented with 5 mM PMSF, during 30 s in an Ultra-Turrax device (IKA-Werke, Staufen, Germany). The fish muscle extracts were then centrifuged at 40 000 g for 20 min at 4°C (J221-M centrifuge; Beckman, Palo Alto, CA). The supernatants were then recovered, filtered using Ultrafree CL (0.22 μm) filters (Millipore, Bedford, MA), and stored at -80°C until gel electrophoresis was performed. Protein concentration in the extracts was determined by the bicinchoninic acid method (Sigma Chemical Co., U.S.A.).

2-DE Analysis. Isoelectric focusing (IEF) was performed at 10°C in a Multiphor II electrophoresis unit (Amersham Biosciences, Sweden) according to previous works.²⁴ Protein extracts (40 mg/mL) were loaded in duplicate on IEF strips with a narrow range of pH 4–6.5 (Amersham Biosciences), using a sample application paper. A mixture of proteins standard in the 2.5–6.5 pH range (low pI standard from Amersham) was included. Isoelectro-focusing conditions were as follows: 1500 V, 50 mA, 30 W until at least 4000 Vh were reached. IEF strips corresponding to individual lanes were cut immediately after the run was completed and kept frozen at -80°C for the second dimension by SDS-PAGE. A duplicate IEF strip was stained with 0.1% Coomassie brilliant blue (Sigma Chemical) according to the Amersham Biosciences staining protocol.

Equilibration of pH 4–6.5 IEF gel strips was carried out at room temperature as described previously,²⁵ and the second dimension was run using gradient (12–14%) SDS-PAGE precast polyacrylamide $245 \times 180 \times 0.5$ mm gradient gels (ExcelGel XL SDS 12–14%, Amersham Biosciences) at 15°C using the

Table 2. Isoelectric Point and Molecular Weight for all NDK Spots from All Hake and Grenadiers Studied

species or subspecies ^a	spot number	pI	M _r (kDa)
<i>M. merluccius</i>	N1	5.42	18.60
<i>M. capensis</i>	N2	5.43	18.59
<i>M. senegalensis</i>	N3	5.43	18.59
<i>M. polli</i>	N4	5.47	18.60
<i>M. paradoxus</i>	N5	5.43	18.59
<i>M. hubbsi</i>	N6	5.29	17.81
<i>M. gayi</i>	N7	5.04	17.80
<i>M. australis polylepsis</i>	N8	5.07	17.81
<i>M. australis australis</i>	N9	5.07	17.80
<i>M. productus</i>	N10	5.07	17.87
<i>M. bilinearis</i>	N11	5.07	17.82
<i>Ma. novaezelandiae novaezelandiae</i>	N12	5.43	16.81
<i>Ma. novaezelandiae magellanicus</i>	N13	5.45	16.81

^a *M.*, *Merluccius* genus; *Ma.*, *Macruronus* genus.

Multiphor II system (Amersham Biosciences) in a Tris-glycine buffer.²⁰ The running conditions were: 1000 V, 40 mA, 40 W during 165 min. A low molecular weight protein standard (14–97 kDa) (LMW Calibration Kit, Amersham Biosciences) was employed as reference. Gels were silver stained using a protocol from Amersham Biosciences and the image analysis was carried out by the PDQuest software (Bio-Rad Laboratories, Ltd, UK).

Preparative 2-DE Gels for Mass Spectrometry Analysis.

Once equilibrated, the IEF strips (pH 4–6.5) were subjected to an electrophoretic separation compatible with a further mass spectrometry analysis, using a preparative vertical SDS-PAGE (10% T and 3% C) gels with Tris-Tricine buffer. The running conditions were: 100 V, 40 mA per gel, and 150 W during 16–18 h. Protein spots were visualized by the Coomassie brilliant blue staining (Amersham Bioscience) according to manufacturer instructions and the PDQuest software (Bio-Rad) was used for the image analysis.

Peptide Sample Preparation for MS and MS/MS Analysis.

NDK spots were localized based on the data (pI and M_r) obtained by previous work,²⁰ and excised from the gel taking care in maximizing the protein-to-gel ratio. Only the most intensely stained region at center of the spot was excised to avoid extracting an excess of gel matrix. Excised pieces were subjected to *in-gel* digestion with trypsin (Promega, Madison, WI) as described.²⁶ Prior to nESI-IT MS analysis, the tryptic peptide samples were individually desalted and concentrated using in-tip reverse-phase resins (ZipTip C18, Millipore, Bedford, MA), according to the manufacturer's recommendations.

Peptide Mapping by MALDI-TOF MS. MALDI-TOF MS was performed using an Autoflex instrument from Bruker Daltonics (Bremen, Germany) equipped with delayed extraction and operating in the reflector mode according to previous work.²⁷ The peptides obtained by the digestion were acidified by the addition of trifluoroacetic acid (TFA) and dried down. Anchor-Chip plates (Bruker Daltonics) were used as MALDI targets. Matrix, 0.5 μ L of saturated DHB (Bruker Daltonics) in 30% acetonitrile containing 0.1% TFA, was first applied to the target and allowed to dry. Aliquots (0.5 μ L) of the extracted peptides were applied to the MALDI target and allowed to dry. External calibration was carried out using a set of synthetic peptides (Bruker Daltonics) added to the target in positions close to the samples. Spectra were taken between 400 and 4000 Da and subsequently, all those masses from trypsin autolysis, matrix clusters (from *m/z* 150 to 1000) and other well-known contaminants peaks, as those of common keratins, were automatically rejected.

Peptide Fragmentation and *de novo* Sequencing by ESI-IT MS. Peptide digests were analyzed online by LC-ESI-IT-MS/MS using a LC system model SpectraSystem P4000 (Thermo-Finnigan, San Jose, CA) coupled to an IT mass spectrometer model LCQ Deca XP Plus (Thermo-Finnigan). The separation was performed on a 0.18 mm $\times$ 150 mm BioBasic-18 RP column (ThermoHypersil-Keystone), using 0.5% acetic acid in water and in 80% acetonitrile as mobile phases A and B, respectively. A 90 min linear gradient from 5 to 60% B, at a flow rate of 1.5–1.7 μ L/min, was used. Peptides were detected using survey scans from 400 to 1600 amu (3 μ scans), followed by a data-dependent ZoomScan (5 μ scans) and MS/MS scan (5 μ scans), using an isolation width of 3 amu and a normalized collision energy of 35%. Fragmented masses were set in dynamic exclusion for 3 min after the second fragmentation event and singly charged ions were excluded from MS/MS analysis.

Analysis by nESI-IT MS were performed off-line using an ion trap mass spectrometer, model LCQ Deca XP Plus from Thermo-Finnigan, equipped with a nanospray interface. Peptides were previously desalted by the use of micro-columns ZipTip C18 and eluted with 5–10 μ L of 70% MeOH/ 0.5% CH₃COOH. PicoTips borosilicate glass needles with 1 μ m orifice (New Objective, Woburn, MA), used as emitters, were filled with 3–5 μ L of sample.

Selected Ion Reaction Monitoring (SIRM). Proteins in 20 μ g of sarcoplasmic extract were denaturated with 8 M urea and reduced with 10 mM DTT in 25 mM ammonium bicarbonate pH 8.25 at 37 °C for 1 h. Iodoacetamide was then added to a final concentration of 50 mM. The resulting mixture was incubated at room temperature in the darkness for 45 min. The mixture was then diluted 4-fold to reduce urea concentration and subjected to *in-solution* digestion with trypsin (Promega, Madison, WI) at 1:50 protease-to-protein ratio during at least 12 h at 37 °C. Digests were individually desalted and concentrated using in-tip reverse-phase resins (ZipTip C18, Millipore). Peptide mixtures (0.05 μ g) were analyzed by capillary RP-HPLC coupled to ESI-IT-MS using the column separation conditions as previously described. Selected ion reaction monitoring (SIRM) was used all along the chromatographic separation as the mass spectrometric scan mode, focused on specific doubly charged precursor ions from selected NDK peptides. Ion chromatograms were represented using the instrument software to show, for each precursor, a selected product ion, which was previously characterized by *de novo* sequencing.

Mass Spectrometry Data Processing. Peak identification and monoisotopic peptide mass assignment of MALDI-TOF MS spectra were automatically performed using FlexAnalysis software (Bruker Daltonics) and the MoverZ program available by Internet (<http://bioinformatics.genomicsolutions.com/>). Database searches were performed, against a NCBI non-redundant protein sequence database (<http://www.ncbi.nih.gov>) using MASCOT (<http://www.matrixscience.co.uk/>), with the following settings: monoisotopic masses, two missed cleavages allowed, 100 ppm mass tolerance, and three potential variable modifications: ACET sites (acetyl N-terminal), CysCAM (carbamidomethylcysteine), and MSO (methionine sulfoxide).

