Charles University in Prague

Faculty of Science

Department of Biochemistry

Studies of NK cell receptors and other proteins using recombinant expressions and mass spectrometry

Daniel Kavan

Ph. D. Thesis

Supervisor: Prof. RNDr. Karel Bezouška, DSc.

Prague 2010

I declare that I have worked on this thesis under the guidance of my supervisor and that all sources of the previous knowledge are properly cited. No part of this work was used and will not be used for obtaining any other academic degree, than PhD from Charles University in Prague.

Prague ....

Acknowledgements:

First I would like to thank to my supervisor Karel Bezouška for giving me the opportunity to work in his laboratory and for encouraging me in any despondent moment, this work is his merit in many aspects.

On the second place I would like to thank to every one of my colleagues in the laboratory for their kind help and support and for pleasant work environment.

My thanks belong also to my family for supporting me and for the tolerance they all have for my work.

The work was supported by grants from Ministry of Education of Czech Republic (MSM_216208083, 1M0505 and LC07017), from Czech Science Foundation (303/09/0477 and 305/09/H008), and by EU Project Spine 2 (contract LSHG-CT-2006-031220).

Contents

Abbreviations

bp	base pair
CD	Cluster of Differentiation
DNA	DeoxyriboNucleic Acid
DTT	DiThioThreitol
EDTA	EthyleneDiamineTetraacetic Acid
ESI	ElectroSpray Ionisation
Fc	Fragment Crystalizable
FPLC	Fast Protein Liquid Chromatography
FT-ICR	Fourier Transform – Ion Cyclotron Resonance
H/D	Hydrogen/Deuterium (exchange)
HSQC	Heteronuclear Single Quantum Coherence
Ig	Immunoglobulin
IPTG	IsoPropyl β-D-1-ThioGalactopyranoside
ITAM	Immunoreceptor Tyrosin-based Activation Motif
ITIM	Immunoreceptor Tyrosin-based Inhibition Motif
KIR	Killer-cell Immunoglobulin-like Receptor
LC	Liquid Chromatography
LRC	Leukocyte Receptor Complex
MHC	Major Histocompatibility Complex
MALDI	Matrix-Assisted Laser Desorption Ionisation
MS	Mass Spectrometry
Ni-NTA	Nickel-NitriloTriacetic Acid
NK	Natural Killer
NKC	Natural Killer gene Complex
NKR-P1	Natural Killer Receptor – Protein 1
NMR	Nuclear Magnetic Resonance
PAMP	Pathogen-Associated Molecular Pattern
PCR	Polymerase Chain Reaction
PRR	Pattern Recognition Receptor

RNA	RiboNucleic acid
SHP	Src Homology Phosphatase
TLR	Toll-like Receptor
Tris	Tris(hydroxymethyl)aminomethane

1 Introduction

1.1 NK cell receptors

1.1.1 NK cells

Natural killer cells (NK cells) are the subpopulation of large granular lymfocytes, which lacks the surface receptors typical for B cells or T cells. They are characterized by the $CD3^{-}CD16^{+/-}CD56^{+}CD161^{+/-}$ phenotype [Trinchieri G. 1989] or by the presence of NKp46 and NKp30 [Moretta L. et. al. 2002], however. They constitute the third kind of cells differentiated from the common lymphoid progenitor (besides B and

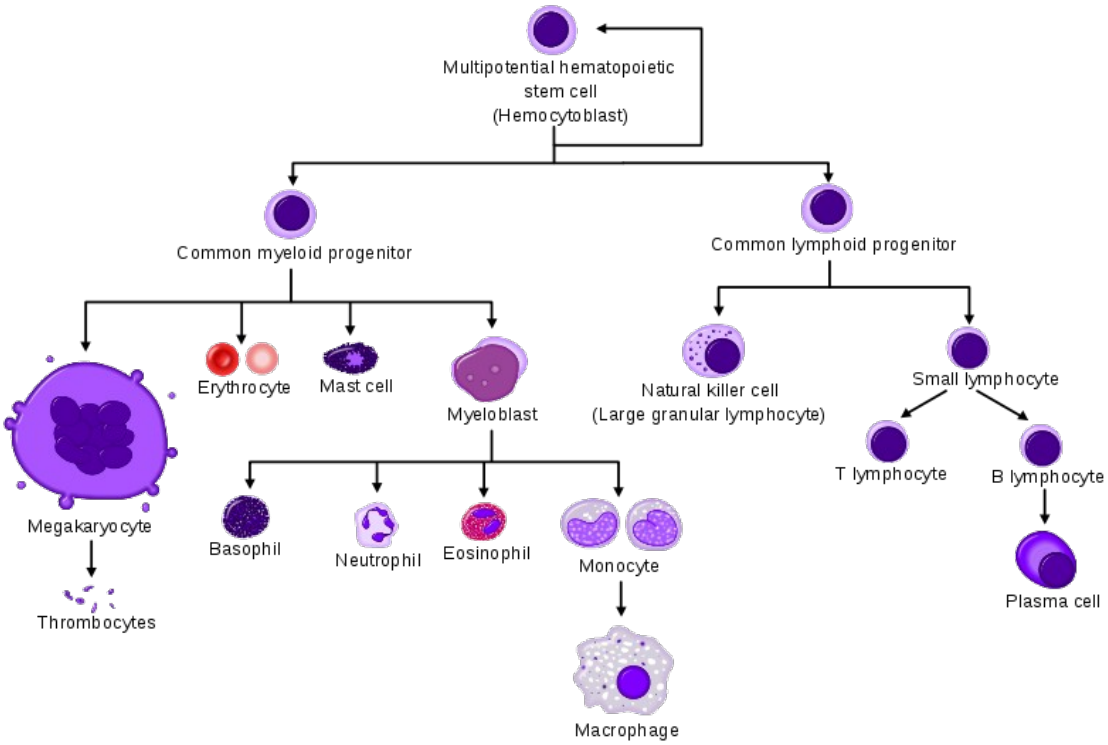


Fig 1: Differentiation of pluripotent cells into various types of blood cells.

T lymphocytes) see Figure 1. They were named natural killers according to their function in the organism, because they do not need any activation and nevertheless they are able to eliminate abnormal (i. e. infected or transformed) cells from the tissue [Kiessling R. et. al. 1975]. This

function is dependent on scanning the major histocompatibility complex (MHC) class I molecules of ambient cells. The resulting action (killing or not killing the target cell) is dependent on the balance of activating or inhibiting signals mediated by the NK cell surface receptors and forwarded to the specific signaling pathway [Raulet D. H. et. al. 2001, Moretta A. et. al. 2001].

Abnormally low MHC class I expression on target cell (common behavior of infected or transformed cells) leads to insufficient stimulation of inhibitory receptors and therefore to the NK cell activation (missing-self recognition) [Ljunggren H. G., Kärre K. (1990), Medzhitov R., Janeway C. A. Jr. (2002)]. However, the presence of ligands for activation receptors is also necessary, therefore erythrocytes, cells without MHC class I, are not attacked by NK cells. [Raulet D. H., Vance R. E. (2006), Oldenborg P. A. et. al. 2000].

Another strategy of target cell recognition is based on fact, that some proteins are expressed only in small amounts on healthy cells and are overexpressed on stressed, transformed or infected cells. These proteins could be recognized by activating receptors (induced-self recognition) as are the stress-induced molecules MICA and MICB by NKG2D receptor [Raulet D. H. 2003, Bauer S. et. al. 1999, Diefenbach A. et. al. 2001].

Another stimulating receptor is Fc-receptor CD16. When the NK cell meets the cell opsonized by IgG, CD16 molecules on the NK cell surface bind the Fc fragments of immunoglobulins. Aggregation of CD16 results in activation of the cytotoxic mechanism and thus in the target cell elimination (antibody-dependent cellular cytotoxicity) [Perussia B. et. al. 1984].

The least explored recognition strategy is based on recognizing of evolutionary conserved pathogen structures (PAMPs) as it occurs in case of LY49H receptor and m157 protein of murine cytomegalovirus (infectious non-self) [Vivier E., Biron C. A. (2002), Brown M. G. 2001].

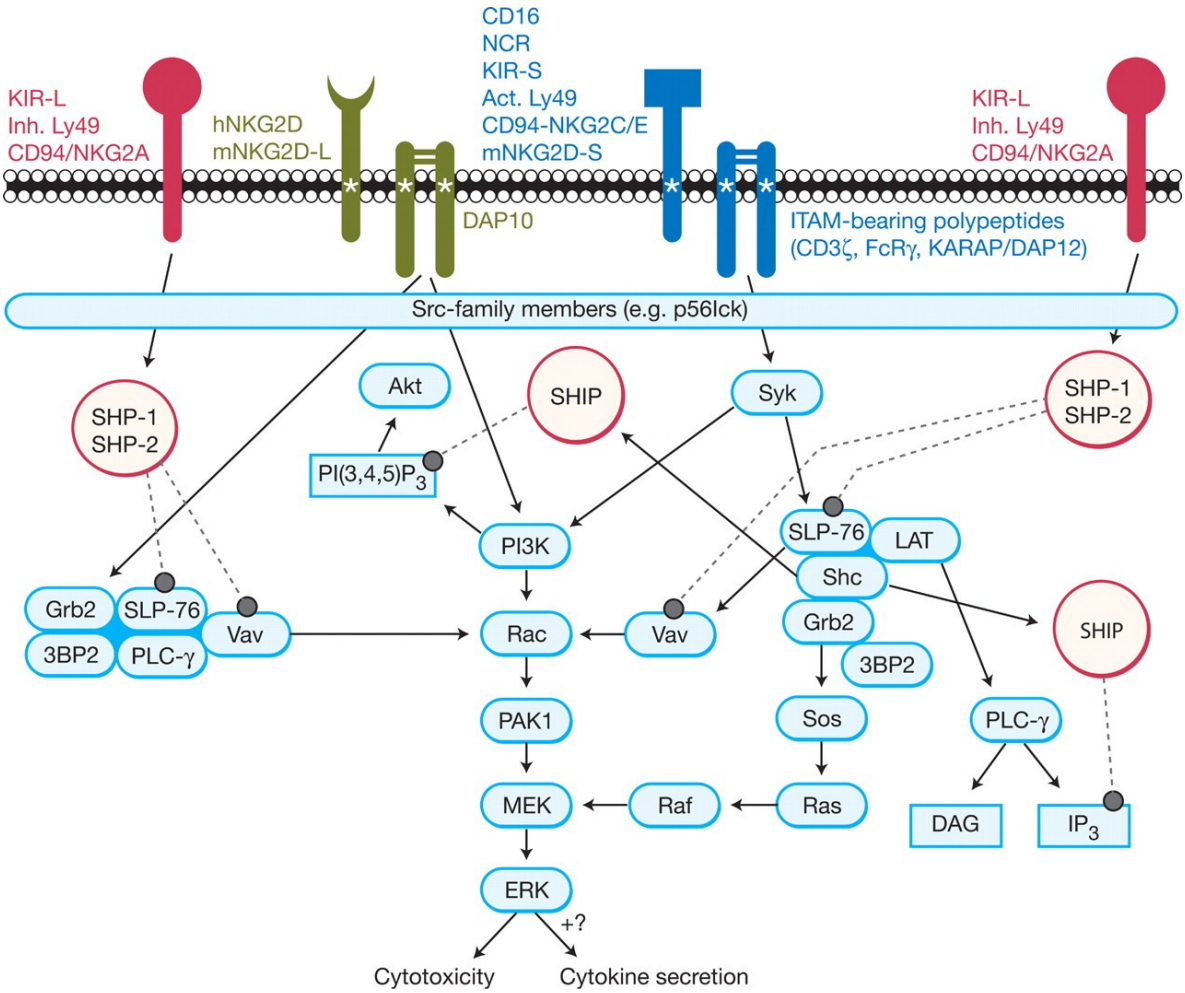


Fig. 2: NK cell signaling pathways. Inhibitory receptors with ITIM in red, activating receptors associated with ITAM bearing adaptor molecules in blue and Syk-independent activation receptor NKG2D associated with DAP10 (containing non-classical ITAM [Billadeau et. al. 2003]) in green. Charged aminoacid (K/R in receptor and D in adaptor) mediating receptor-adaptor contact displayed as asterisk [Vivier E. et. al. 2004].

1.1.2 NK cell receptors

Specificity of NK cells is not based only on one type of antigen receptor as it is in case of T and B cells, but is given by variety of their surface receptors [Lanier L. L. 2005]. These receptors could be divided into several groups according to different parameters. One of these parameters suggested in previous paragraph is the effect of the receptor

stimulation and the division into activation or inhibitory receptors [Moretta A. et. al. 2000].

Activation receptors have very short intracellular part and are usually connected with one or three adaptor molecules (DAP12, FcεRIγ, CD3ζ [Hibbs M. L. et. al. 1989, Anderson P. et. al. 1989, Lanier L. L. et. al. 1998]) containing immunoreceptor tyrosin-based activation motif (ITAM) in the intracellular part of the plasmatic membrane. The signalling occurs through the tyrosine phosphorylation by Src-family kinases followed by Syk-family kinases activation or through G proteins associated with the receptors [Wange R. L. 2000, Jevremovic D. 1999].

Inhibitory receptors have longer intracellular part containing immunoreceptor tyrosin-based inhibition motif (ITIM) [Bix M. et. al. 1991, Vély F., Vivier E. (1997)]. The phosphorylation of ITIM tyrosin by Src-family kinases follows the receptor-ligand interaction and consequently phosphatase (SHP1, SHP2) mediated inhibition of activation signaling cascades [Long E. O. 1999, Valiante N. M. et. al. 1996, Binstadt B. A. 1998] as is described on Figure 2.

Another aspect for the surface receptors classification is their structure. The majority of them could be classified as the lectin-like (CD69, NKR-P1) or immunoglobulin-like (KIR, CD16) receptors, fewer as the Toll-like (CCR7) receptors.

Toll-like receptors recognize structurally conserved molecules derived from microbes and are much more common on other leukocytes (mainly on monocytes, B cells and dendritic cells). They received their name from their similarity to the protein coded by the Toll gene from *Drosophila* and is not a reference to any kind of toll. TLRs are a type of pattern recognition receptors (PRR) recognizing molecules broadly shared by pathogens (PAMPs) [Sivori S. et. al. 2004] but distinguishable from host molecules and are closely related to the interleukin-1 receptors.

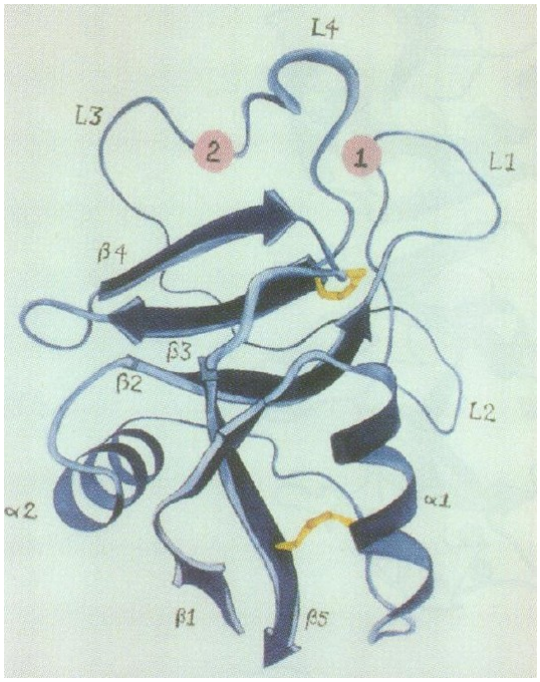


Fig. 3: Ribbon diagram of C-type lectin CRD consensus structure with two α-helices and two anti-parallel β-sheets, disulfide bonds in yellow. [Weis W. et. al. 1991]

Immunoglobulin-like receptors are transmembrane glycoproteins consisting of either two or three immunoglobulin-like extracellular domains. They usually interact with the MHC class I glycoproteins and most of them have longer intracellular part with two ITIMs and thus have the inhibitory effect, i. e. their recognition of MHC molecules supresses the NK cells cytotoxic activity. Only limited number of KIRs have the activating effect [Vilches C., Parham P. (2002)], they have short intracellular part and are associated with DAP12 [Colonna M., Samaridis J. (1995), Wangtmann N. 1995]. Human KIR genes are part of the leucocyte receptor complex (LRC) located on chromosome 19q13.4 [Barten R. et. al. 2001] and are highly polymorphic [Uhrberg M. (2005)].

1.1.3 C-type lectin family receptors

Lectins were first discovered more than 100 years ago in plants (ricin, concanavalin A) and later also in many other organisms from

viruses to mammals with variety of functions (adhesion, immune-recognition). Lectins are defined as multivalent sugar-binding proteins of non-immune origin that agglutinates cells or precipitates glycoconjugates [Goldstein et. al. 1980]. They contain at least two binding sites and are able to bind carbohydrates specifically and reversibly without altering their covalent structure. They could be soluble or bound to the membrane. To distinguish the plant lectins from the animal ones, the term “lectin-like” is used in case of animals [Brewer C. F. 2001]. Lectins could be divided into several groups according to various parameters (specificity, occurrence, …), however, the most useful division is according to their evolutionary relationship as their specificity could evolve independently several times during evolution. According to this, lectins could be diveded into eight groups [Dodd R. B., Drickamer K. (2001), Braakman I. (2001)]: Calnexin, L-types lectins, P-type lectins, C-type lectins, Galectins, I-type lectins, R-type lectins and M-type lectins.

C-type lectins were identified in 1988 [Drickamer K. (1988)] and their characteristic feature is the highly conserved combination of two α-helices and two anti-parallel β-sheets connected by random coils [Weis W. et. al. 1991] with 14 invariant and another 18 highly conserved amino acids [Bezouška et. al. 1991] see Figure 3. This carbohydrate recognition domain (CRD) is internally stabilized by two or three disulfide bonds (and sometimes requires Ca^{2+} ion for proper binding site formation) and consists of around 125 amino acids. Different binding specificity is provided by the participation of variable loops in binding site formation. As this group is defined by amino acid sequence similarity, some members of this group could lose the ability to bind carbohydrates and sometimes even gain another substrate specificity, thus the more convent term for this group is C-type lectin-like proteins [Drickamer K. 1999]. The majority of C-type animal lectins play their role in immune system, other known function imply the cell adhesion, cell-cell recognition generally or even the cryoprotectivity [Ewart K. V. et. al. 1999], nevertheless.

We can divide this structural family into several subgroups, from which mainly C-type lectins type II are expressed on the NK cell surface. These proteins contain one CRD on the extracellular C-terminus. This tertiary structure was initially resolved in case of CD94 molecule [Boyington J. C. et. al. 1999]. CRD is usually glycosylated and receptors occure in form of homodimers (except for heterodimer CD94-NKG2 [Lazetic S. et. al. 1996]).

To name some members of this group we should mention at least CD69, NKR-P1, NKG2 and Ly49. Genes encoding these proteins are located in NK gene complex [Yokoyama W. M., Seaman W. E. (1993)] situated on human chromosome 12, rat chromosome 4 and murine chromosome 6 [Hao L. et. al. 2006]. Original scepticism about similarity of NK cell receptors of human and rodents, when C-type lectins seemed to play the main role in rodents, whereas the immunoglobulin-like receptors play the main role in human organism, is now over, the same receptors could differ in its role from organism to organism [Colucci F. et. al. 2002].

1.1.4 CD69 receptor

CD69 is an important member of C-type animal lectin family closely related to molecules NKR-P1 and Ly49. It was originally described in 1986 as EA 1 on the surface of human T cells after 12-o-tetradecanoyl phorbol-13-acetate treatment [Hara T. et. al. 1986] and later as AIM (activation inducer molecule) [Cebrián M. et. Al. 1988]. It was assigned denomination CD69 at the Fourth International Workshop on Human Leucocyte Differentiation Antigens [Schwarting R. et. al. 1989]. Later it was observed on other immune system cells [Testi R. et. al. 1994] and after cloning [Hamann J. et. al. 1993] it was recognized as C-type lectin.

Human CD69 is constitutively expressed on $CD3^+$ thymocytes, monocytes, granulocytes, neutrophils, epidermal Langerhans cells and platelets and the expression is induced early after NK cells and T cells

activation [Testi R. et. al. 1994], therefore it is considered as the marker receptor of activated lymphocytes. Although it was suggested to participate in process of eliminating the tumor cells *in vitro* [Moretta A. et. al. 1991], later studies in mice indicated its role in silencing the immune reaction [Sancho D. et. al. 2005] and $CD69^-$ mice were more resistant to tumors, probably regarding to the increased apoptosis after contact of $CD69^+$ NK cells with the tumor [Esplugues E. et. al. 2003]. CD69 molecule is a type II transmembrane protein with short intracellular part, transmembrane region, stalk region (containing dimerization cysteine 68) and CRD. It appears as the disulfide bond homodimer of 60 kDa, it is constitutively phosphorylated and contains one typical and one atypical ($N^{111}AC$) possible glycosylation site [Testi R. et. al. 1994, Vance B. A. et. al. 1997, Vance B. A. et. al. 1999].

CD69 stimulation causes various metabolical changes accompanied with increase of the intracellular calcium ions level and cytokine production [Conde M. et. al. 1996]. High levels of CD69 expressions were observed during rheumatoid arthritis, viral hepatitis, atopic reaction [Nopp A. et. al. 2002] and HIV infection [Barboza J. M. et. al. 2002]. Interestingly, the CD69 negative mice have notably higher resistance against some kinds of tumors [Espulgues E. et. al. 2003].

The attempts to characterize the structure of CD69 started shortly after its first description [Gerosa F. et. al. 1991], research in this field continued to computer model [Bajorath J., Aruffo A. 1994] and the structure solving by X-ray crystalography [Natarajan K. et. Al. 2000, Llera A. S. et. al. 2001], that confirmed the typical C-type lectin structure. These works did not find any calcium ions in the molecule, other research revealed structural changes leading to high affinity carbohydrate-binding formation after calcium binding, however [Pavlíček et. al. 2003].

1.2 Introduction to methods

1.2.1 Molecular biology

Molecular biology is the study of biology at the molecular level. It tightly cooperates with genetics and biochemistry and especially biochemists take some specific molecular biology approaches and techniques as a very useful tool – it could be much easier to get the protein of interest by cloning and recombinant expression than by isolation from natural material – the term “molecular biology” will be used in the meaning of techniques in this thesis.

While the origins of molecular biology can be found in 1930's, the real takeoff started in late 1950's and early 1960's. Since then scientists have learned how to isolate and manipulate biomolecules as DNA or RNA and how to exploit proteosynthesis and other enzymatic equipment of the cell.

The basic molecular biology techniques are the expression cloning, polymerase chain reaction (PCR), reverse transcription, restriction digestions, gel electrophoresis, blotting and probing. Other methods may be qualified as advanced, we can mention site-directed mutagenesis, DNA microarrays or quantitative PCR here.

As in other fields of science the new methods and techniques are continuously being developed and improved for achieving faster protocols, higher sensitivity or efficiency and higher yields with fewer starting material. Older methods are being abandoned, however, it is still necessary to know about these older techniques as they could show useful for solving new problems, for which the newer methods may not be suitable.

1.2.2 Protein renaturation *in vitro*

Correct protein function (binding properties, catalytic activity, …) is determined mainly by its structure, and structure is determined mainly

by the amino acid sequence (primary structure). Even modest changes in protein structure can result in malfunctions or even absolute disfunction of the protein. It was shown (by Bruce Merrifield, the Nobel prize winner in chemistry from 1984) that sufficiently simple protein is able to fold in solution spontaneously so we can conclude that the primary structure not only determines the structure, but also more or less directs the folding and thus that the structure could be proclaimed to be the natural consequence of the amino acid sequence.

Although the process of denatured protein refolding is somewhat different from folding of newly synthetized one, the theory of protein folding can be applied to both of these processes. There is a subtle balance between two kinds of forces – the stabilizing (van der Waals, hydrogen bonds, disulfide bonds formation, hydrophobic interactions, salt bridges D-K) and destabilizing (electrostatic repulsion, chain entropy) and the protein folding is the result of this balance and represents the way from unfolded to native conformation, that has lower Gibbs free energy. Unfortunately on the way from unfolded to native conformation the protein could enter local minimums (folding traps) from which could be hard or even impossible to continue to fully native conformation, and the aim of appropriate *in vitro* refolding protocol is to minimize the loses in these minimums.

Generally the process of *in vitro* refolding consists of two steps, the denatured protein (inclusion bodies) solubilization (in chaotropic agents such as urea or guanidinium hydrochloride, extreme pH, detergents, reducing agents or combination of these) and solubilized protein transfer into non-denaturing environment (that could be supplemented with stabilizing chemicals such as arginine or glycerol and redox system such as reduced/oxidized glutathione or cysteamine/cystamine for support of disulfide bonds formation), what is the refolding itself. Three main proceedings are used for refolding: Dialysis, fast dilution method and chromatographic methods. In any of

them the concentration of refolded protein should not exceed 0.1 mg/ml, otherwise the protein could aggregate or even precipitate, as unfolded hydrophobic cores could interact.

Dialysis method is based on slow exchange of solution through the semipermeable membrane, however sometimes it is difficult to keep the protein concentration low enough. As the concentration of denaturing agent is used in high concentration, the dialysis should be done in several steps.

The principle of fast dilution method is that denatured protein is slowly dropped into the excess of stirred refolding buffer, what ensures the lowest possible local concentration of unfolded protein in solution ergo aggregation.

Last widely used refolding techniques are based on chromatographic methods. Gel filtration or affinity chromatography are the most suitable.

1.2.3 Physical biochemistry

Much effort was devoted by biochemists to reveal the relationship between structure and function of biomolecules during last century. Quite a wide range of methods and instrumentation is used for these studies of biomolecules, their complexes and of the processes in which these biomolecules are involved, from classical biochemical approach to above mentioned molecular biology. During last years another type of methods is undergoing rapid development and promises even deeper insight into this problem. These methods are based on physical techniques and with fast technology and informatics development they are likely to impact on the workflow of the most of life scientists.

1.2.3.1 Electrophoresis

While dealing with the physical biochemistry techniques we should start with such a basic procedure as the electrophoresis defined as the motion of dispersed particles relative to a fluid under the influence

of a spatially uniform electric field. This technique exploits the electrical charges associated with most biomolecules. The electrophoretic mobility of the particle depends on many factors the most important of which are charge, molecular mass, shape of the particle and ambient porosity, viscosity and temperature. Some of these factors are unchangeable particle properties (molecular mass), other could be biased by the experimental conditions (shape, charge, ambient temperature and porosity, …). Even very complex mixtures could be separated with high resolution by choice of the convenient electrophoresis format and conditions [Adamson N. J., Reynolds E. C. (1997)].

The first electrophoretic experiment (in 1930's) was performed in free aqueous buffer. Later some kind of solid support (paper, gel, …) allowing to stain and fix separated molecules was introduced into this method. More recently it was discovered that it is possible to apply much higher voltage allowing better resolution when the electrophoresis is performed in thin capillary and the heat dissipation is still ensured. Many capillary electrophoresis systems, zonal as well as free, are available which is an interesting historical flashback to the method's origins.

Many electrophoretic set-ups (denaturing and non-denaturing; analytical and preparative; protein and DNA; plate and capillary; isoelectric focusation) and detection methods (staining, immunodetection, radiolabelling, fluorescence) are known and used.

1.2.3.2 Chromatography

Another method with physical background widely used in biochemistry is chromatography. When we consider hypothetical two-phase system, any chemical substance will partition between these two phases and the concentration ratio is determined by the thermodynamic properties of the phases and substance particles. In liquid-solid system the particles of different thermodynamic properties dilluted in liquid will adsorb differently on the solid phase surface. When we apply these facts in the column, we call it column chromatography,

the most common chromatography in protein science. The thermodynamic properties of the particles are affected by their size, shape, (net) charge, structure, composition, etc., therefore any of these characteristics could be the base for substances separation.

The first chromatographic experiment was the plant pigments separation in the beginning of the last century (the origin of the word chromatography is in Greek Chroma=colour, it is known for a long time that chromatography is applicable to colourless substances as well as coloured, though). Since then many forms of chromatography were introduced from which the column liquid chromatography is used in the protein science the most (solid stationary phase in the column and solubilised protein/peptide mixture as a liquid phase). The development tends to higher efficiency, resolution, sensitivity and reproducibility and shorter separation times, similarly as in other analytical and separation techniques. These goals could be achieved mainly by introducing new stationary phases as the hydrodynamic properties of stationary phase particles play the key role in resolution (peak broadening, peak symetry, plate height, …) and the size of the stationary-phase particles determine the maximal applicable liquid pressure during separation.

We distinguish so called modes of chromatography and formats of chromatography. The mode of chromatography is the type of sample-stationary phase interaction (ion exchange, hydrophobic interactions, affinity, size exclusion, …) and is chosen considering the differences in separated particles. The format of chromatography is determined by the size of stationary-phase particles and limits maximum pressure applicable during chromatography. The largest particles (cca 100 μm) are used in open column chromatography (or low performance LC) whereas the smallest particles (3-5 μm) are used in high performance liquid chromatography (HPLC). Somewhere between these two formats stays fast protein liquid chromatography. The open column chromatography has low resolution and long retention time, however,

stationary-phases are cheap and the separation occurs in atmospheric pressure (and no additional pump is required), therefore it is suitable for large scale pre-separations. In contrary, HPLC has good resolution and short retention times, but HPLC devices and columns are fairly expensive [Kastner M. 2000, Sheehan D. 2009].

1.2.3.3 Mass Spectrometry

Mass spectrometry (MS) is an analytical technique for the determination of molecular mass of sample molecules, moreover it could be used for elucidation of the structure of complex molecules, such as peptides or proteins. The principle of this technique is ionization of sample molecules and determining the mass-to-charge ratio (m/z) of created ions in the gas phase from their behavior in the electromagnetic field. Furthermore the complex molecules could be fragmented and m/z ratio of these fragments could be measured.

Mass spectrometry has originally been used mainly for study of volatile molecules, non-volatile molecules must have been derivatized. Moreover, because of the limitations imposed by ionization MS had only limited utilization in measurement of complex biomolecules till the introduction of novel ionization techniques. Today the most powerful mass spectrometers are able to determine the mass of large proteins (>100 kDa) with resolution of less than 1 Da.

Each mass spectrometer consists of three parts: ion source, mass analyzer and detector. In protein science the most common ion sources are matrix-assisted laser desorption ionization (MALDI), when the sample is co-crystallized on the target plate with matrix, and electrospray ionization (ESI), when the sample is ionized and transferred into the gas phase using electric field. The most useful analyzers are time-of-flight (TOF), quadrupole, quadrupole ion-trap and very powerful but also very expensive Fourier transform ion cyclotron resonance [Graham R. L. et. al. 2007].

Mass spectrometer could be interfaced with other high resolution technique and/or with another mass spectrometer, which greatly enhances the range of analytical problems that could be solved by this methodology. This other high resolution technique could be (among others) gas chromatography, HPLC (the most commonly in protein analytics) or electrophoresis. For instance, in case of HPLC the mixture is separated on chromatography column initially and mass spectrometer is used as a detector (or one of detectors). The most suitable ion source for HPLC is ESI as the chromatograph could be connected with mass spectrometer online, nevertheless it is also possible to spot samples from HPLC onto the pre-prepared MALDI target.

Another useful combination is connecting two (or more) mass spectrometers together [Dongré A. R. et. al. 1997]. It involves introducing just single-mass ions of interest into some kind of collision cell, where this ion could be broken into fragments that are analyzed in the second spectrometer consequently. There is many ways for ion fragmenting, each with its own specificity and characteristics (the most used are collision-induced dissociation, electron-capture dissociation, electron-transfer dissociation, …). This type of analysis is called tandem MS, MS/MS or MS^2 and can be performed also in single appliance equipped with ion trap [Jonscher K. R., Yates J. R. 3rd. (1997)].

1.2.3.4 Nuclear Magnetic Resonance Spectroscopy

Atoms and molecules have a variety of attributes whose values are quantized and variety of spectroscopic techniques exploit the transitions between the energy levels. Nuclear magnetic resonance takes into account the quantization of nuclear magnetic spin. When the nucleus is placed in external magnetic field, the precise energy for transition of the nucleus that depends on the identity of nucleus as well as on its chemical environment, is measured.

The resonance frequency (frequency of absorbed electromagnetic radiation) and consequently the energy of absorption (and consequently

the intensity of the signal) of a particular nucleus is directly proportional to the strength of applied magnetic field. In the magnetic field of strength 21.2 T the protons resonate at 900 MHz, therefore it is common to refer to this strength of magnetic field as to 900 MHz magnet, although various nuclei resonate at different frequencies (dependent on total magnetic spin of the nucleus and on the gyromagnetic ratio) in this field (for comparison, the strength of Earth's magnetic field at our latitude is around 50 μT and protons resonate at cca 2 kHz in this field strength). As the resonance frequency is not influenced only by the identity of nucleus but also by the chemical environment of the atom, it is possible to determine the structure of the complex molecule from resonance shifts. These shifts are relative to some stated standard for each element and their value is small relative to high frequencies so they are expressed in ppm (parts per million). [Sheehan D. 2009].

Another important phenomenon in protein NMR spectroscopy is the molecular rotation. Large molecules rotate slowly and dipolar interactions cause peak broadening. Moreover the peaks are overlapping, therefore in structural studies of protein molecules it is necessary to exploit the correlation effect (through covalent bonds) or nuclear Overhauser effect (through space) and evaluate more-dimensional spectra to retrieve more signals.

The evolution of NMR spectroscopy is based mainly on the increasing of magnetic field strengths, the higher it is the shorter are the times of data acquisition and the better is the resolution. The reasonable NMR appliance for protein structure determination should have at least 600 MHz magnet. Stronger magnets provides better results, but have reasonably higher work costs.

1.2.3.5 Other physical methods

As it has been already mentioned in this thesis, physical methods play very considerable role in modern biochemistry and it is more than probable that this role will become even more important with introducing

new materials and technologies in the future. Single scientist and even single scientific group is not capable of mastering and maintaining all these techniques and since these techniques are supplementary, scientist and scientific groups have to co-operate with each other. Other powerful physical methods not described in this thesis include fluorescence spectroscopy, circular dichroism, Raman and other types of infra-red spectroscopy, surface plasmon resonance, X-ray diffraction by crystals, calorimetry and hydrodynamic methods such as analytical ultracentrifugation and flow cytometry to name at least some of them.

1.2.4 Bioinformatics

Bioinformatics is the application of computer science to the solution of biological problems. Till the massive development and spreading of computers in the end of the last century, the most of information was stored on the paper either in form of books or magazines. Looking for any piece of relevant information was very difficult and time consuming task and data comparison was error prone work. The increasing of processor performance, memory and data storage capacities accompanied with the stunning progress in technology allowed even more data to be easily searchable and accessible. Moreover, rapid Internet connection penetration allowed easier and faster cooperation and huge data collections originated. Originally bioinformatics was domain of only a few of computer scientist, but by the time (at least in the basic form) it became the common part of daily scientific work.

We can weakly distinguish between several areas of bioinformatics (databases, web sites, computer programs, …) that diffuse almost each to other and have one common objective – handling data from scientific experiments. The more powerful the scientific instrumentation is (sensitivity, resolution, shorter experiment times) the more data it generates and the more urgent is the necessity to filter the

output data and select just relevant information – and this is the area, where the bioinformatic specialist plays his role.

2 Aims of the thesis

- To develop the refolding protocol providing sufficient amount of CD69 for NMR structure studies

- To perform NMR titration experiments and examine structural changes during and after GlcNAc binding using NMR

- To prepare CD69 mutants

- To develop DataAnalysis macro for handling of MS data from H/D exchange experiments

- To develop a suite of software tools for visualization of H/D exchange results

- To manage other bioinformatic tasks related to MS data handling

3 Methods

3.1 Molecular biology methods

3.1.1 DNA cloning

DNA cloning is a process of obtaining multiple copies of desired DNA (usually a gene of interest) *in vitro*. It could imply processes such as RNA isolation, reverse transcription, PCR, recombination, various enzymatic DNA manipulations (restriction, ligation, de/phosphorylation, …) and the final step is usually preparation of the expression vector.

Expression vector preparation is a technique, when DNA is cloned into the expression vector, i. e. vector bearing the transcriptional and translational signals necessary for controlling the gene of interest. The aim of this technique is to provide a way for producing large quantities of specific protein, what is usually easier than isolating such a large amount from natural material.

In this thesis DNA cloning usually followed the same workflow: RNA isolation using TRI REAGENT (MRC Inc.), reverse transcription (SuperScript III, Invitrogen), PCR, restriction digestions and ligation. pBlueScript SK+ (Stratagene) or pCR2.1-TOPO (Invitrogen) was used for routine cloning in XL-1 MRF' Blue (Stratagene) or NovaBlue competent cells (Merck) and pRSET B (Invitrogen), pET30a+ or pET28b+ (both Merck) was used as expression vectors. All expression vectors allow to prepend of His-tag on the N-terminus of protein, however, it was utilized just in case of covalently dimeric human CD69 molecule.

3.1.2 Site-directed mutagenesis

Site-directed mutagenesis is a molecular biology technique in which a mutation is created at a defined site in a DNA molecule, usually plasmid. It is useful either for corrections of polymerase errors or for

targeted protein mutations that are integral parts of functional and structural protein research and provide relatively simple and straightforward way to verify supposed role of any particular amino acid in protein tertiary (or quarternary) structure. Thus for instance to ensure if the particular charge of the particular amino acid is involved in ligand binding or structure formation, we prepare mutants with this amino acid mutated to similar neutral amino acid and try to compare binding properties or structure of wild-type (non mutated) and mutant protein.

The procedure starts with synthesis of short oligonucleotide pair according to the sequence of our gene near the mutation site containing the desired mutation somewhere in the center of them, that tend as PCR primers for the second step, which is the PCR using high-fidelity DNA polymerase such as Pfu Turbo DNA polymerase (Stratagene). The error rate of such polymerases is about 1.3×10^{-6} per base per duplication (for comparison the commonly available Taq polymerase has error rate near 8×10^{-6}) and significantly decreases number of undesired mutations, especially in case of long vectors. Another important aspect of this reaction is the polymerization time which should be at least 1 min/1000 base pairs (bp) for plasmids to 10 kbp and even extended for longer templates.

As the last step the template DNA should be eliminated from reaction mixture. Assuming the template DNA has its origin in bacterial replication, one of the easiest way to achieve this is adding restriction endonuclease DpnI recognizing GATC sequence with methylated adenine. PCR mixture could be used for bacterial transformation after this step and all transformants should contain and propagate mutated DNA, nevertheless colonies screening should be used.

3.2 Protein expression and purification

3.2.1 Choice of expression system

The choice of expression system is crucial decision at the beginning of long and sometimes distressful way to final product. Primary problem it this decision is choice of prokaryotic, eukaryotic or even the cell-free expression system.

The main disadvantage of the cell-free system is the lack of experiences and economical aspect as it is fairly novel approach [Cuthbert A. W., Fuller W. (2003)].

Both eukaryotic and prokaryotic expression systems have their advantages and disadvantages, moreover the advantage in one point of view can turn into disadvantage in another one. The main advantage of prokaryotic systems is the simplicity, rapidness and cheapness, whilst the main disadvantage is the lack of posttranslational modifications (turning into advantage in the crystalographer point of view) and sometimes the necessity of *in vitro* refolding.

On the other hand the main advantage of the eukaryotic expression system is possibly properly folded protein with posttranslational modifications and disadvantage is the time consuming production. Other disadvantages, such as low yields, have became less significant recently with introduction of new cell lines and systems.

In this thesis the majority of proteins were produced in *E. coli* BL-21 (DE3) Gold competent cells (Stratagene) in form of aggregated inclusion bodies as prokaryotic expression systems promising soluble protein productions (low temperature optimized cell lines ArcticExpress (Stratagene) or periplasmic production in pMal-p2 (NEB) systems) provided unsatisfactory results. From other works concerning prokaryotic expression of CD69 we can conclude that it is a general finding [Boyington J. C. et. al. 1999, Natarajan K. et. al. 2000, Llera A. S. et. al. 2001].

Transformed competent cells were cultivated in LB medium in Erlenmayer flasks at 37°C until the optical density of 0.6-0.8 (measured at 550 nm), then the production was induced by adding 0.1-0.5 mM IPTG. After 4-8 h of production the bacteria were harvested and freeze-thaw-sonicate lyzed in 25% sucrose buffer. After centrifugation at 5000× g the pellet was resuspended in 0.1% Triton X-100 and centrifuged. The inclusion bodies were washed in 50 mM TrisCl, 150 mM NaCl (pH=7.4).

3.2.2 Protein refolding *in vitro*

The most of proteins considered in this thesis are refolded using fast dilution method. An exception is covalently dimeric human CD69 which is refolded while imobilized on Ni-NTA sepharose column. The only protein, involved here just marginally and indirectly, refolded during dialysis is rhisopuspepsin used in several experiments for protein digestion after H/D exchange [Rey M. et. al. 2009].

Isolated inclusion bodies were dissolved in denaturating buffer. In most cases 6M guanidinium hydrochloride pH=8 with addition of 100mM DTT was used. The refolding itself, was done by fast dilution method by adding dissolved inclusions to appropriate refolding buffer that was slightly different for each protein.

The dimeric CD69 was solubilized in 8M Urea, slowly dropped into 2M Urea with redox buffer and loaded onto Ni-NTA column, where it was consecutively washed with 2M Urea without redox buffer, Tris buffer and finally eluted using 50mM EDTA [Childs R. A. et. al. 1999].

3.2.3 Liquid chromatography

The highest resolution and performance of chromatographic methods used here provides reversed phase HPLC, however the proteins can lose the activity or even denaturate in higher concentrations of organic solvents and under higher pressure used during this type of

chromatography, thus milder FPLC format I more suitable and in most cases sufficient.

Gel filtration on Superdex HR 75 column (GE Healthcare) either as a separation method or at least as a protein homogenity control is used. Ion exchange chromatography is usually used as a first separation step usually on Q-Sepharose FF (Sigma) column.

3.3 Basics of protein NMR spectroscopy

3.3.1 Structural experiments

Unfortunately for protein structure studies via NMR, the most abundant isotopes of carbon (^{12}C) and oxygen (^{18}O) has nuclear spin of 0 and thus it is not visible in NMR experiments, moreover the most abundant nitrogen isotope (^{14}N) has nuclear spin of 1, thus provides useless broad peaks in NMR spectra. So the only usable element in protein remains hydrogen. As we usually need multidimensional heteronuclear spectra, the protein has to be ^{13}C or ^{15}N (or both) labeled. One particularly significant experiment in protein NMR is heteronuclear single quantum coherence (HSQC) providing 2D spectrum with one axis for H and the other for N. Each peak in this spectrum corresponds to one proton attached to nitrogen via covalent bond, therefore majority of peaks belongs to NH in peptide bond and the rest to side chain NH_x groups. In case we are able to assign the peptide bond peaks to particular amino acids (the shifts of proton and nitrogen should be equal in other experiments) we are able to easily observe what particular amino acid is changing its chemical environment (is moving) – its peak moves. Moreover in properly folded protein the peaks are well dispersed and most of peaks are visible, therefore from this type of experiment we are able to conclude if the protein is folded or randomly coiled. If it is not possible to measure a good HSQC spectrum, trying to solve the structure is not a reasonable work.

The process of protein structure solving using NMR spectroscopy is based on restraints generation (distances, angles, orientations) followed by structure calculations usually resulting in more or less converging set of structures.

3.3.2 NMR titrations

Another use of NMR spectroscopy in protein research is in ligand titration experiments based on the fast relaxation of bound ligand that behaves as it would be a part of the large protein molecule. Therefore just free ligand peaks are detectable in hydrogen spectra and the amount of free ligand can be computed after the peak integration (extrapolated to measured free ligand as quantitative standard).

3.4 Mass spectrometry and MS data handling

3.4.1 Protein identification

The classical approach to protein identification is called bottom-up and is based on analysis of protein digest, either enzymatic or chemical. Two strategies or the combination of them could be applied within this approach.

Microsequencing strategy requires tandem mass spectrometry (MS/MS) as the protein is identified on the basis of partial amino acid sequence retrieved from MS/MS data. It is useful to combine the digestion with other chemical methods such as N-terminus sulfonation or ^{18}O tryptic digestion as it facilitates the data processing.

Another strategy is based on analysis of specific enzymatic or chemical protein digests and on the unique set of digested peptide masses and is called peptide mass fingerprinting (PMF).

Both of these strategies would not be possible without databases of known proteins and without the software tools comparing measured experimental values with *in silico* generated theoretical ones. Therefore,

both of these strategies fail in case of unknown protein not included in the database. Although the software tools are freely accessible via Internet, their full potential could be utilized only if you operate your own instance of the software, therefore if you have the possibility to create your own databases, modifications and digest specificity.

Along with introduction of high resolution Fourier transform ion cyclotron resonance mass spectrometry machines a new approach to protein identification called top-down originated. It is based on analysis of intact protein without previous enzymatic or chemical cleavage and its considerable advantage is the possibility to analyze post-translational modification easily.

3.4.2 H/D exchange data handling

Hydrogen deuterium exchange is a chemical process in which a covalently bonded hydrogen atom is replaced by deuterium or vice versa. The examined protons are usually the backbone amides in a protein. The method provides information about the solvent accessibility of various parts of the molecule, and thus the tertiary structure of the protein [Zhang Z., Smith D. L. (1993)].

As the automatic searching for peptides and determination of its exact monoisotopic mass from mass spectra is based on natural abundance of particular isotopes (see Table 1) and/or virtual average amino acid averagine [Senko M. et. al. 1995, Valkenborg D. et. al. 2008] and subsequent calculated shape of isotopic envelope, embedded automatic peptide searching tools in common MS data processing software applications are not able to assign the envelope deformed and shifted by unnatural addition of deuteriums that it belongs to a peptide. Therefore it is necessary to develop either some independent tool or some kind of plug-in for the application to accomplish this task. In order to evaluate MS data from H/D experiments mentioned in this thesis the macro for Data Analysis application (Bruker Daltonics) written in Visual

Element	Isotope	Exact mass (Da)	Natural abundance (%)
H	1	1.0078250320	99.9885000000
	2	2.0141017780	0.0115000000
	3	3.0160492675	
C	12	12.0000000000	98.9300000000
	13	13.0033548378	1.0700000000
	14	14.0032419880	
O	16	15.9949146221	99.7570000000
	17	16.9991315000	0.0380000000
	18	17.9991604000	0.2050000000
N	14	14.0030740052	99.6320000000
	15	15.0001088984	0.3680000000
S	32	31.9720706900	94.9300000000
	33	32.9714585000	0.7600000000
	34	33.9678668300	4.2900000000
	36	35.9670808800	0.0200000000

Table 1: Natural abundances and exact masses of important isotopes found in proteins.

