



HAL
open science

Structural analysis of natural killer cell receptor protein 1 (NKR-P1) extracellular domains suggests a conserved long loop region involved in ligand specificity

Žofie Sovová, Vladimír Kopecký, Tomáš Pazderka, Kateřina Hofbauerová,
Daniel Rozbeský, Ondřej Vaněk, Karel Bezouška, Rüdiger Ettrich

► To cite this version:

Žofie Sovová, Vladimír Kopecký, Tomáš Pazderka, Kateřina Hofbauerová, Daniel Rozbeský, et al.. Structural analysis of natural killer cell receptor protein 1 (NKR-P1) extracellular domains suggests a conserved long loop region involved in ligand specificity. *Journal of Molecular Modeling*, 2010, 17 (6), pp.1353-1370. 10.1007/s00894-010-0837-y . hal-00623308

HAL Id: hal-00623308

<https://hal.science/hal-00623308>

Submitted on 14 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Structural analysis of natural killer cell receptor protein 1 (NKR-P1) extracellular domains suggests a conserved long loop region involved in ligand specificity

Received: 18.07.2010 / **Accepted:** 24.08.2010

Žofie Sovová^{1,2}, Vladimír Kopecký Jr.³, Tomáš Pazderka³, Kateřina Hofbauerová^{3,4}, Daniel Rozbeský^{4,5}, Ondřej Vaněk^{4,5}, Karel Bezouška^{4,5}, Rüdiger Ettrich^{1,2,✉}

¹Laboratory of Structural Biology, Institute of Systems Biology and Ecology, Academy of Sciences of the Czech Republic, Zámek 136, 37333 Nové Hradky, Czech Republic

²Faculty of Sciences, University of South Bohemia, Zámek 136, 37333 Nové Hradky, Czech Republic

³Institute of Physics, Faculty of Mathematics and Physics, Charles University in Prague, Ke Karlovu 5, 12116 Prague 2, Czech Republic

⁴Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, 14220 Prague 4, Czech Republic

⁵Department of Biochemistry, Faculty of Science, Charles University in Prague, Albertov 2030, 12840 Prague 2, Czech Republic

✉Tel: +420-386 361 297, Fax: +420-386 361 279; E-mail: ettrich@nh.usbe.cas.cz

Abstract

Receptor proteins at the cell surface regulate the ability of natural killer cells to recognize and kill a variety of aberrant target cells. The structural features determining the function of natural killer receptor proteins 1 (NKR-P1s) are largely unknown. In the present work, refined homology models are generated for the C-type lectin-like extracellular domains of rat NKR-P1A and NKR-P1B, mouse NKR-P1A, NKR-P1C, NKR-P1F, and NKR-P1G, and human NKR-P1 receptors. Experimental data on secondary structure, tertiary interactions, and thermal transitions are acquired for four of the proteins using Raman and infrared spectroscopy. The experimental and modeling results are in agreement with respect to the overall structures of the NKR-P1 receptor domains, while suggesting functionally significant

local differences among species and isoforms. Two sequence regions that are conserved in all analyzed NKR-P1 receptors do not correspond to conserved structural elements as might be expected, but are represented by loop regions, one of which is arranged differently in the constructed models. This region displays high flexibility but is anchored by conserved sequences, suggesting that its position relative to the rest of the domain might be variable. This loop may contribute to ligand-binding specificity via a coupled conformational transition.

Keywords Molecular dynamics · Two-dimensional correlation analysis · Differential scanning calorimetry · Thermal dynamics · Raman Spectroscopy · FTIR · Topology · RMSF · Cladogram

Introduction

Natural killer (NK) cells are large granular lymphocytes able to recognize and kill a large variety of target cells and to regulate the reactions at the interface of innate and adaptive immunity through secretion of lymphokines and by direct killing [1, 2]. The cytotoxic activity of NK cells is tightly regulated *via* activating and inhibiting cell-surface receptors, one group of them being the NK cell lectin-like receptor proteins, one subtype of which are the NKR-P1s. The NKR-P1s are transmembrane glycoproteins classified as type II due to their external C-terminus, with an extracellular C-type lectin-like domain (CTLCD) and a short cytoplasmic domain. To initiate the NK cell response, activating NK cell receptors recognize a diverse range of ligands including cytokines, antibody Fc domains and other proteins and saccharides presented by target cells. Inhibitory NK cell receptors recognize MHC class I molecules and other proteins serving as a marker of cell health. Tumor, virally infected, stressed or otherwise damaged cells that can resist T cell mediated immunity because of the low levels of these markers expressed at their surface are no longer protected by their inhibitory signals and this may lead to NK cell activation and an elimination of the target cells [3]. Functions of NKR-P1s were enigmatic until their ligands were found to be closely related family of Clr lectin-like receptors [2]. Despite NKR-P1 family include both activating and inhibitory receptors, the CTLCDs of NKR-P1 family members share considerable homology; thus the structural origins of ligand-binding specificity are of interest. In the present work, multiple sequence alignment, protein homology modeling, and molecular dynamics simulations are combined with protein expression, vibrational spectroscopy, and thermal analysis to examine evolutionary and structural divergences within the family of NKR-P1 receptor CTLCDs.

Methods

Protein preparation

DNA coding for the extracellular part of mNKR-P1A/C proteins (mNKR-P1A/C; to specify the origin of the receptor, the first letter of each organism name is used prior to the receptor name, i.e., m for mouse, r for rat, etc.), was amplified from the total mRNA isolated from the spleen of C57BL/6 mouse and subcloned into expression vector pET-30a (Novagen).

Expression plasmids were transformed into *Escherichia coli* BL-21 (DE3) Gold (Stratagene). Bacteria were grown in LB medium, induction was performed with 0.1 mM isopropyl- β -D-thiogalactopyranoside and the induced culture was grown for 2 h. Proteins were refolded from inclusion bodies and were purified by HPLC ion exchange and gel filtration chromatography. The protein samples for experiments were concentrated at 10.4 mg/mL in a 15 mM Tris-HCl buffer with 150 mM NaCl, pH 8.0.

The rNKR-P1A/B proteins were prepared as described previously [4]. Proteins were dissolved in a 10 mM Tris-HCl buffer with 50 mM NaCl, pH 7.4, in concentrations of 9.1 mg/mL and 3.8 mg/mL, respectively. The concentrations were determined by Bradford assay [5].