MS/MS spectra were searched using SEQUEST (Bioworks 3.1 package, Thermo-Finnigan) against the nr.fasta database (NCBI Resources, NIH, Bethesda MD). The following constraints were used for the searches: tryptic cleavage after Arg and Lys, up to two missed cleavage sites, and tolerances $\pm$ 1.8 Da for precursor ions and $\pm$ 0.8 Da for MS/MS fragments ions. The variable

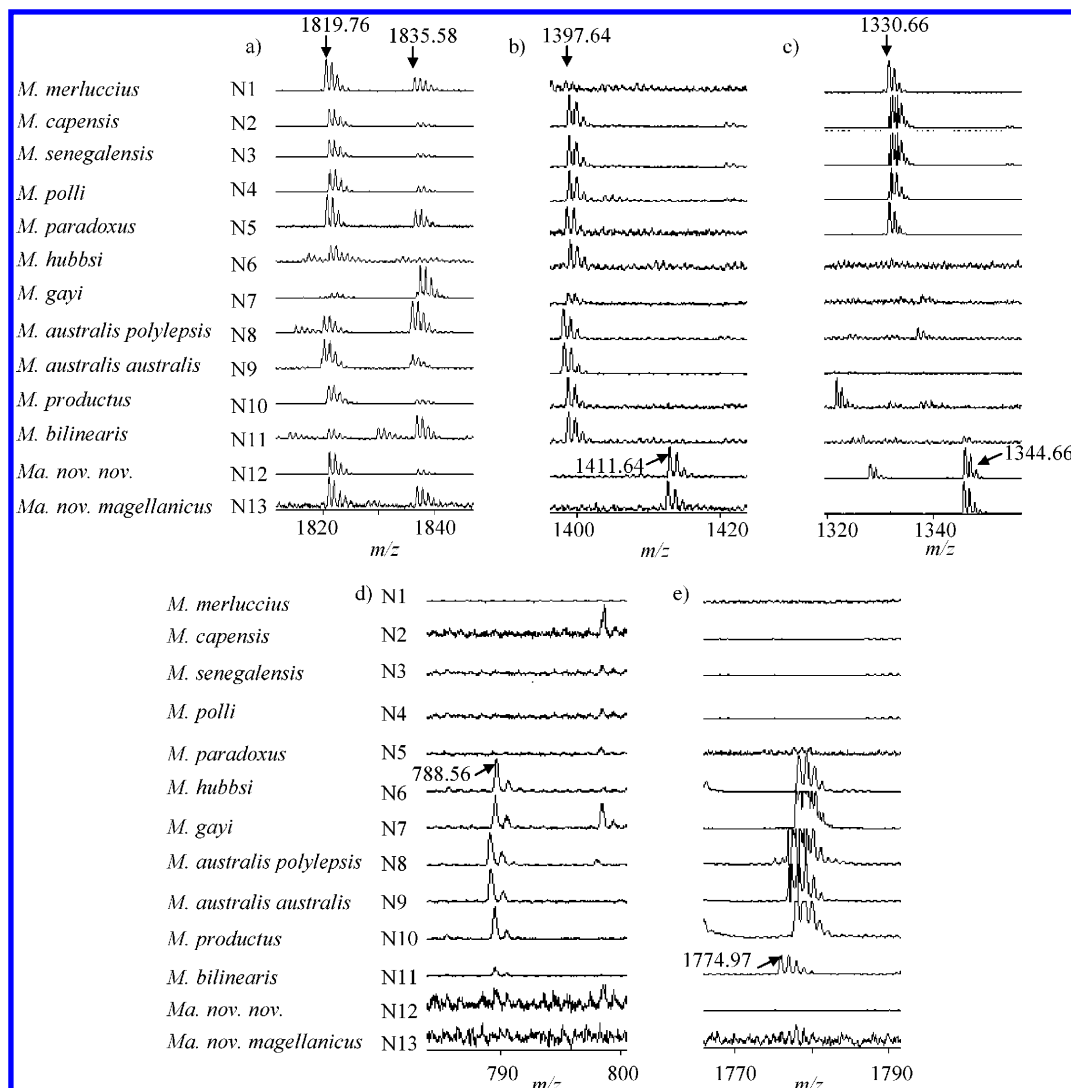


Figure 1. Information rich regions of the MALDI-TOF MS peptide maps from NDK trypsin *in-gel* digested spots. (a) Peaks which are common to all the species (1819.76 and 1835.58 Da), (b, c) those which are specific for either the genus *Merluccius* (1397.64 Da) or the genus *Macruronus* (1411.64 and 1344.66 Da) and for all NDK from Euro-African hakes (1330.66 Da), (d) those which are specific for American hakes (788.56 Da), and (e) those which are specific for *M. bilinearis* species (1774.97 Da).

modifications allowed were methionine oxidation, carbamidomethylation of Cys and acetylation of the N-terminus.

De novo sequencing was performed by manual interpretation of the ion series of the spectra with the aid of the software package DeNovoX (Thermo-Finnigan). Homology between *de novo* obtained sequences and those in the nr-NCBI database was analyzed using BLAST.

Results and Discussion

2-DE Analyses of NDKs. Sarcoplasmic protein extracts of 5 individuals from each of the 11 species (including two subspecies) studied were subjected to 2-DE and the gels silver stained. To ensure reproducibility of the experiments, 3 gels were run per individual (65 specimens). Thus a total of 195 2-DE gels were processed and analyzed by the image software. In a preliminary survey, the study was carried out using strips with a broad pH range (3–10) (data not shown); however, due to the overlapping of some of the proteins under study, strips with a narrower pH range (4–6.5) were chosen. The NDK fraction presented a molecular weight between 16.80 and 18.60 kDa and

a *pI* between 5.04 and 5.47 units, according to the spot identification described previously by ref 20.

The 2-DE NDK patterns for the different hake species and for the two grenadier subspecies are shown in Figure S-1 from the Supporting Information. A compilation of the *pI* and *M_r* of all NDK spots studied, designated as N1–N13, are shown in the Table 2. Qualitative inter-species differences were noticeable. Briefly, the results (*pI* and *M_r*) produced by the 2-DE analysis allowed the classification of the species studied in three groups: (i) Euro-African hakes: *M. merluccius*, *M. capensis*, *M. senegalensis*, *M. polli*, and *M. paradoxus*, showing a spot with a *pI* of 5.42–5.47 and a *M_r* of 18.60 kDa, (ii) American hakes: *M. hubbsi*, *M. gayi*, *M. australis polylepsis*, *M. australis australis*, *M. productus*, and *M. bilinearis*, characterized by a spot with a *pI* of 5.04–5.29 and a *M_r* of 17.80–17.87 kDa, and finally (iii) the grenadier subspecies, which presented a distinctive spot with a *pI* 5.43–5.45 and a *M_r* 16.81 kDa.

The results obtained by 2-DE demonstrated differences in *pI* and *M_r* from this NDK spot, indicating the presence of amino acid substitutions in their sequences. Nevertheless, 2-DE alone

Table 3. Summary of De Novo Peptide Sequence Data Obtained for NDK B from all Hake and Grenadiers Species Studied

<i>m/z</i> (z)	(M+H) ⁺ matched	Peptide Sequence	Modification	NCS	Position	MALDI observed	Spot Number												
							<i>Merluccius</i>						<i>Macrur</i>						
							1	2	3	4	5	6	7	8	9	10	11	12	13
459.03 (1+)	459.26	GVRQ		0	12-15														
361.87 (3+)	1082.63	VAIKPDGVQR		1	6-15	YES (PMF) ^a													
541.74 (2+)	1082.63	VAIKPDGVQR		1	6-15	YES (PMF) ^a													
444.10 (3+)	1330.74	TFVAIKPDGVQR		1	4-15	YES (PMF) ^a													
665.80 (2+)	1330.74	TFVAIKPDGVQR		1	4-15	YES (PMF) ^a													
672.80 (2+)	1344.76	TFIAIKPDGVQR		1	4-15	YES (PMF) ^c													
859.23 (2+)	1718.89	MEQTFVAIKPDGVQR		1	1-15														
880.01 (2+)	1760.93	MEQTFVAIKPDGVQR	N-Acyl	1	1-15	YES													
888.89 (2+)	1776.89	MEQTFVAIKPDGVQR	N-Acyl ₁ ;MSO	1	1-15	YES													
896.90 (2+)	1792.89	MEQTFVAIKPDGVQR	N-Acyl ₁ ;MSO ₂ ;MSO	1	1-15	YES													
887.89 (2+)	1774.95	MEQTFIAIKPDGVQR	N-Acyl	1	1-15	YES													
895.89 (2+)	1790.94	MEQTFIAIKPDGVQR	N-Acyl ₁ ;MSO	1	1-15	YES													
447.01 (2+)	893.42	GLCGEVMK		0	16-23	YES													
893.40 (1+)	893.42	GLCGEVMK		0	16-23	YES													
455.01 (2+)	909.41	GLCGEVMK	MSO	0	16-23	YES													
411.10 (3+)	1231.38	FIQP(468.4)MK		0	37-46	YES	•	•	•	•	•	•	•	•	•	•	•	•	•
616.49 (2+)	1231.38	FIQP(468.4)MK		0	37-46	YES	•	•	•	•	•	•	•	•	•	•	•	•	•
624.10 (2+)	1247.20	FIQP(468.4)MK		0	37-46	YES	•	•	•	•	•	•	•	•	•	•	•	•	•
394.93 (2+)	788.43	HYLDLK	MSO	0	37-46	YES	•	•	•	•	•	•	•	•	•	•	•	•	•
306.05 (3+)	916.48	QHYLDLK		0	48-53	YES (PMF)													
458.75 (2+)	916.48	QHYLDLK		0	47-53	YES (PMF)													
478.38 (2+)	955.47	PFYAGLCK		0	56-63	YES (PMF) ^b													
601.10 (2+)	1201.53	DMPFYAGLCK		0	54-63	YES	•	•	•	•	•	•	•	•	•	•	•	•	•
1201.50 (1+)	1201.53	DMPFYAGLCK	MSO	0	54-63	YES	•	•	•	•	•	•	•	•	•	•	•	•	•
609.10 (2+)	1217.53	DMPFYAGLCK	MSO;Na adduct	0	54-63	YES	•	•	•	•	•	•	•	•	•	•	•	•	•
612.10 (2+)	1217.53	DMPFYAGLCK	MSO;HYDR ₆₃	0	54-63	YES													
617.02 (2+)	1233.53	DMPFYAGLCK		0	54-63	YES													
545.44 (2+)	1089.47	YMSSGPVFAM		0	64-73	YES													
1089.30 (1+)	1089.47	YMSSGPVFAM		0	64-73	YES													
942.38 (1+)	942.43	YMASGPVFA		0	64-72	YES													
958.39 (1+)	958.43	YMASGPVFA	MSO	0	64-72	YES													
1073.39 (1+)	1073.48	YMASGPVFAM		0	64-73	YES													
508.68 (2+)	1016.54	VWEGEGIVK		0	74-82	YES	•	•	•	•	•	•	•	•	•	•	•	•	•
574.12 (2+)	1147.58	MVWEGEGIVK		0	73-82	YES													
582.14 (2+)	1163.57	MVWEGEGIVK	MSO	0	73-82	YES													
609.80 (2+)	1218.61	AMVWEGEGIVK		0	72-82														
683.12 (2+)	1365.68	FAMVWEGEGIVK		0	71-82	YES													
853.43 (2+)	1705.86	SGPVFAMVWEGEGIVK		0	67-82														
696.13 (3+)	2086.99	YMSSGPVFAMVWEGEGIVK		0	64-82	YES													
1043.76 (2+)	2086.99	YMSSGPVFAMVWEGEGIVK	MSO ₆₅	0	64-82	YES													
701.43 (3+)	2102.99	YMSSGPVFAMVWEGEGIVK	MSO ₆₅	0	64-82	YES													
1051.86 (2+)	2102.99	YMSSGPVFAMVWEGEGIVK	Na adduct	0	64-82	YES													
1054.48 (2+)	2108.99	YMSSGPVFAMVWEGEGIVK	MSO ₇₃	0	64-82	YES													
701.72 (3+)	2102.99	YMSSGPVFAMVWEGEGIVK	MSO ₇₃	0	64-82	YES													
1059.79 (2+)	2118.98	YMSSGPVFAMVWEGEGIVK	MSO ₆₅ ;MSO ₇₃	0	64-82	YES													
701.07 (3+)	2118.98	YMSSGPVFAMVWEGEGIVK	MSO ₆₅ ;MSO ₇₃	0	64-82	YES													