Mass (Da)	Relative abundance of isotope						Average composition				
	1	2	3	4	5	6	C	H	N	O	S
500	100	25.6	4.8	0.7	0.1	0	20	33	7	8	0
1000	100	53.6	17.3	4.1	0.8	0.1	43	73	11	16	0
1500	100	87.1	41.5	14.1	3.8	0.9	71	117	15	20	0
2000	90.3	100	64.1	30.1	11.3	3.6	89	141	21	29	1
2500	71.9	100	77.1	42.8	18.8	6.9	111	185	29	34	1
3000	63	100	95.5	67.5	38.6	18.8	124	205	35	43	4
3500	47.3	93.2	100	76.4	46.1	23.3	158	254	38	47	2
4000	36.5	83.2	100	83.8	54.8	29.6	184	278	44	54	1
4500	33.3	79	100	88.8	61.9	35.8	190	301	49	75	1
5000	24.7	68.7	99.6	100	77.9	50	222	342	60	70	1

Table 2: Isotopic envelope (relative abundances of isotopic peaks) for common peptide masses range [left] and average elemental composition of peptides of given mass [right].

Basic for Application (VBScript) was created. The algorithm of this macro is as follows:

- Compute the exact mass and number of maximum exchangeable backbone hydrogens

- Search for peaks at masses corresponding to mass of peptide with exchanged hydrogens and given charge considering the difference of $\Delta m(^{1}H\text{-}^{2}H)=1.0062767$ Da and $\Delta m(^{12}C\text{-}^{13}C)=1.0033548$ Da for eliminating random peptides of similar mass

- If found, find other isotopic peaks belonging to this peptide and calculate peptide's average mass (as monoisotopic mass does not exist in deuterated peptide) as weighted mean of masses with value contribution according to peak intensity

- Repeat for every given peptide from the beginning

The percentage deuterations are calculated from results of this macro (average masses of found peptides) related either to fully deuterated peptide mass (if measured) or to theoretically calculated value of maximal deuteration.

Convenient input peptides are carefully chosen from the results of enzymatic digestion of the protein with demands on sequence coverage, reasonable peptide length, retention time from HPLC and fair intensity in MS spectra.

4 Results

4.1 CD69 related results

4.1.1 Cooperation between Subunits Is Essential for High-Affinity Binding of N-Acetyl-D-hexosamines to Dimeric Soluble and Dimeric Cellular Forms of Human CD69

4060 *Biochemistry* **2010,** *49,* 4060–4067
DOI: 10.1021/bi100181a

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Cooperation between Subunits Is Essential for High-Affinity Binding of *N*-Acetyl-D-hexosamines to Dimeric Soluble and Dimeric Cellular Forms of Human CD69[†]

Daniel Kavan,[‡,§,⊥] Monika Kubíčková,[∥,⊥] Jan Bílý,[‡] Ondřej Vaněk,[‡,§] Kateřina Hofbauerová,[§] Hynek Mrázek,[‡,§] Daniel Rozbeský,[‡,§] Pavla Bojarová,[‡,§] Vladimír Křen,[§] Lukáš Žídek,[∥] Vladimír Sklenář,[∥] and Karel Bezouška*[,‡,§]

[‡]*Department of Biochemistry, Faculty of Science, Charles University, 12840 Prague, Czech Republic,* [§]*Institute of Microbiology v.v.i., Academy of Sciences of Czech Republic, 14220 Prague, Czech Republic, and* [∥]*National Centre for Biomolecular Research, Faculty of Science, Masaryk University, 61137 Brno, Czech Republic* [⊥]*These authors contributed equally to the experiments*

Received July 21, 2009; Revised Manuscript Received April 6, 2010

ABSTRACT: CD69 is an earliest lymphocyte activation antigen and a universal leukocyte triggering molecule expressed at sites of active immune response. The binding of GlcNAc to the dimeric human CD69 was followed by equilibrium dialysis, fluorescence titration, and NMR. Clear cooperation was observed in the high-affinity binding ($K_d = 4.0 \times 10^{-7}$ M) of the carbohydrate to two subunits of the dimeric CD69 (Hill coefficient 1.94). A control monosaccharide ManNAc was not bound by human CD69, and both monosaccharides had no effects on the structure of the receptor. However, a monomeric CD69 obtained by mutating Q93 and R134 at the dimer interface exhibited a much lower affinity for GlcNAc ($K_d = 1.3 \times 10^{-5}$ M) and no cooperativity (Hill coefficient 1.07). Perturbation of the dimer interface resulted in a severe impairment of the signaling ability of cellular CD69 when cross-linked with an antibody or with a bivalent high-affinity *N*-acetylhexosamine dimer-based ligand. The availability of stable preparations of soluble CD69 receptor with well-documented ligand binding properties will be beneficial for immunological experiments evaluating the role of this antigen in the complex environment of the immune system. Moreover, such preparations in combination with efficient ligand mimetics able to both activate $CD69^+$ lymphocytes and to block undesired hyperactivation caused by other cellular ligands will also become indispensable tools in explaining the exact role of the CD69 antigen in the interaction between the tumor cell and the effector natural killer lymphocyte.

CD69 is an early lymphocyte activation marker and a universal leukocyte triggering molecule expressed at sites of active immune response and chronic inflammation (*1*, *2*). Initial *in vitro* studies suggested that CD69 may function as an activating molecule in many leukocyte subsets including γ/δ T-cells and natural killer (NK)[1] cells (*3*, *4*). The *CD69* gene is located within the NK gene complex on human chromosome 12. It codes a type II calcium-dependent membrane lectin, a member of one important family of abundant NK cell surface receptors (*5*, *6*) participating in the formation of the receptor "zipper" at the NK cell−tumor cell interface (*7*). Most receptors of NK cells that recognize target structures at the surface of tumor or virally infected cells mediate their activation or inhibitory effects through their sequentially diverse cytoplasmic domains (*8*). However, the cytoplasmic domain of CD69 is short and lacks the prominent function-associated peptide motifs. Previous studies have shown that the downstream activation processes initiated by CD69 engagement occur through Src-dependent activation of Syk, activation of phospholipase Cγ2 and Vav, and the subsequent transmission of signals through the Rac−ERK pathway (*9*, *10*). Alternatively, CD69 can also propagate activation signals through heterotrimeric G proteins and the subsequent intracellular signaling pathways coupled to these molecular switches (*11*, *12*). Recently, *in vivo* studies in CD69-deficient mice have added yet another dimension into the biology of this receptor revealing its non-redundant role in downregulation of the immune response through the production of the pleiotropic cytokine TGF-β and through interactions with regulatory T-cells (*2*). Using the same experimental model, it has been shown recently that CD69 forms a complex with sphingosine 1-phosphate receptor, negatively regulates its function, and thus inhibits lymphocyte egress from lymphoid organs downstream of interferon-α/β, known mediators of transient egress shut down (*13*). Furthermore, results of many immunological studies indicate that CD69 may be involved in pathogenesis of several diseases including rheumatoid arthritis, chronic inflammatory liver diseases, mild asthma, and acquired immunodeficiency syndrome (*14*).

The identification of the natural ligand for CD69 is a key critical step for further advancement of our knowledge on the biology of this receptor. The initial findings that CD69 binds to calcium and certain *N*-acetyl-D-hexosamines (*15*) could not be later reproduced using a somewhat different expression construct (*16*). Since then, these discrepancies have been at least partially explained by careful structural evaluations of the recombinant proteins used for binding studies, as well as by

[†]Supported by grants from Ministry of Education of Czech Republic (MSM_21620808, 1M0505, and AVOZ50200510 to K.B. and MSM0021622413 and LC0603 to V.S.) and from Czech Science Foundation (303/09/0477 and 305/09/H008 to K.B. and 203/09/P024 to P.B.) and by EU Project Spine 2 (Contract LSHG-CT-2006-02/220 to K.B.).

*To whom correspondence should be addressed at Charles University. Phone: +420-2-2195-1272. Fax: +420-2-2195-1283. E-mail: bezouska@biomed.cas.cz.

[1]Abbreviations: DSS, disucinimidyl suberate; Gal, D-galactose; GlcNAc, *N*-acetyl-D-glucosamine; ManNAc, *N*-acetyl-D-mannosamine; MES buffer, 10 mM MES with 150 mM NaCl and 1 mM NaN_3; MES + C buffer, 10 mM MES, 150 mM NaCl, 1 mM $CaCl_2$, and 1 mM NaN_3; NK, natural killer.

 Published on Web 04/06/2010 © 2010 American Chemical Society

establishing a direct link between the binding of calcium and carbohydrates (*17*). The proper folding of CD69 produced based on the recently suggested constructs encompassing G70−K199 of the entire receptor was established using detailed structural experiments based on both NMR (*18*) and protein crystallography (*19*). The most recent development of efficient structural mimetics of the high-affinity ligand for CD69 opened the way for manipulating with numerous activities of CD69 at the molecular and cellular level (*20*, *21*) and provided efficient compounds for further *in vivo* testing of their immunomodulating properties (*22*−*24*).

Here we report new findings indicating that the binding of *N*-acetyl-D-hexosamines to soluble CD69 is highly cooperative at molecular level, and this cooperativity is not seen for Q93A and R134A mutants with disturbed formation of noncovalent dimers. Similarly at the cellular level, efficient signaling after CD69 cross-linking by antibody or bivalent ligand is diminished for the above mutants with a damaged subunit cross-talk more dramatically than for CD69 bearing C68A mutation, and thus lacking the disulfide bridge forming the covalent dimer identified previously as the critical signaling element.

EXPERIMENTAL PROCEDURES

Materials. All chemicals were analytical grade reagents of the best quality available commercially and were obtained from Sigma unless indicated otherwise. The preparation of dimeric *N*-acetylhexosamine disaccharide with the chemical composition (GalNAcβ1−4GlcNAcβ-NH-CS-NH-CH_2-)$_2$ by a combination of chemical and enzymatic steps has been described previously (*25*).

Preparation of Soluble Dimeric CD69. Preparation of soluble dimeric CD69 using protocol II has been described previously (*18*). For the preparation of a uniformly ^{15}N-labeled form of the receptor, a producing culture of *Escherichia coli* BL-21 Gold (Stratagene) harboring the expression plasmid was used, grown on a standard M9 minimal medium containing $^{15}NH_4Cl$.

Identification of Key Amino Acid Residues Disrupting the Receptor Dimer Interface. We examined the three-dimensional structure of the crystallized soluble dimeric CD69 deposited into the RCSB Protein Databank under accession code 3CCK (*18*). Glutamine Q93 appeared to be involved in two key hydrogen bonds between the amide group of its side chain and two adjacent acidic residues belonging to the opposite subunit, Asp88 and Glu87. With R134, the intertwining with the second subunit is even more profound, and the guanidyl group of this amino acid forms three hydrogen bonds with A136 and Y135 of the opposing subunit and is also involved in the stacking interaction with the phenyl ring of Y135.

Site-Directed Mutagenesis and Expression of the Mutated CD69 Proteins. Mutated forms of CD69 in which the Q93 and/or and R135 were mutated into A were produced using the CD69 expression plasmid in pRSETB (*18*). Mutations were introduced using the QuickChange site-directed mutagenesis kit (Stratagene) in combination with the following oligonucleotide pairs: CD69Q93F, 5′-GAGGACTGGGTTGGCTACGCGAG-GAAATGCTACTTTATT-3′, and CD69Q93R, 5′-AATAAA-GTAGCATTTCCTCGCGTAGCCAACCCAGTCCTC-3′, and CD69R134F, 5′-GACATGAACTTTCTAAAAGCATACGC-AGGTAGAGAGGAA-3′ and CD69R134R, 5′-TTCCTCTCT-ACCTGCGTATGCTTTTAGAAAGTTCATGTC. The introduced mutations were verified by DNA sequencing using an ABI Prism 3130 genetic analyzer (Applied Biosystems). The CD69Q93A/R134A double mutant was prepared sequentially, applying the R134A mutation process on the Q93A mutant. Mutated CD69 proteins were prepared using the same protocol (protocol II) used for the production of the wild-type protein (*18*). Moreover, the proper refolding of the protein was verified using NMR measurement with the homogenously ^{15}N-labeled proteins as described previously (*18*).

Gel Filtration. Gel filtration was performed using a Superdex 200 HR 10/30 column (GE Healthcare) connected to the protein purification system BioSys510 (Beckman Coulter) and equilibrated with MES buffer at room temperature. In order to examine the effect of monosaccharide binding to CD69 on the hydrodynamic volume of the CD69 protein, the protein samples were incubated overnight at 4 °C in the presence of 1 mM ManNAc or 1 mM GlcNAc and then injected onto the gel filtration column equilibrated in MES + C buffer containing 1 mM concentrations of the respective monosaccharides.

Analytical Ultracentrifugation. Sedimentation velocity and sedimentation equilibrium experiments were performed using a ProteomeLab XL-I analytical ultracentrifuge (Beckman Coulter) equipped with an An50Ti rotor and dual absorbance and laser interference optics. Before the experiment, 0.5 mL samples of CD69 proteins diluted to 0.4 mg·mL^{-1} were dialyzed for 20 h against 2 L of MES + C buffer, with or without the monosaccharides, in concentration indicated in the text, and the dialysis buffer was used as a reference and sample dilution buffer. The sedimentation velocity experiment was conducted at 48000 rpm for dimeric CD69 at 20 °C. Data were analyzed with the program SEDFIT (*26*, *27*). Based on buffer composition and amino acid sequence using the program SEDNTERP (www.jphilo.mailway.com), buffer density and CD69 partial specific volume for CD69NG70 were estimated as 1.00309 g·mL^{-1} and 0.7183 mL·g^{-1}, respectively.

Protein Stability Experiments. CD69 proteins were diluted to 0.5 mg/mL, and UV spectra were taken in the 200−300 nm range in a Beckman DU-70 spectrophotometer (Beckman Coulter) equipped with a heated cuvette. The initial UV scan was taken at 25 °C, after which the temperature in the cuvette was increased in 5 °C increments. Experiments were performed routinely in MES + C buffer. Alternatively, protein stability was verified using differential scanning calorimetry and FTIR spectroscopy as has been previously described (*17*, *18*, *28*).

NMR Titrations. All NMR experiments were run at 300 K in a Bruker Avance 600 MHz spectrometer equipped with a cryogenic H/C/N TCI probehead. ^{1}H−^{15}N HSQC spectra of 0.3 mM ^{15}N-labeled wild-type CD69 protein CD69NG70 (*18*) were used as a routine check of protein folding and stability. The sample buffer consisted of 10 mM MES, pH 5.8, with 49 mM NaCl, 1 mM NaN_3, and 10% D_2O. During NMR titration, a 0.1 mM solution of the unlabeled wild-type CD69 protein (*7*) was titrated. In an initial experiment, aliquots of the GlcNAc ligand corresponding to 25%, 50%, 75%, 100%, 200%, and 500% of saturation were added, and signals of the free GlcNAc ligand were observed at 2.2 ppm in the 1D proton spectra and used for the estimation of the free ligand concentration. In a separate experiment aimed at estimating the binding constant, smaller ligand additions were used as the equivalence was appoached. The protein was titrated to 75% of the estimated number of binding sites, after which the amount of ligand was increased in increments of 5% of the estimated number of binding sites until the equivalence point was reached. All spectra were processed

using the software NMRPIPE (*29*). The dissociation constant K_d, defined as $K_d = (c_p - c_L + [L])[L]/(c_L - [L])$, was obtained by a nonlinear fitting of the [L] vs c_L titration curves (Figure 2A,B).

Equilibrium Dialysis. *N*-Acetyl-D-[1-^{3}H]glucosamine (specific activity 500 GBq/mmol) and *N*-acetyl-D-[1-^{3}H]mannosamine (specific activity 650 GBq/mmol) were prepared as described previously (*17*) or purchased from Amersham. To set up equilibrium dialysis experiments, a rotating apparatus with glass blocks containing separate sealable chambers with external access was used as described previously (*17*). Aliquots (200 μL) of 0.1 μM solutions of CD69 proteins in MES + C buffer were incubated with varying amounts of ligand at 27 °C (300 K) for 48 h. After equilibration, 100 μL aliquots were withdrawn from the control and from the protein-containing chambers. The results were calculated and plotted according to Scatchard as described previously (*17*).

Tryptophan Fluorescence Quenching. Tryptophan fluorescence quenching experiments were performed according to the described methodology (*30*) with minor modifications. In initial experiments, 100 nmol aliquots of CD69 protein were pipetted into multiple wells of a UV Star plate (Greiner, Germany) and mixed with 10-fold serial dilutions of the GlcNAc ligand. Incubation proceeded for 1 h at 27 °C (300 K) in the thermostated chamber of a Safire2 plate reader (Tecan, Austria), after which the fluorescence of tryptophan residues was measured in duplicate wells using the bottom fluorescence measurements and the following settings: λ_{ex} = 275 nm, λ_{em} = 350 nm, excitation and emission slits were set to 5 and 20, respectively, and the fluorescence gain was manually set to 66. After finding the lowest concentration of ligand that still caused the quenching of tryptophan fluorescence, detailed dilutions of the ligand by 10% saturation steps were performed, and the concentration of free and bound ligand was calculated as described previously (*17*, *30*).

Preparation of the Eukaryotic Expression Constructs, Transfection into Jurkat Cells, and Selection of the Transfectants. In order to mutate the dimerization cystein C68 (*15*) to A, site-directed mutagenesis was performed using the original expression plasmid (*15*) as described above using oligonucleotide primers CD69C68F, 5′-TCAGTGGGCCAATACAATGCTCCAGGCCAATACACATTC-3′, and CD69C68R, 5′-GAATGTGTATTGGCCTGGAGCATTGTATTGGCCCACTGA-3′, and the pCDA401 plasmid (*15*). Single mutation CD69C68A, double mutations CD69C68A/Q93A and CD69C68A/R134A, and the triple mutation CD69C68A/Q93A/R134A were prepared by applying the mutagenesis protocol onto expression plasmids for wild-type CD69 (*8*) and for the respective dimerization mutants described above. After the mutagenesis and DNA sequencing, DNA fragments coding the C-terminal extracellular segments of CD69 were linked with the DNA fragment coding the N-terminal part of the receptor (*31*) using linking PCR (*8*). The X construct corresponded to the religated pCR3 (mock) and was used as a control (*8*). The eukaryotic expression vectors were sequenced and transfected into a Jurkat T lymphoblastoid cell line maintained in RPMI1640 and supplemented with 10% fetal calf serum (*8*).

Precipitation of Cellular Forms of CD69 Using Antibodies and Dimeric N-Acetylhexosamines. Transfected Jurkat cells (1×10^6) were surface radioiodinated using lactoperoxidase (*31*), washed three times with medium, and then incubated with 1 mM concentrations of dimeric *N*-acetylhexosamine disaccharides for 1 h at room temperature. The incubation was followed by the addition of 100 mM DSS and by cross-linking of the receptors for another 1 h at 4 °C. Thereafter, cells were lysed, and CD69 receptor complexes were immunoprecipitated using G protein beads (GE Healthcare) coated with monoclonal antibodies against CD69, BL-KFB/B1 (*31*). Beads were washed extensively, boiled in sample buffer for SDS−PAGE, and analyzed using 15% SDS−polyacrylamide gels followed by autoradiography.

Cellular Activation Assays and Production of IL-2. Transfected Jurkat cells (10^6) were incubated with dimeric *N*-acetylhexosamine disaccharides as described in the preceding section or with saturating concentrations of monoclonal antibodies against CD69 for 5 min (cellular activation) or 12 h (IL-2 production) and used to determine the free cytoplasmic calcium (*11*) or IL-2 production (*8*).

RESULTS AND DISCUSSION

Evaluation of Calcium and Carbohydrate Binding Activity of Highly Stable CD69 Proteins. We and others have previously generated several constructs optimized for the preparation of highly stable soluble recombinant CD69 proteins suitable for ligand identification experiments (Supporting Information Table S1). Preliminary ligand binding experiments were performed to evaluate the ability of these constructs to bind calcium and monosaccharides shown to be important ligands for the receptor (*15*, *17*). With regard to the binding of calcium, there has been no difference in the ability to bind calcium between the covalent dimeric protein CD69CQ65 and noncovalent dimeric proteins CD69NG70 and CD69NV82 when compared to the monomeric protein CD69MS100: each of these proteins bound 1 mol of calcium/mol of CD69 subunit with K_d of approximately 58 μM (ref *17* and Supporting Information Figure S1). On the other hand, significant differences between these protein constructs were observed with regard to the binding of *N*-acetyl-D-hexosamines. While the IC_{50} values for the soluble monomeric CD69, CD69MS100, with regard to binding of the two active *N*-acetyl-D-hexosamines, D-GlcNAc and D-GalNAc, were each approximately 10^{-5} M, these values were about 10 times lower for the dimeric protein CD69NV82 and about 100 times lower for the other two highly stable dimeric proteins, CD69CQ65 and CD69NG70 (Supporting Information Figure S2). The latter protein has been selected for all of the subsequent binding experiments and will be referred to as soluble dimeric CD69. The homogeneity and monodispersity were routinely evaluated for each batch of the produced soluble dimeric CD69 using SDS electrophoresis under both reducing and nonreducing conditions and gel filtration on a Superdex 200 HR column (ref *18* and Figure 1). Moreover, the identity, quality, and proper refolding of each batch of the produced protein were also verified as described previously (*18*) using high-resolution ion cyclotron resonance mass spectrometry, one-dimensional proton NMR, thermal stability experiments, and tests of the biochemical stability (ref *18* and Table 1).

Cooperativity of GlcNAc Binding Proved by Direct Binding Experiments. The detailed binding studies with soluble dimeric CD69 were performed using D-GlcNAc as the high-affinity carbohydrate ligand, together with D-ManNAc and, in some experiments, D-Gal as negative controls. The initial evidence for the interaction of the soluble dimeric CD69 with GlcNAc was obtained by NMR titration. A 0.1 mM solution of the dimeric receptor was titrated up to equivalence assuming the existence of two high-affinity binding sites per receptor

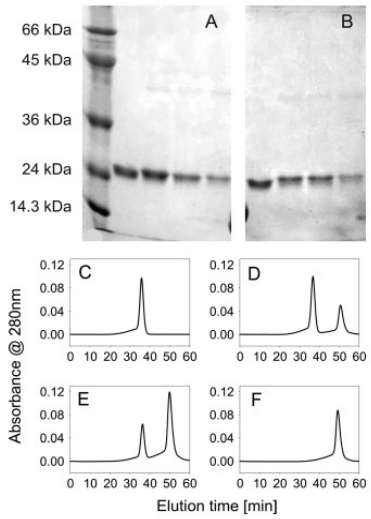


FIGURE 1: Analysis of wild-type and mutant CD69 proteins by SDS−PAGE and gel filtration. In (A) and (B), these proteins were analyzed under reducing and nonreducing conditions, respectively. Analyzed proteins, from left to right, were wild-type CD69, CD69Q93A mutant, CD69R134A mutant, and CD69QRDM. Marker proteins were BSA (65 kDa), ovalbumin (44 kDa), lactoglobulin (18 kDa), lysozyme (14 kDa), and aprotinin (6 kDa). In (C) to (F), these proteins were analyzed by gel filtration on a Superdex 200 HR column, and the four respective panels contain chromatograms for wild-type CD69, CD69Q93A mutant, CD69R134A mutant, and CD69QRDM double mutant.

Table 1: Summary of Stability Properties of Wild-Type Dimeric CD69 and CD69 Dimerization Mutants

protein	characteristics	T_d^a (°C)	T_d^b (°C)	T_d^c (°C)
CD6CD69WT	noncovalent dimers	65	67	65
CD69Q93A	dimer/monomer equilibrium	63	62	64
CD69R134A	dimer/monomer equilibrium	62	60	60
CD69QRDM	monomeric	60	62	61

[a]Determined from thermal UV denaturation measurements. [b]Determined from differential scanning calorimetry. [c]Determined from FTIR spectroscopy.

dimer (*17*). The results of this experiment (Figure 2B and Supporting Information Figure S3B) confirmed the specific binding of GlcNAc to the receptor (2 mol of GlcNAc bound to a receptor dimer) and provided an affinity estimation in the low micromolar range ($K_d = 4.0 \times 10^{-7}$ M). On the other hand, no interaction could be seen with the ManNAc negative control under the same experimental conditions (Figure 2A and Supporting Information Figure S3A). However, NMR did not enable the fraction of the bound ligand to be measured.

In order to confirm the results obtained by NMR titration, additional direct binding experiments were performed. When the bound and the unbound ligands had been separated by dialysis under equilibrium, two binding sites per receptor dimer were detected. Direct binding experiments enabled the degree of saturation at each particular ligand concentration to be calculated. The resulting saturation curve, showing the saturation (fraction bound normalized per receptor subunit) dependence on the ligand concentration (Figure 2C), clearly revealed a striking cooperativity in the highly specific ($K_d = 4.0 \times 10^{-7}$ M) binding

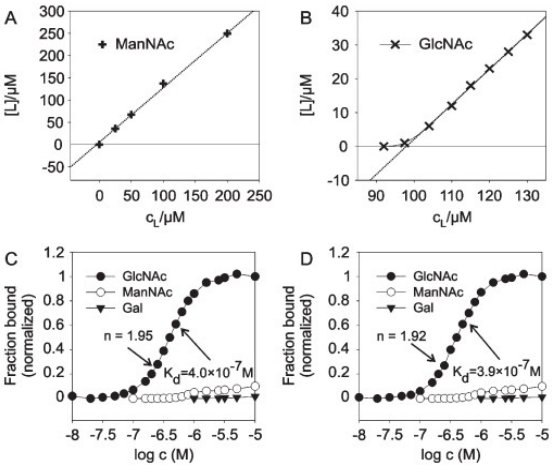


FIGURE 2: Measurements of direct interaction of soluble CD69NG70 with ManNAc and GlcNAc. (A, B) NMR titration of soluble CD69 with ManNAc and GlcNAc, respectively. (C, D) Concentration dependence of receptor saturation measured by equilibrium dialysis and tryptophan fluorescence quenching, respectively, using GlcNAc, ManNAc, and Gal as indicated.

of GlcNAc to the receptor with the Hill coefficient approaching the maximum theoretical value (theory 2.00, experiment 1.95; see Figure 2C). These results were also independently confirmed by the third binding assay, the fluorescent titration, which gave binding parameters essentially identical to those obtained by the equilibrium dialysis (Figure 2D). On the other hand, very little specific binding for both ManNAc and Gal control monosaccharides could be seen in both of the latter assays (Figure 2C,D).

Binding of N-Acetyl-D-hexosamines Did Not Result in Significant Conformational Change in CD69 Protein. To analyze the structural changes of soluble CD69 upon ligand binding, variations in the hydrodynamic properties of the receptor were investigated. The molecular size of the receptor, which had been saturated with an excess of GlcNAc, was studied by gel filtration on Superdex 200 HR and by analytical ultracentrifugation and compared with the size of the receptor preincubated in the ManNAc control. The elution time decrease from 36.5 to 31.7 min in the gel filtration would indicate a change in the molecular size of CD69 upon GlcNAc binding when compared to the presence of ManNAc (Supporting Information Figure S4). However, the detailed analysis of soluble dimeric CD69 in the absence of any ligand, in the presence of 1 mM ManNAc, and in the presence of 1 mM GlcNAc did not reveal any changes in hydrodynamic properties since the value of the experimentally determined sedimentation coefficient was identical (Supporting Information Figure S5).

Binding of N-Acetyl-D-glucosamine to the Stable Monomeric CD69 Follows a Single Site Model and Proceeds with Much Lower Affinity. In the next step, the interactions of GlcNAc with the monomeric subunit of CD69 were studied. Since it proved extremely difficult to prepare the monomeric form of the receptor by dissociation of the CD69 dimer (*18*), we used the available crystal structure of the CD69 dimer and analyzed the dimer interface for critical residues participating in the dimerization. Two such critical residues, namely Q93 and R134, both interacting with two residues of the other subunit, could be identified (Supporting Information Figure S6).

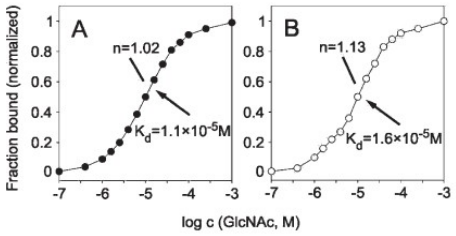


FIGURE 3: Binding of GlcNAc to the monomeric subunit of CD69. (A, B) Binding of GlcNAc to monomeric CD69 analyzed using equilibrium dialysis and fluorescence titration, respectively.

These amino acid residues were mutated to alanine, singly or in combination. All three produced mutated proteins were after their refolding and purification extensively verified using the methodology described previously (*18*) for the wild-type protein CD69NG70 using SDS electrophoresis and ion cyclotron resonance mass spectrometry of the entire protein, as well as nuclear magnetic resonance to check the proper folding of these proteins (Figure 1 and results not shown). The stability of the mutant soluble CD69 proteins was comparable to that of the wild-type protein, indicating that the introduced mutations did not result in any decrease of protein stability. Only the double mutant behaved as a monomeric protein (Figure 1), with stability comparable to that of the dimer receptor (Table 1). This protein was used to analyze the binding of GlcNAc to the monomeric protein. The results clearly showed that binding to the monomeric subunit of CD69 was much weaker and noncooperative (Hill coefficient of 1.07; Figure 3).

Q93/R134/C68 Triple Mutation Is Necessary to Disrupt the Dimerization of the Cellular Form of CD69. The soluble CD69 receptor used in the first part of this study utilized a previously described construct CD69NG70 (*18*) that refolded as a noncovalent soluble dimer from a polypeptide consisting of amino acids G70−K199 (Supporting Information Table S1). This construct contained an extended dimer interface involved in contacts between the ligand binding domains, as well as the neck regions. However, it did not contain the C68 residue that has been shown (*8*, *15*) to participate in the covalent dimerization of the natural form of CD69 found at the surface of leukocytes. It thus appeared interesting to look into the effects of mutations of the critical residues C68, Q93, and R134 at the CD69 dimer interface on the structure and on the well-documented signaling functions of the cellular form of CD69. In order to trigger the CD69-mediated activation of transfected Jurkat cells bearing both wild-type and mutated forms of CD69, an efficient ligand is required for receptor cross-linking. Alternatively, the receptor can be aggregated using specific antibodies. In the experiments presented here, we used both forms of activation using two specific cross-linking monoclonal antibodies against CD69, BL-KFB/B1 and BL-Ac/p26, as well as the *N*-acetylhexosamine disaccharide dimer (Figure 4B), which has been previously described (*25*) as the most efficient carbohydrate ligand at enabling precipitation (and thus cross-linking) of the soluble CD69. The structurally closely related ligand, *N*-acetyllactosamine dimer (Figure 4A), served as a suitable negative control in these experiments.

We transferred inserts coding the CD69 receptors into a eukaryotic expression vector for the transfection of the Jurkat T-lymphocyte leukemic cell line (refs *9* and *10* and Table 2). Those clones displaying identical surface expression of the wild-type and mutated receptors (as shown by flow cytometry) were

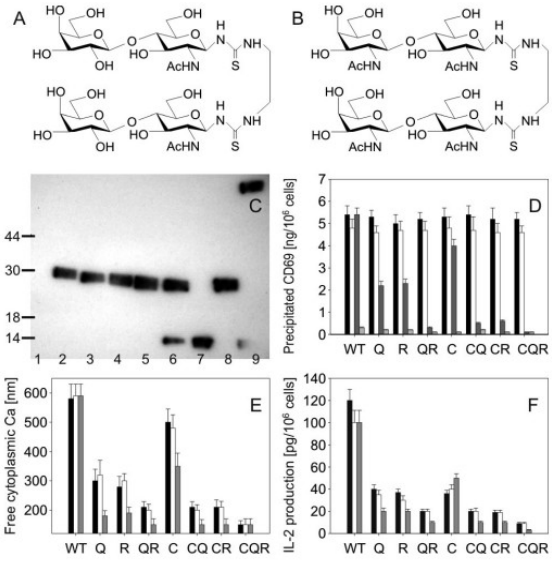


FIGURE 4: Analysis of cellular CD69 on transfected cells and evaluation of binding and signaling efficiency of receptor disturbed at the dimer interface. (A, B) Chemical formulas of *N*-acetyllactosamine and *N*-acetylhexosamine disaccharide dimer, respectively. (C) Surface expression of CD69 immunoprecipitated from control non-transfected cells (lane 1) or from cells containing wild-type CD69 (lane 2), CD69Q93A (lane 3), CD69R134A (lane 4), CD69C68A/Q93A (lane 5), C68A/R134A (lane 6), C68A/Q93A/R134A (lane 7), C68A/Q93A/R134A cross-linked with *N*-acetylhexosamine disaccharide dimer (*8*) (lane 8), and wild-type CD69 cross-linked with *N*-acetylhexosamine disaccharide dimer (lane 9). Receptors were surface-labeled using lactoperoxidase, dimerized using DSS cross-linking, immunoprecipitated, and analyzed by SDS electrophoresis under reducing conditions followed by autoradiography. (D) Precipitation of various forms of wild-type and mutated CD69 using *N*-acetyllactosamine or *N*-acetylhexosamine disaccharide dimer and antibodies against CD69. (E) Cellular activation of Jurkat leukemic T-cells transfected with wild-type or mutated CD69 with mutations at the dimer interface. (F) Production of IL-2 by Jurkat leukemic T-cells transfected with wild-type or mutated CD69 with mutations at the dimer interface. The designation of mutants in (D) to (F) is based on the mutated amino acids and their combinations: C, C68; Q, Q93; R, R134. The cellular CD69 in (D) to (F) was cross-linked using monoclonal antibody BL-KFB/B1 (*2*) (left columns), monoclonal antibody BL-Ac/p26 (*2*) (middle columns), or *N*-acetylhexosamine disaccharide dimer (right columns). The fourth, small column in (D) placed after the three above-mentioned columns indicates precipitation with *N*-acetyllactosamine control.

subcloned, frozen, and used in individual experiments (Figure 5). First, the molecular forms of the surface CD69 were analyzed in both native and ligand cross-linked receptors. In order to determine the precise molecular forms of the receptors in these experiments, native receptors were fixed by di(*N*-hydroxysuccinimido) suberate (DSS), shown previously to produce covalent dimers with soluble CD69 (I. Polakovičová, unpublished observation), before their extraction from the plasma membrane and molecular analysis. Cellular CD69 bearing single mutations in amino acids responsible for the covalent (C68) or noncovalent (Q93, R134) dimerization remained mostly dimeric, as did the double mutants C68A/Q93A, C68A/R134A (Figure 4C, lanes 2−6), and Q93/R134 (not shown). Only the triple mutant C68A/Q93A/R134A was completely monomeric (Figure 4C, lane 7). However, this monomeric protein could still be efficiently dimerized by the *N*-acetylhexosamine disaccharide dimer (lane 8), which caused

Table 2: Characterization of Individual Clones of Jurkat T-Cell Leukemia Transfected with CD69 Isoforms[a]

CD69 isoform	clone	general characteristic	CD69 expression by FACS[b]	used for expts
CD69WT	WT1	good growth	10	no
	WT2	**moderate growth**	**30**	**yes**
	WT3	moderate growth	80	no
CD69Q93A	Q1	good growth	20	no
	Q2	**good growth**	**30**	**yes**
	Q3	moderate growth	70	no
CD69R134A	R1	moderate growth	30	yes
	R2	**good growth**	**90**	**no**
CD69QRDM	RQ1	moderate growth	10	no
	RQ2	**good growth**	**30**	**yes**
	RQ3	good growth	70	no
CD69C68A	**C1**	**good growth**	**30**	**yes**
	C2	good growth	60	no
CD69CQDM	CQ1	good growth	10	no
	CQ2	**moderate growth**	**30**	**yes**
	CQ3	moderate growth	80	no
CD69CRDM	CR1	moderate growth	10	no
	CR2	**moderate growth**	**30**	**yes**
	CR3	slow growth	90	no
CD69CQRTM	CQR1	moderate growth	10	no
	CQR2	good growth	20	no
	CQR3	**moderate growth**	**30**	**yes**
	CQR4	slow growth	60	no

[a]Only the successful transformants with relative fluorescence intensity of CD69 expression between 10 and 100 have been included into this table. [b]Median of relative fluorescence intensity from FACS analyses at the top of the peak.

extensive cross-linking and formation of high molecular weight aggregates in wild-type cellular CD69 (lane 9).

Mutations in Q93 and R134 of Cellular CD69 Result in More Significant Impairment in Its Signaling than Mutation in C68. We further investigated the effect of mutations destabilizing the dimerization of cellular CD69 on the cross-linking of this receptor and its function in cellular activation. While all of the expressed cellular CD69 could be efficiently cross-linked by two monoclonal antibodies against this antigen, the cross-linking by the dimerized ligand was severely impaired in the mutants affecting receptor dimerization even if the dimerization process *per se* was not affected (cf. Figure 4C,D). Interestingly, mutations in the amino acids forming the noncovalent dimer interface had in several instances a more profound effect than the mutation in C68 responsible for covalent dimerization. On the other hand, no precipitation of the receptor occurred when the *N*-acetyllactosamine dimer control compound was used (Figure 4D).

Subsequently, the influence of the mutations affecting the dimerization of cellular CD69 on the ability of this receptor to activate the Jurkat cell line, as documented by the increase in the intracellular calcium levels, was analyzed. Strikingly, even if the cellular CD69 antigens bearing the above mutations could be fully cross-linked by binding to monoclonal antibodies, their ability to confer activation was severely affected (Figure 4E). For instance, single mutations in amino acids at the noncovalent dimer interface (i.e., Q93A or R134A mutations) lowered the efficiency of these receptors in cellular signaling by two-thirds. On the other hand, when using ligand cross-linking with the *N*-acetylhexosamine disaccharide dimer, a very low efficiency of cellular signaling was observed compared to the wild-type

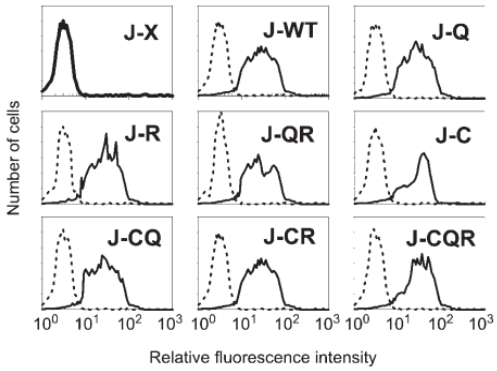


FIGURE 5: Analysis of surface expression of cellular CD69 in transfected Jurkat cell clones as determined by flow cytometry: J-X, insert-free (mock DNA) transfectant; J-WT, cells transfected with wild-type CD69; J-Q, cells transfected with CD69Q93A mutant; J-R, cells transfected with CD69R134A mutant; J-QR, cells transfected with CD69Q93A/R134A double mutant; J-C, cells transfected with CD69C68A mutant; J-CQ, cells transfected with CD69C68A/Q93A double mutant; J-CR, cells transfected with CD69C68A/R134A double mutant; J-CQR, cells transfected with CD69C68A/Q93A/R134A triple mutant. Reactivity with the monoclonal antibody against CD69 (BL-KFB/B1) is shown by the solid line; reactivity with the isotype-matched control antibody against NKR-P1 (HD14) is shown by the dotted line.

controls (Figure 4E). This may be explained by a combination of low efficiency of ligand binding and receptor cross-linking together with a direct effect on signaling efficiency. Similar results were obtained with the production of IL-2 as another measure of cellular activation (Figure 4F).

CONCLUSIONS

The results presented here indicate that the binding of carbohydrate ligands to both the soluble and cellular forms of CD69 is cooperative and is affected by the interaction of the two subunits of this receptor. This is a hitherto unreported observation, since the signaling through CD69 has been so far ascribed exclusively to its association with the signaling adapter proteins in the transmembrane region of the molecule (*8*). Thus, while the previous studies have defined the role of the individual polypeptide segments of CD69 in cellular expression, surface dimerization, and cellular signaling, the present results are unique in emphasizing the role of the intact dimer interface in the receptor molecule that was found to be much larger than the mere cysteine 68 thought previously (*8*, *15*) to be critical for receptor dimerization. Moreover, our findings suggest a novel mechanism for sensitive ligand recognition by CD69 and related immune receptors. Within the C-type lectin family, various alternative strategies have been employed to attain high-affinity binding. Classical hepatic lectins are oligomeric proteins in which the individual carbohydrate-recognition domains within the oligomer cooperate during the recognition of desialylated glycoproteins (*32*). Other members of this protein family employ alternative calcium-dependent processes to achieve this goal (*33*). Among the receptors of the immune system, the soluble mannose-binding protein again employs a multiple carbohydrate-recognition domain to bind specifically to surface polysaccharides of various pathogens and activate their effector functions such as opsonization or complement activation (*34*, *35*). Compared to these lectins, lymphocyte receptors forming group V of

the C-type lectin family are much smaller and mostly dimeric, and thus they seem to have developed alternative strategies for their recognition of specific ligands. These are based either on oligomerization of the receptor within the specialized plasma membrane microdomains of the immune cell (*36*) or on alternative strategies that may be used by some of these receptors (*37*). The molecular mechanism that we propose based on the results of the current work is unique in rapid propagation of the activation signal using the mechanism of rapid cooperative receptor cross-linking and oligomerization based on the positive cooperativity in binding of the multivalent ligands.

ACKNOWLEDGMENT

We thank Angela Risso for assistance and helpful discussions, Iva Polakovičková for help with DSS cross-linking experiments, and Michal Navrátil for help in the development of cellular receptor precipitation assays.