Raman spectroscopy

Raman spectra of aqueous solutions of mNKR-P1A/C proteins were recorded in a standard 90° geometry on a multichannel instrument based on Spex 270M single spectrograph with 1800 grooves/mm grating (Jobin-Yvon), a holographic notch-plus filter (Kaiser Optical Systems) and a liquid nitrogen cooled CCD detection system (Princeton Instruments) measuring 1340 pixels along the dispersion axis. The spectral resolution was approximately 5 cm⁻¹. Samples in a capillary micro-cell (5 μ L inner volume) were excited with a 532.2 nm line (300 mW of radiant power per sample) NdYAG Verdi 2 laser (Coherent) and kept at 4 °C during all experiments using an external water bath (Neslab). The acquisition time for the spectra was 60 minutes. The spectra of temperature dependence each consisted of 5 min exposures. Temperatures in the 5–90 °C region were adjusted with 5 °C increments using the external water bath (Neslab) and equilibrated for 4 minutes before measurement. The wavenumber scale was calibrated with neon glow-lamp lines. Therefore, Raman frequencies of well-resolved bands are accurate to ± 0.5 cm⁻¹.

The solutions of rNKR-P1A/B proteins were excited with a 514.5 nm Ar-ion laser Innova 300 (Coherent) using the same Raman spectrometer with the same experimental setup. Both spectra were accumulated for 600 min to produce traces of the highest quality.

Drop coating deposition Raman spectroscopy

The samples of mNKR-P1A/C (4 μ L) were dialyzed (on Millipore filters 0.025 μ m/white VSWP/13 mm) against deionized distilled water for 35 minutes. A 2 μ L volume of protein

solution, with an approximate concentration of 1 mg/mL, was deposited on a standard DCDR substrate SpectRIM™ (Tienta Sciences) consisting of a polished stainless steel plate coated with a thin layer of Teflon [6]. After air-drying at room temperature, approximately 20 minutes, Raman spectra were collected from “coffee rings” of former droplets [7] using a Raman microspectrometer HR800 (Horiba Jobin Yvon) with a 514.5 nm Ar-ion excitation laser (Melles Griot). A 50× microscope objective (N.A. 0.75, Olympus) was used to focus the 5 mW excitation laser to a diameter, approximately 1.5 μm, on the sample, and the spectra were integrated for 20 min using a 600 grooves/mm grating and liquid nitrogen cooled CCD detector (1024 × 256 pixels, Symphony). The spectrometer was calibrated using a band of Si-vibrations at 520.7 cm⁻¹. The spectral resolution was approximately 5 cm⁻¹.

The samples of rNKR-P1A/B proteins were treated in the same way as the mouse proteins. Nevertheless, the spectra of rNKR-P1A/B protein were integrated for 4 min using a 632.8 nm He-Ne laser excitation in the same setup as mentioned previously.

Infrared spectroscopy

Infrared spectra of mNKR-P1A/C proteins were recorded with a Bruker Vector 33 FTIR spectrometer using a standard MIR source, a KBr beamsplitter and a DTGS detector. 5000 scans were collected with a Blackman-Harris 3-term apodization function at a spectral resolution of 2 cm⁻¹. Aqueous protein solutions were measured at room temperature in a CaF₂-cell with a 10μm path length. Measurements in thermal dynamics were performed using a thermal cell holder BioJACK™ (BioTools). Temperature was adjusted from 5 to 90 °C with an increment of 5 °C using an external water bath (Neslab) and equilibrated for 4 minutes before measurements took place. 1000 scans were performed. Attenuated total reflection (ATR) FTIR measurements, 5000 scans, were realized by using an ATR-MIRacl™AG – single diamond horizontal ATR (Pike Technologies).

The rNKR-P1A/B proteins were measured using a Bruker IFS 66/S FTIR spectrometer equipped with an MCT detector. 4000 scans were collected with a Happ-Genzel apodization function at a spectral resolution of 4 cm⁻¹. The rest of the FTIR setup remained the same as the measurements for mNKR-P1 proteins. Spectral contribution of the buffer was corrected following the standard algorithm [8]. Spectrum of water vapors was subtracted and finally, all spectra were normalized.

Differential scanning calorimetry

The protein's stock solutions were diluted to the desired concentration – i.e. mNKR-P1A to 0.20 mg/mL and mNKR-P1C to 0.52 mg ml⁻¹. Calorimetric measurements were performed using the Model 6100 Nano II Differential Scanning Calorimeter – N-DSC II (Microcalorimetry Sciences Corporation). The samples were scanned from 5 to 90 °C at a heating rate of 1 °C/min under a constant excess pressure of 3 atmospheres. The appropriate DSCRun and CpCalc 2.2 software was used for data acquisition and analysis. After baseline subtraction of the buffer–buffer signal, the molar excess heat capacity function was obtained by dividing the protein concentration and cell volume (0.299 ml).

Sequence and phylogeny analysis

Sequences specified in Table 1 were aligned in ClustalX [21] and used for phylogeny analysis. For sequence analysis purposes, we used only their C-type lectin-like domains (CTLDs). Phylogenetic analysis was originally performed for protein sequences only; resultant trees however revealed too many polytomies. The number of polytomies was decreased by using nucleotide sequences for tree construction. Excluding Bayesian analyses, all protein and DNA sequences analyses were performed using Phylip 3.69 software [22] with the following algorithms: Neighbor-Joining [23], maximum parsimony [24], maximum likelihood [25] with assuming molecular clock and Fitch-Margoliash method assuming molecular clock (Kitsch method) [26]. Sequences were bootstrapped 1000× with the exception of the computationally most expensive method maximum likelihood, where sequences were bootstrapped 100×. To confirm that this value of bootstrapping was high enough, we constructed the trees with the same settings but bootstrapped 1100× or 110×. To calculate the distance matrix, and for the maximum likelihood method, we used the Jones-Taylor-Thornton matrix for protein sequences and the F84 matrix for DNA sequences. Where possible the sequences were jumbled 5×. In the Neighbor-Joining method the order of sequences was randomized. Probabilities of branch occurrence were calculated according to one of the most commonly used tests on the reliability of an inferred tree, Felsenstein's bootstrap test [27], which was evaluated using Efron's [28] bootstrap resampling technique. Resultant consensus cladograms were done using the 50% majority consensus rule. Bayesian analysis was performed using MrBayes 3.1.2 [29] with the same initial sequences used previously. Invariable gamma distribution and a GTR model was used to describe the

Comment [EH1]:
Comment for typesetter:
Ref 9-20 are cited in Table 1

parameters for the likelihood model, 150 000 for protein cycles and 100 000 for DNA Markov chain Monte Carlo cycles, with the state being swapped every 100×.