Table 3 (Continued)

<i>m/z</i> (z)		(M+H) ⁺ matched	Peptide Sequence	Modification	NCS	Position	MALDI observed	Spot Number																
								<i>Merluccius</i>																
								<i>Macrur</i>																
1	2	3	4	5	6	7	8	9	10	11	12	13												
1067.69 (2+)	2134.98	YMSSGPVFAMVWEGEGIVK	MSO ₆₅ ;MSO ₇₃ ;HYDR ₇₅	0	64-82	YES																		
1067.69 (2+)	2134.98	YMSSGPVFAMVWEGEGIVK	MSO ₆₅ ;MSO ₇₃ ;MSO ₇₃	0	64-82	YES																		
1075.74 (2+)	2150.98	YMSSGPVFAMVWEGEGIVK	MSO ₆₅ ;MSO ₇₃ ;HYDR ₇₅ ;HYDR ₇₅	0	64-82	YES																		
1035.98 (2+)	2071.00	YMASGPVFAMVWEGEGIVK		0	64-82	YES																		
1036.01 (3+)	2071.00	YMASGPVFAMVWEGEGIVK		0	64-82	YES																		
1043.93 (2+)	2086.99	YMASGPVFAMVWEGEGIVK	MSO ₆₅	0	64-82	YES																		
1051.93 (2+)	2102.99	YMASGPVFAMVWEGEGIVK	MSO ₆₅ ;MSO ₇₃	0	64-82	YES																		
701.69 (3+)	2102.99	YMASGPVFAMVWEGEGIVK	MSO ₆₅ ;MSO ₇₃	0	64-82	YES																		
1059.95 (2+)	2118.99	YMASGPVFAMVWEGEGIVK	MSO ₆₅ ;MSO ₇₃ ;HYDR ₇₅	0	64-82	YES																		
1067.95 (2+)	2134.99	YMASGPVFAMVWEGEGIVK	MSO ₆₅ ;MSO ₇₃ ;HYDR ₇₅ ;HYDR ₇₅	0	64-82	YES																		
647.10 (2+)	1293.58	MMLGETNPADSK		0	86-97	YES																		
1293.50 (1+)	1293.58	MMLGETNPADSK		0	86-97	YES																		
655.21 (2+)	1309.57	MMLGETNPADSK	MSO ₈₆ or MSO ₈₇	0	86-97	YES (PMF)																		
663.20 (2+)	1325.57	MMLGETNPADSK	MSO ₈₆ ;MSO ₈₇	0	86-97	YES (PMF)																		
902.68 (2+)	1803.87	MMLGETNPADSKPGSIR		1	86-102	YES (PMF)																		
601.90 (3+)	1803.87	MMLGETNPADSKPGSIR		1	86-102	YES (PMF)																		
612.50 (3+)	1835.86	MMLGETNPADSKPGSIR	MSO ₈₆ ;MSO ₈₇	1	86-102	YES (PMF)																		
918.40 (2+)	1835.86	MMLGETNPADSKPGSIR	MSO ₈₆ ;MSO ₈₇	1	86-102	YES (PMF)																		
910.30 (2+)	1819.86	MMLGETNPADSKPGSIR	MSO ₈₆ or MSO ₈₇	1	86-102	YES (PMF)																		
607.98 (3+)	1819.86	MMLGETNPADSKPGSIR	MSO ₈₆ or MSO ₈₇	1	86-102	YES (PMF)																		
913.40 (2+)	1825.87	MMLGETNPADSKPGSIR	Na adduct	1	86-102	YES (PMF)																		
921.30 (2+)	1841.87	MMLGETNPADSKPGSIR	K adduct	1	86-102	YES																		
926.65 (2+)	1851.86	MMLGETNPADSKPGSIR	MSO ₈₆ ;MSO ₈₆ ;MSO ₈₇	1	86-102	YES																		
725.21 (1+)	725.29	GDFCIN		0	103-108																			
497.76 (2+)	994.45	GDFCINIGR		0	103-108																			
526.10 (2+)	1051.49	GDFCINIGR		0	103-111	YES																		
533.16 (2+)	1065.51	GDFCINIGR		0	103-111	YES																		
599.87 (2+)	1198.57	NIHGSDTIVEN		0	112-122																			
699.30 (2+)	1397.70	NIHGSDTIVENAK		0	112-124	YES																		
466.57 (3+)	1397.70	NIHGSDTIVENAK		0	112-125	YES																		
706.30 (2+)	1411.71	NIHGSDTIVENAK		0	112-124	YES (PMF)																		
717.30 (2+)	1433.71	NIHGSDTIVENAK		0	112-124	YES (PMF)																		
748.32 (1+)	748.37	MEIGLW	Na adduct	0	112-124																			
764.31 (1+)	764.36	MEIGLW	MSO	0	125-130																			
512.03 (2+)	1023.56	MEIGLWFK		0	125-130																			
520.20 (2+)	1039.52	MEIGLWFK		0	125-132	YES																		
756.28 (2+)	1511.72	MEVGLWFKPEEF	MSO	1	125-136	YES																		
764.34 (2+)	1527.72	MEVGLWFKPEEF	MSO	1	125-136	YES																		
789.06 (3+)	2365.17	FKPEEFVITYTEKAKAWVYE		3	131-149																			
769.12 (3+)	2305.15	FKPEEFVAYAEKAKAWVYE		3	131-149																			
483.75 (2+)	966.49	FKPEEFVA		1	131-138																			
576.79 (2+)	1152.57	WFKPEEFVA		1	130-138																			
651.10 (1+)	651.20	GEEFVA		0	133-138																			
814.10 (1+)	814.36	GEEFVAY		0	133-137																			
995.32 (2+)	1989.95	GEEFVAYAEKAKAWVYE		2	133-149	YES																		
764.38 (2+)	1527.78	VAYAEKAKAWVYE		2	137-149	YES																		

Table 3 (Continued)