SUPPORTING INFORMATION AVAILABLE

A description of additional experimental procedures with supporting references, binding of calcium to the fourth generation soluble CD69 proteins (Figure S1), carbohydrate inhibition experiments for these proteins (Figure S2), primary data for NMR titrations with ManNAc and GlcNAc (Figure S3), elution profiles for gel filtration of soluble dimeric CD69 (CD69NG70) in the absence of ligand as well as in the presence of 1 mM ManNAc and 1 mM GlcNAc (Figure S4), sedimentation velocity experiments with CD69NG70 in the absence of ligand and in the presence of 1 mM ManNAc and 1 mM GlcNAc (Figure S5), identification of amino acids for the design of the monomeric CD69 mutant (Figure S6), and characterization of CD69 expression constructs (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Testi, R., D'Ambrosio, D., De Maria, R., and Santoni, A. (1994) The CD69 receptor: a multipurpose cell surface trigger for hematopoietic cells. *Immunol. Today 15*, 479–483.
2. Sancho, D., Goméz, M., and Sánchez-Madrid, F. (2005) CD69 is an immunoregulatory molecule induced following activation. *Trends Immunol. 26*, 136–140.
3. Moretta, A., Poggi, A., Pende, D., Tripodi, G., Orengo, A. M., Pella, N., Augugliaro, R., Bortino, C., Ciccone, E., and Moretta, L. (1991) CD69-mediated pathway of lymphocyte activation: anti-CD69 monoclonal antibodies trigger the cytolytic activity of different lymphoid effector cells with the exception of cytolytic T lymphocytes expressing T cell receptor α/β. *J. Exp. Med. 174*, 1393–1398.
4. Borrego, F., Robertson, M. J., Ritz, J., Pena, J., and Solana, R. (1999) CD69 is a stimulatory receptor for natural killer cells and its cytotoxic effect is blocked by CD94 inhibitory receptor. *Immunology 97*, 159–165.
5. Lopez-Cabrera, M., Santis, A. G., Fernandez-Ruiz, F., Blacher, R., Esch, F., Sanchez-Máteos, P., and Sánchez-Madrid, F. (1993) Molecular cloning, expression, and chromosomal localization of the human earlies activation antigen AIM/CD69, a new member of C-type lectin superfamily of signal-transmitting receptors. *J. Exp. Med. 178*, 537–547.
6. Lanier, L. L. (2008) Up on the tightrope: natural killer cell activation and inhibition. *Nat. Immunol. 9*, 495–502.
7. Vivier, E., Tomasello, E., Baratin, M., Walzer, T., and Ugolini, S. (2008) Functions of natural killer cells. *Nat. Immunol. 9*, 503–510.
8. Sancho, D., Santis, A. G., Alonso-Lebrero, J. L., Viedma, F., Tejedor, R., and Sánchez-Madrid, F. (2000) Functional analysis of ligand-binding and signal transduction domains of CD69 and CD23 C-type lectin leukocyte receptors. *J. Immunol. 165*, 3868–3875.
9. Pisegna, S., Zignoni, A., Pirozzi, G., Cinque, B., Cifoni, M. G., Morrone, S., Picolli, M., Frati, L., Palmieri, G., and Santoni, A. (2002) Src-dependent Syk activation controls CD69 mediated signaling and function of human NK cells. *J. Immunol. 169*, 68–74.
10. Zingoni, A., Palmieri, G., Morrone, S., Carretero, M., Lopez-Botet, M., Picolli, M., Frati, L., and Santoni, A. (2000) CD69-triggered ERK activation and functions are negatively regulated by CD94/NKG2A inhibitory receptors. *Eur. J. Immunol. 30*, 644–651.
11. Risso, A., Smilowich, D., Capra, M. C., Baldissarro, I., Yan, G., Bargilessi, A., and Cosulich, M. E. (1991) CD69 in resting and activated T lymphocytes. Its association with a GTP binding protein and biochemical requirements for its expression. *J. Immunol. 146*, 4105–4114.
12. Bikah, G., Pogue-Caley, R. R., and McHeyzer-Williams, M. G. (2000) Regulating T helper cell immunity through antigen responsiveness and calcium entry. *Nat. Immunol. 1*, 402–412.
13. Shiow, R. L., Rosen, D. B., Brdečková, N., Xu, Y., An, J., Langer, L. L., Cyster, J. G., and Matloubian, M. (2006) CD69 acts downstream of interferon-α/β to inhibit $S1P_1$ and lymphocyte egress from lymphoid organs. *Nature 440*, 540–544.
14. Marzio, R., Mauel, J., and Betz-Corradin, S. (1999) CD69 and regulation of the immune functions. *Immunopharmacol. Immunotoxicol. 21*, 565–582.
15. Bezouška, K., Nepovím, A., Horváth, O., Pospíšil, M., Hamann, J., and Feizi, T. (1995) CD69 antigen of human leukocyte is a calcium-dependent carbohydrate-binding protein. *Biochem. Biophys. Res. Commun. 208*, 68–74.
16. Childs, R. A., Galustian, C., Lawson, A. M., Doufán, G., Benwell, K., Frankel, G., and Feizi, T. (2000) Recombinant soluble human CD69 dimer produced in *Escherichia coli*: reevaluation of saccharide binding. *Biochem. Biophys. Res. Commun. 266*, 19–23.
17. Pavlíček, J., Sopko, B., Ettrich, R., Kopecký, V., Baumruk, V., Man, P., Havlíček, V., Vrbacký, M., Martínková, L., Křen, V., Pospíšil, M., and Bezouška, K. (2003) Molecular characterization of binding of calcium and carbohydrates by an early activation antigen of lymphocytes CD69. *Biochemistry 42*, 9295–9306.
18. Vaněk, O., Nálezková, M., Kavan, D., Borovičková, I., Pompach, P., Novák, P., Kumar, V., Vannucci, L., Hudeček, J., Hofbauerová, K., Kopecký, V., Brynda, J., Kolenko, P., Dohnálek, J., Kadeřávek, P., Chmelík, J., Gorčík, L., Žídek, L., Sklenář, V., and Bezouška, K. (2008) Soluble recombinant CD69 receptors optimized to have an exceptional physical and chemical stability display prolonged circulation and remain intact in the blood of mice. *FEBS J. 275*, 5589–5606.
19. Kolenko, P., Skálová, T., Vaněk, O., Štěpánková, A., Dušková, J., Hašek, J., Bezouška, K., and Dohnálek, J. (2009) The high-resolution structure of the extracellular domain of human CD69 using a novel polymer. *Acta Crystallogr. F65*, 1258–1260.
20. Pavlíček, J., Kavan, D., Pompach, P., Novák, P., Lukšan, O., and Bezouška, K. (2004) Lymphocyte activation receptors: new structural paradigma in group V of C-type animal lectins. *Biochem. Soc. Trans. 32*, 1124–1126.
21. Bezouška, K., Šnajdrová, R., Křenek, K., Vančurová, M., Kádek, A., Adámek, D., Lhoták, P., Kavan, D., Hofbauerová, K., Man, P., Bojarová, P., and Křen, V. (2010) Carboxylated calixarenes bind strongly to CD69 and protect $CD69^+$ killer cells from suicidal cell death induced by tumor cell surface ligands. *Bioorg. Med. Chem. 18*, 1434–1440.
22. Křenek, K., Kuldová, M., Hulíková, K., Stibor, I., Lhoták, P., Dudič, M., Budka, J., Pelantová, H., Bezouška, K., Fišerová, A., and Křen, V. (2007) *N*-acetyl-D-glucosamine substituted calix[4]arenes as stimulators of NK cell-mediated antitumor immune response. *Carbohydr. Res. 342*, 1781–1792.
23. Hulíková, K., Benson, V., Svoboda, J., Šíma, P., and Fišerová, A. (2009) *N*-Acetyl-D-glucosamine-coated polyamidoamine dendrimer modulates antibody formation via natural killer cell activation. *Int. Immunopharmacol. 9*, 792–799.
24. Benson, V., Grobárová, V., Richter, J., and Fišerová, A. (2010) Glycosylation regulates NK cell-mediated effector function through PI3K pathway. *Int. Immunol. 22*, 167–177.
25. Bojarová, P., Křenek, K., Wetjen, K., Adamiak, K., Pelantová, H., Bezouška, K., Elling, L., and Křen, V. (2009) Synthesis of LacdiNAc-terminated glycoconjugates by mutant galactosyltransferase—A way to new glycodrugs and materials. *Glycobiology 19*, 509–517.
26. Schuck, P. (2000) Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modelling. *Biophys. J. 78*, 1606–1619.
27. Schuck, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem. 320*, 104–124.
28. Dousseau, F., Therrien, M., and Pézole, M. (1989) On the spectral subtraction of water from the FT-IR spectra of aqueous solution of proteins. *Appl. Spectrosc. 43*, 538–542.

29. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on an UNIX pipes. *J. Biomol. NMR 6*, 277–293.
30. Chipman, D. M., Grisaro, V., and Sharon, N. (1967) The binding of oligosaccharides containing *N*-acetylglucosamine and *N*-acetylmuramic acid to lysozyme. *J. Biol. Chem. 242*, 4388–4394.
31. Hamann, J., Fiebig, H., and Strauss, M. (1993) Expression cloning of the early activation antigen CD69, a type II integral membrane protein with a C-type lectin domain. *J. Immunol. 150*, 4920–4927.
32. Blomhoff, R., Tolleshaug, H., and Berg, T. (1982) Binding of calcium ions to the isolated asialoglycoprotein receptor. Implication for receptor function in suspended hepatocytes. *J. Biol. Chem. 257*, 7456–7459.
33. Kimura, T., Imai, Y., and Irimura, T. (1995) Calcium-dependent conformation of a mouse macrophage calcium-type lectin. Carbohydrate binding activity is stabilized by an antibody specific for a calcium-dependent epitope. *J. Biol. Chem. 270*, 16056–16062.
34. Weis, W. I., and Drickamer, K. (1994) Trimeric structure of a C-type mannose-binding protein. *Structure 15*, 1227–1240.
35. Wallis, R., and Drickamer, K. (1999) Molecular determinant sof oligomer formation and complement fixation in mannose-binding proteins. *J. Biol. Chem. 274*, 3580–3589.
36. Diefenbach, A., and Raulet, D. H. (2001) Strategies for target cell recognition by natural killer cells. *Immunol. Rev. 181*, 170–184.
37. Coombs, P. J., Harrison, R., Pemberton, S., Quintero-Martinez, A., Parry, S., Haslam, S. M., Dell, A., Tailor, M. E., and Drickamer, K. (2010) Identification of novel contributions to high-affinity glycoprotein-receptor interactions using engineered ligands. *J. Mol. Biol. 396*, 685–696.

4.1.2 Soluble recombinant CD69 receptors optimized to have an exceptional physical and chemical stability display prolonged circulation and remain intact in the blood of mice

the FEBS Journal

Soluble recombinant CD69 receptors optimized to have an exceptional physical and chemical stability display prolonged circulation and remain intact in the blood of mice

Ondřej Vaněk[1,2,*], Monika Nálezková[3,*], Daniel Kavan[1,2], Ivana Borovičková[1], Petr Pompach[1,2], Petr Novák[2], Vinay Kumar[2], Luca Vannucci[2], Jiří Hudeček[1], Kateřina Hofbauerová[2,4], Vladimír Kopecký Jr[4], Jiří Brynda[5], Petr Kolenko[6], Jan Dohnálek[6], Pavel Kadeřávek[3], Josef Chmelík[2,3], Lukáš Gorčík[3], Lukáš Žídek[3], Vladimír Sklenář[3] and Karel Bezouška[1,2]

1 Department of Biochemistry, Faculty of Science, Charles University, Prague, Czech Republic
2 Institute of Microbiology, Academy of Sciences of Czech Republic, Prague, Czech Republic
3 National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czech Republic
4 Institute of Physics, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic
5 Institute of Molecular Genetics, Academy of Sciences of Czech Republic, Prague, Czech Republic
6 Institute of Macromolecular Chemistry, Academy of Sciences of Czech Republic, Prague, Czech Republic

Keywords
C-type lectin; leukocyte activation; plasma clearance; refolding; stability

Correspondence
K. Bezouška, Department of Biochemistry, Faculty of Science, Charles University Prague, Hlavova 8, CZ-12840 Praha 2, Czech Republic
Fax: +420 2 4172 1143
Tel: +420 2 4106 2383
E-mail: bezouska@biomed.cas.cz

*These authors contributed equally to this work

(Received 5 June 2008, revised 2 September 2008, accepted 11 September 2008)

doi:10.1111/j.1742-4658.2008.06683.x

We investigated the soluble forms of the earliest activation antigen of human leukocyte CD69. This receptor is expressed at the cell surface as a type II homodimeric membrane protein. However, the elements necessary to prepare the soluble recombinant CD69 suitable for structural studies are a matter of controversy. We describe the physical, biochemical and *in vivo* characteristics of a highly stable soluble form of CD69 obtained by bacterial expression of an appropriate extracellular segment of this protein. Our construct has been derived from one used for CD69 crystallization by further optimization with regard to protein stability, solubility and easy crystallization under conditions promoting ligand binding. The resulting protein is stable at acidic pH and at temperatures of up to 65 °C, as revealed by long-term stability tests and thermal denaturation experiments. Protein NMR and crystallography confirmed the expected protein fold, and revealed additional details of the protein characteristics in solution. The soluble CD69 refolded in a form of noncovalent dimers, as revealed by gel filtration, sedimentation velocity measurements, NMR and dynamic light scattering. The soluble CD69 proved to be remarkably stable *in vivo* when injected into the bloodstream of experimental mice. More than 70% of the most stable CD69 proteins is preserved intact in the blood 24 h after injection, whereas the less stable CD69 variants are rapidly taken up by the liver.

CD69, an earliest activation antigen of lymphocytes and a versatile leukocyte signaling molecule, plays a key role in a large number of immune effector functions. This receptor is constitutively expressed at the surface of $CD3^{bright}$ thymocytes, monocytes, neutrophils, epidermal Langerhans' cells and platelets, and appears very early upon the activation of T-lymphocytes, natural killer (NK) cells and some other cells of

Abbreviations
AUC, analytical ultracentrifugation; CRD, carbohydrate-recognition domain; DLS, dynamic light scattering; FT-ICR, FT-ion cyclotron resonance; NK, natural killer; T_d, temperature of denaturation.

hematopoietic origin [1]. Biochemically, CD69 is a disulfide-linked homodimer with two constitutively phosphorylated and variously glycosylated polypeptides [2]. It belongs to the type II integral membrane proteins possessing an extracellular C-terminal protein motif related to C-type animal lectins [3–5]. Functional studies using a series of CD69/CD23 chimeras clarified the role of individual protein segments in the biology of this receptor [6]. While the transmembrane and cytoplasmic domains are responsible for signaling and cellular expression, the 'stalk' region of CD69 containing the dimerization Cys68 is important for the formation of homodimers and for proper surface expression [7,8]. CD69 is associated with G-proteins, and its rapid surface expression by transition from the intracellular stores can be induced by cellular activation or by heat shock, independently of new RNA and protein synthesis [9].

It has also been shown that in killer lymphocytes, such as cytotoxic T cells and NK cells, CD69 is important for the activation of cytotoxic functions [10] and forms a part of the signalization network involving activating as well as inhibitory (e.g. CD94) receptors on these cells [11]. However, more recent studies using CD69 deficient mice revealed that this receptor may be important in the downregulation of the immune response, mostly through the production of the pleiotropic cytokine transforming growth factor-β [12]. Moreover, $CD69^{-/-}$ mice that could not activate killer cells through an engagement of CD69 receptor were unexpectedly more resistant to experimentally induced tumors [13], probably due to the fact that activated killer lymphocytes were protected from apoptosis. From these experiments, a working hypothesis was proposed suggesting that cross-linking of CD69 on the surface of killer cells by tumor membrane bound ligands may cause hyperactivation of these cells, and their subsequent elimination by apoptosis or other mechanisms [12]. According to this concept, the inhibition of the above cross-linking by either soluble CD69 ligands, or by soluble CD69 receptors might protect $CD69^{+}$ killer cells from apoptosis, and render them more available for killing of the tumors.

Structural and biochemical studies have been performed to define the protein fold of soluble CD69, and to identify its physiological ligands that may become useful as potential modulators of many reactions in the immune system. The globular protein segment corresponding to the carbohydrate recognition domain of C-type lectins (Ser84 to Lys199) mediates the binding of most monoclonal antibodies used for receptor cross-linking. Moreover, this region, which is able to function independently of the rest of CD69 receptor, is assumed to bind physiological ligands [6]. The structure of this part of the molecule has been solved by protein crystallography [14,15] in the crystallized CD69 dimers, and shown to consist of the compact C-type lectin fold stabilized by three disulfides. Two soluble recombinant protein forms used in structural studies and additional forms used previously for ligand identification [8,16–18] comprise potential candidates for testing their immunological activities.

In the present study, we report the results of our physicochemical, biochemical and biological studies of soluble CD69 receptors, which show remarkable *in vitro* and *in vivo* stability that is compatible with their potential use for therapeutical applications.

Results

Design and optimization of the expression construct for soluble CD69

Previous studies using soluble CD69 receptors (for amino acid sequence, see Fig. 1A) have provided some insight into the elements necessary for the stability of these proteins. These studies have emphasized the limited stability of the 'short carbohydrate-recognition domain (CRD)' construct compared to the 'long CRD' variant, and supported the importance of Cys68 for the formation of covalent CD69 dimers [8–13]. We decided to investigate these features systematically, and produced four different expression constructs, starting with Gln65, Gly70, Val82 and Ser84, designated CD69CQ65, CD69NG70, CD69NV82 and CD69NS84, respectively (Fig. 1A).

Only the protein expressed from the first construct contains the interchain dimerization cysteine Cys68, thus predisposing it to occur as a covalent dimer (CD69C). Despite previously published work on the production of disulfide-dimerized soluble CD69 [16], only a very limited amount of this protein could be produced after on-column refolding, removal of the histidine tag and reverse phase separation. SDS/PAGE under nonreducing and reducing conditions (Fig. 1B, lanes 2 and 3, respectively), as well as MS-ESI (Fig. 1C), confirmed the expected characteristics of the protein.

It was observed that, from the remaining three human proteins predicted to occur as monomers or noncovalent dimers (CD69N), the longest construct containing an extended stalk region starting with Gly70 (i.e. CD69NG70) displayed a number of interesting characteristics, even if its initial production using Protocol I led to some problems. Proteins prepared using this protocol appeared homogenous by

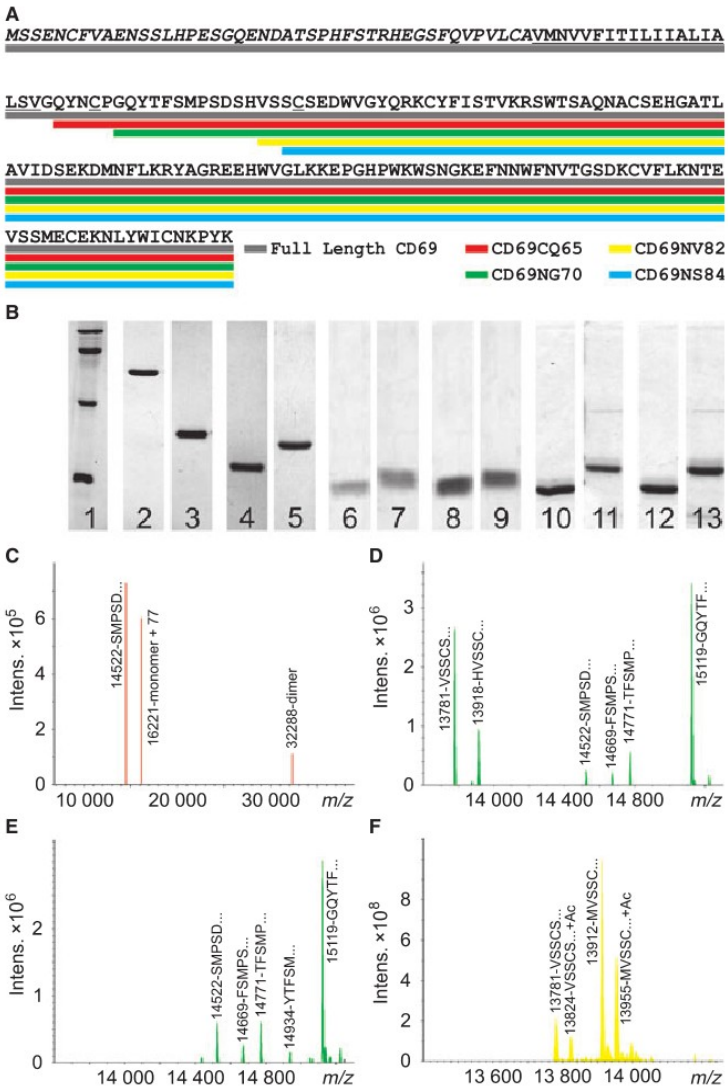


Fig. 1. Amino acid sequences of soluble CD69 proteins used in the present study, and examples of their analyses.
(A) Sequence of the full length human CD69 with the intracellular part (italics), transmembrane domain (underlined) and the extracellular portion including the C-terminal domain homologous to the carbohydrate-recognition domain of C-type lectin family. The extent of CD69 soluble forms is marked by color lines below the full length CD69 sequence. (B) SDS/PAGE of CD68CQ65 (lanes 2 and 3), CD69NG70 (lanes 4 and 5), CD69NV82 (lanes 6 and 7) CD69NS84 (lanes 8 and 9), rat CD69 (lanes 10 and 11) and mouse CD69 (lanes 12 and 13) was performed under nonreducing (even lanes) and reducing (odd lanes) conditions. Lane 1 contains protein size markers: BSA (66 kDa), ovalbumin (44 kDa), trypsinogen (24 kDa) and lysozyme (14 kDa). (C–F) FT-ICR mass spectra are shown for CD69CQ65, CD69NG70 (protocol I), CD69NG70 (protocol II) and CD69NV82, respectively.

SDS/PAGE under reducing conditions (Fig. 1C, lane 5), whereas there was a notable shift in mobility under nonreducing conditions (Fig. 1C, lane 4), most probably because of the more compact arrangements of the protein subunits cross-linked by three disulfide bridges. When examined by high resolution FT-ion cyclotron resonance (ICR) MS, the protein displayed a notable degradation of the N-terminal part of its stalk region as shown by a clear ladder of the degradation products that stopped only at Val82 (Fig. 1D). However, by employing an alternative purification protocol (Protocol II), much more stable preparations predominantly displaying the expected molecular peak at *m/z* 15119 could be obtained (Fig. 1E). The latter molecular form represents the one expected for the protein sequence with the initiation methionine removed, and all three disulfide bonds closed. The complete removal of the initiation methionine during prokaryotic protein production was also confirmed by extensive N-terminal sequencing (up to 45 cycles of automated Edman degradation performed with reduced protein having the cysteine residues modified by acrylamide).

For the last two constructs (CD69NV82 and CD69NS84), homogenous proteins displaying similar molecular characteristics could be prepared in high yield and purity (Fig. 1C, lanes 6–9). High resolution FT-ICR mass spectra of these proteins were very similar and the results for CD69NV82 are shown in Fig. 1F. No extensive N-terminal degradation occurred in these proteins and the minor heterogeneity observed may be assigned to the incomplete removal of the initiation methionine from these proteins during recombinant production.

CD69NG70 has unusual solubility and stability

To assess the solubility and stability of the recombinant preparations of CD69, we concentrated both CD69NG70 and CD69NV82 using a Centricon 10 device, and were able to confirm their very high solubility. Both protein preparations could be concentrated up to 40 $mg{\cdot}mL^{-1}$ without any signs of precipitation or aggregation (these experiments could not be performed with CD69CQ65 and CD69NS84 because of the limited amounts of material available).

To further evaluate the stability of CD69 preparations, we performed thermal denaturation experiments using UV spectroscopy. Upon protein unfolding, many aromatic amino acids forming the protein core become exposed with the concomitant increase in the molar extinction coefficient of the protein, and thus the increase in absorbance in the aromatic region. Shortly thereafter, a gradual unfolding of the protein occurs that results in the increase of turbidity, aggregation and precipitation. Interestingly, when CD69NG70 was tested at moderately high concentration (0.5 $mg{\cdot}mL^{-1}$) in standard Mes buffer at pH 5.8, it displayed unusually high temperature stability, and no unfolding of the protein could be seen, even after 1 h of incubation at temperatures as high as 60 °C (Fig. 2A). To verify the critical role of disulfide bridges in this thermal stability, we performed similar experiments in the presence of dithiothreitol. Exploratory studies employing the mobility shift of the oxidized form in SDS gels revealed that at least 3 mM dithiothreitol is required for a complete and quantitative breakage of all three disulfides in CD69 (results not shown). The addition of 5 mM dithiothreitol during the thermal denaturation experiment indeed caused a significant reduction in the thermal stability with notable unfolding starting already at 44 °C (Fig. 2B). The disulfide-independent unfolding of the protein is also a function of the pH of the reaction buffer and is higher in the alkaline environment. Thus, the unfolding temperatures at pH 6.8 or 7.8 were found to be 40 °C and 30 °C, respectively (Fig. 2C and data not shown). On the other hand, the protein is very stable in the acidic environment and is not denatured or precipitated, even at pH 2.0 in the presence of 40% acetonitrile (i.e. the conditions used during its purification on the reversed phase column).

FTIR spectroscopy represents an alternative method for looking at the thermal stability of CD69 proteins because the changes in the amide I and II bands (Fig. 2D) are sensitive indicators of the change in contents of the individual secondary structure elements. This metodology was therefore employed to investigate the stability of the produced proteins under thermal and pH stress. The content of secondary structure elements upon heating remained constant up until 5 °C below the temperature of denaturation (T_d) determined by differential scanning calorimetry, when the peripheral α-helices started to unfold, and there were less β-turns in some instances (see Table S1). To examine the stability under pH stress, the content of secondary structure elements was measured in buffers with different pH at temperatures set to 5 °C below the T_d. Most of the studied proteins retain their structure under a broad range of pH, except the alkaline (pH 9.0), where they are less stable, in particular CD69NV82 and CD69NS84 (Table S2). Taken together, these investigations support the hierarchy of stability of soluble CD69 proteins in which the somewhat longer proteins (CD69QC65, CD69NG70) appear to be more stable than the shortened ones.

We routinely maintain the stocks of soluble CD69 concentrated to 10 $mg{\cdot}mL^{-1}$ in moderately acidic buffers [10 mM Mes (pH 5.8), with 49 mM NaCl and 1 mM NaN_3] at both 4 °C and 24 °C. Under these conditions of storage, we could not observe any signs of precipitation or biochemical degradation, even after several months. Addition of common salts containing monovalent ions (NaCl, or KCl, up to 1 M concentrations) appeared to have little influence on the stability of the protein. Also, the use of several other common protein stabilizers (mannitol, glycerol, non-ionic detergents) had very little effect on protein stability. From several bivalent ions tested, calcium ion (Ca^{2+}) was the only one with a moderate stabilizing effect. For example, if the stability experiment described in Fig. 2B was performed in the presence of 10 mM $CaCl_2$, the initial unfolding temperature was increased by approximately 2 °C (data not shown). However, calcium bound to CD69 during refolding does not dissociate from the protein at pH up to 5.5, and the protein decalcified in acidic environment can be easily recalcified upon the addition of the external calcium (results not shown).

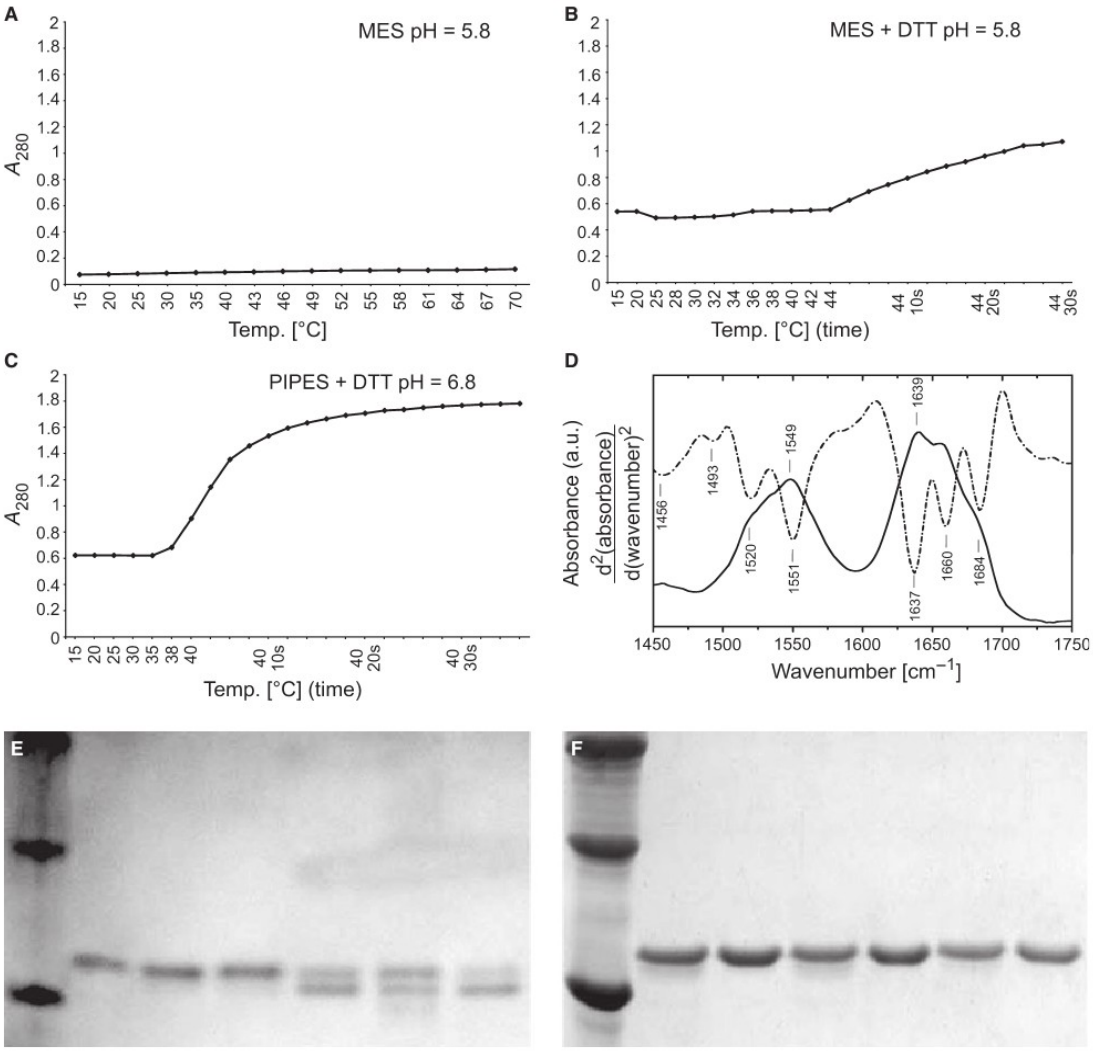


Fig. 2. Physical and biochemical stability of soluble CD69 receptors. (A–C) Thermal denaturation of CD69NG70 was followed by UV spectroscopy. The protein was examined in (A) Mes buffer (pH 5.8) or (B) Mes buffer (pH 5.8) with 10 mM dithiothreitol, or (C) Pipes buffer (pH 6.8) with 10 mM dithiothreitol at 0.5 mg·mL^{-1}, as described in the Exprimental procedures. UV spectra were measured in the termostated cuvette using the Beckman DU-70 spectrophotometer. When the denaturing temperature was reached, the temperature was kept constant, and the spectra were taken in several time intervals (indicated on the right). (D) FTIR spectrum of CD69 protein in the region of the amide I and II bands (the full line). The dash–dot line represents second derivative (smoothed by the Savitski–Golay function at 15 points) of the spectrum. (E, F) Biochemical stability of CD69NG70 purified using protocols I and II, respectively, was observed by SDS/PAGE upon incubation at 37 °C for 1, 2, 3, 4 and 5 days, and compared with the preparation stored at 4 °C (initial lane). Protein markers shown on the left consist of BSA (65 kDa), trypsinogen (24 kDa) and lysozyme (14 kDa).

Because some experiments (NMR, *in vivo* studies) require the long-term use of the protein at elevated temperatures, we decided to follow experimentally the stability at 37 °C. Under these experimental conditions (1 mg·mL^{-1} of protein in 10 mM Mes buffer, pH 5.8), the degradation of the protein depends solely on the production protocol, and thus probably reflects the purity of the final product. For example, as already

Table 1. Summary of the physical and biochemical stability of the investigated proteins.

Protein	Characteristic	T_d[a] (°C)	T_d[b] (°C)	$t_{1/2}$ at 30 °C[c] (days)	$t_{1/2}$ at 37 °C[c] (days)
CD69CQ65	Covalent dimer	67	65	24	9
CD69NG70	Noncovalent dimer	65	63	> 30	> 30
CD69NV82	Noncovalent dimer	56	53	15	4
CD69NS84	Noncovalent dimer	55	52	15	3
Rat CD69	Monomeric	66	63	> 30	24
Mouse CD69	Noncovalent dimer	63	62	> 30	> 30

[a] Determined by differential scanning calorimetry. [b] Determined by FTIR spectroscopy. [c] Calculated from densitometric evaluation of SDS gels.

mentioned, CD69NG70 prepared using Protocol I is degraded by approximately 50% to its lower molecular mass variant, CD69NV82, after 3 days at 37 °C (Fig. 2E). However, the same protein purified using Protocol II is completely stable under these conditions (Fig. 2F).

A summary of all the protein stability data for the four different protein variants under study is provided in Table 1. It is evident that, when purified using Protocol II, CD69NG70 is the best protein from the point of view of both its physical and long-term stability. Protein CD69CQ65 displays an exceptional physical stability upon heating up to 67 °C but it has a much lower long-term biochemical stability. Interestingly, the stability of the short proteins CD69NV82 and CD69NS84 is much lower using these criteria, both from the point of view of their physical stability upon heating and their biochemical stability.

CD69NG70 is a monodisperse, compactly folded protein

Considering the protein stability results as well as the practical aspects such as production yield and complexity of the purification protocol, CD69NG70 appeared to be the best candidate for the stable soluble form of human CD69. To prove its correct fold, we applied NMR analysis as well as protein crystallography.

We produced CD69NG70 in bacteria growing on minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source and purified the uniformly labeled protein (> 95% as judged by FT-ICR MS). The ^{1}H-^{15}N-HSQC spectrum of 0.3 mM solution of this protein is shown in Fig. 3A indicating good dispersion of the backbone and side-chain signals (the latter including those assigned to tryptophane indole groups in the lower left corner of the spectrum and asparagine/glutamine NH_2 signals in the upper right region of the spectrum). When the same sample was analyzed after

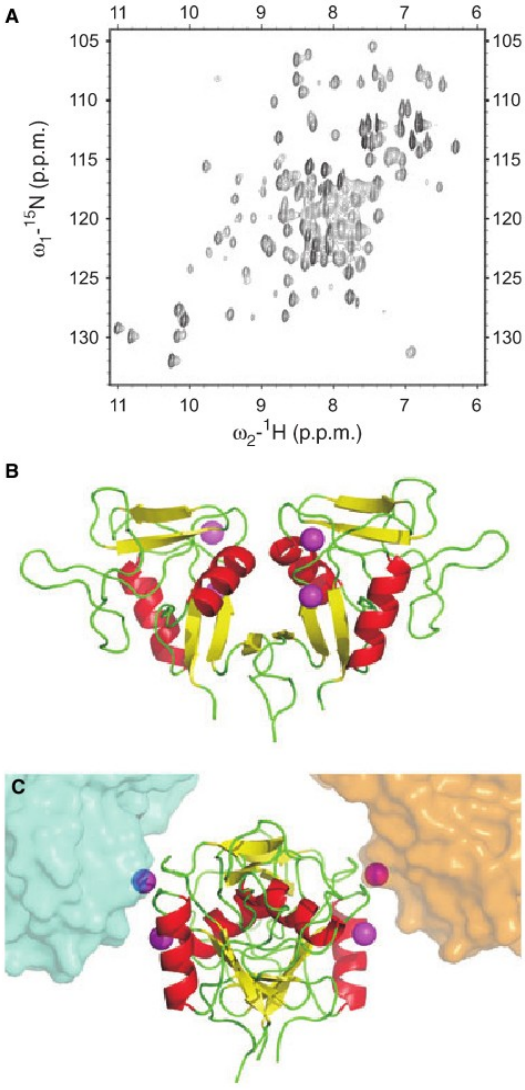


Fig. 3. Structure determination of CD69NG70 protein. (A) ^{1}H-^{15}N HSQC spectra were measured using 0.3 mM CD69NG70 uniformly labeled with ^{15}N at 303 K (30 °C) using the 600 MHz NMR spectrometer Bruker 600 UltraShield. (B) The crystal structure of the CD69 noncovalent dimer (ribbon) with chloride anions (spheres with Van der Waals atomic radius). (C) Showing the same molecule as in (B) rotated by 90° around the vertical axis, with two neighboring molecules shown as cyan and orange transparent molecular surfaces.

6 months, essentially identical results were obtained, again pointing to the high stability of the protein preparation. Even spectra measured using several different batches of the protein looked very similar (data not shown), indicating reproducibility of the refolding and purification protocol.

The crystallization of CD69 has been until now performed in weakly acidic environment (pH of around 4.0) [14,15], supporting both the stability of the protein and its efficient crystallization. At the same time, these conditions prevent the binding of ligands to CD69 because most suggested ligands are at least half-dissociated, even at a slightly acidic pH of around 5.5 [17]. The major incentive of the present study was an attempt to crystallize soluble CD69 in buffers with neutral or slightly alkaline pH under conditions compatible with binding of potential ligands. We succeeded in crystallizing the very stable CD69NG70 protein using L-arginine hydrochloride as buffer and stabilizing agent at pH 7.0 (Fig. 3B). However, in our crystallization trials, we found that attempts to crystallize either the longer CD69CQ65, corresponding to the one used by Natarajan *et al.* [14], or the shorter protein CD69NV92, identical to that used previously by Llera *et al.* [15], produced only small crystals of insufficient quality. The solved structure provided a classical C-type lectin-like protein fold composed of two α-helices and three β-sheets in which the first 11 N-terminal amino acids were not structurally ordered possibly due to their flexibility (see below). CD69NG70 formed noncovalent dimers structurally ordered into the hexagonal crystal lattice. A single dimer can be roughly described as an ellipsoid with three axes extending to 7, 3.8 and 3.1 nm (including the solvation shell), thus indicating the very compact folding of the polypeptide chain (Fig. 3B). The dimer interface is built by short intermolecular β-sheet and hydrophobic aromatic side chains. Both overall fold and dimer arrangement are identical to those described previously [14,15].

NMR ^{15}N relaxation measurements were performed to monitor flexibility of the CD69NG70 backbone. To interpret the data, resonance frequencies of the backbone amides were assigned as described in the Experimental procedures. Chemical shifts of alpha and beta carbons and of backbone amide protons and nitrogens were deposited together with the measured relaxation data in the BioMagResBank (http://www.bmrb.wisc.edu) under Accession No. 15703. The obtained assignment covered 77% of the sequence, with most of the unassigned residues between Glu87 and Phe98. Order parameters calculated from the relaxation data (Fig. 4D) revealed a low flexibility of most residues, with the exception of the N-terminal region, where the order parameter gradually decreased from 0.75 (Val82) to 0.08 (Phe74). This finding is in agreement with the X-ray structure where the residues Gly70 to His81 are missing as disordered.

Because we were interested in co-crystallization of CD69 with its low molecular weight ligands suggested previously [17] in the crystal structure, calcium chloride was added both to the protein and precipitant solution (see Experimental procedures). Based on anomalous Fourier, three structurally ordered anomalously contributing atoms were located in the asymmetric unit of the crystal structure of CD69 (the asymmetric unit comprises one dimer of CD69 and three Cl^- anions). Every monomer binds two Cl^- anions, one in a shallow pocket at the side of the molecule and the second one forming crystal contact with a neighboring dimer in the crystal structure (Fig. 3B,C). Neither of these two binding sites resembled the well-known calcium binding site for classical C-type lectins (such as the mannose binding protein) or the site predicted from our calcium binding data and computer docking experiments [17]. Furthermore, the amino acid neighbourhood of these ions (Ser, Thr, Val, Tyr, Lys) and their distances from the nearest atoms (3.1–3.3 Å) would be rather atypical for the calcium cation, but appropriate for the chloride anion, which has approximately the same intensity of anomalous scattering signal. We therefore assigned these three ions to chlorides.

We also tried to crystallize CD69 in presence of *N*-acetyl-D-glucosamine (in concentrations in the range 1 mM to 1 M), as well as several branched oligosaccharide structures based on GlcNAc that were available in our laboratory [18]. Despite the fact that we were able to collect high resolution data for most of these co-crystals (a total of eight complete datasets with resolution 1.8–2.2 Å), we could not observe any extra electron density corresponding to these potential ligands (data not shown). The crystal structure with best resolution was selected for deposition (accession number Protein Databank code 3CCK).

Examination of the native size of soluble CD69

Because the crystals of CD69NG70 contained molecules packed as noncovalent dimers, we were interested to determine the native size and the monodispersity of the protein in solution. Gel filtration with a Superdex 200 column used for the final purification of the monodisperse proteins strongly suggested that all four proteins examined elute exclusively as dimers (Fig. 4A).

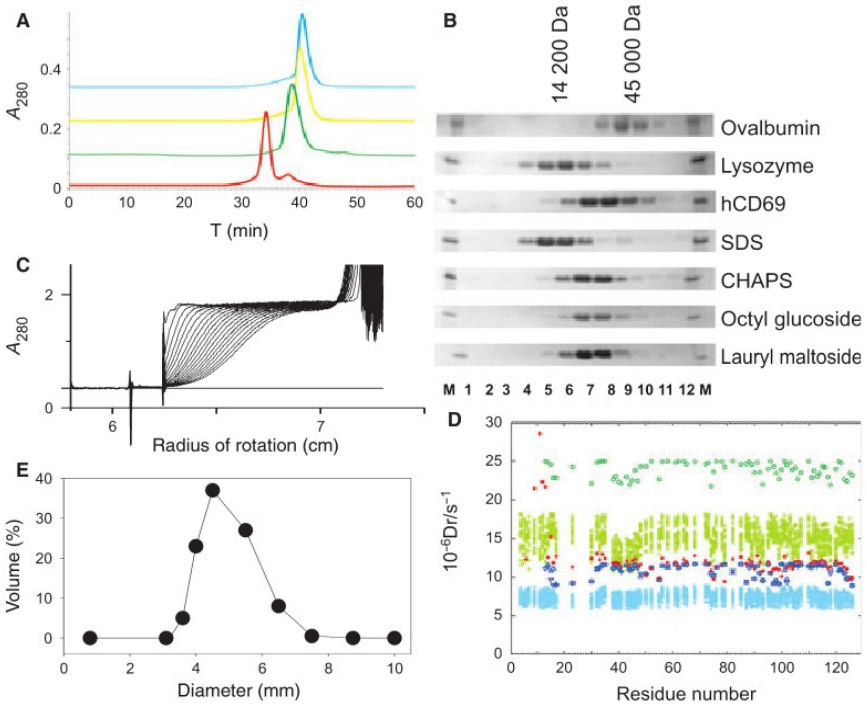


Fig. 4. Estimation of the native size of soluble CD69. (A) The native size of the four different soluble CD69 proteins was determined by gel filtration using a Superdex 200HR column (GE HealthCare) equilibrated in Mes buffer and eluted at 0.4 mL·min^{-1}. From top to bottom: CD69NS84 (blue), CD69NV82 (yellow), CD69NG70 (green) and CD69CQ65 (red). (B) Two hundred microlitres of 0.3 mM solution of CD69NG70 was applied onto the sucrose linear gradient (5–20% sucrose in Mes buffer, pH 5.8) and spun at 392 000 $g_{av.}$ and 30 °C in a SW-60 rotor (Beckman Coulter). In the initial experiment, the optimal time for the separation of the protein markers ovalbumine (44 kDa) and lysozyme (14 kDa) was found to be 15 h. The mobility of CD69NG70 separated under the same conditions, and also in the presence of 0.5% detergents (SDS, Chaps, octyl glucoside or lauryl maltoside, respectively), is shown in the corresponding lanes. (C) Sedimentation velocity measurement. The dialyzed sample was spun at 130 000 $g_{av.}$ and individual scans were recorded at 5 min intervals. (D) Apparent values of rotational diffusion coefficient, obtained from NMR ^{15}N relaxation data fitted separately for each residue (red crosses), are compared with the apparent mean rotational diffusion coefficients calculated by the software HYDRONMR for monomeric (green circles) and dimeric (blue triangles) CD69 structures. Triangles (up and down) distinguish subunits of the dimer; small symbols and light colors refer to individual structures of ensembles with the disordered N-terminal region modeled. (E) DLS measurements were performed as described in the Experimental procedures.

To investigate further the stability and the native size of CD69NG70, we employed hydrodynamic studies, protein NMR and light scattering experiments. When we sedimented CD69NG70 in sucrose density gradients in a preparative ultracentrifuge, it appeared as a single species with a mobility between that of ovalbumin (45 kDa) and lysozyme (14 kDa) (Fig. 4B). Moreover, we used the conditions of this experiment to investigate the chemical factors affecting the dimeric arrangement. The addition of non-ionic detergents such as CHAPS, octyl glucoside or lauryl maltoside did not change the sedimentation behavior of the soluble CD69 receptor but incubation in the presence of the anionic detergent SDS under mild conditions was able to cause dissociation of the dimer into single subunits (Fig. 4B). The single separated subunit remained folded under these experimental conditions because the totally unfolded CD69 obtained by boiling in the identical SDS concentration remained at the top of the centrifugation cuvette (not shown). Moreover, monomeric CD69 subunits remained stable for up to 1 week when stored at 4 °C, displaying an identical sedimentation as in the original experiment. However, upon heating to room temperature, these subunits unfolded with a half-time of several hours, as shown by additional sedimentation analyses not presented here.

Additional experimental techniques confirmed that both CD69NG70 and CD69NV82 are present

exclusively as noncovalent dimers (under the experimental conditions used). Sedimentation velocity measurements (Fig. 4C) in the analytical ultracentrifuge (AUC) provided a value of sedimentation coefficient of 3.51 ± 0.03 S for CD69NG70. When these values were used for molecular mass calculation, a value 30 kDa was obtained, which corresponded very well to the expected mass for the dimer (30.2 kDa). The corresponding values for the CD69NV82 protein were 2.95 ± 0.04 S, and the calculated molecular mass was 27 kDa, again very close to the calculated molecular mass of the dimer (27.5 kDa). The results obtained using sedimentation equilibrium were very similar (data not shown). Moreover, the apparent values of the overall correlation time derived from NMR relaxation measurements (Fig. 4D, see below) are compatible with the dimeric arrangement. Finally, dynamic light scattering (DLS), a modern, fast and versatile experimental technique, confirmed the monodispersity of the CD69 preparation (Fig. 4E), and provided an additional estimation for several of the molecular parameters measured by the previous techniques. These included the radius of gyration [r = 1.91 nm (crystallography) and 2.04 nm (DLS)], the translational diffusion coefficient [$D = 8.53 \times 10^{-7}$ cm^2·s^{-1} (AUC) and 8.47×10^{-7} cm^2·s^{-1} (DLS)], the rotational diffusion coefficient [$D_r = 12 \times 10^6$ s^{-1} (NMR relaxation, see below) and 9.36×10^6 s^{-1} (DLS)], and the sedimentation coefficient [s = 3.51 S (AUC) and 3.02 S (DLS)].

A more detailed picture of the rotational diffusion was derived from the NMR ^{15}N relaxation data. To monitor the effect of the real shape of the molecule on its tumbling, the values of the apparent rotational diffusion coefficient D_r were evaluated for each residue not effected by spectral overlap or slow conformational exchange as described in the Experimental procedures. The apparent D_r values were compared with the values predicted from hydrodynamic calculations of several molecules, including the crystal dimer, its monomeric subunit, and sets of dimeric and monomeric structures with the disordered N-terminal tail modeled in various conformations (Fig. 4D). The comparison clearly showed that largely overestimated D_r values were predicted for the monomeric structures, including those with the N-terminal residues added. On the other hand, values predicted for the X-ray dimer structure closely matched the data obtained form NMR ^{15}N relaxation for the well-ordered portion of the protein. The experimental D_r values for the N-terminal residues deviated from the average apparent D_r, estimated for the rigid core of the protein, and from the values predicted by the rigid-body hydrodynamic calculations. This indicates that motions of the disordered N-terminal residues are largely independent and have a little effect on the rotational diffusion of the well-ordered portion of the protein. In conclusion, NMR ^{15}N relaxation combined with hydrodynamic calculations demonstrated the presence of the dimer. Somewhat higher apparent D_r values (approximately 12×10^6 s^{-1}) compared to those obtained from DLS (see above) reflect the fact that tumbling of the rigid portion of the protein is largely independent of the motions of the disordered N-terminal tail.

Production of soluble rat and mouse CD69

For *in vivo* stability studies in mice, it was desirable to compare the properties of the variant soluble human CD69 proteins with the corresponding rat and mouse orthologs [18,19] that are more compatible with the experimental model used. Therefore, we prepared the corresponding soluble rat and mouse CD69 proteins using the expression constructs having an extended 'stalk' similar to that found in the most stable human CD69, CD69NG70. Thus, in the expression constructs used, there were 15 amino acids before the first cysteine residue defining the 'long' CRD in the human CD69NG70 protein, whereas there were 12 and 15 amino acid residues in the corresponding rat and mouse orthologs, respectively. The rat and mouse CD69 refolded and purified efficiently, giving rise to homogenous proteins on SDS/PAGE (Fig. 1B, lanes 10–13). Moreover, the physical and biochemical stability of the three proteins also appeared to be comparable (Table 1; see also Supporting information, Tables S1 and S2). Interestingly, although the mouse CD69 appeared to form noncovalent dimers similar to human CD69, the rat CD69 protein appeared to be monomeric [18] (Table 1).