Homology modeling

Primary structures of mouse NKR-P1A, mNKR-P1C, rNKR-P1A, and rNKR-P1B extracellular domains were extracted from the database (Table 1). Templates were identified using BLAST [30] with matrix BLOSUM62. The identified homologs were almost the same for all seven models that were built. The template with highest identity to the target sequence was used. If some structures had the same identity, the one with the better resolution was used. Templates and identities used were: 1YPQ (C-type lectin-like domain of human oxidized low density lipoprotein receptor 1 (LOX-1) [31]) for mNKR-P1A/C (identity 31 %/32 %), 1XPH (CD209 antigen-like protein 1 [32]) for rNKR-P1B (identity 33 %), 3HUP (early activation antigen CD69 [33]) for mNKR-P1F (identity 33 %), 1E87 (early activation antigen CD69 [34]) for mNKR-P1G (identity 35 %) and 2BPD (beta-glucan receptor Dectin-1 [35]) for hNKR-P1 (identity 32 %). Only in the case of rNKR-P1A was an alternative technique tried. The neighboring sequence from the Neighbor-Joining phylogenetic tree (1000× bootstrapped) was used, i.e., structure 3CAD (lectin-related NK cell receptor LY49G1 [36]). Alignment of template and modeled sequences used ClustalX. Alignments were checked to ensure that general features in sequence-conserved regions of CTLDs, were maintained as found in sequence multiple alignments of CTLDs. The resulting aligned sequences used for modeling are shown in Fig. 8. Ten models were calculated for every protein and template using Modeller 9v4 [37]. The best model from every group was chosen by the distribution of amino acids in the Ramachandran plot and stereochemical *g*-factor (both calculated by Procheck [38]), Modeller distribution function and by visual inspection using SwissPDBViewer [39]. Secondary structure was determined by Procheck according to Kabsch and Sander [40].

Molecular dynamics

All selected models were minimized in SCP water solution with Gromacs 3.3.3 package [41], using virtual site hydrogens. The modified version of force field Gromos87 (usually called Gromacs) [42, 43] was used. Temperature was held at 300 K by separately connecting the protein and solution to the external temperature bath ($t = 0.1$ ps) while the pressure was held

at 1 bar by connecting to the pressure bath ($t = 0.1$ ps). Algorithms SETTLE (for water) and LINCS (for protein) were used to restrict covalent bond length and long-range electrostatic interactions were calculated using the Particle-Mesh Ewald method. Optimization with steepest descent energy minimization was followed by solvent optimization using a time step of 1 fs for 10 ps. Counter ions were added to neutralize the simulation box and were consequently minimized for 10 ps using a time step of 1 fs. Finally, the protein was minimized for 20 ps using a time step of 2 fs. For production, the simulation runs were started using a time step of 5 fs. Root mean square deviation and radius of gyration analysis was performed every 10 ns to check if the system reached equilibrium. The total production run time was 110 ns for mNKR-P1A, 10 ns for rNKR-P1A, 20 ns for rNKR-P1B and 100 ns for mNKR-P1C. The human and mouse models NKR-P1F/G were equilibrated for 10 ns.

After the dynamics run, all structures were minimized using the steepest descent algorithm. The representation of secondary structure types according to Kabsch and Sander [40] was calculated by Procheck software, and the root mean square fluctuation for each residue was calculated in Gromacs.

Results and discussion

Structural analysis of NKR-P1 proteins

Four NKR-P1 extracellular domains, rNKR-P1A and B and mNKR-P1A and C, were overexpressed in *E. coli*, refolded from inclusion bodies, and purified for spectroscopic analysis as described in Methods. Receptor domains were structurally analyzed by Raman, FTIR, and ATR-FTIR spectroscopy, including a novel drop-coating deposition Raman method that yields native-state spectra from very small solid samples [7]. Spectral assignments are given in Table 2, and secondary structure contents estimated using the pattern recognition least-squares method (LSA) are given in Table 3.

Fig. 1 compares the FTIR spectra of mNKR-P1A and C domains. The similarity of the spectra is immediately obvious, indicating that mouse A and C proteins share the same fold with very similar secondary structure content. The slight differences in the spectra can be explained by

their different amino acid compositions. For example, the second derivative band at 1517 cm^{-1} is connected with Tyr ring vibrations [54], and is more intense in mNKR-P1A that has five Tyr residues than in mNKR-P1C with only one Tyr. ATR-FTIR spectra of proteins were also measured for better resolution of subtle differences, and no further differences were detected (data not shown). The FTIR results are in excellent agreement with Raman spectroscopy data (Fig. 2 and Table 3). Solution Raman spectra (not shown) are also highly similar to drop-coated deposition spectra, which provide better signal-to-noise ratio.

The secondary structure estimates using the LSA method, which analyzes the amide I band in the case of Raman spectroscopy [49], and amide I and II bands in FTIR spectra [50], are not perfect for mNKR-P1A and C domains, suggesting a protein fold whose spectral pattern is not included in the LSA reference set [51]. Thus, the absolute percentages in the secondary structure content probably contain larger errors than estimated in Table 3. The fits for Raman spectral data are better. Nevertheless, both FTIR and Raman spectroscopy predictions differ by no more than 4 % in each method for mouse A and C variants, and for each variant some predictions are in agreement within the margin of error. These results thus suggest only a small difference in the content of α -helices and β strands of about 10 % for the two mouse domains.

Negative bands are observed in the Raman difference spectrum between mouse variants A and C in the region reporting on S-S-bridge conformations (Fig. 2). The presence of these intense and narrow vibration bands reflects a more well-defined conformation of S-S bridges in mNKR-P1C and thus lower flexibility. Thus, the C variant is probably more rigid than the A variant, in agreement with the higher stability detected in calorimetric data as discussed below.