<i>m/z</i> (z)	(M+H) ⁺ Peptide Sequence matched	Modification	NCS	Position	MALDI observed	Spot Number												
						<i>Merluccius</i>												
						N	N	N	N	N	N	N	N	N	N	N	N	N
775.2 (2+)	1549.78 VAYAEKAKAWVE	Na adduct	2	137-149														
794.69 (2+)	1587.80 VTYEKAKAWVE		2	137-149														
763.37 (2+)	1525.73 MEVALWPKPEEF		2	125-136														
841.53 (2+)	1681.83 MEVGLWPKPEEFVA		1	125-138														
944.96 (2+)	1888.91 MEVALWPKPEEFVY	MSO	1	125-139														
952.96 (2+)	1904.91 MEVALWPKPEEFVY		1	125-139														
1086.89 (2+)	2173.06 MEVGLWPKPEEFVAYAEK		1	125-142	YES													
1123.96 (2+)	2247.10 MEVALWPKPEEFVITYTEK		1	125-142														
1131.97 (2+)	2263.09 MEVALWPKPEEFVITYTEK	MSO	1	125-142														
1193.72 (2+)	2386.21 MEVALWPKPEEFVAYAEKAK	MSO	2	125-144														
1201.60 (2+)	2402.21 MEVALWPKPEEFVAYAEKAK		2	125-144														
1007.61 (3+)	3020.49 MEVGLWPKPEEFVAYAEKAKAWVE		3	125-149														
1518.68 (3+)	3036.48 MEVGLWPKPEEFVAYAEKAKAWVE		3	125-149														
667.10 (1+)	667.31 AWVYE	MSO	0	145-149	YES													
866.38 (1+)	866.44 AKAWVE		1	143-149	YES													
1123.49 (1+)	1123.57 EKAKAWVE		2	141-149	YES													

m/z: mass/charge, NCS: non-cleavage sites, MALDI observed: peptide found also by peptide mass fingerprinting (PMF) with *Gillichthys mirabilis* NDKB. ^a PMF sequence coincident with *Prochlorococcus marinus* NDK. ^b PMF sequence coincident with *Scyliorhinus torazame* NDK. ^c PMF sequence coincident with *Xenopus laevis* NDK among others. Peptide sequence: in bold are indicated the punctual amino acids differences between the published NDK sequences and those obtained by *de novo* sequencing in the present work. (Black square with white circle) Common peptides for all species, (■) denotes the presence of a peptide, and (□) the absence. Cysteines are carboxymethylated; CysAcryl, acryloylcysteine; MSO, methionine sulfoxide; N-Acyl, acetyl N-terminal; HYDR, hydroxylation. *Macrur*, genus *Macruronus*; Spot numbers corresponding to: (N1) *M. merluccius*, (N2) *M. capensis*, (N3) *M. bilinearis*, (N4) *M. polli*, (N5) *M. paradoxus*, (N6) *M. hubbsi*, (N7) *M. gayi*, (N8) *M. australis polylophus*, (N9) *M. australis australis*, (N10) *M. productus*, (N11) *M. bilinearis*, (N12) *Ma. nov. nov.*, and (N13) *Ma. nov. magellanicus*.

did not reveal whether NDKs with the same *pI* and *M_r* were identical or not. For that reason, it was necessary to characterize each of the NDK spots by mass spectrometry.

MALDI-TOF MS Characterization of the NDK Spots. Proteins were separated in preparative 2-DE gels, then the excised NDK spots were digested *in-situ* with trypsin and the peptides produced analyzed by MALDI-TOF MS. To increase the reproducibility, peptide fingerprints were taken from each of the samples. Thus, 15 spectra per each of the species were produced, making a total of 195 peptide maps. No intraspecies polymorphic variations at qualitative level were found (data not shown).

The complete list of peaks for each of the NDK digests can be found in Table S-1 in the Supporting Information. Several peptide masses were coincident with those from tryptic peptides of NDKs present in data bases. Seven peaks presented the same masses as peptides from *Gillichthys mirabilis* NDK B (Q9DFL9, Uniprot);¹³ two others presented the same masses as two peptides from an NDK found in the Cyanobacteria *Prochlorococcus marinus* strain MIT 9313 (Q7V425, Uniprot);²⁸ one more was present in the NDK from the cloudy catshark *Scyliorhinus torazame* (Q9YI35, Uniprot);²⁹ and a last one was contained in the NDK sequence from the *Xenopus laevis* (P70010, Uniprot).³⁰ Besides, a total of eight peaks were shared by all the NDKs under study. Only the masses of two of them, 1819.76 and 1835.58 Da (Figure 1a), were coincident with masses of tryptic peptides from the NDKs included in protein databases.

The spectra showed several peaks that were selective for each genus; of them, only one was present in all the species of the genus *Merluccius* 1397.64 Da (Figure 1b) and four were present in both *Macruronus* subspecies: 1023.66, 1411.64 (Figure 1b), 1344.76 (Figure 1c) and 1990.04 Da. Some differences were also obtained among the *Merluccius* species, which may allow the tentative division of hakes into two groups: Euro-Africans, presenting a specific peak at 1330.66 Da (Figure 1c), and Americans, with two other specific peaks at 788.56 (Figure 1d) and 1760.60 Da. Finally, two peaks with an *m/z* of 1774.97 (Figure 1e) and 1807.03 Da were specific for *M. bilinearis*. The information provided by these peaks is enough to clearly differentiate between (i) hakes and grenadiers, (ii) Euro-African hakes of American hakes, and (iii) *M. bilinearis* from the rest of the Merlucciidae family species.

De novo Sequencing of Peptides from NDK. Tryptic digests from all the NDK spots were analyzed using a combination of LC-ESI-IT-MS/MS and nanoESI-IT-MS/MS. The spectra produced were used for manual and computer aided *de novo* amino acid sequence interpretation. Several novel peptide sequences, containing 4–25 amino acids, were assembled. Table 3 shows a compilation of the 108 amino acid sequences obtained, including 17 which were coincident with those previously assigned by PMF, and 76 masses previously found by MALDI-TOF MS analysis. In Figures 2, 3, and 4, several fragmentation spectra as examples of the specific peptides are shown, together with the *m/z* of the precursor ions. The peak with *m/z* of 526.1 (charge 2+) was chosen because it corresponds to one of the 16 sequences common to all the species of hakes and grenadiers (Figure 2). The peak with *m/z* 699.3 (2+) corresponds to one of the peptides which is present in all the *Merluccius* species, being absent in both *Macruronus* subspecies (Figure 3a), whereas 706.3 (2+) was chosen because it is found exclusively found in the *Macruronus* genus (Figure 3b). A substitution of a Val for a Leu in the sequence NIIHGS-

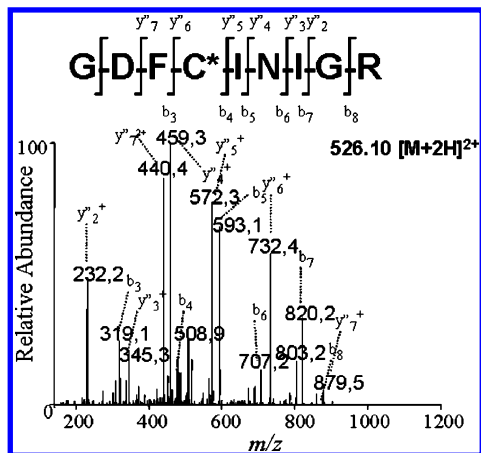


Figure 2. Ion-trap MS/MS spectrum of the peptide GD[F]C*[I]N[I]G[R] (C* carboxymethylated cysteine) which is common to all the isoforms the NDK B sequenced.

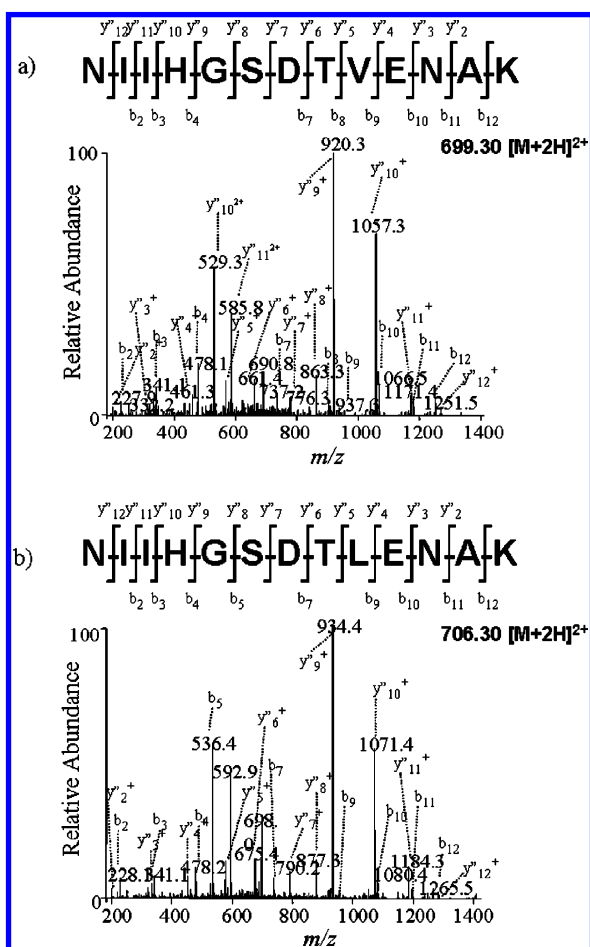


Figure 3. Ion-trap MS/MS spectra of the differential peptides NIIHGSDTVENAK, NIIHGSDTLENAK, which are present in the NDK B of *Merluccius* genus and *Macruronus* genus, respectively.