Stability of soluble CD69 preparations *in vivo*

To assess the suitability of soluble CD69 preparations for *in vivo* therapeutic applications, we radioiodinated these proteins and followed the plasma clearance of these proteins. When injected into the bloodstream of C57BL/6 mice, three of the soluble proteins (CD69CQ65, CD69NG70 and CD69NV82) displayed a prolonged circulation. After the initial dilution caused by binding and retaining in the tissues, the blood level of these proteins stabilized within 4 h, and then remained nearly unchanged for up to 24 h after injection (Fig. 5A). The circulation half-life for these proteins (approximately 40 h) is comparable to that of the endogenous serum proteins (Table 2). Moreover,

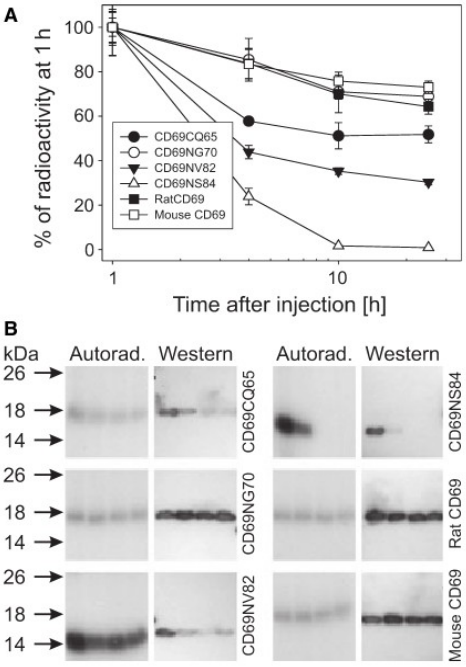


Fig. 5. Plasma clearance of soluble CD69 receptors in the bloodstream of C57BL/6 mice. (A) The ^{125}I-radiolabeled recombinant proteins were injected into the tail vein of the mice and the radioactivity in individual collection times was related to the radioactivity measured 1 h after injection, taken as 100%. (B) Degradation of the radioiodinated proteins CD69CQ65 (upper left panel), CD69NG70 (middle left panel), CD69NV82 (lower left panel), CD69NS84 (upper right panel), rat CD69 (middle right panel) and mouse CD69 (lower right panel), respectively, was determined in mouse serum depleted of serum (glyco)proteins by 15% SDS/PAGE followed by autoradiography, or western blotting. The results in (A) indicate the average values from duplicate radioactivity counting with the range indicated by the error bars.

Table 2. Evaluation of the pharmacokinetics parameters for plasma clearance of soluble CD69 in mice.

Protein	Plasma half life (h)	First order rate constant	Clearance (mL·h^{-1}·kg^{-1})	Apparent volume of distribution (mL·kg^{-1})
CD69CQ65	17.3	0.0509	6.68	71.0
CD69NG70	41.5	0.0268	3.16	48.2
CD69NV82	10.1	0.0803	6.87	59.7
CD69NS84	1.4	0.5047	7.81	71.0
Rat CD69	37.4	0.0291	2.15	41.2
Mouse CD69	47.7	0.0232	1.77	41.2

when we recovered the radiolabeled CD69 proteins from serum samples, and examined the intactness of the protein by SDS/PAGE followed by autoradiography, very little degradation could be seen for these proteins (Fig. 5B). Only the shortest soluble CD69 protein, CD69NS84, was quickly eliminated from the circulation with half-life of approximately 1.4 h (Table 2) concomitantly with the disappearance of this protein (Fig. 5A,B). Both the rat and the mouse CD69 exhibited a prolonged circulation in the blood of mice, which was comparable with the most stable human CD69, CD69NG70 (Fig. 5A), and remained intact and circulating in the blood for up to 24 h (Fig. 5B). Western blot analyses of CD69 proteins extracted from the serum of experimental mice using antibodies recognizing conformation sensitive epitopes on CD69 proteins provide further evidence for the long-term stability of the above mentioned preparations (Fig. 5B). Finally, the best evidence for good *in vivo* stability is provided by the rapid GlcNAc binding test indicating that even the biological (carbohydrate binding) activity of soluble CD69 proteins was preserved under these conditions (Table 3).

Table 3. Evaluation of the biological (carbohydrate-binding) activity of soluble CD69 proteins circulating in the blood of mice for 24 h. ND, not determined.

Protein	Total counts recovered from the serum (c.p.m.)	Total counts bound to GlcNAc matrix (c.p.m.)	Total counts not bound to GlcNAc matrix (c.p.m.)
CD69CQ65	5956	5656	235
CD69NG70	7645	7345	302
CD69NV82	4350	2345	1987
CD69NS84	ND	ND	ND
Rat CD69	6504	5801	657
Mouse CD69	7868	7650	178

Upon killing of the mice 24 h after the injection of the proteins, we collected the most important organs and body fluids for scintillation counting. Interestingly, only approximately 10% of the initial radioactivity was recovered outside the animals, and could be found in urine and faeces (Fig. 6A,B). Otherwise, there were only two major compartments that together accounted for 60–70% of the injected radioactivity, namely liver and blood. The distribution of CD69 radioactivity between these two compartments appeared to be reciprocal. Thus, for long-circulating proteins such as human CD69NG70, and rat and mouse CD69, up to 40% of the injected radioactivity could be recovered in the blood 24 h after injection, whereas the liver took up approximately 20% of the initial dose. On the other hand, CD69NS84, which could serve as an example of a protein rapidly cleared from the blood (Fig. 5A), was taken up predominantly by the liver, which accumulated more than 60% of the initial dose

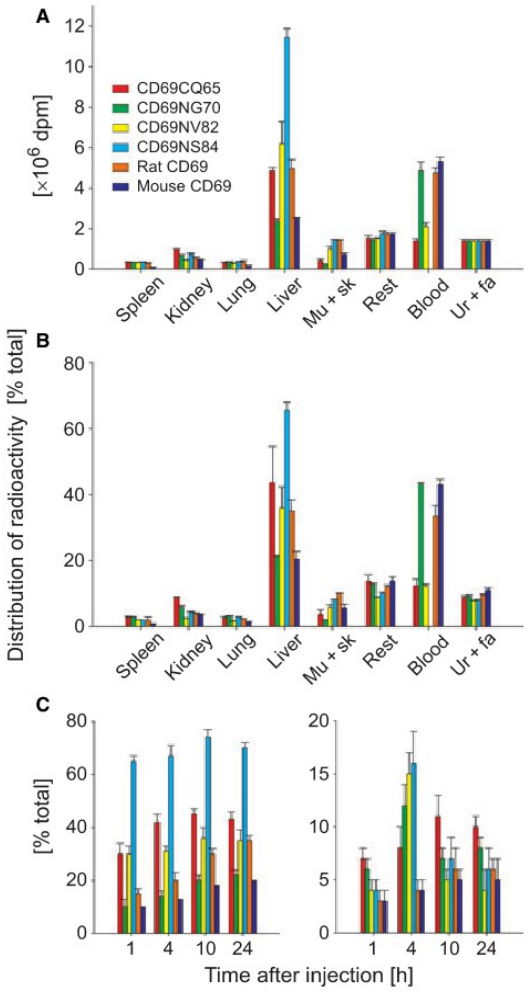


Fig. 6. Distribution of radioactivity in organs, body fluids and excretion of radioactivity in C57BL/6 mice injected with 100 μg of the indicated radiolabeled proteins. The total radioactivity is given in (A), whereas the percentage of the total injected dose is indicated in (B). (C) Accumulation of radioactivity in the liver and kidney, respectively, was followed over 24 h after injection. Mu + sk, muscle plus skin; Rest, rest of the body (see the Experimental procedures); Ur + fa, urine plus faeces. The results show the average values from duplicate radioactivity counting with the range indicated by error bars.

(Fig. 6A,B). A more detailed analysis of the kinetics of accumulation of soluble CD69 receptors in the liver and kidney indicates a fast uptake of the proteins with a short half-life in plasma into these organs, particular into the liver (Fig. 6C,D).

Discussion

Although several soluble CD69 proteins have previously been described in the literature by our group [8,17,18], as well as in other studies [14–16], the physical, biochemical and *in vivo* stabilities of these proteins have not been systematically studied. In the present study, we describe a detailed structure stability investigation of soluble human CD69 receptor using a series of N-terminal deletions. Previous results have demonstrated the critical importance of this N-terminal 'stalk' region and of the three disulfide bonds for CD69 stability [8,16]. However, we now report seminal findings that argue for the importance of the 'extended stalk' region starting just after the dimerization cysteine Cys68, where a short sequence of 11 amino acids (Gly70-His81) appeared to be particularly critical. This short peptide segment is not structurally organized, as shown by a lack of corresponding signals in all the available crystal structures of CD69, as well as by the high mobility of these residues in the NMR relaxation experiments. Yet, despite being structurally unordered, this segment contributes significantly to the physical and biochemical stability of the soluble CD69 receptors, promotes efficiently the formation of noncovalent dimers during *in vitro* refolding, and allows the crystallization of the corresponding protein under conditions compatible with the binding of ligands. Soluble CD69 expressed as covalently linked dimeric protein appeared to be physically even more stable but, on the other hand, posed a number of disadvantages, including a complicated production strategy, difficult purification and low yields. On the other hand, the noncovalent dimeric CD69NG70 protein can be easily purified in high yields (10 $mg \cdot L^{-1}$ of bacterial culture) over a period of 2–3 days using the commonly available equipment in an average biochemical laboratory.

CD69NG70 was therefore selected as the best candidate for the stable and easily available form of soluble CD69 receptor displaying remarkable long-term stability. The biochemical stability of this preparation was even better than that for CD69CQ65 expressed as a covalent dimer, and was superior to that of the shorter proteins CD69NV82 and CD69NS84. In particular, the latter protein, although being just two amino acids shorter, displayed a significantly reduced stability. This corresponds well with our previous findings showing that further reduction of this protein in this area, and particularly the removal of the amino acids forming the third stalk disulfide bridge, is detrimental to the stability of such soluble CD69 proteins [17,18].

The exceptional stability of CD69NG70 is most probably related to its dimeric arrangement, which has

been demonstrated by a number of experimental techniques, including gel filtration, sedimentation velocity analysis, measurement of NMR relaxations and DLS. Our studies attempting to dissociate the dimer in SDS under mild conditions provided further support for such a conclusion. It would appear from these experiments that single globular carbohydrate-recognition domains of CD69, even when stabilized by detergent at the disrupted dimer interface, comprise short-lived molecules that are only moderately stable at low temperatures (4 °C) and start to unfold when heated to ambient temperature. However, the general validity of this conclusion appears to be challenged by the data for rat CD69, which appears to be monomeric.

The remarkable *in vitro* stability of the human soluble CD69 receptor, CD69NG70 protein, makes it a strong candidate for a protein that is potentially useful for therapeutic purposes. Therefore, it was critical to test the *in vivo* stability of this protein and to compare it with the stability of its rat and mouse orthologs that are more compatible with the experimental models in use. The results of these tests revealed both the long circulation and the intactness of the most stable human CD69 protein (together with its rat and mouse orthologs) in the blood of mice. For these proteins, approximately 40% of the initial dose could be still recovered 24 h after injection. Moreover, the exclusion of the protein in urine and faeces, and uptake by the liver, was relatively low. On the other hand, the least stable human CD69NS84 protein displayed a rapid plasma clearance connected with the transfer into the liver that took up more than 60% of the injected dose.

Progress in our knowledge of CD69 biology has advanced rapidly, allowing to propose the individual therapeutic modalities involving the stable soluble CD69 receptors described in the present study. One such protocol may involve the reactivity of the soluble long cirulating CD69 protein with the tumor surface ligands, leading to their blocking or reduced availability for the reaction with the cellular form of CD69 at the surface of the killer lymphocytes. This should result in the protection of these critical cells of antitumor immunity from apoptotic cell death following their hyperactivation by tumor surface ligands [12]. Such a possibility is further supported by recent results obtained by our group as well as in studies by others [20–23]. We have recently shown that mimetics of tumor surface ligands for CD69, when presented in a highly multivalent form, can bind strongly to $CD69^+$ lymphocytes and cause their death by a massive trigger of apoptosis. Graham *et al.* [20] recently reported that hyperstimulation of human T lymphocytes with ligands for CD28 or integrins results in their activation, as demonstrated by high surface expression of CD69, and massive cell death. North *et al.* [22] recently described NK cells that, when primed by incubation with the tumors, acquire the ability to lyse leukemic cell lines and even solid tumors primarily through CD69 receptors, as demonstrated by the ability to inhibit such lysis with soluble recombinant CD69 protein. These results appear to be supported by our histochemical evidence showing that tumor infiltrating lymphocytes significantly upregulate CD69 expression [21]. An increased occurrence of CD69 positive lymphocytes in tumor sites may also be related to its recently described function as a downregulator of lymphocyte egress from lymphoid organs due to the downmodulation of sphingosine 1-phosphate receptor 1 [23]. Konjevic *et al.* [24] reported the predictive value of CD69 expression during the clinical response to chemoimmunotherapy in patients with metastatic melanomas. Collectively, these results appear to support the role of CD69 as one of critical receptors involved in the recognition and killing of tumors, including clinically important solid tumors. In view of all these results, the availability of long-circulating and *in vivo* stable soluble CD69 proteins with a low toxicity for experimental animals would be an advantage for their use in experimental tumor therapy models using rats or mice [21].

In conclusion, our systematic studies of soluble CD69 receptors that have been refolded in the form of covalent or noncovalent dimers demonstrate large variation in the solubility and stability among these proteins and allow us to select the human protein containing amino acids Gly70 to Lys199 (and the rat and mouse orthologs) as the most physically and biochemically stable variant. We have proven an exceptional *in vivo* stability and low toxicity upon injection into the blood of experimental mice in which these proteins remain intact for the prolonged periods of time necessary to elicit their therapeutic effects in our animal tumor therapy models [21]. The availability of such soluble CD69 receptors now opens the way for their testing in various animal experimental models of CD69 related diseases, such as malignant or autoimmune diseases.

Experimental procedures

Materials

All chemicals were analytical grade reagents of the highest commercially available quality. Chemicals used for protein refolding were obtained from Serva (Heidelberg, Germany). Plasmid pCDA401 containing the insert coding

the entire extracellular part of the CD69 was described previously [8].

Expression and purification of recombinant soluble human CD69 receptors

For the production of the covalent dimeric CD69, CD69CQ65, an expression plasmid similar to that described previously [16] was used. DNA was amplified using forward primer 5′-CTCGAGACAATACAATTGTCCAGG-3′, and reverse primer 5′-ACAAAGCTTATTTGTAAGGTTTGTT ACA-3′, and the PCR product was cloned into pBSK + vector using the *Sma*I restriction site, and then into pRSETB vector using *Xho*I and *Hind*III restriction endonucleases. His-Tag was removed using mild tryptic digestion targeted to the two basic amino acids (Arg coded by the *Xho*I site was followed immediately by CD69 sequence Lys) preceding Gln65. For the production of soluble CD69 refolded as noncovalent dimers, three different DNA fragments coding for the extracellular portion of human CD69 were amplified by PCR using Deep Vent DNA polymerase (NEB, Ipswich, MA, USA) as the amplification enzyme, pCDA401 as a template and the following primer pairs: for CD69NG70, 5′-ACATATGGGCCAATACACATTC-3′ and 5′-ACAAAGCTTATTTGTAAGGTTTGTTACA-3′; for CD69NV82, 5′-ACATATGGTTTCTTCATGCTCTG-3′ and 5′- ACAAAGCTTATTTGTAAGGTTTGTTACA-3′; and for CD69NS84, 5′-ACATATGTCATGCTCTGAGGACTGG GTT-3′ and 5′- ACAAAGCTTATTTGTAAGGTTTGTT ACA-3′. PCR products were directly cloned into pBSK + cloning vector (Stratagene, LaJolla, CA, USA) using the *Sma*I restriction site and the desired expression constructs were subsequently cloned into pRSETB expression vector (Invitrogen, Carlsbad, CA, USA) using the *Nde*I and *Hind*III restriction sites introduced by the amplification primers.

Plasmids were transformed into *Escherichia coli* BL-21 (DE3) RIL or Gold strains (Stratagene). Bacteria were grown in 2 L Erlenmeyer flasks with 0.5 L of LB broth at 37 °C with ampicillin and tetracycline (Gold) or chloramphenicol (RIL) as antibiotics. Induction of protein production with isopropyl thio-β-D-galactoside was not necessary, and the culture was left to grow for 16–24 h. Cells were harvested by centrifugation, and inclusion bodies were isolated [25]. Inclusion bodies were dissolved in 50 mM Tris–HCl (pH 8.0) with 6 M guanidine-HCl and 100 mM dithithreitol, centrifuged, and adjusted to 10 $mg{\cdot}mL^{-1}$ protein. *In vitro* refolding of denatured protein was carried out by rapid dilution into the refolding buffer. Clarified protein solution was quickly diluted 100-fold into the refolding buffer composed of 50 mM Tris–HCl (pH 8.5), 0.4 M L-arginine, 2 mM $CaCl_2$, 1 mM NaN_3, 18 mM cysteamine, 1 mM cystamine, 1 mM phenylmethanesulfonyl fluoride, 1 $\mu g{\cdot}mL^{-1}$ leupeptine and 1 $\mu g{\cdot}mL^{-1}$ pepstatine. After slow stirring at 4 °C for 5 h, the refolding mixture was dialyzed against 8 L of 10 mM Tris–HCl (pH 8.5), 0.5 M NaCl and 1 mM NaN_3 for 6 h, and then against 10 L of 10 mM Tris–HCl (pH 8.5), 50 mM NaCl and 1 mM NaN_3 for 12 h at 4 °C.

Two protocols were used further. In Protocol I, protease inhibitors were added to the dialyzed protein, and the pH of the solution was adjusted with acetic acid to 5.5. The insoluble precipitate of misfolded protein was centrifuged at 20 000 ***g*** for 30 min at room temperature. The refolded CD69 protein was then captured on SP-Sepharose FF column (GE Healthcare Europe, Munich, Germany) equilibrated in 20 mM sodium acetate (pH 5.5), 50 mM NaCl and 1 mM NaN_3. The column was eluted by linear gradient of NaCl from 50 mM to 2 M. Fractions containing CD69 protein were pooled and applied onto a reverse phase column Vydac C4 (Dionex, Sunnyvale, CA, USA) with 0.1% trifluoroacetic acid as mobile phase A and eluted by linear gradient of mobile phase B composed of 95% acetonitrile and 5% of 0.1% trifluoroacetic acid from 30% to 40% B over 60 min. Concentrated fractions were further purified by gel filtration on a Superdex 200 HR 10/30 column (GE Healthcare) in 10 mM Mes (pH 5.8) with 100 mM NaCl, 2 mM $CaCl_2$ and 1 mM NaN_3, and concentrated to 10 $mg{\cdot}mL^{-1}$ using a Centriprep and Centricon device (Millipore, Billerica, MA, USA). In Protocol II, dialyzed protein without any acidification was passed through Q-Sepharose FF column (GE Healthcare) equilibrated with the dialysis buffer. The protein passed through the column, and was concentrated by ultrafiltration using PLGC regenerated cellulose membranes (Millipore) with a 10 kDa cut-off. The concentrated protein was finally purified by gel filtration on a Superdex 200 HR 10/30 column, and concentrated as described in Protocol I.

The preparation and analysis of rat CD69 has been described previously [18]. DNA fragment coding for mouse CD69 [5,19] was amplified from RNA prepared from spleens of C57BL/6 mice using RT-PCR with the forward primer 5′-TGCATATGGGCCTTTACGAGAAGTTGGAA-3′ containing the *Nde*I cloning site and the reverse primer 5′-TGA AGCTTCATTATCTGGAGGGCTTGCTGCA-3′ containing the *Hind*III site. The amplified *Nde*I–*Hind*III fragment was transferred into the pRSETB expression vector, and the protein was produced as described above, and purified using Protocol II.

Characterization of the purified soluble proteins

The identity of the prepared protein was confirmed by sequence mapping of tryptic digests with MALDI-TOF MS with a good sequence coverage. The total mass of the protein was measured by means of FT-ICR MS. The size distribution and polydispersity of human CD69 preparation were assessed by DLS at a concentration of 2 $mg{\cdot}mL^{-1}$ in 10 mM Hepes (pH 7.0) with 150 mM NaCl and 1 mM NaN_3. Samples were loaded into a 45 μL quartz cuvette and particle size distribution measurements were performed

repeatedly at 291 K using a Zetasizer Nano (Malvern Instruments, Malvern, UK). Estimates of hydrodynamic radii of the expected molecular species were calculated with the software HYDROPRO [26]. The structures of monomers and dimers of the human CD69 in the present study, and as reported previously [15] (Protein Databank code 1E87), were used as input coordinates for the calculations. The dimer of hCD69 in the Protein Databank record 1E87 was generated using symmetry operators. Calculated Stokes radii were compared with experimental values.

Protein crystallography and data collection

Recombinant human CD69 was crystallized by hanging drop vapor diffusion method using 24-well plates (Hampton Research, Aliso Viejo, CA, USA). Initial crystallization conditions were established at 291 K using selected precipitants from JBScreen kits (Jena Bioscience, Jena, Germany). Each drop was prepared by mixing equal volumes (1 μL) of the protein and the precipitant solution and each drop was equilibrated against 1 mL of precipitant solution. After optimization of poly(ethylene glycol) molecular weight and concentration, and after exchange of precipitant buffer, we determined the suitable crystallization conditions. Needle-shaped, but regular crystals were obtained by mixing 1 μL of protein at a concentration of 5 $mg \cdot mL^{-1}$ in 10 mM Bis-Tris–HCl (pH 6.5), 100 mM NaCl, 2 mM $CaCl_2$ and 1 mM NaN_3 with 1 μL of reservoir solution containing 0.1 M Arginine.HCl (pH 7.0), 20% PEG 3400, 10 mM $CaCl_2$ and 1 mM NaN_3. Crystals appeared within 1 week. For X-ray data collection, crystals were mounted in nylon loops and cryoprotected by soaking in the reservoir solution containing 25% glycerol as a cryoprotectant and then flash frozen in liquid nitrogen. The dataset was collected at 100 K at beamline 19-ID of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory (Argonne, IL, USA). The diffraction data were processed using the HKL-3000 software suite [27] (Table 4). The structure was first determined to 4.0 Å resolution by molecular replacement. A search model constructed from the crystal structure of CD69 (Protein Databank code 1E8I) [15] was used for the simultaneous rotation and translation search of two molecules by the software EPMR [28]. The search yielded an unambiguous solution in the $P6_1$ space group with an initial R_{crys} of 42.2% and a correlation coefficient of 0.57. In the rigid body refinement, both molecules of the dimer were allowed to move independently. During the restrained refinement procedure, noncrystallographic symmetry restraints were applied to the CD69 monomers. Rigid body refinement and subsequent restrained refinement protocol were performed with the software REFMAC 5.1.24 [29] from the CCP4 package [30]. For manual model rebuilding, the software COOT was used [31]. The structure was refined to a crystallographic *R*-factor of 19.0% at 1.8 Å resolution. Further refinement statistics are shown in Table 4.

Table 4. Crystal parameters, data collection and refinement statistics. Values in parentheses represent those obtained for the highest resolution shell.

Space group	$P6_1$
Unit cell parameters (Å, °)	$a = b = 85.69$ and $c = 61.88$ $\alpha = \beta = 90.0$ and $\gamma = 120.00$
Resolution (Å)	75.54–1.80 (1.83–1.80)
Total number of observations to 1.80 Å	422144
No. unique reflections	22276
Completeness	98.03% (76.5%)
$I/\sigma\ (I)$	29.5 (2.6)
R_{sym} [a]	0.045 (0.575)
Refinement resolution (Å)	74.54–1.80
Rmsd bond length from ideal (Å)	0.015
Rmsd bond angles from ideal (°)	1.41
R_{crys} [b]	0.190
R_{free} [c]	0.222

[a] $R_{sym} = \sum|I - \langle I\rangle|/\sum\langle I\rangle$, where I is the observed intensity and $\langle I\rangle$ is the mean intensity of multiple observations of symmetry-related reflections. [b] $R_{crys} = \sum||F_o| - |F_c||/\sum|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes. [c] R_{free} is as for R_{crys} but calculated for a randomly chosen 5% of reflections that were omitted from the refinement.

Structure coordinates

The coordinates and structure factors have been submitted to the RCSB Protein Databank under accession code 3CCK.

Thermal denaturation experiments

Recombinant proteins were diluted to 0.5 $mg \cdot mL^{-1}$, and UV spectra were taken in the 200–300 nm range in Beckman DU-70 spectrophotometer (Beckman Coulter, Fullerton, CA, USA) equipped with the heated cuvette. The initial UV scan was taken at 25 °C, after which the temperature in the cuvette was increased in 5 °C increments up to the denaturation temperature indicated by sudden increase in the absorbance. The experiment was performed in 10 mM Mes (pH 5.8) with 50 mM NaCl and 1 mM NaN_3, in the same buffer containing 5 mM dithiothreitol, and also in 10 mM Pipes buffer (pH 6.8) with 50 mM NaCl, 1 mM NaN_3 and 5 mM dithiothreitol, and in 10 mM Tris buffer (pH 7.8) with 50 mM NaCl, 1 mM NaN_3 and 5 mM dithiothreitol. From the complete spectra, only the absorbance at 280 nm was extracted and plotted because of technical difficulties with buffer subtraction in the far UV region. Alternatively, protein stability was followed using FTIR spectroscopy.

Infrared spectra were recorded with a Bruker IFS-66/S FTIR spectrometer (Bruker, Ettlingen, Germany) using a standard MIR source, a KBr beamsplitter and an MCT detector. Four thousand scans were collected with 4 cm^{-1}

spectral resolution and a Happ–Genzel apodization function. Aqueous protein solution was measured at the indicated temperature in a thermostated CaF_2-BioCell with 10 μm path length (BioTools, Jupiter, FL, USA). The spectral contribution of a buffer was corrected using the standard algorithm [32]. The spectrum of water vapors was subtracted and finally the spectrum was normalized. The fraction content of the secondary structure elements was calculated using the procedure CONTIN with a set of 16 reference proteins [32].

Sedimentation velocity and sedimentation equilibrium measurements

Initial sedimentation measurements were performed in the preparative ultracentrifuge (Beckman Optima LE-80) using the discontinuous sucrose gradient [1 mL each of 40%, 30%, 20% and 10% sucrose in 10 mM Mes buffer (pH 5.8) with 50 mM NaCl and 1 mM NaN_3] placed into the UltraClear polycarbonate tube (Beckman Coulter). The sucrose gradient was overlayed with 0.2 mL of protein samples (10 mg·mL^{-1}) that were spun in SW-60 rotor for 16 h at 392 000 $\boldsymbol{g}_{av}$. Then, 0.35 mL samples were collected from the top to the bottom of the tube, and the protein distribution was examined in 10 μL by 17.5% SDS/PAGE. Further sedimentation velocity and sedimentation equilibrium measurements were performed using an analytical ultracentrifuge ProteomeLabXL-I (Beckman Coulter) using (depending on the sample concentration) absorbance or laser interference optics and an An50Ti rotor. Before the experiment, 0.5 mL samples of recombinant CD69 were dialyzed for 20 h against 2 L of 10 mM Tris–HCl (pH 7.8) with 150 mM NaCl and 1 mM NaN_3. Sedimentation velocity experiments were carried out at 130 000 $\boldsymbol{g}_{av}$ using an Epon aluminium-filled centerpiece (Beckman Coulter). Sample (400 μL) and dialysate (450 μL) were loaded in the sample and reference cells, respectively. Absorbance scans were performed at 280 nm at 5 min intervals and 0.003 cm spatial resolution, and the data were analyzed by the second-moment method using the software provided by the manufacturer. The partial specific volume of CD69NG70 (0.73 mL·g^{-1}) was calculated using the software SEDNTERP (http://www.rasmb.bbri.org). The molecular mass of CD69NG70 was calculated using the approximation for the spherical molecules, as described previously by Lebowitz *et al.* [33], according to the formula:

$$S_{sphere} = 0.012([M^{2/3}(1 - v\rho)]/v^{1/3}).$$

For sedimentation equilibrium experiments, the protein was examined at three different concentrations (0.8, 0.4 and 0.2 mg·mL^{-1}) in three different sample cuvettes placed into an An50Ti rotor. The rotor was spun at 130 000 $\boldsymbol{g}_{av}$ for 2 h, followed by 24 000 $\boldsymbol{g}_{av}$ for 18 h. Thereafter, two consecutive absorbance scans were taken, one at the end of 18 h period, and another after additional 2 h. The equilibrium distribution from three different loading concentrations and up to three different rotor speeds (24 000, 14 000 and 7500 $\boldsymbol{g}_{av}$, respectively) were analyzed using the nonlinear curve fit algorithm in the software package supplied with the centrifuge [33].

NMR measurements

All NMR experiments were run at 300 K on Bruker Avance 600 MHz spectrometer equipped with the cryogenic H/C/N TCI probehead. ^{1}H-^{15}N HSQC spectra were used as a routine check of protein folding and stability during the sample preparation; 0.9 mM ^{13}C/^{15}N-labeled and 0.3 mM ^{15}N-labeled CD69NG70 samples were used for the assignment and relaxation measurements, respectively. The sample buffer consisted of 10 mM Mes (pH 5.8), 50 mM NaCl, 1 mM NaN_3 and 10% D_2O. The standard set of triple resonance experiments [34] and ^{13}C/^{15}N-edited NOESY [35] were used to obtain sequential assignment. The assignment was confirmed by checking side-chain resonances of selected residues in the HCCH-TOCSY spectra [34]. The ^{15}N T1, T2 and steady-state ^{1}H-^{15}N NOE experiments were run as described by Farrow *et al.* [36]. The T1 and T2 relaxation delays were sampled at 11, 56, 134, 235, 381, 560, 896, 1344 ms and 16, 31, 62, 94, 156, 219, 250, 406 ms, respectively. All spectra were processed using the software NMRPIPE [37] and analyzed using the software SPARKY (T. D. Goddard and D. G. Kneller, SPARKY 3; University of California, San Francisco, CA, USA).

Relaxation analysis and hydrodynamic calculations

The backbone amide dynamic parameters were derived in the spirit of the Lipari–Szabo model-free approach [38,39] using the software RELAX [40,41]. The apparent rotational diffusion coefficient (defined as $1/6\tau_m$, where τ_m is the overall correlation time) was fitted separately for each residue for the sake of comparison with the hydrodynamics simulations. The internuclear N-H distance of 0.102 nm and ^{15}N chemical shift anisotropy of 160 p.p.m. were used. The hydrodynamic calculations were performed using software HYDRONMR [42]. Viscosity was set to 0.852 mPa·s. Default effective radius (0.32 nm) and minibeads of six radii in the range 0.15–0.2 nm were used. The structural models were derived from the X-ray structure (Protein Databank code 3CCK). Coordinates of the dimeric and monomeric structures were taken directly from the Protein Databank file. In addition, sets of monomeric and dimeric full-length structures were modeled by adding 11 N-terminal residues and running short molecular dynamics simulations using the software CNS 1.2. The atoms taken from the X-ray structure

were kept fixed during the simulation, whereas the added 11 N-terminal amino acids were restrained by the measured chemical shifts only. The simulation protocol consisted of a 15 ps high temperature (50 000 K) torsion dynamics run, two 15 ps cooling stages (the first one employing torsion dynamics, the second one employing cartesian dynamics) and gradient minimization. Out of 100 structures calculated for each set, 20 structures with the lowest energy were used in hydrodynamics calculations.

NMR data

NMR data were submitted to the BioMagResBank database with an accession number 15703.

Protein distribution *in vivo*

All animal experiments were approved by the Institute Ethical Committee and were performed in accordance with the European Communities Council Directive of 24 November 1986 (86/609/EEC). Recombinant proteins for animal studies were made free of lipopolysaccharide using polymyxin B resin (Bio-Rad, Hercules, CA, USA), radiolabeled using $Na^{125}I$ (GE Healthcare) to a specific activity of 10^6 Bq·μg^{-1}, and repurified by reverse phase chromatography to remove the noncovalently bound iodine. Ten micrograms of each radiolabeled protein was administred in 50 μL of NaCl/P_i into the tail vein of two male C57BL/6 mice, aged 7–8 weeks old (Charles River, Wilmington, MA, USA) that had been accommodated for at least 2 weeks in the conventional housing facility. Blood samples (50 μL) were collected 1, 4, 10 and 24 h after the administration of radioactivity in view of our experience with the rapid effect of the ligands for these receptors on the immune system [21]. From each sample, 20 μL was mixed with 80 μL of 10 mM NH_4Cl and after 1 h at room temperature to lyse the erythrocytes, was used for hemoglobin determination by direct spectrophotometry at 400 nm, and for radioactivity counting. From the remaining 30 μL of sample, the most abundant serum proteins were removed using IgY12 resin (Beckman Coulter), and two 10 μL aliquots of the remaining proteins were resolved by 15% SDS/PAGE, and the gels were exposed to the autoradiographic films Agfa CP-VB (Agfa-Gevaert, Mortsel, Belgium) with intensifying screens, or developed by western blotting using mouse monoclonal antibodies against human CD69 (BL-Ac/p26) [17], rat CD69 [5] and mouse CD69 (Invitrogen), and ECL detection (GE Healthcare). The remaining 10 μL of the clarified protein was used for the solid phase binding using GlcNAc agarose (Sigma, St Louis, MO, USA).

Plasma half-life was calculated using the formula: $c = c_0 \times e^{(-kt)}$, where c is the concentration at the indicated time, c_0 is the initial concentration and k is the first order rate constant [43]. The calculation of the plasma clearance and the apparent volume of distribution was based on a one-compartment kinetic model with instantaneous absorption according to the equation: $C_P = (D/V_d) \times e^{(-CL/Vd)t}$, where C_P is the plasma level at time t, CL is the clearance, D is the administered dose and V_d is the apparent volume of distribution [43].

For organ distribution studies, mice were killed 24 h after injection of the radiolabeled protein, individual organs (spleen, kidney, liver, muscle, skin) were collected, and dissolved completely (60 °C, 48 h) in NCS-II tissue solubilizer, in accordance with the manufacturer's instructions (GE Healthcare) before scintillation counting. The difference between the total initial dose of radioactivity, and the sum of radioactivities recovered in the individual organs, and in the urine and faeces, was designated as the rest of the body (rest), consisting mostly of the bowel, heart, bones, brain and reproductive organs. Because all six mice were housed in a single cage during the experiment, the radioactivity recovered in urine and faeces represents the average value. In an alternative time course experiment, liver and kidney were collected after 1, 4, 10 and 24 h after injection, and processed as described above.

Acknowledgements

This work is dedicated to Professor Danuše Sofrová and Professor Marie Tichá. We thank Pavlína Řezáčová for data collection at the APS synchrotron facility at Argonne National Laboratory; Jan Bílý for his help with the thermal denaturation experiments; Anna Fišerová, Markéta Vančurová and Jozef Hritz for their help with the experiments; and the reviewers for their valuable comments on the manuscript. This work was supported by Ministry of Education of Czech Republic (MSM 21620808, MSM 21620835, MSM 21622413, LC 545, LC 6030, LC 7017 and 1M 4635608802), by the Institutional Research Concept for the Institute of Microbiology (AVOZ 50200510), by the Grant Agency of the Academy of Sciences (IAA5020403 and IAA500-200509), and by the European Commission, Project SPINE2-Complexes (contract LSHG-CT-2006-031220).

References

1 Testi R, D'Ambrosio D, De Maria R & Santoni A (1994) The CD69 receptor: a multipurpose cell surface trigger for hematopoietic cells. *Immunol Today* **15**, 479–483.

2 Gerosa F, Tommasi M, Scardoni M, Accolla RS, Pozzan T, Libonati M, Tridente G & Carra G (1991) Structural analysis of the CD69 early activation antigen by two monoclonal antibodies directed to different epitopes. *Mol Immunol* **28**, 159–168.

3 Hamann J, Fiebig H & Strauss M (1993) Expression cloning of the early activation antigen CD69, a type II

integral membrane protein with a C-type lectin domain. *J Immunol* **150**, 4920–4927.
4 Lopez-Cabrera M, Santis AG, Fernandez-Ruiz E, Blacher R, Esch F, Sanchez-Mateos P & Sanchez-Madrid F (1993) Molecular cloning, expression, and chromosomal localization of the human earliest lymphocyte activation antigen AIM/CD69, a new member of the C-type lectin superfamily of signal-transmitting receptors. *J Exp Med* **178**, 537–547.
5 Ziegler SF, Ramsdell F, Hjerrild KA, Armitage RJ, Grabstein KH, Hennen KB, Fatrách T, Fanslow WC, Shevach EM & Alderson MR (1993) Molecular characterization of the early activation antigen CD69: a type II membrane glycoprotein related to a family of natural killer cell activation antigens. *Eur J Immunol* **23**, 1643–1648.
6 Sancho D, Santis AG, Alonso-Lebrero JL, Viedma F, Tejedor R & Sanchez-Madrid F (2000) Functional analysis of ligand-binding and signal transduction domains of CD69 and CD23 C-type lectin leukocyte receptors. *J Immunol* **165**, 3868–3875.
7 Pisegna S, Zingoni A, Pirozzi G, Cinque B, Cifone MG, Morrone S, Picolli M, Frati L, Palmieri G & Santoni A (2002) Src-dependent Syk activation controls CD69-mediated signaling and function of human NK cells. *J Immunol* **169**, 68–74.
8 Bezouška K, Nepovím A, Horváth O, Pospíšil M, Hamann J & Feizi T (1995) CD69 antigen of human leukocytes is a calcium-dependent carbohydrate-binding protein. *Biochem Biophys Res Commun* **208**, 68–74.
9 Risso A, Smilovich D, Capra MC, Baldissarro I, Yan G, Bargellesi A & Cosulich ME (1991) CD69 in resting and activated T lymphocytes. Its association with a GTP binding protein and biochemical requirements for its expression. *J Immunol* **146**, 4105–4114.
10 Moretta A, Poggi A, Pende D, Tripodi G, Orengo AM, Pella N, Augugliaro R, Bottino C, Ciccone E & Moretta L (1991) CD69-mediated pathway of lymphocyte activation: anti-CD69 monoclonal antibodies trigger the cytolytic activity of different lymphoid effector cells with the exception of cytolytic T lymphocytes expressing T cell receptor α/β. *J Exp Med* **174**, 1393–1398.
11 Borrego F, Robertson MJ, Ritz J, Pena J & Solana R (1999) CD69 is a stimulatory receptor for natural killer cells and its cytotoxic effect is blocked by CD94 inhibitory receptor. *Immunology* **97**, 159–165.
12 Sancho D, Gomez M & Sanchez-Madrid F (2004) CD69 is an immunoregulatory molecule induced following activation. *Trends Immunol* **26**, 136–140.
13 Esplugues E, Sancho D, Vega-Ramos J, Martinez C, Syrbe U, Hamann A, Engel P, Sanchez-Madrid F & Lauzurica P (2003) Enhanced antitumor immunity in mice deficient in CD69. *J Exp Med* **197**, 1093–1106.
14 Natarajan K, Sawicki MW, Margulies DH & Mariuzza RA (2000) Crystal structure of human CD69: a C-type lectin-like activation marker of hematopoietic cells. *Biochemistry* **39**, 14779–14786.
15 Llera AS, Viedma F, Sanchez-Madrid F & Tormo J (2001) Crystal structure of the C-type lectin-like domain from the human hematopoietic cell receptor CD69. *J Biol Chem* **276**, 7312–7319.
16 Childs RA, Galustian C, Lawson AM, Dougan G, Benwell K, Frankel G & Feizi T (2000) Recombinant soluble human CD69 dimer produced in *Escherichia coli:* reevaluation of saccharide binding. *Biochem Biophys Res Commun* **266**, 19–23.
17 Pavlíček J, Sopko B, Ettrich R, Kopecký V, Baumruk V, Man P, Havlíček V, Vrbacký M, Martínková L, Křen V *et al.* (2003) Molecular characterization of binding of calcium and carbohydrates by an early activation antigen of lymphocytes CD69. *Biochemistry* **42**, 9295–9306.
18 Pavlíček J, Kavan D, Pompach P, Novák P, Lukšan O & Bezouška K (2004) Lymphocyte activation receptors: new structural paradigm in group V of C-type animal lectins. *Biochem Soc Trans* **32**, 1124–1126.
19 Ziegler SF, Levin SD, Johnson L, Copeland NG, Gilbert DJ, Jenkins NA, Baker E, Sutherland GR, Feldhaus AL & Ramsdell F (1994) The mouse CD69 gene. Structure, expression, and mapping to the NK gene complex. *J Immunol* **152**, 1228–1236.
20 Graham DB, Bell MP, Huntoon CJ, Griffin MD, Tai X, Singer A & McKean DJ (2006) CD28 ligation costimulates cell death but not maturation of double-positive thymocytes due to defective ERK MAPK signaling. *J Immunol* **177**, 6098–6107.
21 Vannucci L, Fišerová A, Sadalapure K, Lindhorst TK, Kuldová M, Rossmann P, Horváth O, Křen V, Krist P, Bezouška K *et al.* (2003) Effects of *N*-acetyl-D-glucosamine-coated glycodendrimers as biological modulators in the B16F10 melanoma model *in vivo*. *Int J Oncol* **23**, 285–296.
22 North J, Bakhsh I, Marden C, Pittman H, Addison E, Navarrete C, Anderson R & Lowdell MW (2007) Tumor-primed human natural killer cells lyse NK-resistant tumor targets: evidence of a two-stage process in resting NK cell activation. *J Immunol* **178**, 85–94.
23 Shiow LR, Rosen DB, Brdičková N, Xu Y, An J, Langer LL, Cyster JG & Matloubian M (2006) CD69 acts downstream of interferon-α/β to inhibit $S1P_1$ and lymphocyte egress from lymphoid organs. *Nature* **440**, 540–544.
24 Konjevic G, Jovic V, Vuletic A, Radulovic S, Jelic S & Spuzic I (2007) CD69 on CD56+ NK cells and response to chemoimmunotherapy in metastatic melanoma. *Eur J Clin Invest* **37**, 887–896.
25 Valez-Gomez M, Reyburn HT, Mandelboim M & Strominger JL (1998) Kinetics of interaction of HLA-C ligands with natural killer cell inhibitory receptors. *Immunity* **9**, 337–344.
26 de la Torre JG, Huertas ML & Carrasco B (2000) Calculation of hydrodynamic properties of globular

proteins from their atomic-level structure. *Biophys J* **78**, 719–730.
27 Minor W, Cymborowski M, Otwinowski Z & Chruszcz M (2006) HKL-3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes. *Acta Crystallogr D* **62**, 859–866.
28 Kissinger CR, Gehlhaar DK & Fogel DB (1999) Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr D* **55**, 484–491.
29 Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structure by the maximum-likelihood method. *Acta Crystallogr D* **53**, 240–255.
30 Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* **50**, 760–763.
31 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60**, 2126–2132.
32 Dousseau F, Therrien M & Pézole M (1989) On the spectral subtraction of water from the FT-IR spectra of aqueous solution of proteins. *Appl Spectrosc* **43**, 538–542.
33 Lebowitz J, Lewis MS & Schuck P (2002) Modern analytical ultracentrifuge in protein science: a tutorial review. *Protein Sci* **11**, 2067–2079.
34 Sattler M, Schleucher J & Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog NMR Spectrosc* **34**, 93–158.
35 Xia Y, Arrowsmith CH & Gao X (2003) 1HC and 1HN total NOE correlations in a single 3D NMR experiment. 15N and 13C time-sharing in t1 and t2 dimensions for simultaneous data acquisition. *J Biomol NMR* **27**, 193–203.
36 Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD & Kay LE (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984–6003.
37 Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J & Bax A (1995) NMRPipe: a multidimensional spectral processing system based on Unix pipes. *J Biomol NMR* **6**, 277–293.
38 Lipari G & Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* **104**, 4546–4559.
39 Lipari G & Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc* **104**, 4559–4570.
40 d'Auvergne EJ & Gooley PR (2008) Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J Biomol NMR* **40**, 107–119.
41 d'Auvergne EJ & Gooley PR (2008) Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. *J Biomol NMR* **40**, 121–133.
42 Bernadó P, de la Torre JG & Pons M (2000) Interpretation of 15N NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR. *J Biomol NMR* **23**, 139–150.
43 Ruffo S, Messori A, Grasela TH, Longo G, Donati-Cori G, Matucci M, Morfini M & Tendi E (1985) A calculator program for clinical application of the Bayesian method of predicting plasma drug levels. *Comput Programs Biomed.* **19**, 167–177.

Supporting information

The following supplementary material is available:
Table S1. Secondary structure elements measured by FTIR spectroscopy in soluble CD69 proteins subjected to thermal stress.
Table S2. Secondary structure elements measured by FTIR spectroscopy in soluble CD69 proteins subjected to pH stress at temperatures 5 °C below the T_d.

This supplementary material can be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

4.1.3 Lymphocyte activation receptors: new structural paradigms in group V of C-type animal lectins

Lymphocyte activation receptors: new structural paradigms in group V of C-type animal lectins

J. Pavlíček*[1], D. Kavan*†, P. Pompach*†, P. Novák†, O. Lukšan* and K. Bezouška*†

*Department of Biochemistry, Faculty of Science, Charles University, Praha, Czech Republic, and †Institute of Microbiology, Academy of Sciences of Czech Republic, Praha, Czech Republic

Abstract

The structure–function relationship in group V of C-type animal lectins remains incompletely understood despite the new structures of NK (natural killer) cell receptors that have been solved recently. Recombinant, soluble forms of rat and human NKR-P1 and CD69 that we obtained after *in vitro* refolding were analysed by Fourier transform–ion cyclotron resonance MS and heteronuclear NMR (^{1}H-^{15}N correlation). In NKR-P1, calcium may not be removed by chelating agents because of the very high affinity of binding. In CD69, incorporation of calcium causes a structural shift in several amino acids important for the interaction with carbohydrates. Structural studies have also allowed us to understand an interesting preference of these receptors for either linear (NKR-P1) or branched (CD69) carbohydrate sequences.

Introduction

Lectin-type receptors of the C-type lectin family are important antigens at the surface of immune cells. Recently, special attention has been paid to molecules classified as group V of calcium-dependent animal lectins because of their crucial role in the regulation of NK (natural killer) cells. Although research on the inhibitory receptors necessary to keep NK cells under control has made good progress, investigations on the activation receptors have lagged behind. This is somewhat surprising since studies of these receptors are crucial from the point of view of tumour immunotherapies [1,2] and general (non-specific) stimulation of the immune system.