More distinctive differences can be seen between the FTIR spectra of rNKR-P1A and B domains (Fig. 3). The band positions of the rat A variant differ only slightly more than the differences between the two mouse variants, and therefore the rat A variant can be structurally clustered with both mouse variants; rNKR-P1B is the most distinctive protein within the measured group. Raman difference spectroscopy (Fig. 4) adds detail to these distinctions, with significant differences between A and B variants in the amide I region, ca. 1671 cm^{-1} and amide III, ca. 1237 cm^{-1} . In the amide III region, the positive band, ca. 1230 cm^{-1} , corresponds to an excess of β -structures in rNKR-P1A relative to the B variant. The negative band at 1280 cm^{-1} corresponds to lower α -helix content in the A variant when compared with the B

Comment [EH2]:

Comment for typesetter:

Ref 44-47 are cited in Table 2

Ref 48 is cited in figure caption of Fig.2

Ref 49-51 are cited in Table 3

Ref 52-53 is cited in figure caption of Fig.1

variant. The aromatic amino acid composition of rNKR-P1A/B domains are the same; thus the Raman difference spectrum in Fig. 4 reflects mostly secondary structure differences between the proteins. Comparing all Raman spectra, rNKR-P1B protein is again the most distinctive, with mouse A and C variants very similar, and rat variant A being somewhere in between.

Thermal dynamics

Differential scanning calorimetry showed an irreversible denaturation transition for mNKR-P1A and C domains corresponding to a single, noncooperative thermal transition, with melting temperatures of 69 °C for mNKR-P1A and 72 °C for the C variant (data not shown). Thermodynamic parameters for the transitions, determined as described in Methods, were the same for both proteins, $\Delta H = 48 \text{ kcal mol}^{-1}$, $\Delta S = 0.14 \text{ kcal}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$. Therefore the mNKR-P1C domain is slightly more thermostable than the A variant.

Two-dimensional correlation spectroscopy (2DCoS) was applied to FTIR and Raman spectra to add structural detail to the thermal unfolding transitions of mNKR-P1A/C proteins. The aim of this technique is to identify in-phase and out-of-phase correlations between spectral intensity variations occurring at different wavenumbers that are induced by external perturbation of the studied system; a generalized formalism for 2DCoS can be found in ref. 57–59. The main advantage of 2DCoS is that it allows enhancement of spectral resolution by spreading overlapping bands over a second dimension. In addition, sign analysis of the correlation peaks in the 2D maps may improve band assignment and permit establishment of a sequence of events during the perturbation process [57–59].

Synchronous 2D correlation FTIR spectra (Fig. 5 left) suggest slightly lower stability in the A variant relative to C, in agreement with calorimetric data, and suggest this transition is correlated with exposure of hydrophobic regions in β -strands. The Raman synchronous spectrum is significantly less complicated (Fig. 6 left) and reports on the same denaturation process, consistent with the FTIR results. The asynchronous 2DCoS spectra (Figs. 5 and 6 right) reveal sequential, but not coincidental, spectral changes. The asynchronous spectrum has no autopeaks and consists exclusively of crosspeaks that are antisymmetric in regard to the diagonal line. Asynchronous cross peaks develop only if the intensities of two spectral features change and are out of phase with each other. If the asynchronous peak sign becomes

Comment [EH3]:
Comment for typesetter:
Ref 55-56 are cited in figure caption of Fig.2

positive then the intensity change at the given wavenumber on the x -scale occurs predominantly before it is connected with the 2DCoS y -scale and vice versa. This sign rule is reversed if the synchronous correlation intensity at the same coordinate becomes negative.

Application of these so called Noda rules [57–59] to mNKR-P1C second derivative FTIR spectra indicate the following sequential order of structural changes. Changes in β -turns occur first, followed by an increase in β -sheet content, then changes in α -helices followed by changes in loops. These structural changes precede emergence of β -aggregated structures, which is followed by changes in β -turns. The sequential order in the mNKR-P1A variant is the same, although the emergence of β -aggregates was not observable. The sequential order of secondary structure changes from FTIR is in agreement with that from Raman data. Asynchronous analysis of Raman data shows that changes in Tyr/Phe vibration at 1600 cm^{-1} precede changes in β -turns, and the increase of β -sheet content is followed by changes in Trp/Phe vibrations at 1590 cm^{-1} .

Heterospectral 2DCoS was used because it can reveal correlations between different spectral regions or even between two different spectroscopic techniques and may help to detect vibrations of a similar nature or connections with the same process [59]. Only synchronous heterospectral 2DCoS Raman spectroscopy was used to permit direct correlation of secondary structure elements with specific residue types, and because Raman spectra contain more residue-specific spectral information than FTIR spectra. The synchronous spectra display only cross peaks (Fig. 7) because autopeaks are generated only by the correlation of the same bands as in Figs. 5 and 6. Correlations were investigated in the changes of the secondary structure, represented by the Raman amide I region ($1580\text{--}1700\text{ cm}^{-1}$), and the regions where aromatic side chains have intense Raman bands. The strong correlations reported in Fig. 7 suggest that one or more Trp residues experience major changes in their surroundings during the early stages of thermal denaturation.

The structural changes occurring in NKR-P1 proteins during the increase of temperature can be interpreted tentatively as follows. Some large flexible part with high content of β -turns is rearranged first. Trp residues appear to be involved in significant dynamical behavior during these early stages of denaturation. Next, the β -sheet content increases before or during continuous decrease of α -helix content. These changes lead to the exposure of a hydrophobic

region, perhaps containing Tyr and/or Phe residues, leading to β -aggregates followed by changes in β -turns.