DTLENAK is responsible for the mass difference and hence for the different diagnostic fragmentation peaks at m/z 920.3, 1057.3 (699.3 precursor) and 934.4, 1071.4 (706.3 precursor). The N-terminus of the NDK proportionate, as well diagnostic peaks, as four different sequences are possible: position 3 is a Q for the American hakes and a K for Euro-African hakes and both *Macruronus* subspecies, whereas a V is present in position

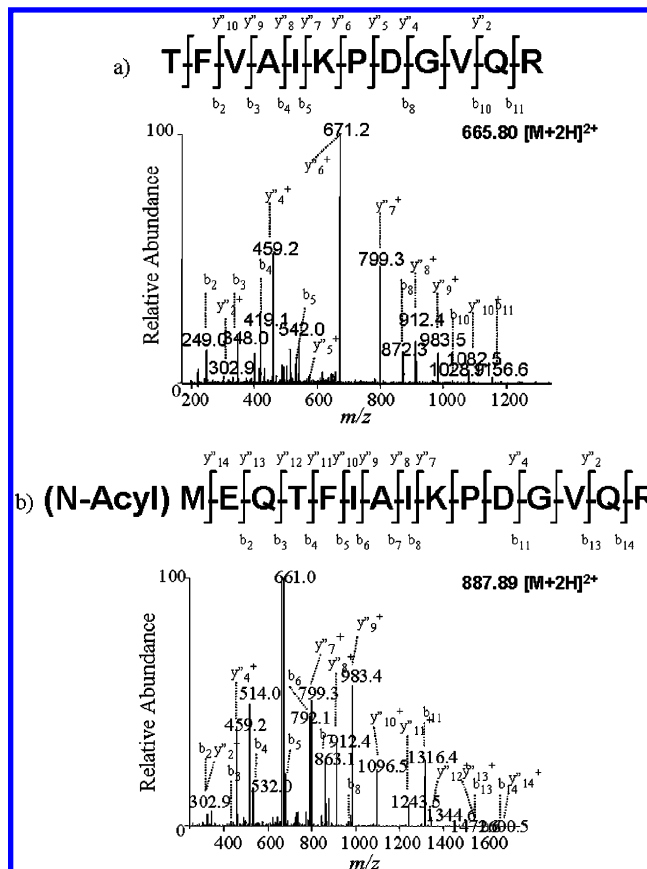


Figure 4. Ion-trap MS/MS spectra of the differential peptides TFVAIKPDGVQR, (N-Acyl) MEQTFAIKPDGVQR, which are present in the NDK B of Euro-African hakes and *M. bilinearis* species, respectively. N-Acyl, acetyl N-terminal.

6 for the American species and for *Macruronus* whereas an I is found in the Euro-African hake species and in *M. bilinearis*. So, the peak at m/z 665.8 (2+) (Figure 4a) corresponds to the sequence TFVAIKPDGVQR which is present exclusively in the Euro-African species, while the peak at 887.89 (2+) (Figure 4b), belonging to the sequence (N-Acyl)-MEQTFAIKPDGVQR, is specific to *M. bilinearis* and differentiates it from all the other hake and grenadiers species.

Alignment of the NDK Sequences. In Figure 5, the NDK amino acid sequences from the different hakes and grenadiers studied were aligned with the nucleic diphosphate kinase B from the long-jawed mudscucker *Gillichthys mirabilis* sequence (Q9DFL9; Uniprot),¹³ which presents the highest homology with those found in the present work. As NDK sequences had been highly conserved throughout the evolution only small differences were observed. The degree of homology between the studied NDKs and that from *Gillichthys mirabilis* (Q9DFL9) was somewhat variable: 72.48% (108/149 residues) for Euro-African hakes, 71.81% (107/149 residues) for *M. bilinearis*, 71.14% (106/149 residues) for the American hakes, and 69.79% (104/149 residues) for both subspecies from the genus *Macruronus*. These small differences may be used to distribute the species into different groups.

Although other modifications, including proteolysis, glycosylation, and O-phosphorylation,³¹ can be found in NDKs from other species, no modified amino acids were detected except for an hydroxylation and the acetylated N-terminus. Acetylated methionine was the N-terminal amino acid in all the American

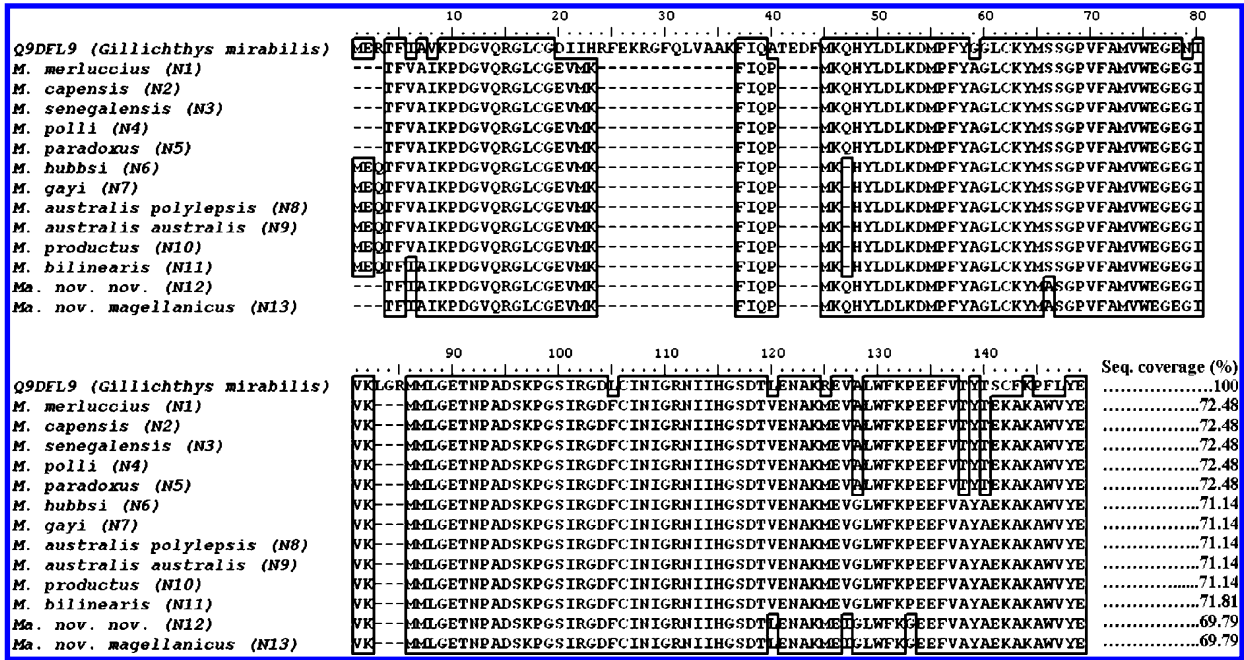


Figure 5. Alignment of NDK B amino acid sequences for all the hakes and grenadiers studied. The sequences, produced by *de novo* MS/MS, were aligned with the NDK B sequence from *Gillichthys mirabilis* (accession number Q9DFL9, Uniprot), which present a high degree of homology with them. Seq. coverage (%), percentage of sequence coverage. (–) not determined amino acid.

hakes matures NDKs. The Met residue in the *P1* position, according to Schecheter's nomenclature,³² has not been cleaved by methionine aminopeptidase. This is expected for the NDKs, given that in their sequence N-terminus Met is followed by Glu. It is known that large side-chains in the second position prevent the activity of methionine aminopeptidase, thus maintaining Met as the N-terminus amino acid in the mature protein.³³ The bioinformatics tool for Met cleavage prediction, available in www.isv.cnrs-gif.fr/terminator2/index.html, leads to the same results.

Unfortunately, not all the peaks produced by MALDI were amenable for MS/MS analysis. Some spectra did not yield enough diagnostic ions to allow the full sequencing of the corresponding peptides, although in some cases, partial sequences were aligned. In these cases, a mass value between brackets is shown to account for the amino acid sequence gaps.

Species Identification using Selected Ion Reaction Monitoring (SIRM). The coupling between the separation capabilities of liquid chromatography and the molecular identification possibilities of mass spectrometry using the SIRM scanning mode provides a fast and reliable method for the differentiation of the species subject of this study. As is shown in Figure 6, unambiguous hake classification was performed by LC–MS/MS, monitoring only four doubly charged ions corresponding to species-specific peptides. Although a complex peptide mixture was used, samples from *in-solution* tryptic digestion of previously unseparated sarcoplasmic proteins, the combination of two variables, i.e., retention time and precursor *m/z*, together with the software representation of specific product ions, made possible the classification of the Merlucciidae family members. The double-charged precursor ions fragmented were: 526.1, 699.3, 665.8 and 887.9. Chromatograms were represented using the ion intensities for fragments at *m/z*: 732.4 (ion y'_{6+} from the precursor 526.1), 1057.3 (ion y'_{10+} from the precursor 699.3), 799.3 (ion y'_{7+} from the precursor 655.8), and 1096.5 (ion y'_{10+} from the precursor 887.9).

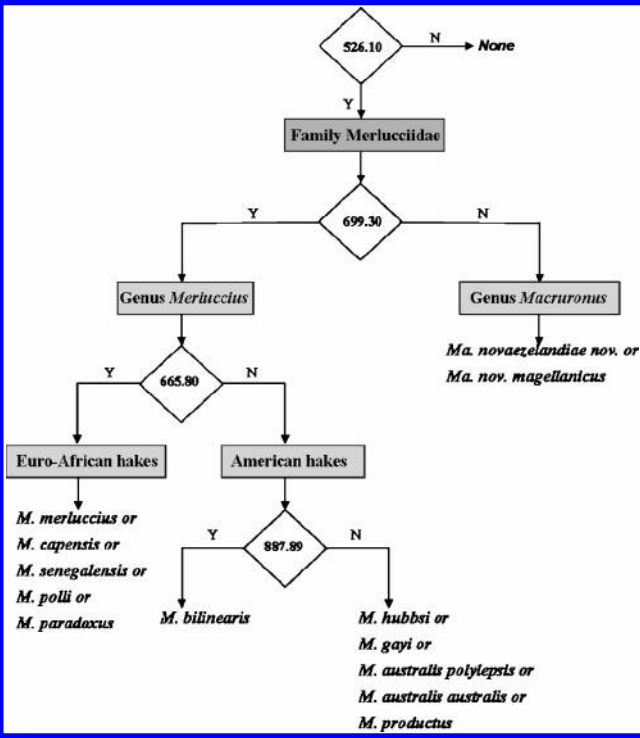


Figure 6. Flow diagram for a systematic discrimination of hakes and grenadiers using only four specific peptides, 526.10, 699.30, 665.80, and 887.89 are the *m/z* of the doubly charged ions of these peptides. Y denotes the presence and N the absence of a particular peptide.