New structural paradigms in group V of C-type animal lectins

A considerable evolutionary divergence of the group V from the classical calcium-dependent animal lectins has resulted in new variations of the basic structural fold accommodating new ligands for these receptors. Inhibitory NK cell receptors such as Ly-49A [3] or CD94 [4] remodelled their ligand-binding loops considerably to recognize their specific inhibitory ligands, MHC class I glycoproteins. Similarly, NKG2D developed a remarkable plasticity in its ligand-binding surface that allows it to bind stress-induced proteins as the specific ligands [5].

Two NK cell activation receptors studied in our laboratory, rat NKR-P1 and human CD69, are unique for their high affinities for the classical ligands of C-type lectins, calcium and carbohydrates. Calcium has very high affinity for the rat NKR-P1, which makes it resistant to removal by chelating agents [6]. In human CD69, a close-up of the binding site for calcium results in structural changes that are essential for the formation of the high-affinity carbohydrate-binding site [7]. Carbohydrates interact with the above proteins over an extensive surface area, but the nature of the oligosaccharide-binding sites is unique for individual receptors. In rat NKR-P1, a binding groove exists that accommodates well the linear oligosaccharides but not branched oligosaccharides [8]. On the other hand, the sugar-binding sites in human CD69 are placed at three distinct locations [7], predicting the preference for branched structures.

New results with rat CD69

To evaluate the validity of the above findings in both rats and humans (two key organisms in which the NK cell receptors have been studied), we initiated structural studies of rat CD69. Using a bacterial expression construct in which the membrane-proximal dimerization cysteine residue has been omitted (Figure 1A), we produced a very stable, pure monomeric rat CD69 suitable for structural and ligand-identification studies. The presence of three disulphide bonds in the protein has been proven by Fourier transform–ion cyclotron resonance MS (Figure 1B). This protein has a stable fold as shown by a spread of the signals in the ^{1}H-^{15}N NMR correlation spectrum (Figure 1C). The exact three-dimensional structure of this protein is being solved by a combination of protein crystallography (Figure 1D) and ^{13}C-^{15}N NMR correlation experiments using the double-labelled (^{15}N/^{13}C) protein.

Rat CD69 binds 1 mol of calcium per mol of protein with an affinity very similar to human CD69, despite the mutation in one of the calcium-chelating amino acids (Figure 2A; [7]). GlcNAc, which is the best ligand among the monosaccharides, binds to three sites in the protein with micromolar affinity (Figure 2B). The oligosaccharide specificity of rat CD69

Key words: lectin-type receptor, ligand identification, lymphocyte, natural killing.

Abbreviation used: NK cells, natural killer cells.

[1]**Correspondence should be addressed at Hlavova 8, CZ-12840 Praha 2, Czech Republic (email pavlicek@biomed.cas.cz).**

©2004 Biochemical Society

Figure 1 | Characterization of the monomeric rat CD69 protein

(**A**) Amino acid sequence of rat CD69 with the initiation methionine (green) and the expected pairing of cysteine residues (red). (**B**) Characterization of the native and reduced proteins by Fourier transform–ion cyclotron resonance MS. (**C**) ^{1}H-^{15}N correlation NMR spectrum obtained after 24 h using a 0.28 mM solution of ^{15}N-labelled protein. (**D**) Crystals of rat CD69 were grown in a hanging drop with a 0.35 mM solution of protein, 24 h after the addition of precipitant.

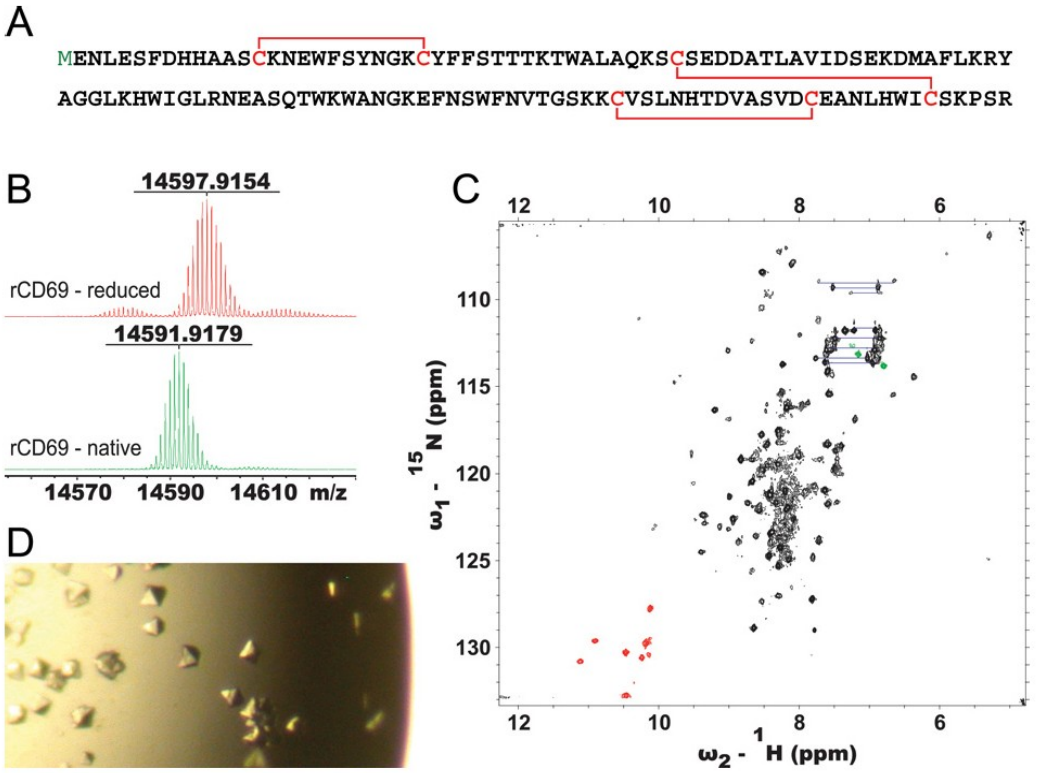
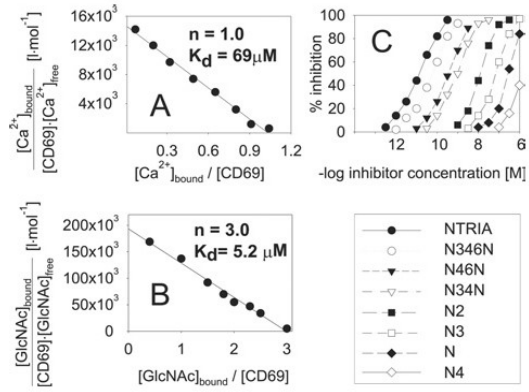


Figure 2 | Binding of ligands to rat CD69

(**A**) Scatchard plot for the binding of calcium. (**B**) Scatchard plot for the binding of GlcNAc. (**C**) Inhibition of the binding of rat CD69 to $GlcNAc_{23}BSA$ by GlcNAc (N), chitobiose (N2), chitotriose (N3), chitotetraose (N4), GlcNAcβ1-3(GlcNAcβ1-4)GlcNAc (N34N), GlcNAcβ1-4-(GlcNAcβ1-6)GlcNAc (N46N), GlcNAcβ1-3(GlcNAcβ1-4)(GlcNAcβ1-6)-GlcNAc (N346N) and by the triantennary oligosaccharide from ovomucoid [GlcNAcβ1-2(GlcNAcβ1-4)Manα1-3](GlcNAcβ1-4)(GlcNAcβ1-2Manα1-6)Manβ1-4GlcNAcβ1-4GlcNAc (NTRIA).



is again very similar to that of its human orthologue: linear oligosaccharides of the chito-oligomer series are not good ligands, whereas the branched oligosaccharides are very-high-affinity ligands with IC_{50} as low as 10^{-10} M (Figure 2C).

Conclusions

Our recent studies revealed the exclusive specificity of lectin-type activation receptors of NK cells for oligosaccharide structures, linear or branched. Our results with rat CD69 support its similarity to the human orthologue with regard to binding of calcium and carbohydrates. Structural investigations of other activation receptors coded by individual genes in the NK gene complex will be important to reveal both natural ligands and their synthetic mimetics. The roles of these

©2004 Biochemical Society

substances as anti-tumour and immunoactivating compounds are being evaluated [2].

We are grateful to M. Nálezková, L. Žídek and V. Sklenář for assistance with NMR spectroscopy and to P. Řezáčová for her help with protein crystallization. This work was supported by Ministry of Education of Czech Republic (MSM113100001), by the Institutional Research Concept AVOZ5020903 and by the Grant Agency of the Academy of Sciences of Czech Republic (IAA5020403).

References

1 Bezouška, K., Yuen, C.-T., O'Brien, J., Childs, R.A., Chai, W., Lawson, A.M., Drbal, K., Fišerová, A., Pospíšil, M. and Feizi, T. (1994) Nature (London) **372**, 150–158
2 Pospíšil, M., Vannucci, L., Horváth, O., Fišerová, A., Krausová, K., Bezouška, K. and Mosca, F. (2000) Int. J. Oncol. **16**, 267–276
3 Boyington, J.C., Riaz, A.N., Patamawenu, A., Coligan, J.E., Brooks, A.G. and Sun, P.D. (1999) Immunity **10**, 75–82
4 Tormo, J., Natarajan, K., Margulies, D.H. and Mariuzza, R.A. (1999) Nature (London) **402**, 623–631
5 Strong, R.K. (2001) Mol. Immunol. **38**, 1029–1037
6 Bezouška, K., Vlahas, G., Horváth, O., Jinochová, G., Fišerová, A., Giorda, R., Chambers, W.H., Feizi, T. and Pospíšil, M. (1994) J. Biol. Chem. **269**, 16945–16952
7 Pavlíček, J., Sopko, B., Ettrich, R., Kopecký, V., Baumruk, V., Man, P., Havlíček, V., Vrbacký, M., Martínková, L., Křen, V. et al. (2003) Biochemistry **42**, 9295–9306
8 Plíhal, O., Byrtusová, P., Pavlíček, J., Mihok, L., Ettrich, R., Man, P., Pompach, P., Havlíček, V., Hušáková, L. and Bezouška, K. (2004) Collect. Czech. Chem. Commun. **69**, 631–644

Received 2 July 2004

©2004 Biochemical Society

4.1.4 Modified electrophoretic and digestion conditions allow a simplified mass spectrometric evaluation of disulfide bonds

Research Article

Journal of MASS SPECTROMETRY

Received: 15 January 2009 Accepted: 21 May 2009 Published online in Wiley Interscience: 23 June 2009

(www.interscience.com) DOI 10.1002/jms.1609

Modified electrophoretic and digestion conditions allow a simplified mass spectrometric evaluation of disulfide bonds

Petr Pompach,[a,b] Petr Man,[a,b]* Daniel Kavan,[a,b] Kateřina Hofbauerová,[a,c] Vinay Kumar,[a,b] Karel Bezouška,[a,b] Vladimír Havlíček[a,d] and Petr Novák[a,b]*

Proper formation of disulfide bonds in proteins is a prerequisite to their stability and function. Information on disulfide pattern may therefore serve as an indication of the proper folding of recombinant proteins, and can also be used in protein homology modeling for the purpose of structure refinement. Protein handling and digestion at basic pH leads to disulfide bond scrambling. That is why the samples are usually treated and digested at low pH where no scrambling occurs. Unfortunately, the specific proteases used in protein research are active at high pH values. Here, we present a complete sample handling protocol, which allows processing of disulfide containing proteins at basic pH. We modified the standard SDS gel electrophoresis and protein digestion conditions by the addition of an oxidative agent, cystamine. This modification prevented disulfide scrambling, which we otherwise observed in the samples handled according to the general protocol. Lysozyme from hen egg was used as a model protein for the development of the method. We then applied our protocol to human leukocyte antigen CD69, for which the disulfide bonding is known, but only for its monomeric form. In addition, the disulfide arrangement was then '*de novo*' identified in the recombinant murine leukocyte receptor NKR-P1A and in the larger glycosylated proteins β-*N*-acetylhexosaminidases from *Aspergillus oryzae* and *Penicillium oxalicum*. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: disulfide bond; FTMS; cystamine; gel electrophoresis; liquid chromatography

Introduction

Disulfide bonds influence protein folding and are also important for the stability of protein structure.[1,2] The knowledge of their arrangement may facilitate protein homology modeling and serve as an indication of proper folding in recombinant protein production.[3,4] Over the past decades, a large number of different techniques for the determination of disulfide bridges were developed. These protocols use chemical modifications of proteins followed by their chemical or enzymatic fragmentation and subsequent analysis by Edman degradation or mass spectrometric techniques. In principle, all methods are aimed to address three main questions. First, how many disulfide bonds and free cysteines are in the protein of interest. Second, how to distinguish disulfide linked peptides in the peptide mixture obtained after proteolysis of the protein. And third, how to identify the sequence of the individual disulfide linked peptides and the exact positions of the linkages. The first question regarding the number of disulfide bonds and cysteine status can be easily answered by differential alkylation. The mass difference between the intact protein and the alkylkated is compared directly and after reduction.[5] The further question, where the exact disulfide bonding is located, requires rather complex techniques that can be roughly divided into three groups.[6]

The first one encompasses techniques using full or partial reduction of a protein, combined with cysteine alkylation. Differential reduction was first used in the early nineties by Smith and colleagues and is still used by many groups.[7,8] Alkylation by 2-nitro-5-thiocyanobenzoic (NTCB) acid followed by chemical cleavage at high pH was introduced in 1973 by Jacobson *et al*.[9] Twenty years later, the protocol was modified and coupled to mass spectrometry by J. T. Watson's group.[10,11] An improvement to this method, called 'negative signature mass', allows automated interpretation, which is based on stepwise elimination of unlikely bonds from the list of all theoretical linkages.[12] Introduction of new reducing agents working at low pH had significant impact on disulfide bond determination as well. Such chemical agents, in particular Tris(2-carboxyethyl)-phosphine (TCEP), have quickly become popular and are still widely used.[13–18]

The second class of techniques is based on the hydrolysis of intact proteins and subsequent analysis by different mass spectrometric techniques. Some research groups determined the disulfide bonds using negative electrospray ionization where characteristic fragmentation pattern or loss of H_2S_2 occured.[19–21] The group of Gorman developed a protocol where isotopic labeling

* *Correspondence to: Petr Novák and Petr Man, Laboratory of Molecular Structure Characterization, Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, Prague, 142 20, Czech Republic. E-mail: pnovak@biomed.cas.cz; pman@biomed.cas.cz*

a *Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague, Czech Republic*

b *Department of Biochemistry, Faculty of Science, Charles University, Prague, Czech Republic*

c *Institute of Physics, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic*

d *Department of Analytical Chemistry, Palacky University, Olomouc, Czech Republic*

Copyright © 2009 John Wiley & Sons, Ltd.

by ^{18}O during enzymatic proteolysis facilitated identification of cystine peptides in complex peptic digests.[22,23]

The third strategy for characterization of disulfide bonds is based on the top down approach.[24] The analysis of native proteins and the absence of digestion and/or protein chemical modifications eliminate the possibility of disulfide bond scrambling. The initial studies used electron capture dissociation (ECD) but later other dissociation techniques, e.g. collision-induced dissociation (CID), infrared multiphoton dissociation (IRMPD) or electron transfer dissociation (ETD) were successfully applied as well.[25–30]

All the abovementioned techniques are capable of solving nearly any disulfide bonding pattern. Despite wide applicability, their drawbacks are mainly in laborious sample preparation and relatively difficult data interpretation. For example, pepsinolysis at low pH prevents scrambling but the non-specific ragged cleavage makes the assignment of disulfide bonds rather complex. Partial reduction and modification of the protein followed by its specific chemical cleavage has to be optimized for each protein separately, which hampers identification of S–S bonds in protein mixtures. The top-down approach that does not require extensive sample preparation or purification is still not easily applicable to large, heavily modified proteins. However, in many cases, highly sophisticated protocols are not necessary as the disulfide bonding pattern is not complicated.

One important area of application is the recombinant protein production (for structure elucidation or therapeutic purposes) where the proper S–S bond formation has to be followed at different stages of the process development. For this purpose, we developed a protocol exploiting the high mass accuracy and resolving power of FTMS. Further simplification of the data analysis was achieved through the use of specific proteases and automated software tools allowing rapid unaided data interpretation. Because specific proteases work at basic pH, where disulfide bond scrambling occurs, an additional protocol that prevents this effect had to be also developed.

We have shown the feasibility of our new methodology on lysozyme, CD69 receptor bearing an interchain dimerization bridge, NKR-P1A receptor and on two larger, glycosylated proteins – fungal β-*N*-acetylhexosaminidases.

The human protein CD69 is a type II transmembrane glycoprotein related to C-type lectins. The transmembrane domain is responsible for cell signaling and cellular activation. The neck region is important for dimerization of the protein and the globular domain mediates the binding of monoclonal antibodies and physiological ligands.[31] The crystal structure of the monomeric extracellular domain of human CD69 recombinant protein was solved by Natarajan and coworkers.[32] Importantly, this molecule exists as a disulfide-linked homodimer and it possesses intermolecular cysteine bridges.[31]

Mouse NKR-P1A (CD161) shares some common features with CD69.[33] NKR-P1A is an activating receptor which mediates cellular cytotoxicity and INF-γ production. Despite the fact that the members of NKR-P1 family were among the first NK cell receptors described their structure still remains unknown.[34–36]

Both receptors mentioned above are produced in bacterial expression systems and refolded '*in vitro*'. Prior to functional and/or structural studies the conditions favoring proper S–S bonding have to be optimized. Thus, there is a need for a method allowing rapid determination of the disulfide status.

β-*N*-acetylhexosaminidases from *Aspergillus oryzae* and *Penicillinum oxalicum* are exoglycosidases which can hydrolyze β-linked GlcNAc or GalNAc residues. These enzymes are widely distributed in organisms from bacteria to humans. They play a crucial role in several biological processes such as degradation of chitin or enzymatic cleavage of terminal GlcNAc or GalNAc.[37] Thanks to their specifities they can be used for the synthesis of unique oligosaccharides.[38] The molecular weight of the hexosaminidases is about 60 kDa. Their sequence contain six cysteine residues and several *N*-glycosylation sites which are occupied by high-mannose structures.[39]

Here, we describe our new fast and sensitive protocol for disulfide bridge characterization which can be applied to proteins in mixtures. The important feature of our approach is that the presence of post-translational modifications does not hamper the assignment of disulfide bonding which we demonstrate on the examples of two highly *N*-glycosylated proteins.

Experimental

Unless stated otherwise, all chemials were purchased from Sigma-Aldrich (Buchs, Switzerland).

Sample preparation

The extracellular portion of human CD69[40] and mouse NKR-P1A were produced in *Escherichia coli* into inclusion bodies. Proteins were purified and folded *in vitro* according to the previously published protocol.[41] CD69 was expressed as two forms, mono- or dimeric, where the monomeric form was shorter by 10 amino acids from the N-terminus (shown in Fig. 1). β-*N*-acetylhexosaminidases were purified from the medium of the producing organism (*A. oryzae* CCF 1066 or *P. oxalicum* strain CCF 1959 and CCF 3438; Czech Collection of Fungi, Department of Botany, Charles University, Prague). Enzyme purification comprised ammonium sulphate precipitation, dialysis and a combination of four different chromatographic steps described elsewhere.[39]

Electrophoresis and in-gel digestion

All proteins were resolved on the 12% SDS polyacrylamide gel electrophoresis (SDS–PAGE) under nonreducing conditions. The highly *N*-glycosylated β-*N*-acetylhexosaminidases were subjected to *'in-gel'* deglycosylation using PNGase F or Endo H endoglycosidases (New England Biolabs, Ipswich, MA). The deglycosylation was carried out for 1 h at 37 °C. The in-gel proteolysis by trypsin or Asp-N proteinases (Sequencing grade, Roche, Mannheim, Germany) was carried out overnight at 37 °C at a protein : enzyme ratio 20/1 (w/w). The digestion buffer consisted of 50 mM 4-ethylmorpholine acetate (pH 8.2) and 10% acetonitrile (LiChrosolv, Merck, Darmstadt, Germany). Cystamine was added to the sample buffer, SDS–PAGE running buffers and to the buffers for deglycosylation and digestion. The final concentration of cystamine was kept at 200 µM throughout.

LC/ESI/FTMS

Five microliters of digestion mixture were used for the analysis. Peptides were desalted on a peptide MacroTrap (Michrom Bioresources, Auburn, CA) and separated on a reversed phase C_{18} column (MAGIC C18 AQ, 0.2 × 150 mm, 5 µm, 200 Å; Michrom Bioresources, Auburn, CA). Mobile phases consisted of 0.2% formic acid in water (solvent A) and 0.16% formic acid in 95% acetonitrile (solvent B). Peptides were eluted under following gradient conditions: 1–15% B in 5 min; 15–40% B in 30 min;

Copyright © 2009 John Wiley & Sons, Ltd.

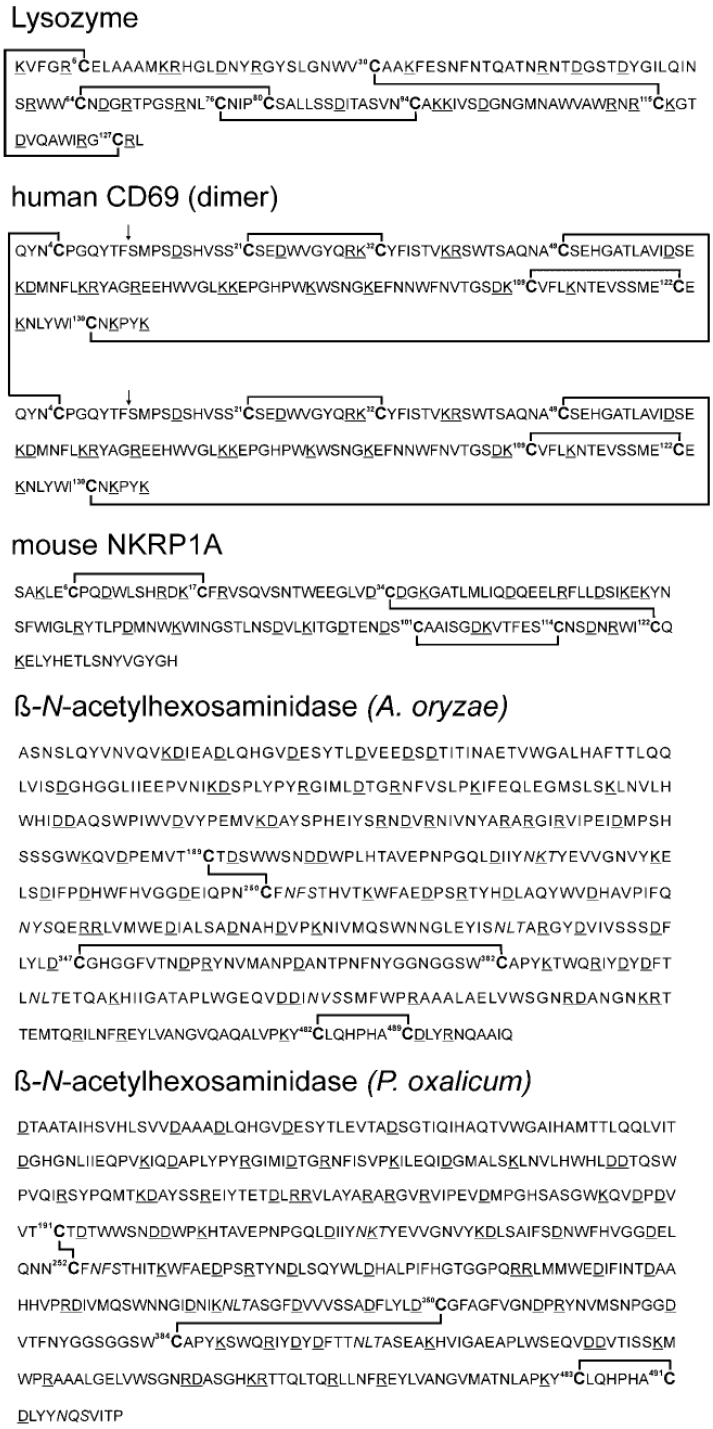


Figure 1. Protein sequences of lysozyme, recombinant human CD69, mouse NKR-P1A, and β-*N*-acetylhexosaminidases from *A. oryzae* and *P. oxalicum* with schematic drawing of the disulfide arrangement. The disulfide arrangement for lysozyme is adopted from the literature,[42] for the other proteins the disulfide arrangement has been determined in this work. The underlined amino acids are the cleavage sites for trypsin (K, R) or Asp-N (D) protease. *N*-glycosylation sequons are in italics. The arrow in CD69 sequence points to the beginning of the monomeric form.

 Copyright © 2009 John Wiley & Sons, Ltd.

40–95% B in 5 min. The flow rate was 4 μl min^{-1} and the column was directly connected to the mass spectrometer. Mass spectra were acquired on a APEX-Qe FTMS instrument equipped with a 9.4 T superconducting magnet (Bruker Daltonics, Billerica, MA). The cell was opened for 4 msec, accumulation time was set at 0.5 sec, and 256 experiments were collected for one LC run where one experiment consisted of the average of five spectra. The acquisition data set size was set to 512 000 points with the mass range starting at m/z 300 amu, resulting in a resolution of 200 000 at m/z 400. The instrument was externally calibrated using triply and doubly charged ions of angiotensin I, and quadruply, quintuply and sextuply charged ions of bovine insulin resulting in mass accuracy below 2 ppm. Prior to the data processing, all experiments were internally post calibrated using masses of unmodified peptides which lowered the mass assignment errors bellow 1 ppm.

Data analysis

The acquired spectra were apodized with a sine bell function, Fourier transformed with one zero-fill and run through a data reduction macro written in TCL/TK for the Bruker Daltonics XMASS software package.[43] The macro produced a list of unique monoisotopic masses which was matched to a theoretical library of regular and cross-linked (disulfide bonded) peptides within the 1 ppm maximum allowed error interval. The library was created by the program Automated Spectrum Assignment Program (ASAP).[44–46] The ASAP was set as follows: increment for formation of a disulfide bond (cross-link) was set to −2.01565 amu; the protease specificity was set to -K, -R for trypsin and D- for Asp-N; the degree of incomplete digestion was up to two missed cleavages. The ASAP algorithm was also set to consider the possible single oxidation of methionine (+15.9949 amu) and modification of glycosylated asparagines by *N*-acetyl glucosamine (+203.0794 amu). For samples deglycosylated by PNGase F, the conversion of Asn to Asp was set as shift by +0.9840 amu.

Results and Discussion

In the first step, we focused on development of a protocol preventing disulfide bond scrambling at high pH. Lysozyme from hen egg was selected as the model protein because its structure was solved by several groups leading to the following disulfide pattern: C6/C127, C30/C115, C64/C80 and C76/C94 (Fig. 1).[43,47]

When we digested lysozyme in-gel using the standard protocol, we were not able to get results matching exactly with the known pattern. Among disulfide linked peptides, signals corresponding to the correct linkages were still the most abundant but other, less intense, signals pointing on the scrambling were found as well. Ions 2101.9871 amu and 2117.9834 amu can be mentioned as examples of disulfide bond scrambling. The masses correspond to a dipeptide 6-13/22-33 and its oxidized form linked through Cys6 and Cys30.

To overcome this problem, we tested the addition of a compound preventing disulfide bond opening. There are several chemicals which can serve this purpose. For instance, cysteamine or glutathione are, in a mixture with their oxidized forms, used in protein *in vitro* folding.[48,49] Our experience as well as observation of other groups showed that irrespectively of their quite similar standard redox potentials, the cysteamine/cystamine system is more effective in promoting disulfide bond formation.[48] This can be probably attributed to its smaller size allowing better penetration into the protein structure. Because we wanted to prevent disulfide bond opening, we tested addition of cystamine.

Results obtained from digests of lysozyme performed in the presence of cystamine improved with increasing cystamine concentration. Starting from 150 μM concentration they were in good agreement with the published data but still not totally unambiguous. Therefore, we included cystamine also in the electrophoresis (sample and running buffers) and raised its concentration to 200 μM. Only such sample handling completely prevented disulfide scrambling.

All results obtained in this study are summarized in Table 1 and in graphical form in Fig. 1. Lysozyme digest prepared according to modified protocol provided us with the following results. The first disulfide bridge Cys6/Cys127 was represented by the tryptic cystine peptide with *m/z* 1168.5273 and the link was confirmed in Asp-N digest by a disulfide linked peptide with *m/z* 3200.6851. The second disulfide bridge was assigned based on the tryptic cystine peptide with *m/z* 1515.7083 and Asp-N cystine peptide with *m/z* 5474.4785. Deciphering the status of the last two pairs of Cys proved to be more difficult and we were not able to clearly assign the exact bonding. However, we limited the number of possibilities and confirmed that all cysteines are involved in disulfide bridges. We detected disulfide containing peptide with *m/z* 3268.4642 corresponding to the sequence $WW^{64}CNDGR/NL^{76}CNIP^{80}CSALLSSDITASVN^{94}CAK$. It is evident that all four Cys in this peptide are involved in disulfide bonds but we were not able to distinguish the exact linkages. From the Asp-N digest, we detected another cystine peptide with two disulfide links, $DYGILQINSRWW^{64}CN/DGRTPGSRNL^{76}CNIP^{80}CSALLSS/DITASVN^{94}CAKKIVS$ (*m/z* 5371.6104). Presence of this peptide and the fact that no signal for two individual peptides corresponding to the links Cys64/Cys94 and Cys76/Cys80 was found, indicated that Cys64 and Cys94 are not linked together but either to Cys76 or Cys80.

This last example also points to a weakness of the method. If the protein does not have enough cleavage sites for efficient 'separation' of individual cysteines, the accurate mass measurement is not sufficient for the assignment of the exact bonding. In such cases tandem mass spectrometric techniques, preferentially ECD, must be used.[50]

This optimized protocol was further applied to three different proteins – the monomeric or disulfide linked homodimer of human CD69 receptor, mouse NKR-P1A and two fungal β-*N*-acetylhexosaminidases. All protein sequences and their important features are shown in Fig. 1.

The second protein sample analyzed was the leukocyte receptor CD69. As mentioned earlier, this protein has been crystalized as a monomer but it dimerizes via an intermolecular disulfide bond in its biologically active conformation. The monomeric form was digested with trypsin which provided us with the information on the intramolecular disulfide bonds. The following cystine peptides were detected. The tryptic dipeptide at *m/z* 3214.3905 indicates a presence of a bond between Cys21 and Cys32. The other signal at *m/z* 3843.8115 matches the peptide with a link between Cys49 and Cys130. And the third peptide (*m/z* 1862.8338) confirms a bond between Cys109/Cys122 (Table 1 and Fig. 1), which corresponds to results previously obtained by an X-ray analysis.[32]

To elucidate the intermolecular disulfide bonds in the homodimeric form, the protein was digested by Asp-N protease. Here, we have detected the N-terminal cystine peptide $QYN^{4}CPGQYTFSMPS/QYN^{4}CPGQYTFSMPS$ (*m/z* 3242.2971) with

 Copyright © 2009 John Wiley & Sons, Ltd.

Journal of MASS SPECTROMETRY

Table 1. Identification of lysozyme, human CD69, mouse NKR-P1A and β-*N*-acetylhexosaminidases disulfide-linked peptides

Protein name S–S bonds	Peptides	m/z (Monoisotopic values) Theoretical	Experimental	Error (ppm)
Lysozyme				
Cys6/Cys127 (t)	6-13/126-128	1168.5276	1168.5273	0.2
Cys6/Cys127[a] (t)	6-13/126-128	1184.5161	1184.5171	0.8
Cys6/Cys127 (a)	1-17/119-129	3200.6867	3200.6851	0.5
Cys6/Cys127[a] (a)	1-17/119-129	3216.6819	3216.6816	0.1
Cys30/Cys115 (t)	22-33/115-116	1515.7083	1515.7082	0.1
Cys30/Cys115 (a)	18-47/101-118	5474.4775	5474.4785	0.2
Cys64/Cys80, Cys76/Cys94[&] (t)	62-68/74-96	3268.4642	3268.4656	0.4
Cys64/Cys80, Cys76/Cys94[&] (a)	52-65/66-86/87-100	5371.6083	5371.6104	0.4
Human CD69 (D – dimer, M – monomer)				
M: Cys21/Cys32 (t)	11-30/32-39	3214.3915	3214.3905	0.3
M: Cys49/Cys130 (t)	41-63/125-135	3843.8108	3843.8115	0.2
M: Cys109/Cys122 (t)	109-113/114-124	1862.8333	1862.8338	0.3
M: Cys109/Cys122[a] (t)	109-113/114-124	1878.8282	1878.8285	0.2
D: Cys4/Cys4 (a)	1-14/1-14	3242.3003	3242.2971	0.9
D: Cys4/Cys4[a] (a)	1-14/1-14	3258.2952	3258.2925	0.8
D: Cys21/Cys32, Cys49/Cys130 Cys109/Cys122 (a)	15-23/24-59/107-135	8530.9532	8530.9462	0.8
Mouse NKR-P1A				
Cys6/Cys17 (t)	4-14/17-19	1805.8210	1805.8202	0.4
Cys6/Cys17 (a)	1-8/15-32	2969.4132	2969.4136	0.1
Cys34/Cys122 (t)	20-37/120-124	2640.2067	2640.2078	0.3
Cys34/Cys122 (a)	33-34/117-139	3059.3411	3059.3396	0.5
Cys101/Cys114 (t)	92-108/109-119	2965.2422	2965.2405	0.6
Cys101/Cys114 (a)	99-106/107-116	1849.7579	1849.7571	0.4
β-N-acetylhexosaminidase *(Aspergillus Oryzae)* – deglycosylated with PNGase F (Asn bearing N-glycosylation converted to Asp)				
Cys189/Cys250 (t)	181-218/229-259	8029.6113	8029.6089	0.3
Cys189/Cys250 (a)	183-190/244-251	1857.7344	1857.7348	0.2
Cys189/Cys250[a] (a)	183-190/244-251	1873.7292	1873.7300	0.4
Cys347/Cys382 (t)	332-358/359-386	5952.6069	5952.6029	0.7
Cys347/Cys382[a] (t)	332-358/359-386	5968.6016	5968.5999	0.3
Cys347/Cys382 (a)	346-355/366-392	4083.7562	4083.7544	0.4
Cys482/Cys489 (intra chain) (t)	481-493	1616.7102	1616.7095	0.5
Cys482/Cys489 (intra chain) (a)	446-489	4995.4806	4995.4767	0.8
β-*N*-acetylhexosaminidase *(Penicillium Oxalicum)* – deglycosylated with Endo H				
Cys191/Cys252[b] (t)	183-203/231-261	6263.7612	6263.7658	0.7
Cys191/Cys252[b] (a)	187-192/246-265	3281.4399	3281.4392	0.2
Cys350/Cys384[b] (t)	330-361/362-388	6381.8427	6381.8385	0.7
Cys350/Cys384 (a)	349-358/371–394	3725.6216	3725.6208	0.2
Cys483/Cys491[b] (intra chain) (t)	483-502	2566.1382	2566.1377	0.2
Cys483/Cys491 (intra chain) (a)	448-491	4964.4981	4964.5024	0.9
Cys483/Cys491[a] (intra chain) (a)	448-491	4980.4930	4980.4897	0.7

[a] Oxidation of Methionine.
[b] Modification of asparagine by HexNAc. (t), tryptic digestion. (a), Asp-N digestion.
[&] the exact linkage could not be assigned based on the accurate mass.

an intermolecular disulfide bridge (Cys4/Cys4). In addition, the large disulfide linked peptide comprising all other intramolecular cysteine bridges (Cys21/Cys32, Cys49/Cys130, Cys109/Cys122) was detected at *m/z* 8530.9462 (Table 1 and Fig. 1).

The next sample was mouse NKR-P1A receptor. Tryptic in-gel digestion followed by LC-FTMS analysis led to identification of dipeptides with *m/z* 1805.8202, 2640.2078 and 2965.2405 which provided us with the following pattern: Cys6/Cys17, Cys34/Cys122 and Cys101/Cys114. To verify those S–S linkages, NKR-P1A was also digested by Asp-N. The observation of cystine peptides with *m/z* 2969.4136, 3059.3396, and 1849.7571 clearly confirmed the S–S bonding suggested by the results of tryptic digest (Table 1 and Fig. 1). The same protein but of rat origin was studied by Kogelberg *et al.*[51] They showed identification of disulfide bridges in rNKR-P1A after Asp-N digestion. The bridging pattern described in their work nicely correlates with the one we found for mouse NKR-

 Copyright © 2009 John Wiley & Sons, Ltd.

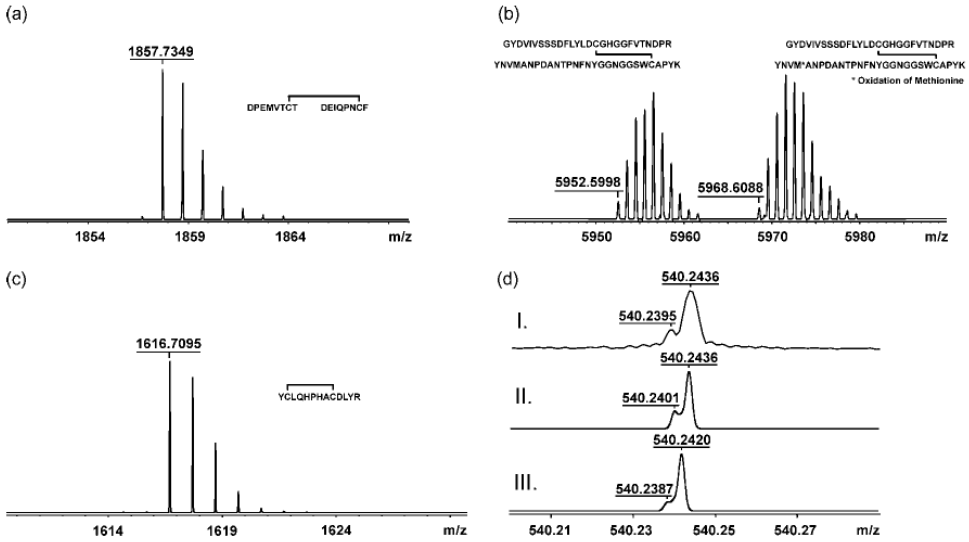


Figure 2. Deconvoluted spectra of cystine peptides from β-*N*-acetylhexosaminidase (*A. oryzae*). (a) Asp-N cystine peptide Cys189/Cys250 at m/z 1857.7349. (b) Tryptic cystine peptide Cys347/Cys382 at m/z 5952.5998 and its oxidized form. (c) intrachain tryptic cystine peptide Cys482/Cys489 at m/z 1616.7095. All peptides were measured with the error less than 1.0 ppm. (d) The contribution of ^{34}S visible on the third isotopic peak – I.) experimental data – cystine peptide Cys482/Cys489 at m/z 540.2436 ($z = 3$); II.) theoretical spectrum of the same sequence; III.) theoretical spectrum of a peptide (SSDFLYLDCGHGGFV) which fits into the window (3 ppm) of accurate mass search, but has only one sulfur.

P1A. Interestingly, they used rather standard protocol for Asp-N digestion in solution (Tris-Cl pH 7.0, 30 °C, 12 h) but claimed to see only the correct linkages. Absence of scrambling is surprising but there are several reasons why they did not detect any incorrect bridges. First, they digested the protein at lower pH (7.0) and temperature (30 °C). Both factors are undoubtedly affecting the rate at which the disulfides are opening and re-associating. Second, they monitored the HPLC separation of digestion products by UV and only selected fractions were subjected to mass spectrometric analysis. From their published chromatogram, one can see some minor peaks which may contain scrambled peptides. And finally, the tendency to disulfide scrambling is likely not just a result of physicochemical factors but can be also influenced by the intrinsic properties of a protein. Our results showed that, for example, lysozyme is prone to scrambling at slightly basic pH but hexosaminidases do not scramble so easily.

Finally, we applied our protocol to complex β-*N*-acetylhexosaminidases from *A. oryzae* and *P. oxalicum*. The hexosaminidase from *A. oryzae* bears five classical potential glycosylation sites which were shown to be glycosylated.[39] To reduce the complexity of the sample, and also to decrease the molecular weight of cystine peptides bearing the *N*-glycosylation, we decided to perform full deglycosylation. The hexosaminidase was treated in-gel using PNGase F. Cystamine was included in the deglycosylation buffer. After deglycosylation, the protein was digested either by trypsin or by Asp-N. We successfully identified three tryptic dipeptides with *m/z* 8029.6216, 5952.5998, and 1616.7095, showing the disulfide bridges Cys189/Cys250, Cys347/Cys382, and Cys482/Cys489, respectively. In the Asp-N digest, we found peptide with *m/z* 1857.7348 which clearly confirmed the bond Cys189/Cys250, whereas the *m/z* 4083.7544 and 4995.4767 supported the previous finding on the links Cys347/Cys382 and Cys482/Cys489 (Table 1 and Fig. 1). The identified disulfide bond arrangement was further used for optimization of the three dimensional model of β-*N*-acetylhexosaminidase from *A. oryzea*. Interestingly, it clarified the role of the disulfide bond in the stabilization of the protein loop and active site of the enzyme.[3]

Figure 2 shows representative mass spectra of disulfide bonded peptides from β-*N*-acetylhexosaminidase from *A. oryzae*. It also nicely shows the benefit of high resolution due to which the isotope pattern, even for data collected in broad band mode during LC-MS run, could be used as a verification criterion for disulfide linked peptides.[52] Because the isotope ^{34}S has natural abundance of 4.29% and the mass of isotopic peak corresponding to $^{34}S_1$ species differes from the mass of isotopic peak corresponding to $^{13}C_2$, they can be easily resolved and based on the intensity of $^{34}S_1$ an estimation of the number of sulfur atoms within a peptide can be done. In our case, the accurate mass plus enzyme specificity filter provided us with a single match to each m/z. Thus, if we want to demonstrate the utility of isotope pattern matching, broader mass tolerance window and no enzyme specificity had to be used for the mass search. Indeed, when we searched our data against individual sequences with 3 ppm tolerance and no enzyme specificity, two possible matches occurred in one case (^{481}YCLQHPHACDLY^{493}R, M_{mono} 1615.7024; ^{339}SSDFLYLDCGHGGF^{353}V M_{mono} 1615.6977). The intesity ratios and the shape of the third isotope of a peak identified as a peptide ^{481}YCLQHPHACDLY^{493}R (Cys481/Cys493) (Fig. 2d I.) nicely match the theoretical envelope of a peptide YCLQHPHACDLYR having two sulfur atoms (Fig. 2d II.), but not the one with a single sulfur atom modeled for peptide ^{339}SSDFLYLDCGHGGF^{353}V (Fig. 2d III.)

The last protein in this study was β-*N*-acetylhexosaminidase isolated from *P. oxalicum*. This enzyme contains five *N*-glycosylation sites and our initial experiments confirmed presence of a glycosylation. The deglycosylation was in this case done by Endo H cleaving the bond between two *N*-acetylhexosamines and leaving the anchoring HexNAc attached to asparagine. Unlike PNGase F, it is not inhibited by the presence of SDS but it has a narrower speci-

Copyright © 2009 John Wiley & Sons, Ltd.

Journal of MASS SPECTROMETRY

ficity. After EndoH deglycosylation and subsequent proteolysis, we found the tryptic cystine peptide (*m/z* 6263.7658) and Asp-N cystine peptide (*m/z* 3281.4392) indicating that the disulfide bridge is between Cys191 and Cys252. Both cystine peptides carried a single HexNAc residue. Likewise in the case of hexosaminidase from *A. oryzae*, we found the second pair of cysteines (Cys350/Cys384) linked together. This link was identified based on a tryptic cystine peptide bearing a single HexNAc residue (*m/z* 6381.8385) and in the Asp-N digest as a peptide with *m/z* 3725.6208. We also found tryptic cystine peptide Y^{483}CLQHPHA^{491}CDLYYNQSVITP, modified by a single HexNAc (*m/z* 2566.1377) showing that the last pair of cysteines forms the third disulfide bridge. Finally, we detected the Asp-N dipeptide at *m/z* 4964.5024 verifying the bond Cys483/Cys491 (Table1 and Fig. 1).

The disulfide bond arrangement of β-*N*-acetylhexosaminidase from *P. oxalicum* is highly similar to the pattern found in other β-*N*-acetylhexosaminidases where the first cysteine is linked to the second cysteine, third cysteine to the fourth and the fifth cysteine is coupled with the sixth cysteine.[53]

Conclusions

A new approach for determination of disulfide bonds in proteins was developed and successfully applied to the analysis of different proteins. The method is based on the separation of proteins by SDS–PAGE and consecutive digestion with specific proteases carried under conditions preventing disulfide bond scrambling. FTMS analysis followed by an automated data processing is used for the rapid assignment of disulfide linked peptides. Because of the high sensitivity and mass accuracy of FTMS, several micrograms of protein are enough for this structural analysis. The major advantage of this approach is its use for the characterization of disulfide bonds of proteins from protein mixtures, and fact that post-translational modifications do not influence the analysis. The method does not require any reduction or modification of proteins, except for deglycosylation, which is advisable in the case of glycosylated proteins. This approach should be very helpful in '*de novo*' analysis of disulfide bonds in proteins and we have shown its utility for control of the correct folding of the recombinant proteins.

Acknowledgements

The authors thank Dr Gary Kruppa from Bruker Daltonics for the macro reducing LC FTMS data and Dr Malin Young from Sandia National Laboratory for data interpretation algorithm. Financial support from the Grant Agency of the Academy of Sciences of the Czech Republic (B400200501, A5020403, KJB500200612); Institutional research concept of the Institute of Microbiology (AVOZ50200510), and from the Ministry of Education, Youth and Sports of the Czech Republic (LC 545, LC 07017 and MSM 21620808) and European Commission (MKTD-CT-2004-014407) is gratefully acknowledged.