Modeling

Homology modeling and molecular dynamics

Candidate template sequences were identified using BLAST [30] and pair-aligned with modeling target sequences using ClustalX as described in Methods. The template structure with highest sequence identity to each target was used for modeling; identities ranged from 31 to 35%, and the resulting aligned sequences used for modeling are shown in Fig. 8. Template structures used were: 1YPQ (C-type lectin-like domain of human oxidized low density lipoprotein receptor 1 (LOX-1) [31]) for mNKR-P1A/C, 1XPH (CD209 antigen-like protein 1 [32]) for rNKR-P1A/B, 3HUP (early activation antigen CD69 [33]) for mNKR-P1F, 1E87 (early activation antigen CD69 [34]) for mNKR-P1G, and 2BPD (beta-glucan receptor Dectin-1 [35]) for hNKR-P1. These template structures are similar to each other as judged by $C\alpha$ root-mean-square deviations ranging from 0.94 to 1.50 among them, and all have functions close to those of the NKR-P1 proteins [61]. C-type lectin-like NK receptors are dimeric in their nature, but as the mode of dimerization is not conserved within this receptor family, we did not attempt to model NKR-P1s as dimers. Moreover, in native receptors covalent linkage by one or more disulfide bridges in the so-called "stalk region" close to the cell membrane is involved in dimerization. This region is usually omitted from soluble recombinant receptor domain constructs and thus is also missing in available crystal structures. Finally, all recombinant NKR-P1s prepared in this study behaved like a monomeric species during their purification (data not shown), pointing out for rather weak dimer formation within the NKR-P1 family, however, we cannot exclude that cooperativity between monomers will occur upon ligand binding. For consistency, the numbering of sequences begins at residue 89 in all sequences, and specific residue numbers discussed in the text are those of the mNKR-P1A sequence. Ten models were calculated for every sequence-template pair using Modeller 9v4 [37]. For further analysis the best model from every group was chosen as described in Methods, and refined by at least 10 ns of molecular dynamic simulations at room temperature in explicit solvent.

Comment [EH4]:
Comment for typesetter:
Ref 60 is cited in figure caption of Fig.7

All models share the same basic α/β fold represented in Fig. 9 by rNKR-P1B. Helices surround a beta core composed of long antiparallel β -strands 2 and 8 that form a central 'pillar' flanked at one end by short strands 1 and 3 and at the other end by a small antiparallel sheet formed by strands 4-7. Although β -strand 3 is very short, its residues are highly conserved and they adopt very similar positions in all models (Table 4). The lengths of secondary structure elements (Table 5) reveal that a conserved number of residues forms strands 2 and 8 of the central pillar except in mNKR-P1F that is 1-2 residues shorter. These strands define the height of the receptor on the outside surface of the cell membrane. The models contain the three disulfide bonds identified in experimental data [62], except for mNKR-P1C where Cys122 is replaced by Ser.

Although the core of the protein is nearly the same for all seven models ($C\alpha$ rmsd $< 2 \text{ \AA}$ for the beta core defined above), the number and orientation of helices as well as the topological organization of secondary elements vary somewhat among the models, defining four groups (Fig. 10). Mouse NKR-P1A/C models are more similar to each other than rat models are to each other. In addition to the topology differences, the prominent loops anchored by small antiparallel sheets differ in their arrangement among models, making rmsd comparisons meaningless for these segments. In all cases one small sheet (residues Ala127, Tyr128 and Leu129 in mNKR-P1A) anchors one loop of ~ 5 to 7 residues (residues Met130 to Gln135 in mNKR-P1A, green loop in Fig. 10, first half of region III in Fig. 12) and another small sheet (residues Trp165, Lys166 and Trp167 in mNKR-P1A) anchors one long loop of ~ 19 residues (e.g., residues Arg168 to Asp187 in mNKR-P1A, blue loop between sheets 5 and 6 in Fig. 10, region V until beginning of region IVb in Fig. 12). In rNKR-P1B and hNKR-P1 the long loop contains an additional short helix that packs on one side of the core together with helix B. In rNKR-P1B the connectivity of the loops to the small sheet also differs.

These differences in the spatial arrangement of the loops among the models are reflected also in the behavior of each model in molecular dynamics simulations. In all simulations the protein core is stable, with $C\alpha$ rmsds reaching plateau values of ~ 1.3 to 2.5 \AA (data not shown). The root mean square fluctuations for $C\alpha$ calculated from the last nanosecond of the equilibrated part of each trajectory clearly identify the long loop region as very flexible (Fig. 11). Both ends of the loop, immediately after the N- and C-terminal anchoring residues, are extremely flexible. The root mean square fluctuations of the other residues in the loop are slightly higher than the average for the whole protein, although the structure within the loop is stable in the time period that could be simulated. The N-terminal anchoring residues are the

highly conserved WKW sequence motif (residues Trp165, Lys166 and Trp167 in mNKR-P1A). The loop occasionally folds back onto the protein surface, where it forms a hydrophobic interaction with the two tryptophan residues. Together these results suggest that the loop region is stably anchored to the core, but can adopt alternative positions relative to the core.

Sequence analysis

The 33 available NKR-P1 CTLD domain sequences (Table 1) were aligned in ClustalX [21]. Seven conserved regions are identifiable (Fig. 12). Among these regions, conservation among orthologs is highest in the beta core and lowest in the two loops, the smaller loop corresponding to region III and the large loop corresponding to region V. In region III the chemical properties of loop residues are preserved despite the sequence variations. In contrast, substitutions in extended region V (L160 to T183) cause significant changes in the chemical character of some loop residues among natural NKR-P1 variants, leading to two groups typified by NKR-P1 subfamilies A-D and F/G. The NKR-P1A/D group presents L160, N164, T171, K178 and T183, whereas the NKR-P1F/G group substitutes a polar residue (Glu or Gln) for Leu160, a beta-branched residue (Thr, Val, or Ile) for Asn164, a more hydrophobic residue (Ile or Val) for Thr171, a Ser or Lys residue for Arg178, and a more polar residue (Asn, Asp, or Glu) for Thr183. The results of *in vivo* binding experiments with C-type lectin receptor (Clr) isoforms [11] suggest that these sequence differences may be related to ligand specificity: rNKRPIA and rNKRPIB bind to Clr11 only; rNKRPIF and rNKRPIG bind to Clr2, Clr6, and Clr7 but not to Clr11, and rNKRPIF also binds to Clr3 and Clr4. Thus, the long adaptable loop of the C-type lectin-like domain may encode its ligand specificity.