In the sarcoplasmic extracts from the *Macruronus* genus, only the transition 526.1 → 732.4 appeared with a high S/N ratio (Figure 7a), whereas no significant signal was produced for the other three transitions. Conversely, for all the extracts belonging to the *Merluccius* genus (Figure 7b, c), the well-defined transitions 526.1 → 732.4 and 699.3 → 1057.3 were

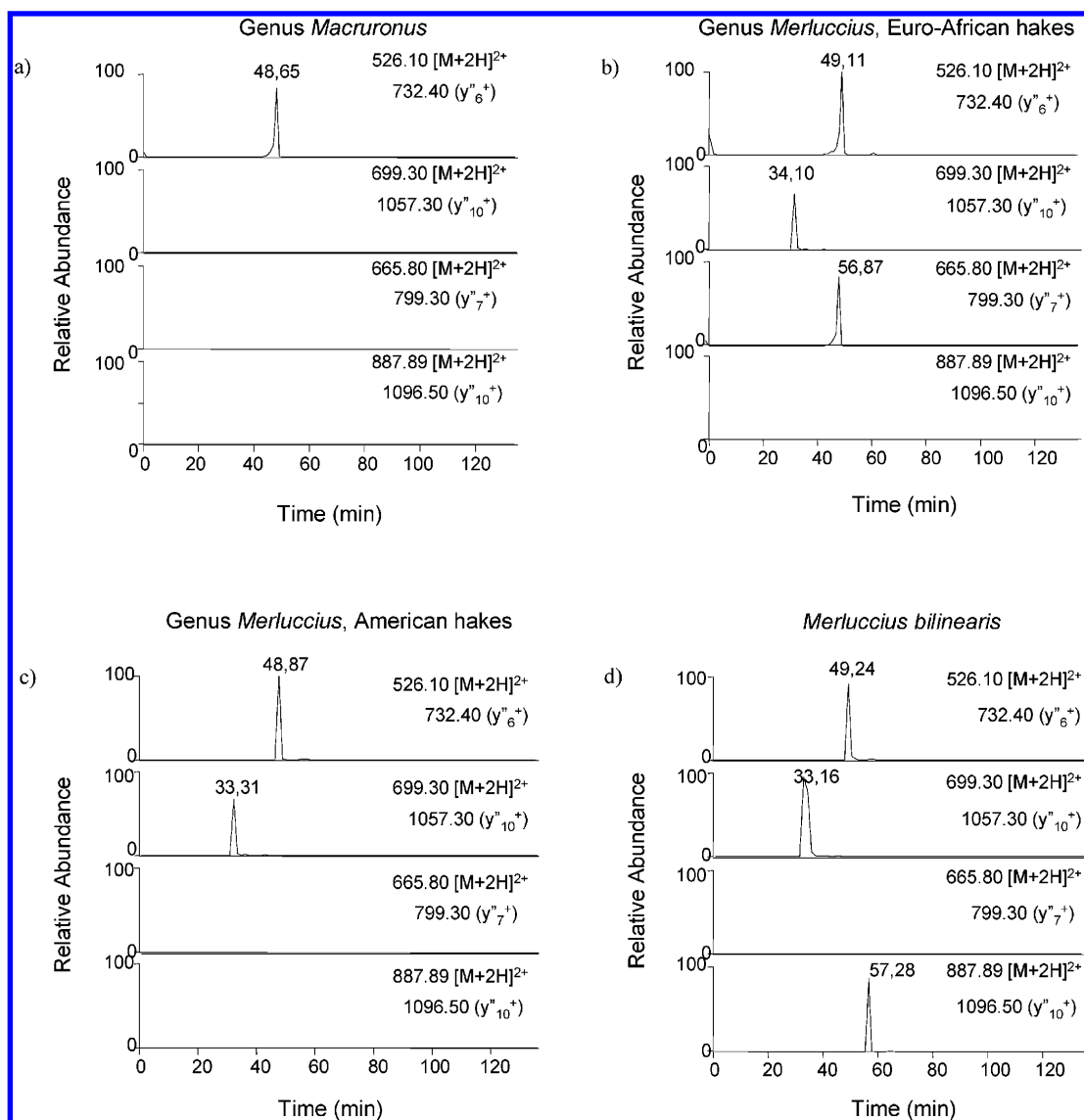


Figure 7. Discrimination of the Merlucciidae species by the monitorization of four NDK B species-specific tryptic peptides. LC–MS/MS using the SIRM scanning mode was used. The IT detector was set to perform a continuous operation in the MS/MS mode, being fragmented only the doubly charged ions corresponding to specific peptides at 526.10, 699.30, 665.80, and 887.89 [M + 2H]²⁺. Total ion counts for each selected products ions as a function of retention time were plotted for the discrimination of (a) genus *Macruronus*, (b) genus *Merluccius*, American hakes, (c) genus *Merluccius*, EuroAfrican hakes, and (d) genus *Merluccius*, *M. bilinearis* species. Note that the relative abundance scale is the same in all the examples presented.

observed, making possible the unequivocal genus differentiation. Within the *Merluccius* genus, the presence or absence of the transition 665.8 → 799.3 allowed the classification of hake species in two groups: American and Euro-African (Figure 7b, c). Finally, as can be seen in Figure 7d, extracts from *M. bilinearis* presented the unique and specific transition 887.9 → 1096.5, which could be used for the unambiguous identification of this species. As expected, the MS/MS spectra obtained in the SIRM experiments exactly matched those represented in Figures 2–4. These results show how the discrimination of close-related species is possible using a SIRM configuration and suggest that this strategy can be used as a new, simple, fast, straightforward, sensitive, and effective means of conducting fish species identification in a single LC–MS/MS experiment.

Conclusions

In this work, the *de novo* sequencing of NDK B from all the commercial hake and grenadier species belonging to the family

Merlucciidae has been reported. No data are yet available for the amino acid sequence of this protein for the specimens of this family. *De novo* sequencing was carried out by manual and computer aided interpretation of the spectra obtained by LC–MS/MS and nano ESI–MS/MS for the peptides produced after tryptic *in-gel* digestion of NDK B spots separated by 2-DE. A high degree of homology was found among the NDK sequences found for each of the species and the NDK B from *Gillichthys mirabilis*. Specific NDK tryptic peptides can be useful for differentiation among members of the family Merlucciidae, providing: (i) a selective differentiation between the genus *Merluccius* and *Macruronus*, (ii) a classification of the hake species into two groups according to their geographic precedence, American or Euro-African hakes, and (iii) an unequivocal identification of *M. bilinearis* species from the others. Finally, the coupling of these peptide biomarkers together with the SIRM configuration in mass spectrometry provides a simple,

fast, and straightforward method for the routine species discrimination.

Acknowledgment. We thank Mrs. Lorena Barros for her excellent technical assistance and Dr. Anabel Marina, Mr. Alberto Jorge García, and the members of the Proteomics Facility of the Centro de Biología Molecular Severo Ochoa for their helpful assistance and suggestions. We also acknowledge Freiremar SA and Cetmar for their assistance in the collection of hakes and grenadiers used in this study. This work was supported by the Comisión Interministerial de Ciencia y Tecnología (CICYT) (Project AGL2000-0440-P4-02). B.C. is supported by the *Ramón y Cajal* program (RYC-2003) under the auspices of the Spanish Ministry of Science and Technology.