References

[1] M. Matsumura, G. Signor, B. W. Matthews. Substantial increase of protein stability by multiple disulphide bonds. *Nature* **1989**, *342*, 293.
[2] W. J. Wedemeyer, E. Welker, M. Narayan, H. A. Scheraga. Disulfide bonds and protein folding. *Biochemistry* **2000**, *39*, 4216.
[3] R. Ettrich, V. Kopecky, K. Hofbauerova, V. Baumruk, P. Novak, P. Pompach, P. Man, O. Plihal, M. Kuty, N. Kulik, J. Sklenar, H. Ryslava, V. Kren, K. Bezouska Jr. Structure of the dimeric *N*-glycosylated form of fungal beta-*N*-acetylhexosaminidase revealed by computer modeling, vibrational spectroscopy, and biochemical studies. *BMC Structural Biology* **2007**, *7*, 32.
[4] P. K. Jensen, A. K. Harrata, C. S. Lee. Monitoring protein refolding induced by disulfide formation using capillary isoelectric focusing-electrospray ionization mass spectrometry. *Analytical Chemistry* **1998**, *70*, 2049.
[5] J. Pavlicek, B. Sopko, R. Ettrich, V. Kopecky, V. Baumruk, P. Man, V. Havlicek, M. Vrbacky, L. Martinkova, V. Kren, M. Pospisil, K. Bezouska Jr. Molecular characterization of binding of calcium and carbohydrates by an early activation antigen of lymphocytes CD69. *Biochemistry* **2003**, *42*, 9306.
[6] J. J. Gorman, T. P. Wallis, J. J. Pitt. Protein disulfide bond determination by mass spectrometry. *Mass Spectrometry Review* **2002**, *21*, 216.
[7] D. L. Smith, Z. R. Zhou. Strategies for locating disulfide bonds in proteins. *Methods in Enzymology* **1990**, *193*, 389.
[8] J. J. Pitt, E. Da Silva, J. J. Gorman. Determination of the disulfide bond arrangement of Newcastle disease virus hemagglutinin neuraminidase. *Journal of Biological Chemistry* **2000**, *275*, 6478.
[9] G. R. Jacobson, M. H. Schaffer, G. R. Stark, T. C. Vanaman. Specific chemical cleavage in high yield at the amino peptide bonds of cysteine and cystine residues. *Journal of Biological Chemistry* **1973**, *248*, 6591.
[10] J. Wu, D. A. Gage, J. T. Watson. A strategy to locate cysteine residues in proteins by specific chemical cleavage followed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Analytical Biochemistry* **1996**, *235*, 174.
[11] J. Wu, J. T. Watson. A novel methodology for assignment of disulfide bond pairings in proteins. *Protein Sciences* **1997**, *6*, 398.
[12] J. Qi, W. Wu, C. R. Borges, D. Hang, M. Rupp, E. Torng, J. T. Watson. Automated data interpretation based on the concept of "negative signature mass" for mass-mapping disulfide structures of cystinyl proteins. *Journal of the American Society for Mass Spectrometry* **2003**, *14*, 1038.
[13] M. E. Levison, A. S. Josephson, D. M. Kirschenbaum. Reduction of biological substances by water-soluble phosphines: Gamma-globulin (IgG). *Cellular and Molecular Life Sciences* **1969**, *25*, 127.
[14] J. A. Burns, J. C. Butler, J. Moran, G. M. Whitesides. Selective reduction of disulfides by tris(2-carboxyethyl)phosphine. *Journal of Organic Chemistry* **1991**, *56*, 2650.
[15] W. R. Gray. Disulfide structures of highly bridged peptides: A new strategy for analysis. *Protein Science* **1993**, *2*, 1748.
[16] W. R. Gray. Echistatin disulfide bridges: Selective reduction and linkage assignment. *Protein Science* **1993**, *2*, 1755.
[17] T. Y. Yen, H. Yan, B. A. Macher. Characterizing closely spaced, complex disulfide bond patterns in peptides and proteins by liquid chromatography/electrospray ionization tandem mass spectrometry. *Journal of Mass Spectrometry* **2002**, *37*, 30.
[18] N. A. Horn, G. B. Hurst, A. Mayasundari, N. A. Whittemore, E. H. Serpersu, C. B. Peterson. Assignment of the four disulfides in the N-terminal somatomedin B domain of native vitronectin isolated from human plasma. *Journal of Biological Chemistry* **2004**, *279*, 35878.
[19] D. Bilusich, V. M. Maselli, C. S. Brinkworth, T. Samguina, A. T. Lebedev, J. H. Bowie. Direct identification of intramolecular disulfide links in peptides using negative ion electrospray mass spectra of underivatised peptides. A joint experimental and theoretical study. *Rapid Communications in Mass Spectrometry* **2005**, *19*, 3074.
[20] D. Chelius, M. E. Huff Wimer, P. V. Bondarenko. Reversed-phase liquid chromatography in-line with negative ionization electrospray mass spectrometry for the characterization of the disulfide-linkages of an immunoglobulin gamma antibody. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 1598.
[21] M. Zhang, I. A. Kaltashov. Mapping of protein disulfide bonds using negative ion fragmentation with a broadband precursor selection. *Analytical Chemistry* **2006**, *78*, 4829.
[22] T. P. Wallis, J. J. Pitt, J. J. Gorman. Identification of disulfide-linked peptides by isotope profiles produced by peptic digestion of proteins in 50% (18)O water. *Protein Science* **2001**, *10*, 2271.

 Copyright © 2009 John Wiley & Sons, Ltd.

[23] T. P. Wallis, C. Y. Huang, S. B. Nimkar, P. R. Young, J. J. Gorman. Determination of the disulfide bond arrangement of dengue virus NS1 protein. *Journal of Biological Chemistry* **2004**, *279*, 20741.
[24] F. W. McLafferty, E. K. Fridriksson, D. M. Horn, M. A. Lewis, R. A. Zubarev. Techview: biochemistry. Biomolecule mass spectrometry. *Science* **1999**, *284*, 1290.
[25] R. A. Zubarev, N. A. Kruger, E. K. Fridriksson, M. A. Lewis, D. M. Horn, B. K. Carpenter, F. W. McLafferty. Electron capture dissociation of gaseous multiply-charged proteins is favored at disulfide bonds and other sites of high hydrogen atom affinity. *Journal of the American Chemical Society* **1999**, *121*, 2862.
[26] R. A. Zubarev. Electron-capture dissociation tandem mass spectrometry. *Current Opinion in Biotechnology* **2004**, *15*, 16.
[27] S. L. Wu, I. Jardine, W. S. Hancock, B. L. Karger. A new and sensitive on-line liquid chromatography/mass spectrometric approach for top-down protein analysis: the comprehensive analysis of human growth hormone in an *E. coli* lysate using a hybrid linear ion trap/Fourier transform ion cyclotron resonance mass spectrometer. *Rapid Communications in Mass Spectrometry* **2004**, *18*, 2207.
[28] S. A. Raspopov, A. El-Faramawy, B. A. Thomson, K. W. Siu. Infrared multiphoton dissociation in quadrupole time-of-flight mass spectrometry: top-down characterization of proteins. *Analytical Chemistry* **2006**, *78*, 4577.
[29] L. M. Mikesh, B. Ueberheide, A. Chi, J. J. Coon, J. E. Syka, J. Shabanowitz, D. F. Hunt. The utility of ETD mass spectrometry in proteomic analysis. *Biochimica et Biophysica Acta* **2006**, *1764*, 1822.
[30] Y. Ge, B. G. Lawhorn, M. ElNaggar, E. Strauss, J. H. Park, T. P. Begley, F. W. McLafferty. Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *Journal of the American Chemical Society* **2002**, *124*, 678.
[31] D. Sancho, A. G. Santis, J. L. Alonso-Lebrero, F. Viedma, R. Tejedor, F. Sanchez-Madrid. Functional analysis of ligand-binding and signal transduction domains of CD69 and CD23 C-type lectin leukocyte receptors. *Journal of Immunology* **2000**, *165*, 3875.
[32] K. Natarajan, M. W. Sawicki, D. H. Margulies, R. A. Mariuzza. Crystal structure of human CD69: a C-type lectin-like activation marker of hematopoietic cells. *Biochemistry* **2000**, *39*, 14786.
[33] A. Mesci, B. Ljutic, A. P. Makrigiannis, J. R. Carlyle. NKR-P1 biology: from prototype to missing self. *Immunological Research* **2006**, *35*, 26.
[34] L. Glimcher, F. W. Shen, H. Cantor. Identification of a cell-surface antigen selectively expressed on the natural killer cell. *Journal of Experimental Medicine* **1977**, *145*, 1.
[35] W. Chambers, N. Vujanovic, A. DeLeo, M. Olszowy, R. Herberman, J. Hiserodt. Monoclonal antibody to a triggering structure expressed on rat natural killer cells and adherent lymphokine-activated killer cells. *Journal of Experimental Medicine* **1989**, *169*, 1389.
[36] R. Josien, M. Heslan, J. P. Soulillou, M. C. Cuturi. Rat spleen dendritic cells express natural killer cell receptor protein 1 (NKR-P1) and have cytotoxic activity to select targets via a Ca2+-dependent mechanism. *Journal of Experimental Medicine* **1997**, *186*, 472.
[37] E. Cabib. The synthesis and degradation of chitin. *Advances in Enzymology and Related Areas of Molecular Biology* **1987**, *59*, 101.
[38] L. Weignerova, P. Vavruskova, A. Pisvejcova, J. Thiem, V. Kren. Fungal beta-N-acetylhexosaminidases with high beta-N-acetylgalactosaminidase activity and their use for synthesis of beta-GalNAc-containing oligosaccharides. *Carbohydrate Research* **2003**, *338*, 1008.
[39] O. Plihal, J. Sklenar, K. Hofbauerova, P. Novak, P. Man, P. Pompach, D. Kavan, H. Ryslava, L. Weignerova, A. Charvatova-Pisvejcova, V. Kren, K. Bezouska. Large propeptides of fungal beta-N-acetylhexosaminidases are novel enzyme regulators that must be intracellularly processed to control activity, dimerization, and secretion into the extracellular environment. *Biochemistry* **2007**, *46*, 2734.
[40] R. A. Childs, C. Galustian, A. M. Lawson, G. Dougan, K. Benwell, G. Frankel, T. Feizi. Recombinant soluble human CD69 dimer produced in Escherichia coli: reevaluation of saccharide binding. *Biochemical and Biophysical Research Communications* **1999**, *266*, 23.
[41] O. Vanek, M. Nalezkova, D. Kavan, I. Borovickova, P. Pompach, P. Novak, V. Kumar, L. Vannucci, J. Hudecek, K. Hofbauerova, V. Kopecky, J. Brynda, P. Kolenko, J. Dohnalek, P. Kaderavek, J. Chmelik, L. Gorcik, L. Zidek, V. Sklenar, K. Bezouska Jr Soluble recombinant CD69 receptors optimized to have an exceptional physical and chemical stability display prolonged circulation and remain intact in the blood of mice. *FEBS Journal* **2008**, *275*, 5589.
[42] J. Jolles, P. Jauregui-Aolles. Preliminary contribution to the study of the disulfide bonds in hen's egg-white lysozyme. *Biochimica et Biophysica Acta* **1963**, *71*, 490.
[43] G. H. Kruppa, J. Schoeniger, M. M. Young. A top down approach to protein structural studies using chemical cross-linking and Fourier transform mass spectrometry. *Rapid Communications in Mass Spectrometry* **2003**, *17*, 162.
[44] R. B. Jacobsen, K. L. Sale, M. J. Ayson, P. Novak, J. Hong, P. Lane, N. L. Wood, G. H. Kruppa, M. M. Young, J. S. Schoeniger. Structure and dynamics of dark-state bovine rhodopsin revealed by chemical cross-linking and high-resolution mass spectrometry. *Protein Science* **2006**, *15*, 1317.
[45] K. A. Kellersberger, E. Yu, G. H. Kruppa, M. M. Young, D. Fabris. Top-down characterization of nucleic acids modified by structural probes using high-resolution tandem mass spectrometry and automated data interpretation. *Analytical Chemistry* **2004**, *76*, 2445.
[46] M. M. Young, N. Tang, J. C. Hempel, C. M. Oshiro, E. W. Taylor, I. D. Kuntz, B. W. Gibson, G. Dollinger. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *The Proceedings of the National Academy of Sciences U. S. A.* **2000**, *97*, 5806.
[47] W. R. Rypniewski, H. M. Holden, I. Rayment. Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: an X-ray analysis at 1.8-A resolution. *Biochemistry* **1993**, *32*, 9858.
[48] J. R. Huth, F. Weijun, R. W. Ruddon. Redox conditions for stimulation of in vitro folding and assembly of the glycoprotein hormone chorionic gonadotropin. *Biotechnology and Bioengineering* **1994**, *44*, 72.
[49] K. J. Woycechowsky, R. T. Raines. Native disulfide bond formation in proteins. *Current Opinion in Chemical Biology* **2000**, *4*, 539.
[50] P. Novak, W. E. Haskins, M. J. Ayson, R. B. Jacobsen, J. S. Schoeniger, M. D. Leavell, M. M. Zouny, G. H. Kruppa. Unambiguous assignment of intramolecular chemical cross-links in modified mammalian membrane proteins by Fourier transform-tandem mass spectrometry. *Analytical Chemistry* **2005**, *77*, 5106.
[51] H. Kogelberg, A. M. Lawson, F. W. Muskett, R. A. Carruthers, T. Feizi. Expression in Escherichia coli, folding in vitro, and characterization of the carbohydrate recognition domain of the natural killer cell receptor NKR-P1A. *Protein Expression and Purification* **2000**, *20*, 10.
[52] T. Solouki, M. R. Emmett, S. Guan, A. G. Marshall. Detection, number, and sequence location of sulfur-containing amino acids and disulfide bridges in peptides by ultrahigh-resolution MALDI FTICR mass spectrometry. *Analytical Chemistry* **1997**, *15*, 1168.
[53] C. G. Schuette, J. Weisgerber, K. Sandhoff. Complete analysis of the glycosylation and disulfide bond pattern of human beta-hexosaminidase B by MALDI-MS. *Glycobiology* **2001**, *11*, 556.

 Copyright © 2009 John Wiley & Sons, Ltd.

4.2 Other proteins related results

4.2.1 Accessibility changes within diphtheria toxin T domain when in the functional molten globule state, as determined using hydrogen/deuterium exchange measurements

the FEBS Journal

Accessibility changes within diphtheria toxin T domain when in the functional molten globule state, as determined using hydrogen/deuterium exchange measurements

Petr Man[1,2,*], Caroline Montagner[3,*], Heidi Vitrac[3], Daniel Kavan[2], Sylvain Pichard[4], Daniel Gillet[4], Eric Forest[1] and Vincent Forge[3]

1 Laboratoire de Spectrométrie de Masse des Protéines, Institut de Biologie Structurale (CEA, CNRS, UJF, UMR 5075), Grenoble, France
2 Laboratory of Molecular Structure Characterization, Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, Prague 4, Czech Republic
3 CEA; DSV; iRTSV; Laboratoire de Chimie et Biologie des Métaux (UMR 5249); CEA-Grenoble, Grenoble, France
4 Commissariat à l'Energie Atomique (CEA), Institut de Biologie et Technologies de Saclay (iBiTecS), Service d'Ingénierie Moléculaire des Protéines (SIMOPRO), F-91191 Gif sur Yvette, France

Keywords
diphtheria toxin; hydrogen/deuterium exchanges; mass spectrometry; protein/membrane interactions; translocation domain

Correspondence
D. Gillet, Commissariat à l'Energie Atomique (CEA), Institut de Biologie et Technologies de Saclay (iBiTecS), Service d'Ingénierie Moléculaire des Protéines (SIMOPRO), F-91191 Gif sur Yvette, France
Fax: +33 1 69 08 94 30
Tel: +33 1 69 08 76 46
E-mail: daniel.gillet@cea.fr
E. Forest, Laboratoire de Spectrométrie de Masse des Protéines, Institut de Biologie Structurale (CEA-CNRS-UJF), 41 rue Jules Horowitz, 38027 Grenoble, France
Fax: +33 4 38 78 54 94
Tel: +33 4 38 78 34 03
E-mail: eric.forest@ibs.fr
V. Forge, CEA; DSV; iRTSV; Laboratoire de Chimie et Biologie des Métaux (UMR 5249); CEA-Grenoble, 17 rue des martyrs, 38054 Grenoble, France
Fax: +33 4 38 78 54 87
Tel: +33 4 38 78 94 05
E-mail: vincent.forge@cea.fr

*These authors contributed equally to this work

(Received 7 August 2009, revised 6 November 2009, accepted 23 November 2009)

doi:10.1111/j.1742-4658.2009.07511.x

The translocation domain (T domain) of diphtheria toxin adopts a partially folded state, the so-called molten globule state, to become functional at acidic pH. We compared, using hydrogen/deuterium exchange experiments associated with MS, the structures of the T domain in its soluble folded state at neutral pH and in its functional molten globule state at acidic pH. In the native state, the α-helices TH5 and TH8 are identified as the core of the domain. Based on the high-resolution structure of the T domain, we propose that TH8 is highly protected because it is buried within the native structure. According to the same structure, TH5 is partly accessible at the surface of the T domain. We propose that its high protection is caused by the formation of dimers. Within the molten globule state, high protection is still observed within the helical hairpin TH8–TH9, which is responsible for the insertion of the T domain into the membrane. In the absence of the lipid bilayer, this hydrophobic part of the domain self-assembles, leading to the formation of oligomers. Overall, hydrogen/deuterium-exchange measurements allow the analysis of interaction contacts within small oligomers made of partially folded proteins. Such information, together with crystal structure data, are particularly valuable for using to analyze the self-assembly of proteins.

Structured digital abstract

- MINT-7298382, MINT-7298394: *diphtheria toxin* (uniprotkb:Q6NK15) and *diphtheria toxin* (uniprotkb:Q6NK15) *bind* (MI:0407) by *molecular sieving* (MI:0071)

Abbreviations
C domain, catalytic domain; ESI-TOF, electrospray ionization-time of flight; H/D, hydrogen/deuterium; MG domain, molten globule domain; N, native; T domain, translocation domain.

Introduction

Diphtheria toxin is a protein secreted by *Corynebacterium diphtheriae* as a single polypeptide chain of 58 kDa [1]. During cell intoxication, diphtheria toxin is cleaved by furin into two fragments: the A chain corresponding to the catalytic domain (C domain); and the B chain corresponding to the translocation domain (T domain) and the receptor-binding domain. The C and T domains remain covalently linked by a disulfide bond. Following binding to its cell-surface receptor, diphtheria toxin is internalized through the clathrin-coated pathway. The acidic pH in the endosome triggers a conformational change, leading to insertion of the toxin in the membrane. The C domain is then translocated across the endosomal membrane into the cytosol. The C domain ADP-ribosylates the elongation factor 2, blocking protein translation and leading to cell death.

At neutral pH, the T domain is refolded and soluble, and possesses a globin fold containing nine α-helices (TH1–TH9) [2,3] The activation of the T domain requires the formation of a molten globule (MG) state propitious to membrane interaction [4,5]. The MG state is a partially folded state that occurs transiently during the folding reaction of many proteins [6]. However, some proteins, such as the T domain, acquire an MG state for functional purpose [4,7–9]. The MG state is highly dynamic. Thus, high-resolution structural methods for analyzing the MG state are not applicable. The method of choice for analyzing MG states at amino acid resolution is based on hydrogen/deuterium (H/D) exchange experiments coupled to NMR or MS [10–15]. In the case of the T domain, NMR spectra were not of sufficient quality (A. Chenal, V. Forge, D. Gillet, unpublished data). Indeed, high-quality NMR spectra are usually recorded at acidic pH to minimize fast proton-exchange effects, a pH that cannot be used to study the T domain in the native (N) state. Thus, the MS approach used in this work offers a valuable alternative.

The data enabled us to identify the core of the protein in the N- and MG states, the regions of moderate and high accessibility, and regions involved in the oligomerization of both states of the T domain in solution.

Results

Monitoring H/D-exchange kinetics within the T domain under various conditions

We compared H to D exchange kinetics of the T domain at pD 7.0 (N state) and pD 4.0 (MG state) (where pD is pH in D_2O). The protein was placed in D_2O solvent at the studied pD and in the presence or absence of NaCl (see the Materials and methods). The H/D exchanges were allowed to proceed for various periods of time, from 30 s to 3 days ($\sim 2.6 \times 10^5$ s). For each time-point, the exchange was quenched by a jump to pH 2.3 and rapid freezing. For monitoring the extent of H/D exchange throughout the protein, samples were thawed and submitted to proteolysis. The mass of the generated peptides was measured using electrospray ionization-time of flight (ESI-TOF) MS.

We first digested the T domain with pepsin. This resulted in full sequence coverage but provided poor resolution in the N-terminal region, namely helices TH1–3, for which large fragments of 38-73 amino acids were obtained. In order to achieve higher resolution we digested the protein with acidic fungal protease type XVIII [16]. When used alone, acidic fungal protease type XVIII did not yield satisfactory results because the digestion was incomplete and quick verification using MALDI-TOF MS showed that large fragments (10–13 kDa) were undigested. This remained unchanged regardless of the protein/protease ratio tested. However, when acidic fungal protease type XVIII was used in combination with pepsin, no large fragments were found and satisfactory spatial resolution over the whole protein sequence was achieved (Fig. 1). Therefore, we employed pepsin and protease XVIII digestion in the analysis of local exchange kinetics. Changes of isotopic profiles as a function of exchange time are shown in Fig. 2 for representative peptides. The initial isotopic profiles are those of the nondeuterated peptides (Fig. 2; black line) and the final isotopic profile is that of the fully deuterated peptides (Fig. 2; grey line). Various behaviors were observed, depending on the peptide and the pD. For peptide 230-236, the exchange was complete for both pDs at the shortest exchange time (30 s). This peptide was fully accessible to the solvent regardless of the experimental conditions. For peptide 278-284, the isotopic profiles evolved towards that of the fully deuterated peptide as the exchange time increased. Therefore, exchange kinetics could be monitored for this peptide. For peptide 351-355, different behaviors were observed depending on the pD. A continuous change of the isotopic profile was measured at pD 7, whereas the peptide remained nondeuterated at pD 4. This result indicated that this peptide was fully protected against H/D exchange at pD 4, while a kinetic could be monitored at pD 7.

To perform correction for back-exchange occuring during digestion and analysis, fully deuterated and

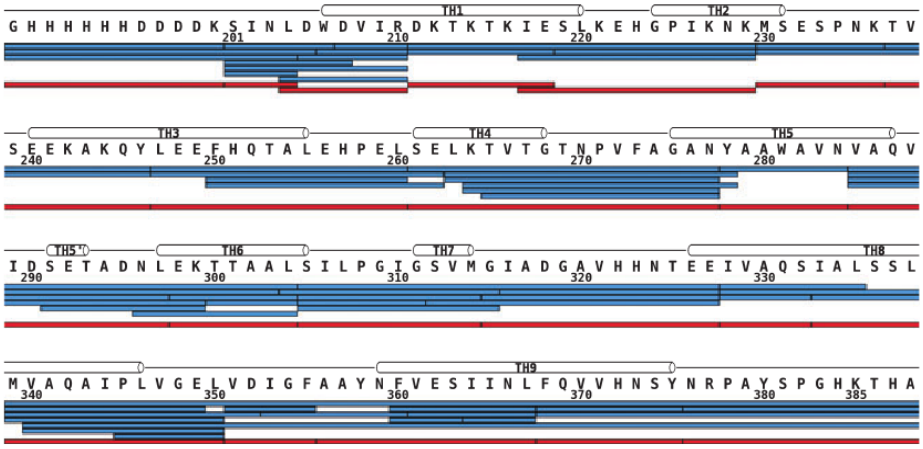


Fig. 1. Peptide mapping of the T domain after digestion with a mixture of pepsin and protease type XVIII. All identified peptides are shown as blue bars. Red bars are peptides used for recording H/D exchange in this study. They cover the entire sequence of the T domain. Native sequence numbering is shown below the sequence, and schematic drawings of secondary structure elements, including their names, are shown above the sequence.

nondeuterated samples were digested and analyzed under the same conditions as the samples collected during the H/D-exchange kinetics [17]. In general, the amount of back-exchange was between 15 and 25%, except for the N-terminal HisTag (residues 1-12), which had a back-exchange of 85%. This exceptional behavior is undoubtedly a result of the amino acid composition of this region [18]. For each peptide the average mass of the nondeuterated and fully deuterated forms was used to correct the extent of H/D exchange during the experiments with the various states of the T domain.

Because the intrinsic rates of H/D exchange are highly sensitive to pH, it is necessary to take into account the intrinsic pD effect on the time-dependence of exchange to compare the results obtained at various pD values [11,19]. Depending on the pH range, the H/D exchange is either acid-catalyzed or base-catalyzed [20]. As a consequence, the dependence of $\mathrm{Log}(k_{exch})$ (the logarithm of the exchange rate) as a function of the pH is a chevron plot with a minimum of around pH 3 for oligopeptides. Between pH 4 and pH 7, the H/D exchange is base-catalyzed and the $\mathrm{Log}(k_{exch})$ increases linearly with pH; the k_{exch} is proportional to 10^{pH}. Exchange times were normalized to pD 4.0 by multiplying the times by 1000 for pD 7.0, a factor corresponding to the 1000-fold increase in the intrinsic H/D-exchange rate found at pH 7.0 compared with pH 4.0, for fully exposed amide protons of the backbone of the protein. Time dependencies of exchange are shown in Fig. 3 for representative peptides. Three types of peptide behaviors were found with respect to their H/D-exchange rates under the different conditions tested (pD 7.0, pD 4.0, with and without NaCl). A first type corresponded to peptides with fast exchange rates regardless of the experimental conditions, such as peptide 230-236 (the loop between helices TH2 and TH3; Fig. 3B). This revealed the regions of the protein exposed to the solvent regardless of the conditions. The large majority of the peptides belonged to a second type for which some protection against H/D exchange was detected at both pH values, but with exchange rates slower at pD 7.0 than at pD 4.0. These included, for example, peptides 211-218 (from TH1), 237-246 (from TH3), 278-284 (from TH5) and 328-332 (from TH8) (Fig. 3A, C, D, E, respectively). This indicated the regions of the protein that were more accessible to the solvent at pD 4.0 than at pD 7.0. Finally, some peptides corresponded to a third type with exchange rates slower at pD 4.0 than at pD 7.0, as illustrated with peptide 351-355 (the loop between α-helices TH8 and TH9) (Fig. 3F). In this case, the corresponding region of the protein was less accessible to the solvent at pD 4.0.

H/D exchange-profile of the T domain in the N-state (pD 7.0)

To follow exchange kinetics, in more detail, over the whole T domain we created exchange profiles summarizing the extent of H/D exchange throughout the entire

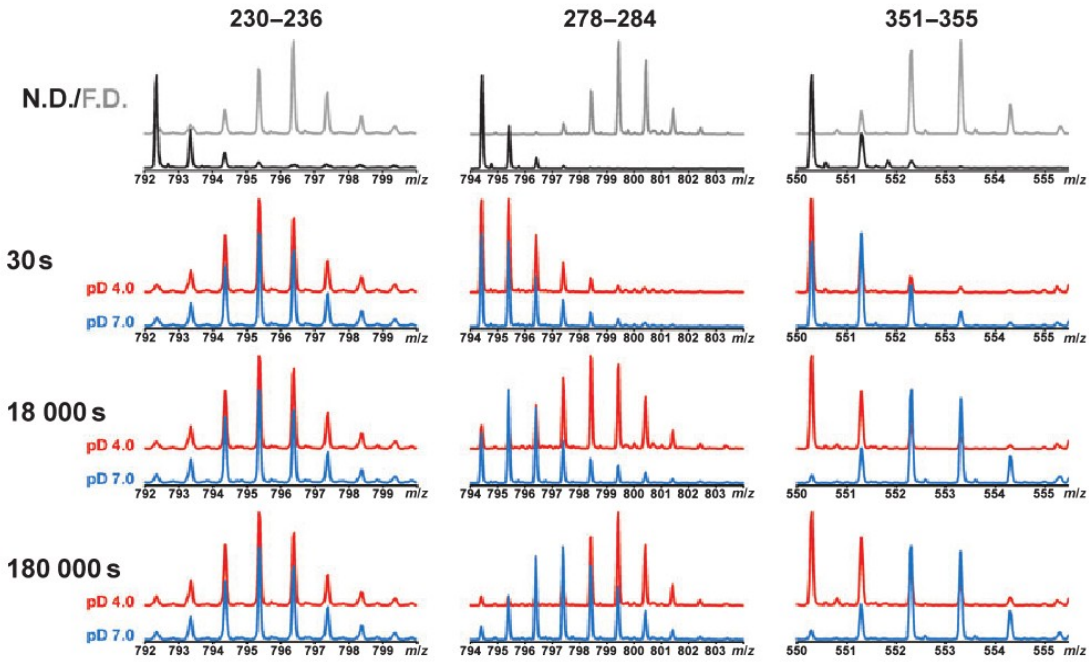


Fig. 2. Selected examples of raw MS data. The top row represents isotopic profiles of nondeuterated (N.D., black) and fully deuterated (F.D., grey) peptides (the native sequence numbering is shown at the top of each column). The three rows below show changes in the isotopic distribution of each peptide during the time-course of the experiment. Examples are shown for conditions without NaCl [pD 4.0 (red) and pD 7.0 (blue)] and for three distinct time-points (30, 18 000 and 180 000 s) of the monitored kinetics. The data are shown without correction for different intrinsic rates of H/D exchange at pD 4.0 and 7.0.

sequence, at a given time. The blue line in Fig. 4A shows the exchange profile found at pD 7.0 (the N state) at 180 s (1.8×10^5 s on the timescale normalized to pD 4.0), at the start of the exchange kinetics (Fig. 3, blue circles). We found a good correlation between α-helices and regions of lower exchange. This indicated that these segments of the protein were significantly protected against H/D exchange, highlighting the regions of the protein with either higher stability and/or lower accessibility to the solvent. The regions exhibiting the lowest exchange (below 40% D occupancy), in other words, the highest protection, define the core of the protein's N state [21]. For the T domain, these were the center of TH5 and TH8, and the N-terminal half of TH9. All α-helices (TH1, 3, 4, 5', 6, 7 and the remaining parts of TH5 and TH9) except for TH2 showed intermediate protection (between 40 and 70% of exchange). It is noteworthy that connecting loops TL1-2, TL3-4, TL4-5, TL5-5', TL5'-6, TL6-7 and TL8-9 were found in this category. Finally, the N-terminus, TH2, loops TL2-3, TL7-8 and the C-terminus were poorly protected (more than 70% of exchange).

When the H/D exchange was allowed to proceed for a much longer time (24 h) (8.6×10^7 s on the timescale normalized to pD 4.0), the exchange profile was drastically changed (Fig. 4B, blue line). Only the centers of TH5 and TH8 showed significant protection. Thus, these regions define the core of the protein (i.e. the most stable part of the protein). From this result, TH9 could be excluded from the core.

Electrostatic interactions between the T domain and the membrane were shown to play an important role in the pH regulation of the protein's function [5,22,23]. These interactions were detected by analyzing the effect of ionic strength on the membrane penetration of the T domain. For this reason, we investigated the effect of NaCl on the solvent accessibility of the domain. The cyan circles of Fig. 3 show that NaCl had a marginal effect on the H/D-exchange kinetics. However, the exchange profiles shown by the cyan line in Fig. 4A revealed that NaCl had a tendency to increase the exchanges within the TH8-TH9 region.

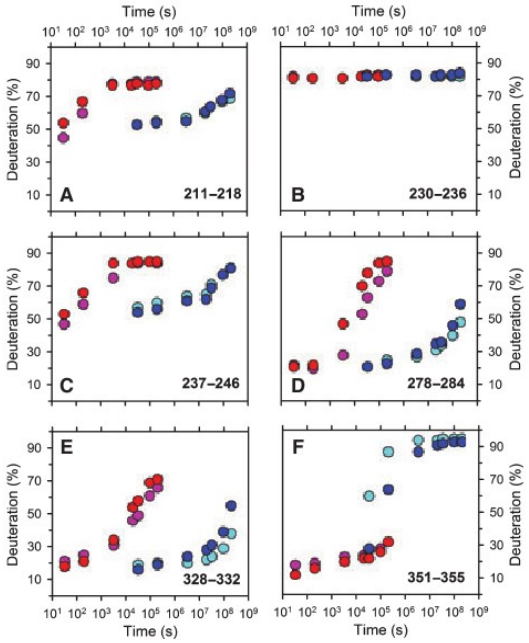


Fig. 3. H/D-exchange kinetics of representative peptides (native sequence numbering). The plots show the corrected percentage of deuteration versus time. Red circles, pD 4.0; pink circles, pD 4.0 with 200 mM NaCl; blue circles, pD 7.0; cyan circles, pD 7.0 with 200 mM NaCl. The plotted exchange times are normalized to pD 4.0; the real exchange times at pH 7 are multiplied by 10^3 to take into account the pH effect on the exchange rates.

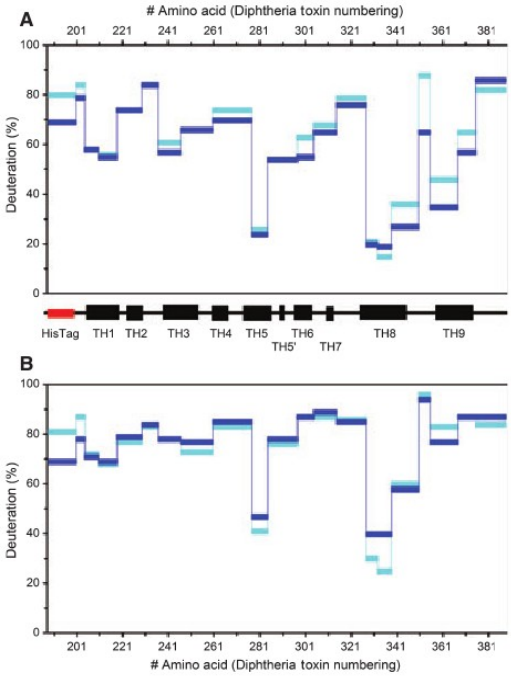


Fig. 4. H/D-exchange profiles of the T domain at pD 7.0 (N state) in the absence (blue) or presence (cyan) of 200 mM NaCl after 3 min (A) or 24 h (B) of exchange. The times given here are the real exchange time (i.e. without correction for the pH effect as those presented in Fig. 3). Localizations of the α-helices within the amino acid sequence (native numbering) of the T domain are shown on a scheme. The percentages of deuteration correspond to the corrected values (see the main text).

H/D-exchange profile of the T domain in the MG state (pD 4.0)

Figure 5A (red line) shows the exchange profile found at pD 4.0 (MG state) at 180 s, a time-point corresponding to the start of the exchange kinetics (Fig. 3, red circles). Although the T domain was in the MG state, two regions of the protein were significantly protected. Regions spanning helices TH5 to TH5' and TH8 to TH9 were found to have less than 30% of exchange. This correlated fairly well with the core detected in the N state. Helices TH1, TH3 and TH4, loop TL4-5, helices TH6 and TH7, and loops TL6-7 and TL7-8 showed intermediate protection (between 35 and 70% of exchange). Altogether, even in the MG state, three categories of regions could be distinguished with respect to solvent accessibility/stability. Surprisingly, after an extended exchange time (8.6×10^4 s, i.e. 24 h) (Fig. 5B, red line), the TH8-TH9 regions still displayed low exchange. The region encompassing the C-terminus of TH8, loop TL8-9 and the N-terminus of TH9 had less than 40% of exchange. This result will be investigated below. The presence of NaCl had a marginal effect on the T domain, with a small tendency to stabilize helices TH5 to TH6, including loops TL5-5' and TL5'-6 (Fig. 5B, pink line).

Comparison of the N and MG states

When comparing exchange profiles for both pD conditions at the same time-point (1.8×10^5 s, ~ 2 days) (Fig. 6), it was obvious that the extents of exchange were much higher in the MG state (red line) than in the N state (blue line). This reflected the expected lower stability of protein structures in the MG state [10,24]. Nevertheless, there was a noticeable exception for loop TL8-9 and the N-terminal part of helix TH9, which were even more protected in the MG state. A possible explanation for this was that this region was involved in the formation of multimers at pD 4.0.

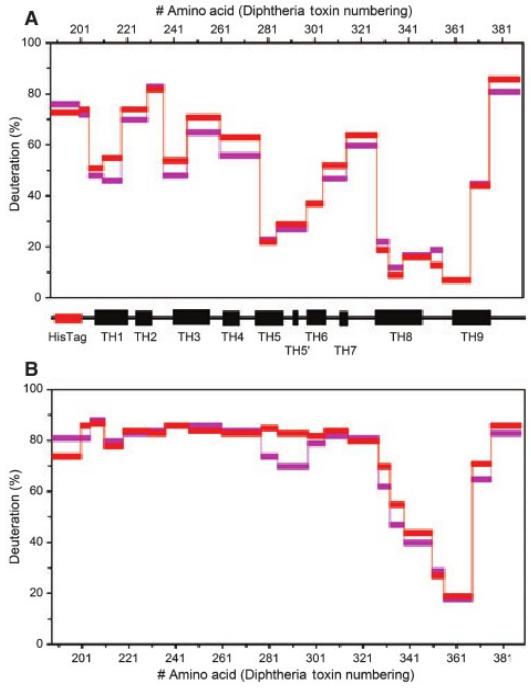


Fig. 5. H/D-exchange profiles of the T domain at pD 4.0 (MG state) in the absence (red) or presence (pink) of 200 mM NaCl after 3 min (A) or 24 h (B) of exchange.

Size-exclusion chromatography was used to test this hypothesis (Fig. 7). The results showed that the elution volume of the T domain was increased at pH 4.0 compared with pH 7.0. Comparison with molecular mass markers showed that the protein was eluted as a dimer at pH 7.0; the estimated molecular mass was around 37 kDa (Fig. 7B), which is close to the theoretical molecular mass of a dimer (44.6 kDa). At pH 4.0, the elution volume of the T domain was quite similar to that of the dead volume (Fig. 7A). According to a general estimation, the oligomers formed at pH 4.0 are at least 10-mers with an apparent molar weight of around 200 kDa (Fig. 7B). It was clear from the exchange profiles (Fig. 5) that the TH8-TH9 region was involved in the formation of oligomers at pH 4.0. By contrast, dimer formation at pH 7.0 may involve helix TH5 (see the Discussion).

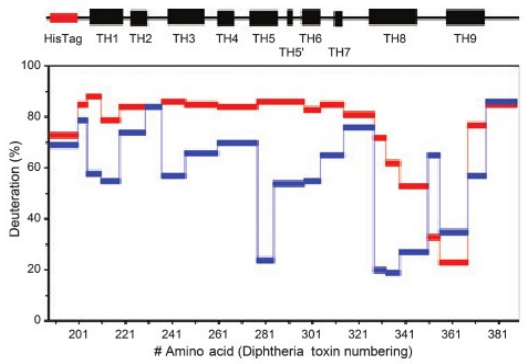


Fig. 6. Comparison of H/D-exchange profiles of the T domain at pD 7.0 (blue) and pD 4.0 (red) after 3 min of exchange at pD 7.0, which corresponds to 1.8×10^5 s at pD 4.0 and on the timescale of Fig. 3.

Discussion

In the present work, we showed that MS can be an alternative to NMR for characterizing the structure of partially folded states of proteins in H/D-exchange experiments. This is particularly helpful when NMR spectra are not of sufficiently high quality. This may

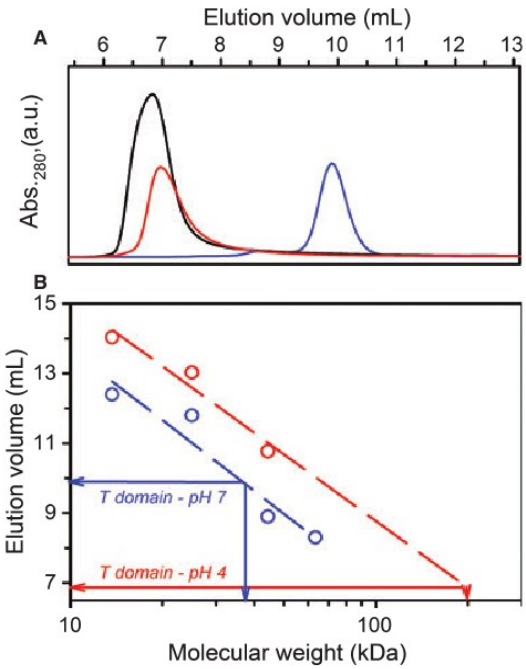


Fig. 7. Size-exclusion chromatography experiments on the T domain at pH 7.0 and pH 4.0 (A) Elution profiles of the T domain at pH 7.0 (blue line) and pH 4.0 (red line). The dead volume of the column is shown with the elution profile of dextran (black line). (B) Estimation of the size of the T domain at pH 7.0. At pH 4 this estimation is highly approximated because the elution volume of the T domain is close to the void volume of the column.

be the case for proteins with conformational exchanges leading to peak broadening in the NMR spectra, for proteins that cannot be stabilized in the N state at the pH conditions propitious for NMR experiments, for proteins with a tendency to aggregate at high concentrations, etc. In the case of the T domain, the MG state corresponds to the functional state, which initiates the translocation of the catalytic domain. Here, the data allowed identification of the core of the protein in the N state and the evolution of the overall structure of the protein in the MG state. This degree of resolution is unprecedented for the T domain of the diphtheria toxin. Three levels of protection were defined, based on our results, corresponding to strong, intermediate and absence of protection. The protection pattern along the sequence of the T domain correlates with the localization of α-helices and loops, with the exception of TH2, which is barely protected within the MG state (Fig. 5A).

The N state appears as a dimer (Fig. 7). This dimer is probably relevant to the isolated T domain but not to the whole toxin, which can also form dimers, but through domain swapping [25]. The most protected region, the core of the domain, corresponds to helices TH5 and TH8 (Fig. 4). According to the crystal structure [2,3], it is not surprising that TH8 is in the core because it is buried in the structure. Within the whole toxin, TH5 is partly covered by the C-terminal part of the receptor (R) domain [2,3]. Within the isolated T domain, this helix is likely to have one face at the protein surface and, as a consequence, should be at least partly accessible to the solvent. The high protection against H/D exchange in TH5 suggests that this helix is buried because of the dimer formation. For illustration only, an attempt of T-domain docking within a dimer is shown in Fig. 8. In this putative dimer structure, TH8 is protected against H/D exchange because it is buried within the native structure of the monomeric T domain and TH5 is protected because it is involved in the dimer interface. TH9 can be considered as involved in the protein core but to a lesser extent. Interestingly, from these results, the current view of the core of the T domain evolves, as it was previously thought to involve its most hydrophobic part, the helical hairpin TH8-TH9 [2,3]. This is not really surprising because TH9 appears to be relatively exposed in the crystal structure (Fig. 8) [2,3]. In the MG state, as expected for a partially folded state, the overall protection against H/D exchange is much lower (Fig. 6). However, the core of the dimeric T domain can still be recognized (Fig. 5A). TH5 and TH8 are still more protected than the rest of the protein, with the exception of TH9, which is discussed below. If one assumes that the protection of TH5 is a result of dimer formation, the dimer may be still present in the MG state.

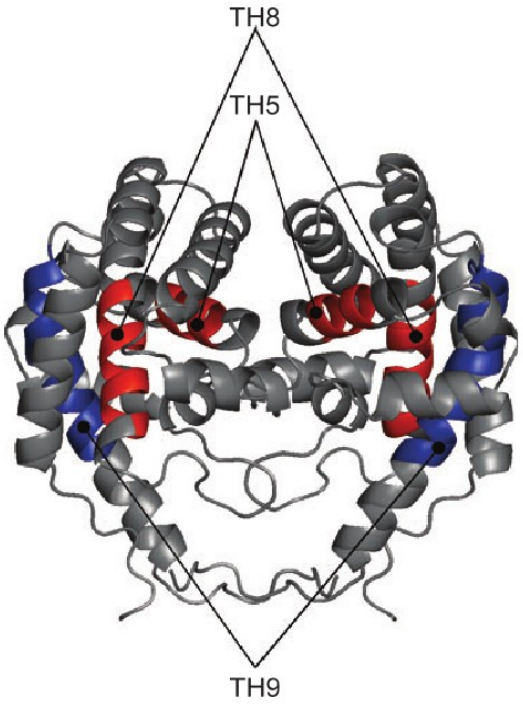


Fig. 8. Putative backbone structure of the T-domain dimer prepared with T domain isolated from the whole toxin crystal structure (PDB: 1F0L) (see the Materials and methods). The parts coloured in red are those with the highest protection against H/D exchange at pH 7.0 (N state), and the regions coloured in blue correspond to those with the highest protection at pH 4.0 (MG state).

The most protected region of the protein in the MG state is TH9 (Fig. 5B). Indeed, after the longest time of exchange, the N-terminal part of TH9 is still highly protected, while the H/D exchange within TH5 and TH8 is almost complete (Fig. 5B). Such a level of protection is abnormal for an MG state in solution. Therefore, we propose that TH9 is involved in oligomer formation in the MG state. The fact that TH9 is highly hydrophobic [2], but loosely involved in the core, renders it available for membrane interaction. In the absence of a phospholipid bilayer, oligomerization is the alternative to bury this hydrophobic region of the protein. Previous work show the tendency of the T domain to form oligomers at acidic pH in the absence of membrane [26,27]. In the event that the soluble T domain is a dimer (Fig. 8), there are two sites for intermolecular interactions on each dimer (Fig. 8). This

can result in the formation of large oligomers similar to those detected at acidic pH (Fig. 7). These oligomers are formed when the N-terminal part of TH9 is available for intermolecular interactions (i.e. when the tertiary structure is lost and the domain is stabilized in the MG state).

In conclusion, for most proteins, the core is tightly correlated with the hydrophobic regions [21]. The core is preserved in partially folded states. This implies that these regions still interact with one another in the MG state. As illustrated here, when hydrophobic parts of the protein are only loosely involved in the core (TH9), they are available for self-interaction. This may lead to oligomerization or aggregation. At last, we provide data on the usefulness of H/D-exchange measurements to analyze interaction contacts within small oligomers made of partially folded proteins. Such information, together with crystal structure data, is particularly valuable in analyzing the self-assembly of proteins.

Materials and methods

All chemicals, proteases and solvents were from Sigma-Aldrich (St Louis, MO, USA) unless otherwise stated. The protein construct corresponding to the T domain of diphtheria toxin (residues 201-386), bearing mutation C201S, was expressed in *Escherichia coli* and purified as described previously [5,28].