Phylogenetic analysis

Phylogenetic analysis was originally performed with only the protein sequences of the C-type lectin-like domains, starting at position 89 for mouse NKR-P1A. Although five methods gave similar results, cladograms contained a large number of polytomies. DNA nucleotide sequences were then applied to bring in additional information for tree construction. Cladograms based on DNA sequences were constructed with several different approaches as described in Methods, leading to nearly identical results. In every case, the major branches of the cladogram were the same, as confirmed by bootstrapping. Bootstrapping failed only when

the arrangement of species within the clade was different. This robustness of the final result supports the principal correctness of the tree. The resultant cladogram obtained by the maximum-likelihood method is shown in Fig. 13. Where some methods suggest polytomy or further branching, both possibilities are drawn.

The cladogram displays three major branches. The first is composed of human and chicken NKR-P1s. The second main branch is composed of rodent NKR-P1s from the telomeric part of the NK gene complex. In this branch the sequences are first branched according to the receptor subfamily and then according to the species. The third major group consists of rodent receptors from the centromeric part of the NK gene complex. Here, the sequences are ordered first according to species and then according to receptor subfamily. The most undefined part of the cladogram includes the rat NKR-P1As, as some methods even create pentatomy.

Comparison of modeling and spectroscopic results

The extent of predicted α -helical structure in rNKR-P1B agrees closely with experiments, especially with the data from FTIR. In the case of mNKR-P1A/C, more α -helical structure is detected in the spectra than is present in the models. One likely explanation that would be in quantitative agreement with the data is that short α -helix C, which is present in human and rat NKR-P1B models, is also present in both mouse structures. In the case of β -strands, the algorithm used to determine hydrogen bonds in models uses an average cutoff value, whereas in nature this cutoff is not so sharp, and it also depends on the β -strand environment and other factors. The agreement between the spectroscopic data and the models is particularly strong for beta turns, especially when measured by FTIR, whereas Raman spectra give slightly higher turns content. The phylogeny depicted in Fig. 13 is also consistent with the vibrational spectra, which group mouse A/C variants together as most distinct from rNKR-P1B, and with rat variant A being somewhere in between.

Disulfide bond conformations for Cys122-210 and Cys94-105 are unambiguously in a GGG conformation in all computational models. Therefore these two bridges can be assigned to the two clear GGG conformations determined from the band at 509 cm^{-1} (Fig. 2) representing the disulfide bridges region of stretching vibrations. The third disulfide bridge between Cys189-202 in the models adopts conformations in between the more rare GGT and TGT conformations, and could be assigned to the TGT conformation observed experimentally (Fig. 2).

The inferred conformational adaptability of the long loop region finds support in the spectroscopic data as well. Opening of the loop as suggested by its behavior in molecular dynamics simulations would expose the two anchoring tryptophan residues to solvent, in agreement with the early shifts observed for tryptophan residues in the experimental thermal dynamics data.

Conclusions

MD-refined homology models of rat NKR-P1A/B, mouse NKR-P1A/C/F/G and human NKR-P1 receptor C-type lectin-like extracellular domains were generated in order to understand evolution of the structural features determining their functions. The rat NKR-P1A/B and mouse NKR-P1A/C domains were structurally analyzed by Raman and FTIR spectroscopy. Combined protein and DNA sequence analysis divided NKR-P1s into three major phylogenetic groups.

The derived models agree well with the Raman and infrared spectra not only in terms of overall secondary structure content, but additionally the models capture key structural and dynamic features consistent with thermal dynamics and MD simulations. Two sequence regions that are conserved in all analyzed NKR-P1s, and thus under strong evolutionary pressure, define the major phylogenetic branches, including the distinction between the NKR-P1A/D and the NKR-P1F/G subfamilies. One of these regions is the long loop present in nearly every CTLD; it is alternatively arranged in the different models, and is predicted to contain only a minor portion of secondary structure that is variably present. The experimental data from vibrational thermal analysis would be consistent with the facile dynamics of this region observed in MD simulations, in which an underlying hydrophobic surface becomes exposed. Sequence and phylogenetic analysis suggests that this loop region evolved in concert with target specificity, adapting the chemical properties of the residues to specific target ligands while embedding them in a structurally stable framework. The results thus suggest that the long loop plays a key role in ligand specificity, and that a conformational equilibrium is involved in ligand binding.

Homology modeling is based on the presumption that proteins with similar primary structures (> ~30% identity) have the same tertiary structure. Even though exceptions are known [e.g.

63, 64], the use of templates with functions close to those of the modeled targets in the present case suggested that this basic assumption should be correct, and that the core folds are likely to be similar. Local deviations can occur, and these may be detected by combining vibrational spectroscopy with computational modeling [65], as shown by the results presented here.

The resulting molecular models of NKR-P1 are available upon request.

Acknowledgments

The Grant Agency of the Academy of Sciences of the Czech Republic and the Ministry of Education of the Czech Republic are gratefully acknowledged for support (LC 06010, No. KJB101120805, Nos. MSM6007665808, MSM 0021620835, MSM 0021620808, 1M0505, AVOZ60870520, AV0Z50200510 respectively). Additionally, Z.S. was supported by the University of South Bohemia, grant GAJU 170/2010/P. Access to the METACentrum computing facilities provided under the research intent MSM6383917201 is highly appreciated.