Supporting Information Available: Results and Discussion, including Figure S-1 and Table S-1. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Postel, E. H. NM23-NDP kinase. *Int. J. Biochem. Cell Biol.* **1998**, 30(12), 1291–1295.
- Muñoz-Dorado, J.; Intuye, S.; Intuye, M. Nucleoside diphosphate kinase from *Myxococcus xanthus*. II. Biochemical characterization. *J. Biol. Chem.* **1990**, 265(5), 2707–2712.
- Lacombe, M. L.; Wallet, V.; Troll, H.; Véron, M. Functional cloning of a nucleoside diphosphate kinase from *Dictyostelium discoideum*. *J. Biol. Chem.* **1990**, 265(17), 10012–10018.
- Kimura, N.; Shimada, N.; Fukuda, M.; Ishijama, Y.; Miyazaki, H.; Ishii, A.; Yakagi, Y.; Ishikawa, N. Regulation of cellular functions by nucleoside diphosphate kinases in mammals. *J. Bioenerg. Biomembr.* **2000**, 32(3), 309–317.
- Bosnar, M. H.; Gunzburg, J.; Bago, R.; Brečević, L.; Weber, I.; Pavelić, J. Subcellular localization of A and B Nm23/NDPK subunits. *Exp. Cell Res.* **2004**, 298(1), 275–284.
- Lacombe, M. L.; Milon, L.; Munier, A.; Mehus, J. G.; Lambeth, D. O. The human Nm23/nucleoside diphosphate kinases. *J. Bioenerg. Biomembr.* **2000**, 32(3), 269–275.
- Wagner, P. D.; Steeg, P. S.; Vu, N. D. Two-component kinase-like activity of nm23 correlates with its motility-suppressing activity. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, 94(17), 9000–9005.
- Ma, D.; McCorkle, J. R.; Kaetzel, D. M. The metastasis suppressor NM23-H1 possesses 3'-5' exonuclease. *J. Biol. Chem.* **2004**, 279(17), 18073–18084.
- Timmons, L.; Shearn, A. Role of AWD/nucleoside diphosphate kinase in *Drosophila* development. *J. Bioenerg. Biomembr.* **2000**, 32(3), 293–300.
- Janin, J.; Dumas, C.; Moréra, S.; Xu, Y.; Meyer, P.; Chiadmi, M.; Cherfils, J. Three-dimensional structure of nucleoside diphosphate kinase. *J. Bioenerg. Biomembr.* **2000**, 32(3), 215–225.
- Lascu, I.; Giartosio, A.; Ransac, S.; Erent, M. Quaternary structure of nucleoside diphosphate kinases. *J. Bioenerg. Biomembr.* **2000**, 32(3), 227–236.
- Lee, J. S.; Lee, S. H. Cloning and characterization of cDNA encoding zebrafish *Danio rerio* NM23-B gene. *Gene* **2000**, 245(1), 75–79.
- Gracey, A. Y.; Troll, J. V.; Somero, G. N. Hypoxia-induced gene expression profiling in the euryoxic fish *Gillichthys mirabilis*. *PNAS* **2001**, 98(4), 1993–1998.
- Couture, P.; Kumar, P. R. Impairment of metabolic capacities in copper and cadmium contaminated wild yellow perch (*Perca flavescens*). *Aquat. Toxicol.* **2003**, 64(1), 107–120.
- Audet, D.; Couture, P. Seasonal variations in tissue capacities of yellow perch (*Perca flavescens*) from clean and metal-contaminated environments. *Can. J. Fish. Aquat. Sci.* **2003**, 60(3), 269–278.
- Rajotte, J. W.; Couture, P. Effects of environmental metal contamination on the condition, swimming performance, and tissue metabolic capacities of wild yellow perch (*Perca flavescens*). *Can. J. Fish. Aquat. Sci.* **2002**, 59(8), 1296–1304.
- Couture, P.; Dutil, J. D.; Guderley, H. Biochemical correlates of growth and condition in juvenile Atlantic cod (*Gadus morhua*) from Newfoundland. *Can. J. Fish. Aquat. Sci.* **1998**, 55(7), 1591–1598.
- Martínez, M.; Gunderley, H.; Dutil, J. D.; Winger, P. D.; He, P.; Walsh, S. J. Condition, prolonged swimming performance and muscle metabolic capacities of cod *Gadus morhua*. *J. Exp. Biol.* **2003**, 206(3), 503–511.
- Bélanger, F.; Blier, P. U.; Dutil, J. D. Digestive capacity and compensatory growth in Atlantic cod (*Gadus morhua*). *Fish Physiol. Biochem.* **2002**, 26(2), 121–128.
- Piñeiro, C.; Vázquez, J.; Marina, A. I.; Barros-Velázquez, J.; Gallardo, J. M. Characterization and partial sequencing of species-specific sarcoplasmic polypeptides from commercial hake species by mass spectrometry following two-dimensional electrophoresis. *Electrophoresis* **2001**, 22(8), 1545–1552.
- Jaillon, O.; Aury, J. M.; Brunet, F.; Petit, J. L.; Stange-Thomann, N.; Mauceli, E.; Bouneau, L.; Fischer, C.; Ozouf-Costaz, C.; Bernot, A.; Nicaud, S.; Jaffe, D.; Fisher, S.; Lutfalla, G.; Dossat, C.; Segurens, B.; Dasilva, C.; Salanoubat, M.; Levy, M.; Boudet, N.; Castellano, S.; Anthouard, V.; Jubin, C.; Castelli, V.; Katinka, M.; Vacherie, B.; Biemont, C.; Skalli, Z.; Cattolico, L.; Poulain, J.; de Berardinis, V.; Cruaud, C.; Duprat, S.; Brottier, P.; Coutanceau, J. P.; Gouzy, J.; Parra, G.; Lardier, G.; Chapple, C.; McKernan, K. J.; McEwan, P.; Bosak, S.; Kellis, J. N.; Volff, J. N.; Guigo, R.; Zody, M. C.; Mesirov, J.; Lindblad-Toh, K.; Birren, B.; Nusbaum, C.; Kahn, D.; Robinson-Rechavi, M.; Laudet, V.; Schachter, V.; Quetier, F.; Saurin, W.; Scarpelli, C.; Wincker, P.; Lander, E. S.; Weissenbach, J.; Roest Crolius, H. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **2004**, 431(7011), 946–957.
- Aparicio, S.; Chapman, J.; Stupka, E.; Putnam, N.; Chia, J.; Dehal, P.; Christoffels, A.; Rash, S.; Hoon, S.; Smit, A.; Gelpke, M. D. S.; Roach, J.; Oh, T.; Ho, I. Y.; Wong, M.; Detter, C.; Verhoeve, F.; Predki, P.; Tay, A.; Lucas, S.; Richardson, P.; Smith, S. F.; Clark, M. S.; Edwards, Y. J. K.; Doggett, N.; Zharkikh, A.; Tavtigian, S. V.; Pruss, D.; Barnstead, M.; Evans, C.; Baden, H.; Powell, J.; Glusman, G.; Rowen, L.; Hood, L.; Tan, Y. H.; Elgar, G.; Hawkins, T.; Venkatesh, B.; Rokhsar, D.; Brenner, S. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **2002**, 297(5585), 1301–1310.
- Carrera, M.; Gallardo, J. M.; Piñeiro, C.; Cañas, B.; López, D.; Vázquez, J. Procedure for the identification of the commercial species from the Merlucciidae family, necessary elements and applications. Patent approval pending, CSIC registration number 200603287.
- Piñeiro, C.; Barros-Velázquez, J.; Pérez-Martín, R. I.; Gallardo, J. M. Specific enzyme detection following isoelectric focusing as a complimentary tool for the differentiation of related Gadoid fish species. *Food Chem.* **2000**, 70(2), 241–245.
- Piñeiro, C.; Barros-Velázquez, J.; Sotelo, C. G.; Pérez-Martín, R. I.; Gallardo, J. M. Two-dimensional electrophoretic study of the water-soluble protein fraction in white muscle of Gadoid fish species. *J. Agric. Food Chem.* **1998**, 46(10), 3991–3997.
- Jensen, O. N.; Wilm, M.; Shevchenko, A.; Mann, M. Sample preparation methods for mass spectrometric peptide mapping directly from 2-DE gels. *Methods Mol. Biol.* **1999**, 112, 513–530.
- Carrera, M.; Cañas, B.; Piñeiro, C.; Vázquez, J.; Gallardo, J. M. Identification of hake and grenadier species by proteomic analysis of the parvalbumin fraction. *Proteomics* **2006**, 6(19), 5278–5287.
- Rocap, G.; Larimer, F. W.; Lamerdin, J.; Malfatti, S.; Chain, P.; Ahlgren, N. A.; Arellano, A.; Coleman, M.; Hauser, L.; Hess, W. R.; Johnson, Z. I.; Land, M.; Lindell, D.; Post, A. F.; Regala, W.; Shah, M.; Shaw, S. L.; Steglich, C.; Sullivan, M. B.; Ting, C. S.; Tolonen, A.; Webb, E. A.; Zinser, E. R.; Chisholm, S. W. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **2003**, 424(6952), 1042–1047.
- Kim, Y. T.; Cho, J. J. Submitted the nucleotide sequence to the EMBL / GenBank / DDBJ databases (Jan. 1998).
- Ouatas, T.; Abdallah, B.; Gasmi, L.; Mazabraud, A. Submitted the nucleotide sequence mRNA to the EMBL / GenBank / DDBJ databases (May 1996).
- Bominnar, A. A.; Tepper, A. D.; Véron, M. Autophosphorylation of nucleoside diphosphate kinase on non-histidine residues. *FEBS Lett.* **1994**, 353(1), 5–8.
- Schechter, I.; Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, 27(2), 157–162.
- Frottin, F.; Martínez, A.; Peynot, P.; Mitra, S.; Holz, R. C.; Giglione, C.; Meinel, T. The proteomics of N-terminal methionine cleavage. *Mol. Cell Proteomics* **2006**, 5(12), 2336–2349.

PR0701963

SUPPORTING INFORMATION

RESULTS AND DISCUSSION

2-DE analyses of NDK

As shown in the Figure S-1 and Table 1 of the article, a major NDK spot, with an apparent pI between 5.42-5.47 and a M_r around 18.60 kDa, was detected in the Euro-African hakes: *M. merluccius* (N1), *M. capensis* (N2), *M. senegalensis* (N3), *M. polli* (N4) and *M. paradoxus* (N5). A different spot, with a pI between 5.04-5.29 and a M_r around 17.80 kDa, was detected only in the American hakes: *M. hubbsi* (N6), *M. gayi* (N7), *M. australis polylepsis* (N8), *M. australis australis* (N9), *M. productus* (N10) and *M. bilinearis* (N11). Finally, both grenadier subspecies: *Ma. novaezelandiae* (N12) and *Ma. nov. magellanicus* (N13) showed a distinctive spot with a pI between 5.43-5.45 and a M_r around 16.80 kDa.