H/D exchange

The T domain was dissolved in 4 mM citrate-phosphate buffer, pH 7.0, at a concentration of 70 μM. One part of the solution was kept at pH 7.0 and in the other the pH was lowered to 4.0 by the addition of citric acid. Protein solutions were prepared with or without 200 mM NaCl.

H/D exchange was initiated by dilution of the protein mixture 20-fold in deuterated 4 mM citrate-phosphate buffer, with or without 200 mM NaCl. The exchange was carried out at pD 4.0 and 7.0, and the temperature was kept at 6 °C. The exchange was quenched at the selected time-points by the addition of precooled phosphoric acid and freezing in liquid nitrogen. Samples were stored at −80 °C until analysis. Totally deuterated T domain was prepared by incubation of the protein in D_2O at 30 °C for 6 h followed by concentration on a speed-vac. The cycle of incubation in D_2O and concentration was repeated three times.

Protein digestion

Protein after exchange was digested by a mixture of pepsin and rhizopuspepsin (protease type XVIII). The protein/protease ratios were 1 : 1 (w/w) for pepsin and 1 : 14 (w/w) for protease type XVIII. The digestion was carried out in an ice-bath for 2 min.

LC-MS and LC-MS/MS analysis

Samples after digestion were injected onto the system comprising injection and switching valves (Rheodyne, IDEX Health & Science, Oak Harbor, WA, USA), peptide MacroTrap (MichromBioresources, Auburn, CA, USA) and a reversed-phase column (Jupiter C18, 1 × 50 mm; Phenomenex, Torrance, CA, USA) immersed in an ice-bath. All samples were desalted by solvent A and the peptides were separated by a gradient elution of 15–51% solvent B in 20 min on a reverse-phase column equilibrated in 15% solvent B. The HPLC solvents were: A, 0.03% trifluoroacetic acid in water; and B, 95% CH_3CN/0.03% trifluoroacetic acid. The column was interfaced to a mass spectrometer via an electrospray ion source.

Peptide mapping (MS/MS) was carried out on a quadrupole ion trap (Bruker Esquire 3000+; Bruker Daltonics, Bremen, Germany). Tandem mass spectra were interpreted using MASCOT software (MatrixScience, London, UK) and the assignments were further confirmed by accurate mass measurement on ESI-TOF (Agilent 6210 Time-of-Flight LC/MS; Agilent Technologies, Santa Clara, CA, USA). H/D-exchange kinetics were analyzed on an ESI-TOF instrument. Spectra for each peptide were averaged and exported to MAGTRAN software [29]. The corrections for back-exchange were made according to methods described previously [17].

Size-exclusion chromatography

Protein samples were prepared as for the H/D-exchange experiments, either at pH 7.0 or at pH 4.0. The NaCl concentration was either 0 or 200 mM at pH 7.0 and either 50 or 200 mM at pH 4.0. At acidic pH, at least 50 mM NaCl is necessary for the column. The samples were loaded onto a Superdex 200 10/300GL (Amersham, Piscataway, NJ, USA), equilibrated with the same buffer as for the incubation, which had been calibrated using the following protein standards (Amersham): RNase (13.7 kDa), chymotrypsinogen (25 kDa), ovalbumin (43 kDa), BSA (67 kDa) and Blue Dextran (2000 kDa). BSA was removed from the standards at pH 4.0 because of oligomerization, which leads to an abnormal molecular mass value.

Building of putative dimer structure

In order to find out how the T domain can interact at pH 7.0, a region corresponding to the T domain was taken from the structure 1F0L. It was then loaded onto the ClusPro server (http://nrc.bu.edu/cluster/) and homo-multimeric docking using the DOT algorithm was performed. This approach performs docking based only on the shape complementarity

[30,31]. Out of the fifteen dimeric structures, only one matched our observation of dimer formed though the helices TH5 from each monomer. It is worth mentioning that this represents a general approximation of how the dimer should be formed because the overall packing of the T-domain might be different from that in the model of the whole diphtheria toxin.

Acknowledgements

This work was supported by the Commissariat à l'Energie Atomique (Programme: Signalisation et transport membranaires).

References

1 Chenal A, Nizard P & Gillet D (2002) Structure and function of diphtheria toxin: from pathology to engineering. *J Tox-Tox Rev* **21**, 321–359.
2 Choe S, Bennett MJ, Fujii G, Curmi PM, Kantardjieff KA, Collier RJ & Eisenberg D (1992) The crystal structure of diphtheria toxin. *Nature* **357**, 216–222.
3 Bennett MJ & Eisenberg D (1994) Refined structure of monomeric diphtheria toxin at 2.**3** **Å** resolution. *Protein Sci* **3**, 1464–1475.
4 Zhan H, Choe S, Huynh PD, Finkelstein A, Eisenberg D & Collier RJ. (1994) Dynamic transitions of the transmembrane domain of diphtheria toxin: disulfide trapping and fluorescence proximity studies. *Biochemistry* **33**, 11254–11263.
5 Chenal A, Nizard P, Forge V, Pugnière M, Roy MO, Mani JC, Guillain F & Gillet D (2002) Does fusion of domains from unrelated proteins affect their folding pathways and the structural changes involved in their function? A case study with the diphtheria toxin T domain *Protein Eng* **15**, 383–391.
6 Arai M & Kuwajima K (2000) Role of the molten globule state in protein folding. *Adv Protein Chem* **53**, 209–282.
7 Miller WL (2007) Mechanism of StAR's regulation of mitochondrial cholesterol import. *Mol Cell Endocrinol* **265–266**, 46–50.
8 Romero P, Obradovic Z & Dunker AK (2004) Natively disordered proteins: functions and predictions. *Appl Bioinformatics* **3**, 105–113.
9 Greene LH, Wijesinha-Bettoni R & Redfield C (2006) Characterization of the molten globule of human serum retinol-binding protein using NMR spectroscopy. *Biochemistry* **45**, 9475–9484.
10 Forge V, Wijesinha RT, Balbach J, Brew K, Robinson CV, Redfield C & Dobson CM (1999) Rapid collapse and slow structural reorganisation during the refolding of bovine alpha-lactalbumin. *J Mol Biol* **288**, 673–688.
11 Man P, Montagner C, Vernier G, Dublet B, Chenal A, Forest E & Forge V (2007) Defining the interacting regions between apomyoglobin and lipid membrane by hydrogen/deuterium exchange coupled to mass spectrometry. *J Mol Biol* **368**, 464–472.
12 Chenal A, Vernier G, Savarin P, Bushmarina NA, Gèze A, Guillain F, Gillet D & Forge V (2005) Conformational states and thermodynamics of alpha-lactalbumin bound to membranes: a case study of the effects of pH, calcium, lipid membrane curvature and charge. *J Mol Biol* **349**, 890–905.
13 Krishna MM, Hoang L, Lin Y & Englander SW (2004) Hydrogen exchange methods to study protein folding. *Methods. Sep.* **34**, 51–64.
14 Maier CS, Schimerlik MI & Deinzer ML (1999) Thermal denaturation of *Escherichia coli* thioredoxin studied by hydrogen/deuterium exchange and electrospray ionization mass spectrometry: monitoring a two-state protein unfolding transition. *Biochemistry* **38**, 1136–1143.
15 Mazon H, Marcillat O, Forest E, Smith DL & Vial C (2004) Conformational dynamics of the GdmHCl-induced molten globule state of creatine kinase monitored by hydrogen exchange and mass spectrometry. *Biochemistry* **43**, 5045–5054.
16 Rey M, Man P, Brandolin G, Forest E & Pelosi L (2009) Recombinant immobilized rhizopuspepsin as a new tool for protein digestion in H/D exchange mass spectrometry. *Rapid Commun Mass Spectrom* **23**, 3431–3438.
17 Zhang Z & Smith DL (1993) Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci* **2**, 522–531.
18 Rand KD & Jørgensen TJ (2007) Development of a peptide probe for the occurrence of hydrogen (1H/2H) scrambling upon gas-phase fragmentation. *Anal Chem* **79**, 8686–8693.
19 Wang L, Lane LC & Smith DL (2001) Detecting structural changes in viral capsids by hydrogen exchange and mass spectrometry. *Protein Sci* **10**, 1234–1243.
20 Bai Y, Milne JS, Mayne L & Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86.
21 Li R & Woodward C (1999) The hydrogen exchange core and protein folding. *Protein Sci* **8**, 1571–1590.
22 Chenal A, Savarin P, Nizard P, Guillain F, Gillet D & Forge V (2002) Membrane protein insertion regulated by bringing electrostatic and hydrophobic interactions into play. A case study with the translocation domain of diphtheria toxin. *J Biol Chem* **277**, 43425–43432.
23 Montagner C, Perier A, Pichard S, Vernier G, Ménez A, Gillet D, Forge V & Chenal A (2007) Behavior of the N-terminal helices of the diphtheria toxin T domain during the successive steps of membrane interaction. *Biochemistry* **46**, 1878–1887.
24 Schulman BA, Redfield C, Peng ZY, Dobson CM & Kim PS (1995) Different subdomains are most protected from hydrogen exchange in the molten globule

and native states of human alpha-lactalbumin. *J Mol Biol* **253**, 651–657.
25 Bennett MJ, Choe S & Eisenberg D (1994) Refined structure of dimeric diphtheria toxin at 2.0 A resolution. *Protein Sci* **3**, 1444–1463.
26 Palchevskyy SS, Posokhov YO, Olivier B, Popot JL, Pucci B & Ladokhin AS (2006) Chaperoning of insertion of membrane proteins into lipid bilayers by hemifluorinated surfactants: application to diphtheria toxin. *Biochemistry* **45**, 2629–2635.
27 Bell CE, Poon PH, Schumaker VN & Eisenberg D (1997) Oligomerization of a 45 kilodalton fragment of diphtheria toxin at pH 5.0 to a molecule of 20-24 subunits. *Biochemistry* **36**, 15201–15207.
28 Perier A, Chassaing A, Raffestin S, Pichard S, Masella M, Ménez A, Forge V, Chenal A & Gillet D (2007) Concerted protonation of key histidines triggers membrane interaction of the diphtheria toxin T domain. *J Biol Chem* **282**, 24239–24245.
29 Zhang Z & Marshall AG (1998) A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J Am Soc Mass Spectrom* **9**, 225–233.
30 Comeau SR, Gatchell DW, Vajda S & Camacho CJ (2004) ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res* **32**, W96–W99.
31 Comeau SR & Camacho CJ (2005) Predicting oligomeric assemblies: N-mers a primer. *J Struct Biol* **150**, 233–244.

4.2.2 mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data

Anal. Chem. **XXXX**, *xxx,* 000–000

mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data

Martin Strohalm,*[,†,‡] Daniel Kavan,[†,§] Petr Novák,[†,§] Michael Volný,[†] and Vladimír Havlíček[†,∥]

Institute of Microbiology, Academy of Sciences of the Czech Republic, v.v.i., Vídeňská 1083, 142 20 Prague, Czech Republic, Institute of Chemical Technology, Technická 5, 166 28 Prague, Czech Republic, Department of Biochemistry, Faculty of Science, Charles University, Hlavova 8, 128 40 Prague, Czech Republic, and Department of Analytical Chemistry, Palacky University, 17.listopadu 12, 771 46 Olomouc, Czech Republic

While tools for the automated analysis of MS and LC-MS/MS data are continuously improving, it is still often the case that at the end of an experiment, the mass spectrometrist will spend time carefully examining individual spectra. Current software support is mostly provided only by the instrument vendors, and the available software tools are often instrument-dependent. Here we present a new generation of *mMass*, a cross-platform environment for the precise analysis of individual mass spectra. The software covers a wide range of processing tasks such as import from various data formats, smoothing, baseline correction, peak picking, deisotoping, charge determination, and recalibration. Functions presented in the earlier versions such as in silico digestion and fragmentation were redesigned and improved. In addition to Mascot, an interface for ProFound has been implemented. A specific tool is available for isotopic pattern modeling to enable precise data validation. The largest available lipid database (from the LIPID MAPS Consortium) has been incorporated and together with the new compound search tool lipids can be rapidly identified. In addition, the user can define custom libraries of compounds and use them analogously. The new version of *mMass* is based on a stand-alone Python library, which provides the basic functionality for data processing and interpretation. This library can serve as a good starting point for other developers in their projects. Binary distributions of *mMass*, its source code, a detailed user's guide, and video tutorials are freely available from www.mmass.org.

Acquisition of a single mass spectrum is still an important analytical tool in many areas such as process control, analysis of a dominant reaction or extraction products, or top down proteomics. In addition, many software tools for mass spectrometry imaging assemble the final images from independent spectra files obtained for each individual pixel. The primary focus of academic open software developers is on LC/MS experiments, which leaves most researchers dealing with single-spectrum experiment analysis to rely on the proprietary solutions of instrument vendors. Tight software−instrument relationships in such cases also pose a significant complication for data sharing between collaborating laboratories. There are many excellent tools focused on a specific task (Mascot,[1] ProFound,[2] GPMAW,[3] massXpert,[4] etc.), but no coherent software environment is available.[5] In many cases, one has to switch between multiple software and data formats to analyze a single spectrum. Another key reason for developing free open source software for analysis of mass spectra is mainly the users who need to analyze data on their computers, but purchasing a sufficient number of commercial licenses can be prohibitively expensive for them (e.g., graduate students and central facility customers).

With the first release of *mMass*,[6] we have provided a simple solution for the analysis of the single-spectrum experiments. The positive users' response was the main motivation to continue the development. To reflect the current state of mass spectrometry, the new version of *mMass* has been completely redesigned and many new tools have been added to make the software suitable for a wider range of possible applications.

EXPERIMENTAL SECTION

mMass is written in the Python programming language and uses wxPython libraries for the graphical user interface (Figure 1). The NumPy module is used for faster computing of mathematical tasks, and part of the code is written in C, particularly to accelerate the drawing of high-resolution mass spectra.

From the programming point of view, the most important change since the previous version is a complete separation of the core functionality and the user interface of the new *mMass*. The core library, named *mspy*, consists of several modules that cover a wide range of data processing tasks and in silico experimental simulations. The *mspy* library provides an import function from popular data formats, smoothing, baseline correction, automatic peak picking, deisotoping, charge determination, and data recalibration. Using Python's object-oriented nature, *mspy* defines a

* To whom correspondence should be addressed. E-mail: strohalm@biomed.cas.cz.

† Academy of Sciences of the Czech Republic.

‡ Institute of Chemical Technology.

§ Charles University.

∥ Palacky University.

(1) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551–3567.

(2) Zhang, W.; Chait, B. T. *Anal. Chem.* **2000**, *72*, 2482–2489.

(3) Peri, S.; Steen, H.; Pandey, A. *Trends Biochem. Sci.* **2001**, *26*, 687–689.

(4) Rusconi, F. *Bioinformatics* **2009**, *25*, 2741–2742.

(5) Matthiesen, R. *Mass Spectrometry Data Analysis in Proteomics*; Humana Press: Totowa, NJ, 2008.

(6) Strohalm, M.; Hassman, M.; Košata, B.; Kodicek, M. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 905–908.

10.1021/ac100818g © XXXX American Chemical Society

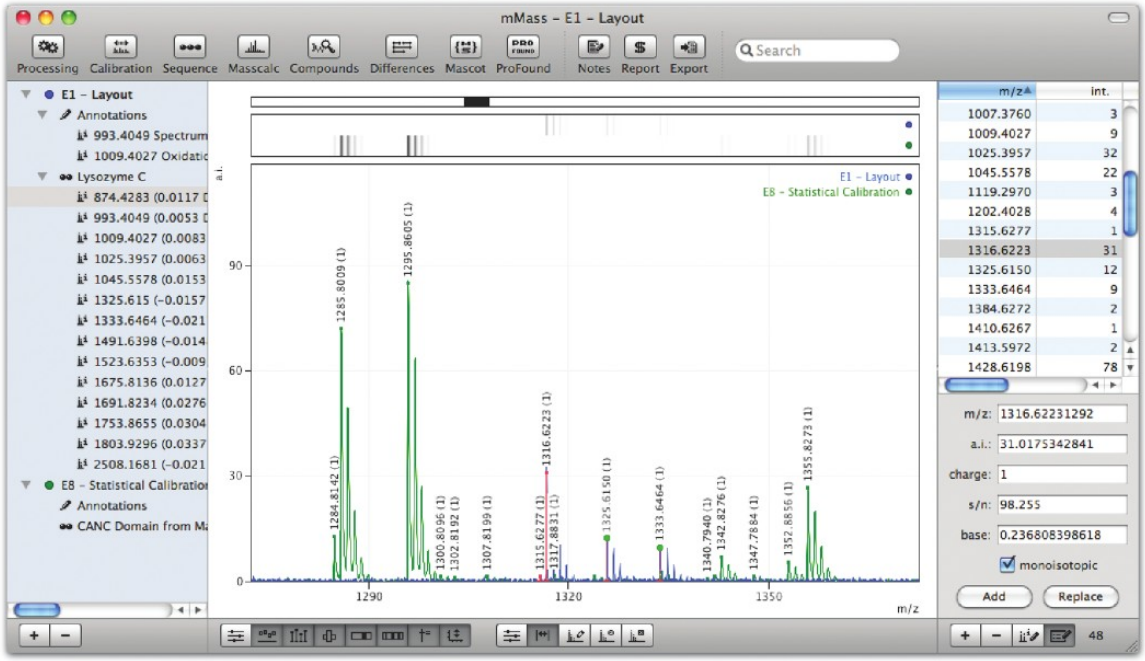


Figure 1. New user interface of *mMass* showing documents panel (left), spectrum viewer (middle), and peak list panel (right).

couple of basic objects such as *scan*, *peak*, *compound*, and *sequence*, each with its own advanced methods. Together with common enzymes and modification definitions, these objects are used for in silico experimental simulations starting from isotopic pattern modeling and going further to protein digestion or peptide fragmentation. In addition, a spectrum widget presents a powerful object for data visualization with advanced manipulation and viewing capabilities.

RESULTS

Data Processing. To support modern high-resolution instruments, an automatic peak picking functionality and subsequent charge determination and deisotoping have become essential. As reviewed by Young et al.,[7] current peak detection algorithms share a common procedure consisting of smoothing, baseline correction, and peak picking. Charge determination and deisotoping can be performed as additional steps. In *mMass*, spectra can be smoothed by moving average or Savitzky−Golay filters, depending on the mass range. If the baseline needs to be corrected, it is determined from the spectrum noise, and its precision can be specified by the user as a number of segments to be calculated separately. Peak picking starts with a search for local maxima followed by estimation of a peak centroid to yield the final m/z value. If the resolution of a particular data set is good enough to resolve isotopes, the charge state can be determined and monoisotopic peaks are labeled. To reveal possible overlaps, the isotope intensities are also taken into account and compared with theoretical profiles calculated from the averagine formula.[8] Common criteria such as the signal-to-noise ratio and intensity thresholds can be specified by the user to ignore low-intensity noise peaks. This automatic peak picking feature with its possibility to define user presets provides an easy way to label the peaks. The tools previously available for manual peak labeling are retained as well.

Data Recalibration. Using *mMass*'s recalibration tool (Figure 2A), acquired data can be corrected by external or internal calibration, or in the case of protein digests, statistical calibration can be utilized. Once reference masses are assigned, calibration constants are calculated using least-squares fitting and a linear or quadratic model and corresponding error plot is immediately shown. The resulting calibration can then be applied to any opened spectrum.

For every *mMass* tool, where some theoretical masses are compared with acquired data, an error plot is shown to reveal miscalibrations. Such tools provide an easy way to use identified peaks as reference masses for data recalibration. Although this feature must be used carefully, it can improve the data quality and reveal false positive matches in many cases.

Mass Calculation and Isotopic Pattern Simulation. A mass calculator tool provides a utility for calculation of molecular weights, compositions, ion series, and isotopic pattern simulations (Figure 2B). Molecular formulas can be manually typed, but the true advantage of this calculator is the ability to obtain the data from other *mMass* tools. Isotopic patterns for any in silico-generated item such as a peptide, a fragment, or any other compound can be modeled with a single mouse click and overlaid with acquired data. With this general tool, one can reveal peak overlaps and false positive matches or validate compounds with specific isotopic profiles.

(7) Yang, C.; He, Z. *BMC Bioinf.* **2009**, 10.
(8) Senko, M.; Beu, S.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229–233.

Figure 2. Examples of *mMass* processing and interpretation tools. All the tools are available in the form of floating panels to retain the user's focus on acquired data: (A) calibration tool, (B) isotopic pattern modeling, (C) compound search tool, and (D) protein digest tool.

In contrast to the previous version of *mMass*, mostly focused on singly charged MALDI-TOF data, any charge state can now be calculated throughout the software. Together with the automatic peak picking feature and its charge determination, *mMass* can now be used for the interpretation of electrospray data as well.

Compound Search. Searching for specific compounds in acquired data still remains one of the most common tasks in mass spectrometry. This is especially true for metabolomic or lipidomic projects, as well as for laboratories doing service analysis in addition to their own research. Small compounds are often ionized in the form of different adducts, and interpretation of such complex spectra is often a laborious task. *mMass* provides a dedicated tool to automate and accelerate this process (Figure 2C). The users can define their own database of compounds of interest that can be searched and compared with the acquired data. Different adducts, charge states, and mass tolerances can be specified by the user. All the matched peaks can be automatically annotated, and an error plot is shown. To validate the data, calculation of the theoretical isotopic pattern of a particular compound (and its overlay with the acquired data) is a matter of a single click.

This simple tool is essential when a long list of compounds has to be searched. Different types of lipids might be a good example, since thousands of lipids are currently known according to the LIPID MAPS Consortium.[9] With their permission, this lipid database has been incorporated into *mMass*. Combination of the database with the *mMass* functionalities such as automatic peak annotations, isotopic pattern modeling, and data recalibration makes the *Compound Search* a powerful tool for lipidomic analyses. Once the matched peaks are annotated, a detailed report provides direct links to the LIPID MAPS Consortium via specific IDs for gaining more information about a particular lipid.

Sequence Tools. In the *Sequence Editor*, an unlimited number of modifications can be applied to a protein or peptide sequence, each with the fixed or variable status. While corresponding peptides (using the *Protein Digest* tool) or fragments (using the *Peptide Fragmentation* tool) are generated, all the possible variants of modifications are calculated and can be matched with acquired data. The error plot, automatic annotation, isotopic pattern modeling, and data recalibration functionalities are available with these tools as well (Figure 2D).

Protein Identification. In addition to the popular Mascot search tools, another interface for protein identification has been added. Thanks to a support from the ProFound original authors, we were able to implement an interface for their ProFound tool to be accessible directly from *mMass*.

Analysis Report. According to the feedback obtained from the users, *mMass* is often used in the laboratories that perform service analysis in addition to research tasks. That is why a new analysis report tool is now available. Beside an operator's name, contact, instrument description, and spectrum metadata, it includes a spectrum image, analysis notes/results, peak annotations, and all the sequences together with corresponding peptide or fragment matches. The analysis report is generated in the form of a HTML

(9) Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill, A. H.; Murphy, R. C.; Raetz, C. R.; Russell, D. W.; Subramaniam, S. *Nucleic Acids Res.* **2007**, *35*, D527–D532.

page, which opens in a web browser and can be easily printed, saved, or converted into PDF format to be sent to a client.

CONCLUSION

At the beginning of the *mMass* project, its main objective was to provide a simple tool for bypassing a tight software−instrument relationship created by instrument vendors. It enabled basic interpretation of the data originating from different instruments on a single software platform. While this basic nature has been retained, many new features have been added, and those previously available were redesigned and improved, which significantly extends the range of possible *mMass* applications.

mMass remains a single-spectrum analysis tool and a general-usage software that is useful in many areas of mass spectrometry. We do not present a new algorithm but rather a unique combination of different utilities. Together with the powerful spectrum viewer, the combination of *mMass* tools in a single environment is the true advantage of the software. The possibility of generating detailed analysis reports and the fact that *mMass* is free and supports various data formats (mzXML,[10] mzData,[11,12] and ASCII XY) appear to be beneficial for laboratories providing service analysis and facilitate data sharing in general. Since *mMass* provides its own implementation of the basic processing functions, raw data can be analyzed without any prior processing.

An especially strong new feature is the implementation of the lipid database obtained from LIPID MAPS Consortium. This tool allows fast identification of lipid candidates according to their exact masses and is cross-linked with the online database that provides more detailed information about each potential lipid species.

The new version of *mMass* is based on a stand-alone Python library called *mspy*. We believe that this library could provide a good starting point for other developers and helps them focus on their own topics rather than on the basic tasks. Some of its functionality is also provided by OpenMS/TOPP, one of the well-known open-source projects written in C++.[13,14] Python enables fast software development and prototyping and therefore is very popular in the bioinformatics community.

Although *mMass* is primarily focused on single-spectrum analyses, it can be benefical for LC/MS experiments as well. Different *mMass* utilities can be used to validate results obtained from other software platforms such as Trans-Proteomic Pipeline (TPP)[15] via mzXML support. In addition, the *mspy* library has been initially developed to support our different projects and therefore is capable of handling LC/MS data as well.[16]

mMass remains free, open-source, and platform-independent software. It has been released under GNU General Public License, and since the source code is also available, it has good potential to be easily modified or extended by additional modules for specific needs. Binary distributions, source code, a user's guide, and video tutorials are freely available from www.mmass.org.

(10) Pedrioli, P. G.; Eng, J. K.; Hubley, R.; Vogelzang, M.; Deutsch, E. W.; Raught, B.; Pratt, B.; Nilsson, E.; Angeletti, R. H.; Apweiler, R.; Cheung, K.; Costello, C. E.; Hermjakob, H.; Huang, S.; Julian, R. K.; Kapp, E.; McComb, M. E.; Oliver, S. G.; Omenn, G.; Paton, N. W.; Simpson, R.; Smith, R.; Taylor, C. F.; Zhu, W.; Aebersold, R. *Nat. Biotechnol.* **2004**, *22*, 1459–1466.

(11) Orchard, S.; Taylor, C.; Hermjakob, H.; Zhu, W.; Julian, R.; Apweiler, R. *Expert Rev. Proteomics* **2004**, *1*, 179–183.

(12) Orchard, S.; Hermjakob, H.; Taylor, C. F.; Potthast, F.; Jones, P.; Zhu, W.; Julian, R. K., Jr.; Apweiler, R. *Proteomics* **2005**, *5*, 3552–3555.

(13) Sturm, M.; Bertsch, A.; Gröpl, C.; Hildebrandt, A.; Hussong, R.; Lange, E.; Pfeifer, N.; Schulz-Trieglaff, O.; Zerck, A.; Reinert, K.; Kohlbacher, O. *BMC Bioinf.* **2008**, *9*, 163–174.

(14) Sturm, M.; Kohlbacher, O. *J. Proteome Res.* **2009**, *8*, 3760–3763.

(15) Keller, A.; Eng, J.; Zhang, N.; Li, X. J.; Aebersold, R. *Mol. Syst. Biol.* **2005**, *1*, 1–8.

(16) Strohalm, M.; Novak, P.; Pompach, P.; Man, P.; Kavan, D.; Witt, M.; Dzubak, P.; Hajduch, M.; Havlicek, V. *J. Mass Spectrom.* **2009**, *44*, 1565–1570.

ACKNOWLEDGMENT

We thank to the members of the LIPID MAPS Consortium for providing us the lipid database and the developers of the ProFound tool for their support. This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (LC07017 and MSM6198959216) and Institutional Research Concept (AV0Z50200510).

SUPPORTING INFORMATION AVAILABLE

A detailed user's guide describing software installation and usage. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review March 30, 2010. Accepted May 3, 2010.

AC100818G

4.2.3 Utilization of high-accuracy FTICR-MS data in protein quantitation experiments

Research Article

Journal of MASS SPECTROMETRY

Received: 22 December 2008 Accepted: 12 May 2009 Published online in Wiley Interscience: 17 June 2009

(www.interscience.com) DOI 10.1002/jms.1602

Utilization of high-accuracy FTICR-MS data in protein quantitation experiments

Martin Strohalm,[a] Petr Novak,[a] Petr Pompach,[a] Petr Man,[a] Daniel Kavan,[a] Matthias Witt,[b] Petr Dzubak,[c] Marian Hajduch[c] and Vladimir Havlicek[a,d]*

Human acute T-lymphoblastic leukemia cell line (CEM) treated with cisplatin, and the stable isotope labeling by amino acids in cell culture (SILAC) strategy were used to present an improved method of data processing in high-accuracy mass spectrometry (MS). By using peptide mass fingerprinting with low mass tolerance, we were able to utilize far more data retained in MS scans which would normally be missed by a standard processing method. This new way of data interpretation results in an improvement of the relevance of quantitation experiments and enabled us to search and quantify different types of posttranslational modifications. Furthermore, we used this technique to distinguish among different protein isoforms, commonly returned by Mascot search engine. **Copyright © 2009 John Wiley & Sons, Ltd.**

Keywords: mass spectrometry; protein quantitation; SILAC; workflow; software

Introduction

Mass spectrometry (MS) has become a very popular tool for protein quantitation and there are many different approaches currently available.[1,2] One of the most widely used methods applied to quantitation and screening of protein expression changes utilizes stable isotope labeling by amino acids in cell culture (SILAC).[3] In this technique, two cell populations are grown on two separate media containing either normal or isotopically labeled amino acids such as arginine or lysine. These amino acids can be chosen individually with respect to the particular proteomic task but, in general, the use of essential amino acids represents the best choice. After several cell doublings, a mixture of total cell lysates of labeled and nonlabeled culture is often fractionalized by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), digested and analyzed by liquid chromatography coupled to MS. This approach results in the detection of peptides as pairs derived from nonlabeled and labeled culture and the light-to-heavy intensity ratio (L/H) then reflects the expression changes of a corresponding protein. Although this method was originally developed for mammalian cells,[4] it has also been adapted for bacteria,[5] yeasts[6] and plants.[7]

Though online liquid chromatography coupled to tandem mass spectrometry with electrospray ionization (LC-ESI-MS) is mostly preferred for the analysis, offline liquid chromatography coupled to tandem mass spectrometry with matrix-assisted laser desorption/ionization (LC-MALDI-MS) is becoming popular as well.[8] Both techniques have their own benefits and limitations. In the case of online LC-ESI-MS, the speed and better ability to automate the whole process is the major benefit. On the other hand, the resulting mass spectra are more complex as multiply charged peptides are generated, and interpretation is thus more difficult. In contrast, MALDI-MS generates singly charged peptides only and often with higher accuracy compared to commonly used ion traps in LC-ESI-MS experiments. The benefit is also the possibility to reanalyze samples preserved on matrix assisted laser desorption/ionization (MALDI) targets to increase the number of identified peptide pairs or to subsequently search for any possible modification(s).

The general strategy behind the MS analysis in SILAC quantitation experiments is rather simple.[3,9,10] In the case of LC-ESI-MS, as the sample is flowing out of an LC system, mass spectrometer is continuously switching between full MS and MS/MS scanning. Typically, from every MS scan up to three or four precursors are selected for subsequent MS/MS experiments. During interpretation of the data, peptides/proteins are identified by the MS/MS scans while the MS scans are used only for quantitation calculations (Fig. 1). Since only peaks previously identified by MS/MS are used for quantitation, most of the information from full MS scans is not utilized at all. In addition, precursor candidates are often selected from the most intense peaks in a particular scan. Altogether, this approach has a significant impact on protein sequence coverage and therefore on quantitation results. It also dramatically decreases a chance to quantify post-translational modifications. Although this procedure is often used with very good results, there is still a lot of information that can be utilized, especially if the data were measured with high accuracy that can be achieved by modern instruments such as

* *Correspondence to: Vladimir Havlicek, Laboratory of Molecular Structure Characterization, Institute of Microbiology, Academy of Sciences of the Czech Republic, Videnska 1083, Prague, 142 20, Czech Republic. E-mail: vlhavlic@biomed.cas.cz*

a *Laboratory of Molecular Structure Characterization, Institute of Microbiology, Academy of Sciences of the Czech Republic, Videnska 1083, Prague, 142 20, Czech Republic*

b *Bruker Daltonics GmbH, Bremen, Germany*

c *Laboratory of Experimental Medicine, Department of Pediatrics, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czech Republic*

d *Palacky University, Department of Analytical Chemistry, Olomouc, Svobody 8, Czech Republic*

Copyright © 2009 John Wiley & Sons, Ltd.

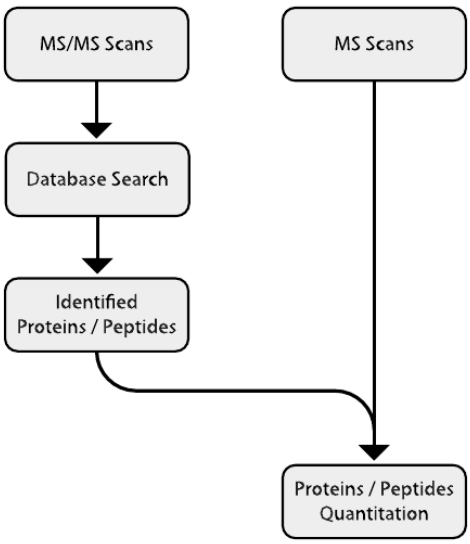


Figure 1. Schematic diagram of a standard data processing workflow in protein quantitation experiments.

Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS).

Experimental

Cell culture and SILAC labeling

Human acute T-lymphoblastic leukemia cell line CEM (ATTC, Manassas, VA) was grown on a regular nonlabeled cell culture medium (RPMI 1640 with 2 mM glutamine, 100 U/ml of penicillin, 100 µg/ml of streptomycin and 10% of heat inactivated fetal bovine serum and treated with at 5 × 50% growth inhibitory concentration of cisplatin corresponding to 0.757 µg/ml. Control cell culture was grown on a medium containing isotopically labeled arginine and lysine ($^{13}C_6$) for five cell doubling times. The optimal timepoint for cell harvesting was set up at 150 min after cisplatin treatment, which corresponds to the half time to caspase 3 and/or 7 activation. Initial concentration of T-lymphoblastic leukemia cells at cultivation time zero was adjusted to 1×10^6 cells per 4 ml of a medium and kept constant until harvesting. Caspase activation was measured in vital cells treated with *versus* without cisplatin using the Magic Red Caspase Detection Kit (Immunochemistry Technologies, LLC, Bloomington, MN).

Before the cell lysis, both the control and treated cultures were mixed together in a 1 : 1 ratio and washed twice with ice cold phosphate buffered saline (PBS) containing the mixture of protease and phosphatase inhibitors (Complete Mini tabs, Roche, 1 mM β-glycerolphosphate, Sigma, 5 mM sodium fluoride, Sigma, and 1 mM sodium orthovanadate, Sigma). The lysis was carried out in a buffer containing 20 mM Tris-HCl, 6 M urea, 100 mM dithiothreitol, 1% Triton X-100, 0.5% SDS and 10 units of benzonase endonuclease.

Gel electrophoresis and digestion

The mixture of total lysates of labeled and nonlabeled cultures was loaded onto a preparative SDS-PAGE with 7–15% gradient gel (7 mm × 100 mm tube). Hundred microliters of the sample with 50 µl of loading buffer was used. After separation, gel tube was sliced to strips of 2 mm in height (37 fractions) for in-gel reduction (by 40 mM tris(2-carboxyethyl)phosphine, 90 °C, 15 min), alkylation (by 40 mM iodoacetamide, room temperature, 90 min) and trypsin digestion (1 µg/fraction).

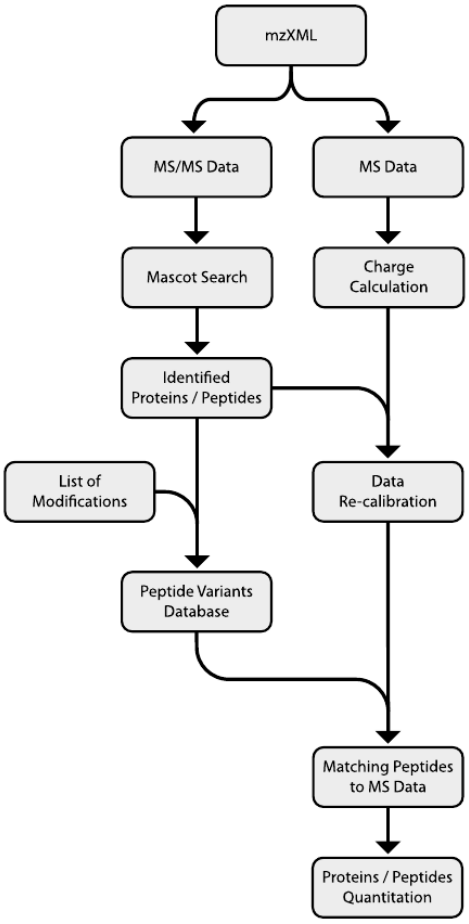


Figure 2. Schematic diagram of our improved data processing workflow in protein quantitation experiments.

Liquid chromatography and mass spectrometry

UltiMate 1000 capillary LC system (LC-Packings) was used for sample separation. The peptide mixture was loaded onto a reversed phase Magic C18AQ column (5 µ, 200 Å, 0.2 mm × 150 mm) for gradient elution (A: 0.2% formic acid in 5% MeCN/water and B: 0.2% formic acid in 95% MeCN/water). For online LC-ESI-MS analysis, gradient from 5 to 35% of B within 40 min was used and eluted peptides were continuously measured by APEX Ultra 9.4 T mass spectrometer equipped with Apollo II ESI/MALDI ion source (Bruker Daltonics, Billerica MA). One full MS and three MS/MS spectra were collected for each experiment cycle. An external calibration was performed on signals of arginine clusters prior to the analysis. Raw data

 Copyright © 2009 John Wiley & Sons, Ltd.

Journal of MASS SPECTROMETRY

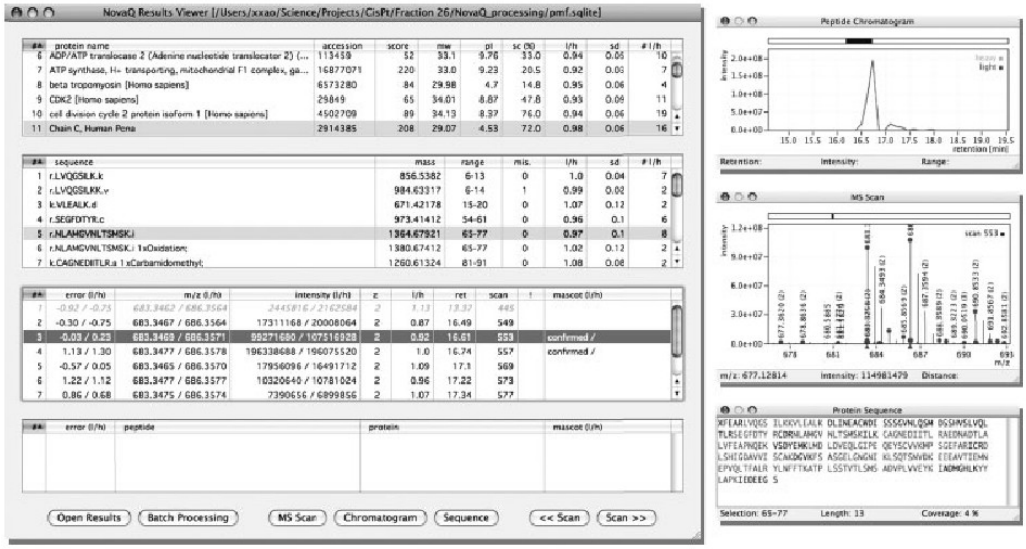


Figure 3. Screenshot of the software. The first part (the topmost) of the main window contains a summary for all identified proteins. The second part lists matched peptides for a selected protein, the third part lists the corresponding matched peaks and the fourth part lists possible co-matched peptides of a selected peak pair. In addition, three small windows with MS scan of a selected peak pair, chromatogram of a selected peptide and protein sequence can also be shown.

were converted to centroided mzXML format using ApexControl and CompassXport software (Bruker Daltonics). The Apex peak-picking algorithm was used. Data were processed by in-house Mascot server and additional software developed in our laboratory. Part of the sample was also used for offline LC-MALDI-MS. After separation on the same LC system by using gradient from 5 to 40% of B within 85 min, peptides were eluted on pre-spotted AnchorChip target (PAC384, Bruker Daltonics) using PROTEINEER-LC robot (Bruker Daltonics). All of the 384 positions were utilized. MALDI-TOF-MS spectra were acquired on ULTRAFLEX III MALDI-TOF-TOF instrument (Bruker Daltonics) with external calibration. Data were analyzed by automatic workflow using WARP-LC 1.1 and BioTools 3.1 software (Bruker Daltonics) and in-house Mascot server.

Results and Discussion

Utilization of high-accuracy mass data

As mentioned earlier, in a classical data processing workflow the peptides identified by preceding MS/MS scans are used for protein identification and quantitation only. This is obvious with such complex mixtures as whole cell lysates, where the measured m/z value usually corresponds to many different peptides within a relevant tolerance (typically 10 ppm for TOF-TOF or 100 ppm for ion-trap instruments). In the case of FTICR-MS, however, we can reach an accuracy even below 1 ppm. For such a high accuracy, the number of theoretical peptides corresponding to a particular m/z value decreases dramatically.[11–13] Unfortunately, in this type of experiments, the complexity of the samples is still too high to identify proteins simply by peptide mass fingerprint (PMF) alone, even with this very high mass accuracy. Once the proteins are successfully identified by MS/MS scans, however, we can use the MS scans to search only against these proteins. By this approach, protein sequence coverage can be increased, which consequently not only improves the relevance of the protein quantitation but also allows us to search for a number of different types of posttranslational modifications.

On the basis of the idea mentioned above and using Python programming language,[14] we have developed a software tool for high-accuracy data interpretation in protein quantitation experiments. The overall application workflow is illustrated in Fig. 2. Raw data are imported as mzXML[15–17] files with centroided peaks. This format is open source and well-documented and appropriate converter exists for almost all manufacturers' native files. These data, as well as all the processing results, are then stored locally in a simple SQLite database file.[18] Once the data are imported, charge state is calculated for all peak clusters in MS scans using mass differences and averagine intensity distribution. In the next step, MS/MS scans are used to generate a Mascot query (in Mascot Generic Format[19]) and are sent to a local Mascot server. Identified proteins are stored and a list of predefined modifications is applied to all the sequences. These modifications can be set as fixed (e.g. carbamidomethylation) or variable (e.g. acetylation). Different modifications can be set for Mascot search and subsequent PMF search. In addition, theoretical m/z values of positively identified peptides are used to recalibrate all the MS scans to reduce the average mass error.[20] According to the sequences of identified proteins and specified digestion enzyme, a database of theoretical peptides is generated with all the possible variants of modifications included. Recalibrated MS scans are then searched against this database within a specified tolerance. Matching peptides, where both light and heavy forms were found, are then used for subsequent quantitation calculations. The L/H ratio is calculated from areas of corresponding extracted ion chromatograms.

The software consists of two main parts – processing part and results viewer. Typical quantitation experiment is usually

 Copyright © 2009 John Wiley & Sons, Ltd.

Table 1. Summary of identified and quantified proteins from a single '35 kDa' gel fraction. Proteins are ordered by decreasing Mascot Score based on LC-ESI-MS experiment

Protein name	Accession number	Mascot Score	Average L/H		Number of peptides		Sequence coverage		LC-MALDI-MS validation Peptides/Unique
			St.	PMF	St.	PMF	St.	PMF	
Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	5 174 447	321	0.94	0.94	11	19	41.5	68.4	8/2
Tyrosine 3/tryptophan 5 -monooxygenase activation protein, epsilon polypeptide	5 803 225	249	0.95	0.97	10	18	33.5	49.2	8/2
Porin 31HM	238 427	238	0.98	0.96	6	13	22.7	62.4	7/2
S3 ribosomal protein	7 765 076	222	0.92	0.93	8	17	41.7	74.0	11/6
ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	16 877 071	220	0.94	0.92	5	7	18.2	20.5	3/1
Chain C, Human Pcna	2 914 385	208	1.00	0.98	8	16	31.4	72.0	6/3
Ribosomal protein L7a	4 506 661	175	0.94	0.95	7	13	22.6	40.8	6/2
Tropomyosin 3 isoform 2	24 119 203	175	0.91	0.92	10	13	31.6	42.1	9/1
Histone cluster 1, H1c	4 885 375	138	0.97	1.00	6	11	23.6	28.3	5/0
Histone cluster 1, H1d	4 885 377	138	0.97	0.99	6	10	22.7	27.3	6/1
3-hydroxyacyl-CoA dehydrogenase	1 483 511	118	0.95	0.96	5	12	16.0	32.3	4/3
Acidic (leucine-rich) nuclear phosphoprotein 32 family, member A	5453880	117	1.00	0.96	4	12	16.5	37.1	4/2
Ribosomal protein S6	6 677 809	102	0.96	0.95	5	6	21.4	23.8	4/0
Solute carrier family 25 member 3 isoform b precursor	4505775	99	0.93	0.93	4	15	12.5	35.3	8/5
Cell division cycle 2 protein isoform 1	4 502 709	89	0.97	0.94	5	19	18.6	76.0	9/4
Ribosomal protein S2	15 055 539	88	0.95	0.94	10	17	37.7	51.4	8/3
Human pre-mRNA splicing factor SF2p32	338 043	87	0.91	0.94	2	7	10.4	33.5	2/2
Beta tropomyosin	6 573 280	84	0.93	0.95	3	4	11.3	14.8	2/0
Purine nucleoside phosphorylase	387 033	77	0.98	0.98	4	12	13.5	38.2	6/5
Porin	190 200	76	0.97	0.93	3	9	13.3	30.3	2/1
CDK2	29 849	65	0.95	0.93	3	11	11.1	47.8	4/1
Ribosomal protein L7	35 903	62	0.95	0.96	3	12	14.2	47.8	4/2
Enolase-phosphatase 1	10 864 017	61	0.98	0.97	2	6	9.2	36.5	1/0
Spermidine synthase	531 202	61	0.92	0.97	2	6	7.6	23.6	2/1
Histone cluster 1, H1b	4 885 381	59	1.15	1.06	2	10	10.2	26.7	4/3
Microtubule-associated protein, RP/EB family, member 1	6 912 494	57	0.96	0.96	6	15	25.8	58.1	5/3
EF-hand domain family, member D2	20 149 675	55	0.97	0.94	4	11	15.9	35.1	5/3
F-actin capping protein beta subunit	4 826 659	55	0.97	0.96	4	10	18.8	26.9	3/2
ADP/ATP translocase 2	113 459	52	0.93	0.94	4	10	14.8	33.0	2/1
Splicing factor	337 926	51	0.95	0.95	3	5	18.6	24.5	3/1
Voltage-dependent anion channel 3	25 188 179	49	0.96	0.95	3	4	12.1	19.1	2/1
Actin related protein 2/3 complex subunit 2	5 031 599	48	1.04	0.97	3	18	8.7	60.5	7/6
Esterase D	182 265	47	1.05	1.02	1	9	4.0	27.9	2/1
Ribosomal protein small subunit	306 553	45	0.97	0.95	2	20	4.9	41.8	7/6
2-oxoglutarate carrier	23 844	41	0.92	0.99	4	15	12.8	48.2	5/3
Enoyl Coenzyme A hydratase 1, peroxisomal	16 924 265	41	0.90	0.87	1	7	4.9	14.1	4/3
PHAPI2b protein	1 498 227	41	1.00	1.01	3	7	17.0	27.8	2/2
THO complex 4	55 770 864	41	0.91	0.99	1	6	4.3	27.3	3/1

St., standard approach; PMF, peptide mass fingerprinting data acquired by our improved approach. LC-MALDI data are related to TOF-TOF standard approach.

 Copyright © 2009 John Wiley & Sons, Ltd.

represented by several LC-MS runs, therefore it is efficient to process the data in a batch mode. Many of the parameters can be specified and stored as a method in a human readable XML file. After processing the data, results are shown in the viewer, where a summary for all identified proteins and matched peptides can be evaluated (Fig. 3). It is generally a good practice to manually validate the results in quantitation experiments. For this purpose, all the matched peaks are listed in the viewer together with each peptide and can be seen in a context of the corresponding MS scan. Even if it is not usually the case, one *m/z* value can match more than one peptide. If so, all the co-matched peptides are also listed after the peak is highlighted. Using a small window showing peptide chromatogram, the results can also be validated within the meaning of peptide's sequence and retention time. Any peptide or even individual peak pair can be discarded and overall quantitation is recalculated on the fly.

Data processing example

To demonstrate this new approach of data processing, we used human acute T-lymphoblastic leukemia CEM cells treated with cisplatin and SILAC quantitation strategy. A single electrophoretic fraction corresponding to proteins with molecular weight about 35 kDa was analyzed by the LC-ESI-MS as well as by the LC-MALDI-MS.

The raw data from LC-ESI-MS experiment were converted to mzXML format and processed by the software developed in our laboratory. Initially, the data were analyzed by using a standard method, where only those peptides successfully identified by MS/MS were used for subsequent quantitation (Fig. 1). Peptides' identifications were achieved by sending the MS/MS scans to the in-house Mascot server. Data were searched against human proteome in NCBInr database. A tolerance of 5 ppm was used for the precursor mass and 0.01 Da for the fragments. The precursor charge state was set to 1+, 2+ and 3+, and the instrument type to ESI-FTICR. Furthermore, SILAC_KR_13C(6)15N(0) (KR), Oxidation (M), Acetyl (Protein N-term), Methyl (K) and Dimethyl (K) were set as variable modifications and carbamidomethyl (C) as fixed modification. Once the peptides were identified, their theoretical masses were used to recalibrate the MS data, and corresponding light-heavy peak pairs were searched within 1.5 ppm tolerance. As summarized in Table 1, we have successfully identified and quantified 38 different proteins with an average L/H ratio of 1 : 1. The number of identified peptides for a particular protein varied from 1 to 11 (5 as an average) and the corresponding sequence coverage varied from 4 to 40% (18% as an average). In addition, in six proteins we have identified acetylation of their *N*-terminal amino acids (Table 2).

In the next step, we have processed the data in the way of our improved approach using peptide mass fingerprinting to match also the peaks not selected for MS/MS experiment (Fig. 2). All the other parameters remained the same. As the identification itself is based solely on MS/MS scans, identified proteins are the same; however, the number of matched peptides increased significantly, varying from 4 to 20 (11 as an average). We can see the same tendency also in the sequence coverage, varying from 14 to 76% (40% as an average). Furthermore, we have successfully found *N*-terminal acetylation in four different proteins, while another four proteins were found to be methylated (Table 2).

To validate the peptides identified by PMF, the sample was also analyzed using offline LC-MALDI-MS. By this method, we have identified quite a similar set of proteins. However, the corresponding peptides were partially different compared to those identified previously by the LC-ESI-MS experiment (Table 1). These unique peptides were in a good agreement with the data we have obtained by PMF. Moreover, *N*-terminal acetylation and lysine methylation of Histone cluster 1, H1b were also confirmed successfully. To validate the acetylations in other proteins, we used the annotated SwissProt sequence database, in which these modifications were assigned for the remaining three proteins. Unfortunately, only one of the four methylations was confirmed by LC-MALDI-MS. Nevertheless, two methylations were identified in the proteins belonging to the histone family, which is in agreement with literature data.[21]

It is a common experience that more than one isoform of a protein is returned from Mascot search, especially when the number of identified peptides is limited. We have used those peptides acquired by PMF to distinguish the isoforms sharing the same set of peptides identified by MS/MS. For instance, in this particular experiment, the protein with the highest Mascot Score (Table 1) has been returned together with four additional proteins. These protein 'isoforms' differed in their *N*-terminal parts that have not been covered by MS/MS data. Using PMF we were able to found additional peptide even with *N*-terminal acetylation.

Table 2. List of proteins with identified posttranslational modification

Protein name	Modification	Peptide	ESI	MALDI	SwissProt
Acidic (leucine-rich) nuclear phosphoprotein 32 family, member A	Methylation	14–27	–	–	–
ADP/ATP translocase 2	Acetylation	1–9	–	–	+
Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	Acetylation	1–7	–	–	+
Histone cluster 1, H1b	Acetylation	1–16	–	+	+
	Methylation	67–81	–	+	–
Histone cluster 1, H1c	Acetylation	1–16	+	+	+
	Methylation	33–45	–	–	–
Histone cluster 1, H1d	Acetylation	1–16	+	+	+
	Methylation	34–46	–	–	–
Microtubule-associated protein, RP/EB family, member 1	Acetylation	1–16	+	+	+
Porin 31HM	Acetylation	1–11	+	+	+
Splicing factor	Acetylation	1–16	+	+	+
THO complex 4	Acetylation	1–13	–	–	+
Tropomyosin 3 isoform 2	Acetylation	1–10	+	+	+

Confirmation of particular modification with respective method is marked by 'plus' sign.

 Copyright © 2009 John Wiley & Sons, Ltd.

In one of the sequences only, this peptide is *N*-terminal and can be acetylated and therefore other sequences can be discarded. For this particular protein, *N*-terminal acetylation is also mentioned in the SwissProt database. We have processed all the proteins in a same way and even if we were not fully successful in all the cases, we were at least able to discard some isoform entries from the final protein list.

Conclusions

In this paper, we used human acute T-lymphoblastic leukemia CEM cell line treated with cisplatin and SILAC strategy to present a new method of high-accuracy quantitative MS data processing. By using peptide mass fingerprinting with 1.5 ppm mass tolerance, we were able to utilize far more data retained in MS scans which would normally be missed by a standard processing method. We have significantly increased the number of peptides used for protein quantitation and also the sequence coverage. Altogether, it enabled us to perform a more accurate quantitation as well as to search and identify some interesting posttranslational modifications. In addition, we used this technique to distinguish among closely related protein isoforms, commonly returned by Mascot search engine. Nonetheless, it has to be mentioned, that even with the high mass precision and accuracy of FTICR-MS instrument, false positive matches can also appear as the sample complexity increases. If the powerful and joint combination of online LC-ESI-FTICR-MS and offline LC-MALDI-TOF-TOF-MS is applied, our new approach of data processing can be successfully used in straightforward candidate selection process for subsequent offline detailed mass analysis.

Acknowledgements

This paper is dedicated to our friend and colleague, Dr Josef Chmelik. The work was financially supported by the Ministry of Education, Youth and Sports of the Czech Republic (LC07017 and MSM6198959216) and Institutional Research Concept (AV0Z50200510). We wish to express our gratitude to all people involved in the excellent Python programming language development.

References

[1] A. Bachi, T. Bonaldi. Quantitative proteomics as a new piece of the systems biology puzzle. *Journal of Proteomics* **2008**, *71*, 357. DOI: 10.1016/j.jprot.2008.07.001.
[2] C. H. Chen. Review of a current role of mass spectrometry for proteome research. *Analytica Chimica Acta* **2008**, *624*, 16. DOI: 10.1016/j.aca.2008.06.017.
[3] S. E. Ong, M. Mann. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature Protocols* **2007**, *1*, 2650. DOI: 10.1038/nprot.2006.427.
[4] S. E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey, M. Mann. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular and Cellular Proteomics* **2002**, *1*, 376. DOI: 10.1074/mcp.M200025-MCP200.
[5] M. J. Kerner, D. J. Naylor, Y. Ishihama, T. Maier, H. C. Chang, A. P. Stines, C. Georgopoulos, D. Frishman, M. Hayer-Hartl, M. Mann, F. U. Hartl. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **2005**, *122*, 209. DOI: 10.1016/j.cell.2005.05.028.
[6] H. Jiang, A. M. English. Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine. *Journal of Proteome Research* **2002**, *1*, 345.
[7] A. Gruhler, W. X. Schulze, R. Matthiesen, M. Mann, O. N. Jensen. Stable isotope labeling of Arabidopsis *thaliana* cells and quantitative proteomics by mass spectrometry. *Molecular and Cellular Proteomics* **2005**, *4*, 1697. DOI: 10.1074/mcp.M500190-MCP200.
[8] F. Schmidt, H. K. Hustoft, M. Strozynski, C. Dimmler, T. Rudel, B. Thiede. Quantitative proteome analysis of cisplatin-induced apoptotic Jurkat T cells by stable isotope labeling with amino acids in cell culture, SDS-PAGE, and LC-MALDI-TOF/TOF MS. *Electrophoresis* **2007**, *28*, 4359. DOI: 10.1002/elps.200700119.
[9] S. E. Ong, M. Mann. Mass spectrometry-based proteomics turns quantitative. *Nature Chemical Biology* **2005**, *1*, 252. DOI: 10.1038/nchembio736.
[10] S. E. Ong, L. J. Foster, M. Mann. Mass spectrometric-based approaches in quantitative proteomics. *Methods* **2003**, *29*, 124. DOI: 10.1016/S1046-2023(02)00303-1.
[11] B. Spengler. De novo sequencing, peptide composition analysis, and composition-based sequencing: a new strategy employing accurate mass determination by Fourier transform ion cyclotron resonance mass spectrometry. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 703. DOI: 10.1016/j.jasms.2004.01.007.
[12] B. Spengler. Accurate mass as a bioinformatic parameter in data-to-knowledge conversion: Fourier transform ion cyclotron resonance mass spectrometry for peptide de novo sequencing. *European Journal of Mass Spectrometry* **2007**, *13*, 83. DOI: 10.1255/ejms.840.
[13] R. A. Zubarev, P. Håkansson, B. Sundqvist. Accuracy Requirements for Peptide Characterization by Monoisotopic Molecular Mass Measurements. *Analytical Chemistry* **1996**, *68*, 4060. DOI: 10.1021/ac9604651.
[14] Python Programming Language. Available: http://www.python.org/ (accessed November 2008).
[15] P. G. Pedrioli, J. K. Eng, R. Hubley, M. Vogelzang, E. W. Deutsch, B. Raught, B. Pratt, E. Nilsson, R. H. Angeletti, R. Apweiler, K. Cheung, C. E. Costello, H. Hermjakob, S. Huang, R. K. Julian, E. Kapp, M. E. McComb, S. G. Oliver, G. Omenn, N. W. Paton, R. Simpson, R. Smith, C. F. Taylor, W. Zhu, R. Aebersold. A common open representation of mass spectrometry data and its application to proteomics research. *Nature Biotechnology* **2004**, *22*, 1459. DOI:10.1038/nbt1031.
[16] S. M. Lin, L. Zhu, A. Q. Winter, M. Sasinowski, W. A. Kibbe. What is mzXML good for?. *Expert Review of Proteomics* **2005**, *2*, 839.
[17] Seattle Proteome Center. Available: http://tools.proteomecenter.org/ (accessed November 2008).
[18] SQLite Homepage. Available: http://www.sqlite.org/ (accessed November 2008).
[19] Matrix Science. http://www.matrixscience.com/ (accessed November 2008).
[20] R. Zubarev, M. Mann. On the proper use of mass accuracy in proteomics. *Molecular and Cellular Proteomics* **2007**, *6*, 377.
[21] C. A. Johnson, J. P. Goddard, R. L. Adams. The effect of histone H1 and DNA methylation on transcription. *Biochemical Journal* **1995**, *305*, 791.

 Copyright © 2009 John Wiley & Sons, Ltd.

5 Discussion

About 30% of all recently introduced pharmaceuticals are of a protein nature, as it allows higher specificity and thus lower probability of various side-effects. It is possible due to the rapid progress in technology resulting in deeper understanding of the processes in living organisms. Crucial role in this research play the protein structures and functions, therefore many scientific teams aim towards these aspects.

The exact role of CD69 molecule in the process of antitumor immunity is still not fully understood. Although CD69 was initially suggested to be an activation receptor [Moretta A. et. al. 1991], more recently it has been shown that the exact role is more likely in immune response downregulation and it seems that some tumors are able to trigger apoptosis of NK cells by stimulating this surface receptor [Esplugues E. et. al. 2003]. This mechanism could be very important for the organism in order to keep NK cells under control and it could be very beneficial to take control over this process during the tumor therapies. Furthermore, our laboratory is also focused on other NK cell surface receptors (such as NKR-P1 family, NKG2D, etc.) and therefore we hope to deduce the consequences of stimulating these receptors complexly.

The preparation of homogenous populations of NK cells is extremely difficult and the isolation of transmembrane protein from natural material is problematic and the behavior and properties of the native transmembrane protein in a solution can be very confusing. Moreover the glycosylation can constitute an obstacle in structural studies. As it was also proved, that CD69 glycosylation is not the key point in CD69 signaling [Vance B. A. et. al. 1997 and 1999], we can study the binding properties and structure of CD69 produced in prokaryotic expression systems. We need to keep on mind the difference and any *in vivo* experiments should be verified using material isolated from living organisms. Moreover, other members of our research team orientate

their efforts to eukariotyc systems (*Pichia pastoris* – wild-type and engineered to produce basic mammal type of glycosylation, and HEK293 cell line).

To provide sufficient amount of a protein of interest is very challenging task indeed. It is necessary to examine several constructs and to optimize the refolding protocol. Moreover, NMR experiments require relatively high concentrations (1mM) of stable, expensively ^{13}C and ^{15}N labeled samples, as the experiments run for long times at the room temperature. The searching for the optimal refolding protocol becomes (not only) economically very important in this point of view.

The exact sequence of the protein indicated to be another very important aspect. Many various constructs were cloned and produced in our laboratory. There were significant differences in *in vitro* refolding protocols as well as in long term stability of protein constructs just slightly differing from each other, whereas N-terminus of the protein accomplished to be the most significant part. The construct G70-K199 showed to be the most stable, it could be concetrated up to 20 mg/ml and was stable for several days at room temperature, while several others precipitated at concentrations as low as 5 mg/ml and a degradation visible on SDS gel electrophoresis occurred even after 24 hours at room temperature.

The majority of C-type lectin constructs produced in prokaryotic expression systems in our laboratory is precipitated into the inclusion bodies during production. We have tried the periplasmic production in pMAL-p2x system and production with assistance of low-temperature active chaperons in ArcticExpress cells, nevertheless none of this system produced reasonable amount of soluble protein, possibly because of three disulfide bonds in the molecule. Another possibility to try is the production in Origami 2 host cells with thioredoxin reductase and glutathione reductase mutations which should greatly enhance the

cytoplasmic disulfide bonds formation. Moreover the constructs could be cloned into the pET-32 vectors with thioredoxin fusion tag.

Not absolutely convenient, but sufficient way to obtain soluble CD69 is thus production in form of inclusion bodies and consequent *in vitro* refolding. The majority of proteins produced in our laboratory could be refolded with better or worse results using fast dilution method. Some of them seem to refold well in almost any non-denaturing solution, nevertheless we use at least 0.4M arginine as a stabilizing agent, Tris as a buffer at pH 8 and 5mM cysteamine and 1mM cystamine as a redox environment. CD69 belongs to medium-difficult group of proteins that are not very sensitive to concentration of stabilizing agent, but sensitive to cysteamine:cystamine ratio that is 18:1 for human CD69. Interestingly the rat and mouse CD69 renaturate well if this ratio is about 5:1. We have started attempts of bovine CD69 production recently in our laboratory and it seems to renaturate well in buffers similar to that of murine and rat receptors. The only form of human CD69 that could not be successfully renaturated using fast dilution method is the form containing the odd dimerizating cysteine residue. This form is produced with N-terminal fusion histidine tag and the refolding is performed while the protein is immobilized on the Ni-NTA column. This immobilization probably ensures the proper orientation and distance of dimerization thiol groups.

Before performing the ligand binding experiments we need to verify structure and stability of the protein. Besides classical biochemical techniques as SDS electrophoresis both under reducing and under non-reducing conditions, thermal stability experiments, long-term stability experiments and conventional MALDI-TOF PMF identification, also modern techniques as disulfide bonds characterization using FT-ICR MS or HSQC NMR experiments were used. Nicely dispersed sharp signals typical for well refolded protein were observed in HSQC spectra.

This type of experiments was also used for routine control of every new batch of protein before any NMR measurement.

Our proteins had all cysteines bound the same way as in published crystal structure (if it exists) and in accordance with consensus structure of this protein group. The disulfide bonds were checked using mass spectrometry.

Any binding study should be started only after obtaining a proof that the protein is stable and in native conformation. The binding studies can be performed in several ways and the better conception of the reality can be achieved through their combination. The binding experiments were performed using equilibrium dialysis, by the plate binding assays with either radioisotopical or fluorescent labels on examined proteins and some of them also by NMR titrations. The results from all of these methods were in a good accordance with each other and showed the K_d for GlcNAc to be in µM order and for branched saccharides even in 10^{-10} M order.

The binding studies concerning structural aspect could be performed also by H/D exchange. This kind of experiments were not held for C-type NK lectins within this thesis, however.

Another powerful tool of any structural biologist is molecular cloning and especially site-directed mutagenesis that has been used several times during the work on this thesis for instance for creation of non-dimerizing human CD69 molecules.

Majority of above mentioned problems and solutions are also generally applicable to other proteins and many of these problems can be solved with help of mass spectrometry.

We can take disulfide bonds as an example. They represent one of the most important structural element in protein fold and verification of their arrangement serves as an indicator of proper folding in recombinant protein production. Over past decades a large number of

techniques and protocols for disulfide bridges determination was developed. The disadvantages of almost all of these methods are laborious sample preparation, relatively difficult data interpretation and possible disulfide bonds scrambling. As it is very important tool for our research, we needed to develop simple, fast and reliable method. The method is derived from standard enzymatic digest of samples in polyacrylamide SDS gels, to prevent the disulfide bonds scrambling we add 0.2mM cystamine as a compound that prevents the bond opening. The use of cystamine provided better results than oxidized glutathione, possibly due to its small size and consequent easier penetration into the protein structure and it should be added also to the SDS sample buffer and to the electrophoresis running buffer. The digestion mixture is desalted and injected on reversed phase column connected online with mass spectrometer. Resulting peptide masses are compared to *in silico* digest data. Currently the work on even easier protocol is in progress. It is based on pepsinolysis and exceptional mass and abundance of ^{34}S.

The conventional MS data processing software provided by instrument vendors contains just the basic data handling functions common for almost any protein MS usage such as peak picking, deisotoping, charge calculation, baseline smoothing etc., and unfortunately has not the advanced functions to cover special issues. On the other hand it provides some sort of programming interface and scripting environment. DataAnalysis, the Bruker Daltonics software, uses Visual Basic for Application language for user scripting and provides access to some of its internal data structures – objects, their attributes and methods. Unfortunately these data structures are accessible just partially and Visual Basic for Applications is just a simple language, therefore writing of complex data handling programs is inconvenient and resulting scripts are rather confusing. Therefore almost every complex data handling program starts as DataAnalysis macro and at particular level of complexity it is necessary to rewrite it as a standalone

application. Such a situation is an ideal background for existence of free and open source MS data handling application that could be improved by anyone and adjusted to cover any needs. An example of such an application is mMass, currently in version 3. It provides the same basic functionality as the software from instrument vendor and in addition also many extra functions available in other applications such as *in silico* digestions and fragmentations or Mascot server interaction. mMass 3 is written in Python programming language which is very popular in scientific environment and uses wxPython libraries for its graphical interface, numPy module for some of its internal data structures and simple C language module for time-critical computations. mMass is platform independent and thus can be used also on minority operating systems. Its current main disadvantage is inability to handle LC-MS data.

The standard algorithms for protein/peptide peaks detection in mass spectrum are based on the shape of isotopic envelope, ergo on average elemental composition and on the natural isotope abundances, thus it fails to detect peptides with abnormal isotopic composition such as ^{13}C or ^{15}N labeled, or deuterium enriched peptides from H/D exchange experiments. Especially for H/D data evaluation it needs to be bypassed by user scripting because mass spectrometry has became an essential method for this technique.

Evaluation of H/D exchange experiments by NMR provides unique resolution as every single amino acid (except for proline) can be distinguished. Unfortunately, large amounts of proteins are required and the peaks have to be assigned to amino acids first. This task alone is complicated and some proteins may even not provide spectra of sufficient quality. By contrast the resolution of MS in H/D exchange depends on cleavage efficiency and is usually between five and ten amino acids. Nevertheless the amount of protein necessary for this experiment is in order of micrograms and it is possible to evaluate H/D exchange of flexible structures such as molten globules.

The current H/D exchange MS analysis software is a DataAnalysis macro. The main disadvantage of this form is that it could be used only for Bruker Daltonics instruments, therefore we plan to rewrite it as a standalone application processing mzXML data format. Current input to the macro is the peptide sequence, monoisotopic $[M+H]^+$ mass, charge and optionally scan limits for peptide occurrence. The output of first round of processing is the average mass of deuterated peptide of the defined charge from the scan where it was the most abundant. It would be better if the software analyzed all charge states (or at least defined interval of charge states) from all scans between defined limits and output should be the weighted average of these values. The point is, that one particular charge state in one particular scan can be randomly shifted. We plan to implement this new evaluation algorithm in the next version of the software which will probably still be the DataAnalysis macro. Another improvement planed for the next version is omitting the peptide mass from input file as it could be easily calculated from the sequence by the software.

After this first round follows the second one that calculates the percentage of deuteration related either to full deuteration (deuteration during long-time period) or to total theoretical deuteration (if the long-term experiment is not evaluated). The final result is 2D matrix of deuteration percentages of each peptide in each time period.

The key point of H/D exchange experiments are not the deuteration levels themselves, however. The full potential of this method is in comparison of two or more datasets with each other (for instance comparison of deuteration of protein itself and deuteration of protein with ligand can reveal the position of binding site as the ligand would mask the protein surface and thus slow down the deuteration).

Since comparing of two (or even more) datasets just in form of tables is inconvenient and non-descriptive, we have developed a software suite for H/D exchange data visualization accompanied with tools for

facilitating the experiment planning. It is web server based set of PHP scripts and is freely available via Internet on http://ms.biomed.cas.cz/MSTools/ (server in Prague) and mirrored on http://www.hxms.com/mstools/ (server in Seattle). The form of PHP scripts was chosen regarding the accessibility and adaptability as it is much easier to modify PHP script than C or Java language source code. However, as the tools evolve we plan to rewrite it as a standalone application with adjustable configuration. The modes of visualizing differ from compact heatmap, with all times and conditions in one image, to set of deuteration plots with one plot for each peptide. One of these tools also allows to create scripts for Jmol 3D structure viewer showing the progress of deuteration directly on the pdb structure. There are formats for quick overview as well as for presenting in scientific publications.

References

Adamson N. J., Reynolds E. C. (1997): Rules relating electrophoretic mobility, charge and molecular size of peptides and proteins. J Chromatogr B, 699, 133-47.

Anderson P., Caligiuri M., Ritz J., Schlossman S. F. (1989): CD3-negative natural killer cells express zeta TCR as part of a novel molecular complex. Nature, 341, 159-62.

Bajorath J., Aruffo A. (1994): Molecular model of the extracellular lectin-like domain in CD69. J Biol Chem, 269, 32457-63.

Barboza J. M., Salmen S., Cova J. A., Albarrán B., Goncalves L., Borges L., Hernández M., Berrueta L. (2002): Uncoupling activation-induced modulation of CD16 and CD69 in CD56+ cells during AIDS. APMIS, 110, 415-22.

Barten R., Torkar M., Haude A., Trowsdale J., Wilson M. J. (2001): Divergent and convergent evolution of NK-cell receptors. Trends Immunol, 22, 52-7.

Bauer S., Groh V., Wu J., Steinle A., Phillips J. H., Lanier L. L., Spies T. (1999): Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. Science, 285, 727-9.

Bezouska K, Crichlow G.V., Rose J.M., Taylor M.E., Drickamer K. (1991): Evolutionary conservation of intron position in a subfamily of genes encoding carbohydrate-recognition domains. J Biol Chem, 266, 11604-9.

Billadeau D. D., Upshaw J. L., Schoon R. A., Dick C. J. Leibson P. . (2003): NKG2D-DAP10 triggers human NK cell-mediated killing via a Syk-independent regulatory pathway. Nat Immunol, 4, 557-64.

Binstadt B. A., Billadeau D. D., Jevremović D., Williams B. L., Fang N., Yi T., Koretzky G. A., Abraham R. T., Leibson P. J. (1998): SLP-76 is a direct substrate of SHP-1 recruited to killer cell inhibitory receptors. J Biol Chem, 273, 27518-23.

Bix M, Liao NS, Zijlstra M, Loring J, Jaenisch R, Raulet D. (1991): Rejection of class I MHC-deficient haemopoietic cells by irradiated MHC-matched mice. Nature, 349, 329-31.

Boyington J. C., Riaz A. N., Patamawenu A., Coligan J. E., Brooks A. G., Sun P. D. (1999): Structure of CD94 reveals a novel C-type lectin fold:

implications for the NK cell-associated CD94/NKG2 receptors. Immunity, 10, 75-82.

Braakman I. (2001): A novel lectin in the secretory pathway. An elegant mechanism for glycoprotein elimination. EMBO Rep, 2, 666-8.

Brewer C. F. (2001): Lectin cross-linking interactions with multivalent carbohydrates. Mol Immunol Complex Carbohydr 2 (491), 17-25.

Brown M. G., Dokun A. O., Heusel J. W., Smith H. R., Beckman D. L., Blattenberger E. A., Dubbelde C. E., Stone L. R., Scalzo A. A., Yokoyama W. M. (2001): Vital involvement of a natural killer cell activation receptor in resistance to viral infection. Science, 292, 934-7.

Cebrián M., Yagüe E., Rincón M., López-Botet M., de Landázuri M. O., Sánchez-Madrid F. (1988): Triggering of T cell proliferation through AIM, an activation inducer molecule expressed on activated human lymphocytes. J Exp Med, 168, 1621-37.

Childs R. A., Galustian C., Lawson A. M., Dougan G., Benwell K., Frankel G., Feizi T. (1999): Recombinant soluble human CD69 dimer produced in Escherichia coli: reevaluation of saccharide binding. Biochem Biophys Res Commun, 266, 19-23.

Colonna M., Samaridis J. (1995): Cloning of immunoglobulin-superfamily members associated with HLA-C and HLA-B recognition by human natural killer cells. Science, 268, 405-8.

Colucci F., Di Santo J. P., Leibson P. J. (2002): Natural killer cell activation in mice and men: different triggers for similar weapons? Nat Immunol, 3, 807-13.

Conde M., Montaño R., Moreno-Aurioles V. R., Ramirez R., Sanchez-Mateos P., Sanchez-Madrid F., Sobrino F. (1996): Anti-CD69 antibodies enhance phorbol-dependent glucose metabolism and Ca2+ levels in human thymocytes. Antagonist effect of cyclosporin A. J Leukoc Biol, 60, 278-84.

Cuthbert A. W., Fuller W. (2003): Investigation of folding and degradation of in vitro synthesized mutant proteins in microsomes. Methods Mol Biol, 232, 265-83.

Diefenbach A., Jensen E. R., Jamieson A. M., Raulet D. H. (2001): Rae1 and H60 ligands of the NKG2D receptor stimulate tumour immunity. Nature, 413, 165-71.

Dodd R. B., Drickamer K. (2001): Lectin-like proteins in model organisms: implications for evolution of carbohydrate-binding activity. Glycobiology 11(5), 71R-9R.

Dongré A. R., Eng J. K., Yates J. R. 3rd. (1997): Emerging tandem-mass-spectrometry techniques for the rapid identification of proteins. Trends Biotechnol, 15, 418-25.

Drickamer K. (1988): Two distinct classes of carbohydrate-recognition domains in animal lectins. J Biol Chem 263 20, 9557-60.

Drickamer K. (1999): C-type lectin-like domains. Curr Opin Struct Biol, 9, 585-90.

Esplugues E., Sancho D., Vega-Ramos J., Martínez C., Syrbe U., Hamann A., Engel P., Sánchez-Madrid F., Lauzurica P. (2003): Enhanced antitumor immunity in mice deficient in CD69. J Exp Med, 197, 1093-106.

Ewart K. V., Rubinsky B., Fletcher G. L. (1999): Structural and functional similarity between fish antifreeze proteins and calcium-dependent lectins. Biochem Biophys Res Commun, 185, 335-40.

Gerosa F., Tommasi M., Scardoni M., Accolla R. S., Pozzan T., Libonati M., Tridente G., Carra G. (1991): Structural analysis of the CD69 early activation antigen by two monoclonal antibodies directed to different epitopes. Mol Immunol, 28, 159-68.

Goldstein I.J., Hughes R.C., Monsigny M., Osawa T. And Sharon N. (1980): What should be called a lectin. Nature 285, 66.

Graham R. L. , Graham C., McMullan G. (2007): Microbial proteomics: a mass spectrometry primer for biologists. Microb Cell Fact, 6 26.

Hamann J., Fiebig H., Strauss M. (1993): Expression cloning of the early activation antigen CD69, a type II integral membrane protein with a C-type lectin domain. J Immunol, 150, 4920-7.

Hao L., Klein J., Nei M. (2006): Heterogeneous but conserved natural killer receptor gene complexes in four major orders of mammals. Proc Natl Acad Sci USA, 103, 3192-7.

Hara T., Jung L. K., Bjorndahl J. M., Fu S. M. (1986): Human T cell activation. III. Rapid induction of a phosphorylated 28 kD/32 kD disulfide-linked early activation antigen (EA 1) by 12-o-tetradecanoyl phorbol-13-acetate, mitogens, and antigens. J Exp Med, 164, 1988-2005.

Hibbs M. L., Selvaraj P., Carpén O., Springer T. A., Kuster H., Jouvin M. H., Kinet J. P. (1989): Mechanisms for regulating expression of membrane isoforms of Fc gamma RIII (CD16). Science, 246, 1608-11.

Jevremovic D., Billadeau D. D., Schoon R. A., Dick C. J., Irvin B. J., Zhang W., Samelson L. E., Abraham R. T., Leibson P. J. (1999): Cutting edge: a role for the adaptor protein LAT in human NK cell-mediated cytotoxicity. J Immunol, 162, 2453-6.

Jonscher K. R., Yates J. R. 3rd. (1997): The quadrupole ion trap mass spectrometer--a small solution to a big challenge. Anal Biochem, 244, 1-15.

Kastner M. (2000): Protein Liquid Chromatography Journal of Chromatography Library v. 61. Elsevier Ltd.

Kiessling R., Klein E., Pross H., Wigzell H. (1975): "Natural" killer cells in the mouse. II. Cytotoxic cells with specificity for mouse Moloney leukemia cells. Characteristics of the killer cell. Eur J Immunol, 5, 117-21.

Kiessling R., Klein E., Wigzell H. (1975): "Natural" killer cells in the mouse. I. Cytotoxic cells with specificity for mouse Moloney leukemia cells. Specificity and distribution according to genotype. Eur J Immunol, 5, 112-7.

Lanier L. L. (2005): NK cell recognition. Annu Rev Immunol, 23, 225-74.

Lanier L. L., Corliss B. C., Wu J., Leong C., Phillips J. H. (1998): Immunoreceptor DAP12 bearing a tyrosine-based activation motif is involved in activating NK cells. Nature, 391, 703-7.

Lazetic S., Chang C., Houchins J. P., Lanier L. L., Phillips J. H. (1996): Human natural killer cell receptors involved in MHC class I recognition are disulfide-linked heterodimers of CD94 and NKG2 subunits. J Immunol, 157, 4741-5.

Ljunggren H. G., Kärre K. (1990): In search of the 'missing self': MHC molecules and NK cell recognition. Immunol Today, 11, 237-44.

Llera A. S., Viedma F., Sánchez-Madrid F., Tormo J. (2001): Crystal structure of the C-type lectin-like domain from the human hematopoietic cell receptor CD69. J Biol Chem,276 10, 7312-9.

Long E. O. (1999): Regulation of immune responses through inhibitory receptors. Annu Rev Immunol, 17, 875-904.

Medzhitov R., Janeway C. A. Jr. (2002): Decoding the patterns of self and nonself by the innate immune system. Science, 296, 298-300.

Moretta A., Biassoni R., Bottino C., Moretta L. (2000): Surface receptors delivering opposite signals regulate the function of human NK cells. Semin Immunol, 12, 129-38.

Moretta A., Bottino C., Vitale M., Pende D., Cantoni C., Mingari M.C., Biassoni R., Moretta L. (2001): Activating receptors and coreceptors involved in human natural killer cell-mediated cytolysis. Annu Rev Immunol, 19, 197-223.

Moretta A., Poggi A., Pende D., Tripodi G., Orengo A. M., Pella N., Augugliaro R., Bottino C., Ciccone E., Moretta L. (1991): CD69-mediated pathway of lymphocyte activation: anti-CD69 monoclonal antibodies trigger the cytolytic activity of different lymphoid effector cells with the exception of cytolytic T lymphocytes expressing T cell receptor alpha/beta. J Exp Med, 174, 1393-8.

Moretta L., Bottino C., Pende D., Mingari M. C., Biassoni R., Moretta A. (2002): Human natural killer cells: their origin, receptors and function. Eur J Immunol, 32, 1205-11.

Natarajan K., Sawicki M. W., Margulies D. H., Mariuzza R. A. (2000): Crystal structure of human CD69: a C-type lectin-like activation marker of hematopoietic cells. Biochemistry, 39, 14779-86.

Nopp A., Stridh H., Grönneberg R., Lundahl J. (2002): Lower apoptosis rate and higher CD69 expression in neutrophils from atopic individuals. Inflamm Res, 51, 532-40.

Oldenborg P. A., Zheleznyak A., Fang Y. F., Lagenaur C. F., Gresham H. D., Lindberg F. P. (2000): Role of CD47 as a marker of self on red blood cells. Science, 288, 2051-4.

Pavlícek J., Sopko B., Ettrich R., Kopecký V. Jr., Baumruk V., Man P., Havlícek V., Vrbacký M., Martínková L., Kren V., Pospísil M., Bezouska K. (2003): Molecular characterization of binding of calcium and carbohydrates by an early activation antigen of lymphocytes CD69. Biochemistry, 42, 9295-306.

Perussia B., Trinchieri G., Jackson A., Warner N. L., Faust J., Rumpold H., Kraft D., Lanier L. L. (1984): The Fc receptor for IgG on human natural killer cells: phenotypic, functional, and comparative studies with monoclonal antibodies. J Immunol, 133, 180-9.

Raulet D. H. (2003): Roles of the NKG2D immunoreceptor and its ligands. Nat Rev Immunol, 3, 781-90.

Raulet D. H., Vance R. E. (2006): Self-tolerance of natural killer cells. Nat Rev Immunol, 6, 520-31.

Raulet D.H., Vance R.E., McMahon C.W. (2001): Regulation of the natural killer cell receptor repertoire. Annu Rev Immunol, 19, 291-330.

Rey M., Man P., Brandolin G., Forest E., Pelosi L. (2009): Recombinant immobilized rhizopuspepsin as a new tool for protein digestion in hydrogen/deuterium exchange mass spectrometry. Rapid Commun Mass Spectrom, 23, 3431-8.

Sancho D., Gómez M., Sánchez-Madrid F. (2005): CD69 is an immunoregulatory molecule induced following activation. Trends Immunol, 26, 136-40.

Schwarting R., Niedobytek G., Stein H. (1989): Leucocyte Typing IV. White Cell Differentiation Antigens (editors W. Knapp, B. Dörken, W.R. Gilks, E.P. Rieber, R.E. Schmidt, H. Stein, A.E.G. Kr. von dem Borne). p. 428. Oxford University Press, Oxford 1989

Senko M., Beu S., McLaffertyF. W. (1995): Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. J Am Soc Mass Spectrom, 6, 229-33.

Sheehan D. (2009): Physical Biochemistry Principles and Applications Second Edition. John Wiley & sons Ltd.

Sivori S., Falco M., Della Chiesa M., Carlomagno S., Vitale M., Moretta L., Moretta A. (2004): CpG and double-stranded RNA trigger human NK cells by Toll-like receptors: induction of cytokine release and cytotoxicity against tumors and dendritic cells. Proc Natl Acad Sci USA, 101, 10116-21.

Testi R., D'Ambrosio D., De Maria R., Santoni A. (1994): The CD69 receptor: a multipurpose cell-surface trigger for hematopoietic cells. Immunol Today, 15, 479-83.

Trinchieri G. (1989): Biology of natural killer cells. Adv Immunol, 47, 187-376.

Uhrberg M. (2005): The KIR gene family: life in the fast lane of evolution. Eur J Immunol 35 (1), 10–15.

Valiante N. M., Phillips J. H., Lanier L. L., Parham P. (1996): Killer cell inhibitory receptor recognition of human leukocyte antigen (HLA) class I blocks formation of a pp36/PLC-gamma signaling complex in human natural killer (NK) cells. J Exp Med, 184, 2243-50.

Valkenborg D., Jansen I., Burzykowski T. (2008): A model-based method for the prediction of the isotopic distribution of peptides. J Am Soc Mass Spectrom, 19, 703-12.

Vance B. A., Bennett M. J., Ward Y., Gress R. G., Kearse K. P. (1999): Distinct but dispensable N-glycosylation of human CD69 proteins. Arch Biochem Biophys, 368, 214-20.

Vance B. A., Wu W., Ribaudo R. K., Segal D. M., Kearse K. P. (1997): Multiple dimeric forms of human CD69 result from differential addition of N-glycans to typical (Asn-X-Ser/Thr) and atypical (Asn-X-cys) glycosylation motifs. J Biol Chem, 272, 23117-22.

Vély F., Vivier E. (1997): Conservation of structural features reveals the existence of a large family of inhibitory cell surface receptors and noninhibitory/activatory counterparts. J Immunol, 159, 2075-7.

Vilches C., Parham P. (2002): KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. Annu Rev Immunol 20, 217–51.

Vivier E., Biron C. A. (2002): Immunology. A pathogen receptor on natural killer cells. Science, 296, 1248-9.

Vivier E., Nunès J. A., Vély F. (2004): Natural killer cell signaling pathways. Science, 306, 1517-9.

Wagtmann N., Biassoni R., Cantoni C., Verdiani S., Malnati M. S., Vitale M., Bottino C., Moretta L., Moretta A., Long E. O. (1995): Molecular clones of the p58 NK cell receptor reveal immunoglobulin-related molecules with diversity in both the extra- and intracellular domains. Immunity, 2, 439-49.

Wange R. L. (2000): LAT, the linker for activation of T cells: a bridge between T cell-specific and general signaling pathways. Sci STKE, 2000 63, re1.

Weis W. I., Kahn R., Fourme R., Drickamer K., Hendrickson W. A. (1991): Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. Science 254 5038, 1608-15.

Yokoyama W. M., Seaman W. E. (1993): The Ly-49 and NKR-P1 gene families encoding lectin-like receptors on natural killer cells: the NK gene complex. Annu Rev Immunol, 11, 613-35.

Zhang Z., Smith D. L. (1993): Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. Protein Sci, 2, 522-31.

Appendices

List of publications

Strohalm M, Kavan D, Novák P, Volný M, Havlíček V.: mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data. Anal Chem. 2010 May 13.

Kovalová A, Ledvina M, Šaman D, Zyka D, Kubíčková M, Žídek L, Sklenář V, Pompach P, Kavan D, Bílý J, Vaněk O, Kubínková Z, Libigerová M, Ivanová L, Antolíková M, Mrázek H, Rozbeský D, Hofbauerová K, Křen V, Bezouška K.: Synthetic N-Acetyl-d-glucosamine Based Fully Branched Tetrasaccharide, a Mimetic of the Endogenous Ligand for CD69, Activates CD69(+) Killer Lymphocytes upon Dimerization via a Hydrophilic Flexible Linker. J Med Chem. 2010 Apr 30;53(10):4050-65.

Renaudet O, Krenek K, Bossu I, Dumy P, Kádek A, Adámek D, Vanek O, Kavan D, Gazák R, Sulc M, Bezouska K, Kren V.: Synthesis of multivalent glycoconjugates containing the immunoactive LELTE peptide: effect of glycosylation on cellular activation and natural killing by human peripheral blood mononuclear cells. J Am Chem Soc. 2010 May 19;, 132(19):6800-8.

Kavan D, Kubícková M, Bílý J, Vanek O, Hofbauerová K, Mrázek H, Rozbeský D, Bojarová P, Kren V, Zídek L, Sklenár V, Bezouska K.: Cooperation between subunits is essential for high-affinity binding of N-acetyl-D-hexosamines to dimeric soluble and dimeric cellular forms of human CD69. Biochemistry. 2010 May 18;49(19):4060-7

Bezouska K, Snajdrová R, Krenek K, Vancurová M, Kádek A, Adámek D, Lhoták P, Kavan D, Hofbauerová K, Man P, Bojarová P, Kren V.: Carboxylated calixarenes bind strongly to CD69 and protect CD69(+) killer cells from suicidal cell death induced by tumor cell surface ligands. Bioorg Med Chem. 2010 Feb 15;18(4):1434-40.

Man P, Montagner C, Vitrac H, Kavan D, Pichard S, Gillet D, Forest E, Forge V.: Accessibility changes within diphtheria toxin T domain when in the functional molten globule state, as determined using hydrogen/deuterium exchange measurements. FEBS J. 2010 Feb;277(3):653-62.

Pompach P, Man P, Kavan D, Hofbauerová K, Kumar V, Bezouska K, Havlícek V, Novák P.: Modified electrophoretic and digestion conditions

allow a simplified mass spectrometric evaluation of disulfide bonds. J Mass Spectrom. 2009 Nov;44(11):1571-8.

Strohalm M, Novak P, Pompach P, Man P, Kavan D, Witt M, Dzubak P, Hajduch M, Havlicek V.:Utilization of high-accuracy FTICR-MS data in protein quantitation experiments. J Mass Spectrom. 2009 Nov;44(11):1565-70.

Kuzelová K, Grebenová D, Pluskalová M, Kavan D, Halada P, Hrkal Z.: Isoform-specific cleavage of 14-3-3 proteins in apoptotic JURL-MK1 cells. J Cell Biochem. 2009 Mar 1;106(4):673-81.

Man P, Kovár V, Sterba J, Strohalm M, Kavan D, Kopácek P, Grubhoffer L, Havlícek V.: Deciphering Dorin M glycosylation by mass spectrometry. Eur J Mass Spectrom (Chichester, Eng). 2008;14(6):345-54.

Vanek O, Nálezková M, Kavan D, Borovicková I, Pompach P, Novák P, Kumar V, Vannucci L, Hudecek J, Hofbauerová K, Kopecký V Jr, Brynda J, Kolenko P, Dohnálek J, Kaderávek P, Chmelík J, Gorcík L, Zídek L, Sklenár V, Bezouska K.: Soluble recombinant CD69 receptors optimized to have an exceptional physical and chemical stability display prolonged circulation and remain intact in the blood of mice. FEBS J. 2008 Nov;275(22):5589-606.

Stodůlková E, Novák P, Deininger SO, Man P, Capková J, Kavan D, Ivasková E, Flieger M.: LC MALDI-TOF MS/MS and LC ESI FTMS analyses of HLA-B27 associated peptides isolated from peripheral blood cells. Immunol Lett. 2008 Feb 15;116(1):79-85.

Susor A, Ellederova Z, Jelinkova L, Halada P, Kavan D, Kubelka M, Kovarova H.: Proteomic analysis of porcine oocytes during in vitro maturation reveals essential role for the ubiquitin C-terminal hydrolase-L1. Reproduction. 2007 Oct;134(4):559-68.

Plíhal O, Sklenár J, Hofbauerová K, Novák P, Man P, Pompach P, Kavan D, Ryslavá H, Weignerová L, Charvátová-Pisvejcová A, Kren V, Bezouska K.: Large propeptides of fungal beta-N-acetylhexosaminidases are novel enzyme regulators that must be intracellularly processed to control activity, dimerization, and secretion into the extracellular environment. Biochemistry. 2007 Mar 13;46(10):2719-34.

Kaplan O, Vejvoda V, Plíhal O, Pompach P, Kavan D, Bojarová P, Bezouska K, Macková M, Cantarella M, Jirků V, Kren V, Martínková L.: Purification and characterization of a nitrilase from Aspergillus niger K10. Appl Microbiol Biotechnol. 2006 Dec;73(3):567-75.

Kristová V, Martínková L, Husáková L, Kuzma M, Rauvolfová J, Kavan D, Pompach P, Bezouska K, Kren V.: A chemoenzymatic route to mannosamine derivatives bearing different N-acyl groups. J Biotechnol. 2005 Jan 26;115(2):157-66.

Pavlícek J, Kavan D, Pompach P, Novák P, Luksan O, Bezouska K.: Lymphocyte activation receptors: new structural paradigms in group V of C-type animal lectins. Biochem Soc Trans. 2004 Dec;32(Pt 6):1124-6.