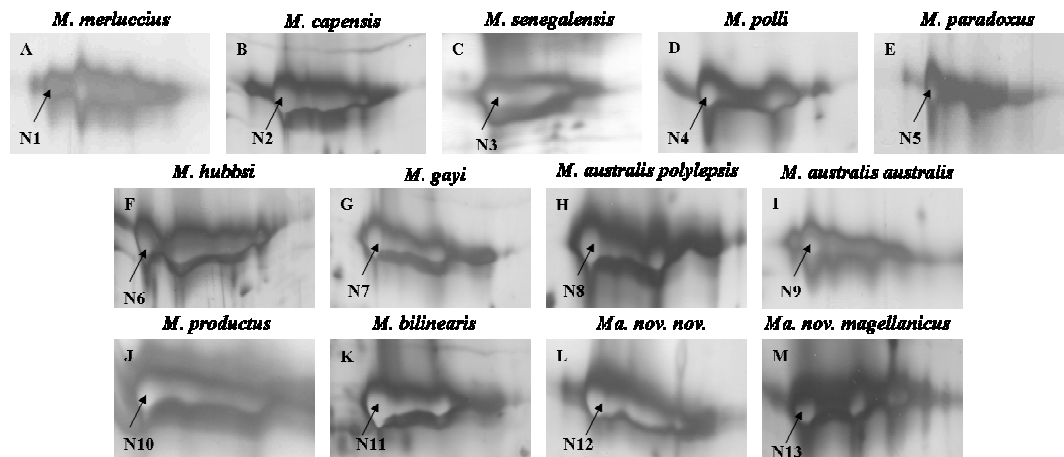


Figure S-1: Cropped regions of representative gradient (12-14 %) silver-stained 2-DE gels showing the separated NDK from all hakes or grenadiers studied. Each of the NDK isoforms studied is indicated with different Arabic numbers. *M.* (*Merluccius* spp.); *Ma. nov.* (*Macruronus novaezelandiae*).

Table S 1. Complete information of the dataset for MS analysis of HPLC.

O.mass ^a	C. mass ^b	Delta	Peptide	Organism	Pos.	Modification	Spot Number												
							Merluccius										Macrur		
							1	2	3	4	5	6	7	8	9	10	11	12	13
707.19							•	•	•	•	•	•	•	•	•	•	•	•	•
788.43	788.56	0.13	HYLDLK	<i>Gillichthys mirabilis</i>	48-53		•	•	•	•	•	•	•	•	•	•	•	•	•
866.14							•	•	•	•	•	•	•	•	•	•	•	•	•
893.39							•	•	•	•	•	•	•	•	•	•	•	•	•
899.68							•	•	•	•	•	•	•	•	•	•	•	•	•
909.60			(893.39)			MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
916.48	916.37	-0.11	QHYLDLK	<i>Gillichthys mirabilis</i>	47-53		•	•	•	•	•	•	•	•	•	•	•	•	•
942.56							•	•	•	•	•	•	•	•	•	•	•	•	•
955.47	955.50	0.03	PFYAGLCK	<i>Scyliorhinus torazame</i>	56-63		•	•	•	•	•	•	•	•	•	•	•	•	•
958.66			(942.56)			MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
960.64							•	•	•	•	•	•	•	•	•	•	•	•	•
994.43							•	•	•	•	•	•	•	•	•	•	•	•	•
1015.86							•	•	•	•	•	•	•	•	•	•	•	•	•
1023.66							•	•	•	•	•	•	•	•	•	•	•	•	•
1039.46			(1023.66)			MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
1051.42			(994.43)				•	•	•	•	•	•	•	•	•	•	•	•	•
1053.39			(1015.86)			K	•	•	•	•	•	•	•	•	•	•	•	•	•
1065.41			(994.43)			CysAcryl	•	•	•	•	•	•	•	•	•	•	•	•	•
1067.47							•	•	•	•	•	•	•	•	•	•	•	•	•
1073.66							•	•	•	•	•	•	•	•	•	•	•	•	•
1082.63	1082.19	-0.44	VAIKPDGVQR	<i>P. marinus</i>	6-15		•	•	•	•	•	•	•	•	•	•	•	•	•
1089.06							•	•	•	•	•	•	•	•	•	•	•	•	•
1105.46	1104.70	-0.76	YMSSGPVFAM	<i>Gillichthys mirabilis</i>	64-73	MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
1123.30							•	•	•	•	•	•	•	•	•	•	•	•	•
1144.56							•	•	•	•	•	•	•	•	•	•	•	•	•
1146.90							•	•	•	•	•	•	•	•	•	•	•	•	•
1164.50			(1146.90)			MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
1174.49							•	•	•	•	•	•	•	•	•	•	•	•	•
1201.45			(1144.56)				•	•	•	•	•	•	•	•	•	•	•	•	•
1217.44			(1144.56)			MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
1233.52			(1144.56)			MSO;HYDR	•	•	•	•	•	•	•	•	•	•	•	•	•
1239.40			(1144.56)			MSO;Na	•	•	•	•	•	•	•	•	•	•	•	•	•
1247.45			(1231.38)			MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
1293.58	1293.50	-0.08	MMLGETNPADSK	<i>Gillichthys mirabilis</i>	86-97		•	•	•	•	•	•	•	•	•	•	•	•	•
1309.58	1308.55	-1.03	MMLGETNPADSK	<i>Gillichthys mirabilis</i>	86-97	MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
1330.74	1330.66	-0.08	TFVAIKPDGVQR	<i>P. marinus</i>	4-15		•	•	•	•	•	•	•	•	•	•	•	•	•
1344.76	1344.70	-0.06	TFIAIKPDGVQR	<i>Xenopus laevis</i> , others	4-15		•	•	•	•	•	•	•	•	•	•	•	•	•
1352.74	1352.63	-0.11	TFVAIKPDGVQR	<i>P. marinus</i>	4-15	Na	•	•	•	•	•	•	•	•	•	•	•	•	•
1365.50							•	•	•	•	•	•	•	•	•	•	•	•	•
1366.76	1366.68	-0.08	TFIAIKPDGVQR	<i>X. laevis</i> and others	4-15	Na	•	•	•	•	•	•	•	•	•	•	•	•	•
1368.74	1368.63	-0.11	TFVAIKPDGVQR	<i>P. marinus</i>	4-15	K	•	•	•	•	•	•	•	•	•	•	•	•	•
1376.02							•	•	•	•	•	•	•	•	•	•	•	•	•
1397.64							•	•	•	•	•	•	•	•	•	•	•	•	•
1411.71	1411.64	-0.07	NIIHGSDTLENK	<i>Gillichthys mirabilis</i>	112-124		•	•	•	•	•	•	•	•	•	•	•	•	•
1419.70	1419.67	-0.03	(1397.64)		112-124	Na	•	•	•	•	•	•	•	•	•	•	•	•	•
1435.70	1435.27	-0.43	(1397.64)		112-124	K	•	•	•	•	•	•	•	•	•	•	•	•	•
1475.65							•	•	•	•	•	•	•	•	•	•	•	•	•
1487.98							•	•	•	•	•	•	•	•	•	•	•	•	•
1511.74							•	•	•	•	•	•	•	•	•	•	•	•	•
1527.80			(1511.74)			MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
1760.60							•	•	•	•	•	•	•	•	•	•	•	•	•
1774.97							•	•	•	•	•	•	•	•	•	•	•	•	•
1775.96			(1760.60)			ACET;MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
1791.72			(1774.97)			ACET;MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
1792.82			(1760.60)			ACET;MSO;MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
1798.07			(1760.60)			ACET;MSO;Na	•	•	•	•	•	•	•	•	•	•	•	•	•
1803.87	1803.79	-0.08	MMLGETNPADSKPGSIR	<i>Gillichthys mirabilis</i>	86-102		•	•	•	•	•	•	•	•	•	•	•	•	•
1807.03							•	•	•	•	•	•	•	•	•	•	•	•	•
1814.21			(1760.60)			ACET;MSO;K	•	•	•	•	•	•	•	•	•	•	•	•	•
1819.87	1819.76	-0.11	MMLGETNPADSKPGSIR	<i>Gillichthys mirabilis</i>	86-102	MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
1828.92			(1774.97)			ACET;MSO;K	•	•	•	•	•	•	•	•	•	•	•	•	•
1835.87	1835.57	-0.30	MMLGETNPADSKPGSIR	<i>Gillichthys mirabilis</i>	86-102	MSO;MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
1851.87	1851.59	-0.28	MMLGETNPADSKPGSIR	<i>Gillichthys mirabilis</i>	86-102	MSO;MSO;MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
1913.05							•	•	•	•	•	•	•	•	•	•	•	•	•
1990.04							•	•	•	•	•	•	•	•	•	•	•	•	•
2070.01							•	•	•	•	•	•	•	•	•	•	•	•	•
2086.20							•	•	•	•	•	•	•	•	•	•	•	•	•
2102.51			(2086.20)			MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
2107.78			(2086.20)			Na	•	•	•	•	•	•	•	•	•	•	•	•	•
2118.30			(2086.20)			MSO;MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
2134.30			(2086.20)			MSO;MSO;MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
2150.10			(2086.20)			MSO;MSO;MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
						MSO	•	•	•	•	•	•	•	•	•	•	•	•	•
2210.91							•	•	•	•	•	•	•	•	•	•	•	•	•
2382.60							•	•	•	•	•	•	•	•	•	•	•	•	•

a) Monoisotopic masses in Da; b) Observed mass-calculated mass. O. mass (observed mass) and C. mass (calculated mass); Pos. (position); (■) peptides common to all the species, (■) denotes the presence and (□) the absence of a peak; Cysteine are carboxy-methylated, MSO: methionine sulfoxide; Na (sodium adduct); K (potassium adduct); HYDR (hydroxylation).

Macrur: genus *Macruronus*; Spot numbers corresponding to: (1) *M. merluccius*, (2) *M. capensis*, (3) *M. senegalensis*, (4) *M. polli*, (5) *M. paradoxus*, (6) *M. hubbsi*, (7) *M. gayi*, (8) *M. australis polylepsis*, (9) *M. australis australis*, (10) *M. productus*, (11) *M. bilinearis*, (12) *Ma. nov. nov.*, (13) *Ma. nov. magellanicus*.