References

1. Lanier LL (1998) NK Cell receptors. *Annu Rev Immunol* 16:359–393
2. Vivier E, Tomasello E, Baratin M, Walzer T, Ugolini S (2008) Function of natural killer cells. *Nat Immunol* 9:503–510
3. Ljunggren HG, Karre K (1990) In search of 'missing self': MHC molecules and NK cell recognition. *Immunol Today* 11:237–244
4. Plíhal O., Byrtusová P., Pavlíček J., Mihók L., Etrich R., Man P., Pompach P., Havlíček V., Hušáková L., Bezouška K. (2004) The isoforms of rat natural killer cell receptor NKR-P1 display a distinct binding of complex saccharide ligands. *Collect Czech Chem Commun* 69:631–644
5. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248–254
6. Zhang D, Xie Y, Mrozek MF, Ortiz C, Davisson VJ, Ben-Amotz D (2003) Raman detection of proteomic analytes. *Anal Chem* 75:5703–5709
7. Kopecký V Jr, Baumruk V (2006) Structure of the ring in drop coating deposited proteins and its implication for Raman spectroscopy of biomolecules. *Vibr Spectrosc* 42:184–187
8. Dousseau F, Therrien M, Pézolet M (1989) On the spectral subtraction of water from FT-IR spectra of aqueous solutions of proteins. *Appl Spectrosc* 43:538–542
9. Voigt S, Mesci A, Ettinger J, Fine JH, Chen P, Chou W, Carlyle JR (2007) Cytomegalovirus Evasion of Innate Immunity by Subversion of the NKR-P1B:Clr-b Missing-Self Axis. *Immunity* 26:617–627
10. Giorda R, Rudert WA, Vavassori C, Chambers WH, Hiserodt JC, Trucco M (1990) NKR-P1, a signal transduction molecule on natural killer cells. *Science* 249(4974):1298–1300
11. Kveberg L, Dai KZ, Westgaard IH, Daws MR, Fossum S, Naper C, Vaage JT (2009) Two major groups of rat NKR-P1 receptors can be distinguished based on chromosomal localization, phylogenetic analysis and Clr ligand binding. *Eur J Immunol* 39: 541–551
12. Dissen E, Ryan JC, Seaman WE, Fossum S (1996) An autosomal dominant locus, Nka, mapping to the Ly-49 region of a rat natural killer (NK) gene complex, controls NK cell lysis of allogeneic lymphocytes. *J Exp Med* 183:2197–2207

13. Appasamy PM, Kenniston TW, Brissette-Storkus CS, Chambers WH (1996) NKR-P1dim/TCR alpha beta + T cells and natural killer cells share expression of NKR-P1A and NKR-P1D. *Nat Immunol* 15(5):259–268
14. Giorda R, Trucco M (1991) Mouse NKR-P1. A family of genes selectively coexpressed in adherent lymphokine-activated killer cells. *J Immunol* 147(5):1701–1708
15. Carlyle JR, Mesci A, Ljutic B, Belanger S, Tai LH, Rousselle E, Troke AD, Proteau MF, Makrigiannis AP (2006) Molecular and genetic basis for strain-dependent NK1.1 alloreactivity of mouse NK cells. *J Immunol* 176(12):7511–7524
16. Kung SK, Su RC, Shannon J, Miller RG (1999) The NKR-P1B gene product is an inhibitory receptor on SJL/J NK cells. *J Immunol* 162(10):5876–5887
17. Carninci P, Hayashizaki Y (1999) High-efficiency full-length cDNA cloning. *Methods Enzymol* 303:19–44
18. Plougastel B, Matsumoto K, Dubbelde C, Yokoyama WM (2001) Analysis of a 1-Mb BAC contig overlapping the mouse Nkrp1 cluster of genes: cloning of three new Nkrp1 members, Nkrp1d, Nkrp1e, and Nkrp1f. *Immunogenetics* 53(7):592–598
19. Lanier LL, Chang C, Phillips JH (1994) Human NKR-P1A. A disulfide-linked homodimer of the C-type lectin superfamily expressed by a subset of NK and T lymphocytes. *J Immunol* 153(6):2417–2428
20. Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401(6756):923–925
21. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948
22. Felsenstein J (1989) PHYLIP – phylogeny inference package (version 3.2). *Cladistics* 5:164–166
23. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstruction phylogenetics trees. *Mol Biol Evol* 4:406–425
24. Eck RV, Dayhoff MO (1966) Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, MD
25. Felsenstein J (1973) Maximum-likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst Zool* 22:240–249
26. Fitch WM, Margoliash E (1967) Construction of phylogenetics trees. *Science* 155:279–284

27. Felsenstein J. (1985) Confidence-limits on phylogenies – an approach using bootstrap. *Evolution* 39(4):783–791
28. Efron B (1982) Maximum-likelihood and decision-theory. *Ann Stat* 10(2):340–356
29. Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574
30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
31. Park H, Adsit FG, Boyington JC (2005) The 1.4 Å crystal structure of the human oxidized low density lipoprotein receptor lox-1. *J Biol Chem* 280:13593–13599
32. Snyder GA, Colonna M, Sun PD (2005) The structure of DC-SIGNR with a portion of its repeat domain lends insights to modeling of the receptor tetramer. *J Biol Chem* 347:979–989
33. Kolenko P, Skalova T, Vanek O, Stepankova A, Duskova J, Hasek J, Bezouska K, Dohnalek J (2009) The high-resolution structure of the extracellular domain of human CD69 using a novel polymer. *Acta Crystallogr F* 65: 1258–1260
34. 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(10):7312–7319
35. Brown J, O’Callaghan CA, Marshall ASJ, Gilbert RJC, Siebold C, Gordon S, Brown GD, Jones EY (2007) Structure of the fungal beta-glucan-binding immune receptor dectin-1: Implications for function. *Protein Sci* 16: 1042–1052
36. Deng L, Cho S, Malchiodi EL, Kerzic MC, Dam J., Mariuzza RA (2008) Molecular architecture of the major histocompatibility complex class I-binding site of Ly49 natural killer cell receptors. *J Biol Chem* 283:16840–16849
37. Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A (2006) Comparative protein structure modeling with MODELLER. *Current Protocols in Bioinformatics*, Wiley, Supplement 15, 5.6.1–5.6.30
38. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283–291
39. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18:2714–2723
40. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* 22:2577–2637

41. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26 (16):1701–1718
42. van Buuren AR, Marrink SJ, Berendsen HJC (1993) A molecular dynamics study of the decane/water interface. *J Phys Chem* 97:9206–9212
43. Mark AE, van Helden SP, Smith PE, Janssen LHM, van Gusteren WF (1994) Convergence properties of free energy calculations: α -cyclodextrin complexes as a case study. *J Am Chem Soc* 116:6293–6302
44. Tensmeyer LG, Kaufman EW (1996) Protein structure as revealed by nonresonance Raman spectroscopy. In: Havel HA (ed) *Spectroscopic methods for determining protein structure in solutions*. VCH Publishers Inc, New York, pp 69–93
45. Overman SA, Thomas GJ Jr (1999) Raman markers of nonaromatic side chains in an α -helix assembly: Ala, Asp, Glu, Gly, Ile, Leu, Lys, Ser, and Val residues of phage *fd* subunits. *Biochemistry* 38:4018–4027
46. Tuma R (2005) Raman spectroscopy of proteins: from peptides to large assemblies. *J Raman Spectrosc* 36:307–319
47. Miura T, Thomas GJ Jr (1995) Raman spectroscopy of proteins and their assemblies In: Biswas BB, Roy S (eds) *Subcellular Biochemistry*, vol. 24, *Proteins: Structure, Function, and Engineering*. Plenum Press, New York, pp 55–99
48. Miura T, Takeuchi H, Harada I (1989) Tryptophan Raman bands sensitive to hydrogen bonding and side-chain conformation. *J Raman Spectrosc* 20:667–671
49. Williams RW (1986) The secondary structure analysis using Raman amide I and amide III spectra. *Method Enzymol* 130:311–331
50. Dousseau F, Pézolet M (1990) Determination of the secondary structure content of proteins in aqueous solutions from their amide I and amide II infrared bands. Comparison between classical and partial least-squares methods. *Biochemistry* 29:8771–8779
51. Baumruk V, Pancoska P, Keiderling TA (1996) Predictions of secondary structure using statistical analyses of electronic and vibrational circular dichroism and Fourier transform infrared spectra of proteins in H₂O. *J Mol Biol* 259:774–791
52. Fabian H, Mäntele W (2002) Infrared spectroscopy of proteins. In: Chalmers JM, Griffiths PR (eds) *Handbook of vibrational spectroscopy*, vol. 5. Wiley, Chichester, pp 3399–3425
53. Barth A, Zscherp C (2002) What vibrations tell us about proteins. *Q Rev Biophys* 35:369–430

54. Barth A (2000) The infrared absorption of amino acid side chains. *Prog Biophys Mol Biol* 73:1–49
55. Miura T, Takeuchi H, Harada I (1988) Characterization of individual tryptophan side chains in proteins using Raman spectroscopy and hydrogen–deuterium exchange kinetics. *Biochemistry* 27:88–94
56. Siamwiza MN, Lord RC, Chen MC (1975) Interpretation of the doublet at 850 and 830 cm^{-1} in the Raman spectra of tyrosyl residues in proteins and certain model compounds. *Biochemistry* 14:4870–4876
57. Noda I (1990) Two-dimensional infrared (2D IR) spectroscopy: Theory and applications. *Appl Spectrosc* 44:550–561
58. Noda I (1993) Generalized two-dimensional correlation method applicable to infrared, Raman, and other types of spectroscopy. *Appl Spectrosc* 47:1329–1336
59. Noda I, Ozaki Y (2004) Two-dimensional correlation spectroscopy: Applications in vibrational and optical spectroscopy. Wiley, Chichester
60. Miura T, Takeuchi H, Harada I (1991) Raman spectroscopic characterization of tryptophan side chains in lysozyme bound to inhibitors: Role of the hydrophobic box in the enzymatic function. *Biochemistry* 30:6074–6080
61. Weis WI, Taylor ME, Drickamer K (1998) The C-type lectin like superfamily in the immune system. *Immunol Rev* 163:19–34
62. Pompach P, Man P, Kavan D et al (2009) Modified electrophoretic and digestion conditions allow a simplified mass spectrometric evaluation of disulfide bonds. *Journal of Mass Spectrometry* 44: 1571–1578
63. Kabsch W, Sander C (1985) Identical pentapeptides with different backbones. *Nature* 317:19–25
64. Sternberg MJE, Islam SA (1990) Local protein sequence similarity does not imply a structural relationship. *Protein Eng* 4:125–131
65. Kopecký V Jr, Etrich R, Hofbauerová K, Baumruk V (2004) Vibrational spectroscopy and computer modeling of proteins: solving structure of α_1 -acid glycoprotein. *Spectrosc Int J* 18:323–330

Tables

Table 1 Protein and nucleotide sequences used for phylogenetic and sequence analyses

protein	protein sequence	DNA sequence	Reference
rNKR-P1A WAG	ABO15817	EF100677	Voigt et al., 2007 [9]
rNKR-P1A SD	ABO15823	EF100683	Voigt et al., 2007 [9]
rNKR-P1A RNK	ABO15821	EF100681	Voigt et al., 2007 [9]
rNKR-P1A F344	AAA41710	M62891	Giorda et al, 1990 [10]
rNKR-P1A PVG	ACJ47804	FJ416340	Kveberg et al., 2009 [11]
rNKR-P1A WAG	ABO15818	EF100678	Voigt et al., 2007 [9]
rNKR-P1B BS	AAQ08908	AF525533	Voigt et al., 2007 [9]
rNKR-P1B PVG	ABA40404	NM_001040189	Voigt et al., 2007 [9]
rNKR-P1B SD	ABO15824	EF100684	Voigt et al., 2007 [9]
rNKR-P1B RNK	ABO15822	EF100682	Voigt et al., 2007 [9]
rNKR-P1B F344	AAB01986	U56936	Dissen et al., 1996 [12]
rNKR-P1B TO	AAQ11375	AF541943	Voigt et al., 2007 [9]
rNKR-P1B WAG	ABO15819	EF100679	Voigt et al., 2007 [9]
rNKR-P1F SD	ABO15825	EF100685	Voigt et al., 2007 [9]
rNKR-P1F F344	CAA66111	X97477	Appasamy et al., 1996 [13]
rNKR-P1F WAG	ABO15820	EF100680	Voigt et al., 2007 [9]
rNKR-P1G PVG	ABA40405	DQ157011	Kveberg et al., 2009 [11]
rNKR-P1G DA	ABA40406	DQ157012	Kveberg et al., 2009 [11]
rNKR-P1G B6	AAA39822	M77676	Giorda et al., 1991 [14]
mNKR-P1A BALB	ABB72025	DQ237927	Carlyle et al., 2006 [15]
mNKR-P1A SJL	AAK39101	AF354261	Kung et al., 1999 [16]
mNKR-P1A B6	AAK08512	AF338321	Carlyle et al., 2006 [15]
mNKR-P1B SJL	AAK39099	AF354259	Kung et al., 1999 [16]
mNKR-P1B NOD	BAE41390	AK169820	Carninci et al., 1999 [17]
mNKR-P1B BALB	ABB72026	DQ23728	Carlyle et al., 2006 [15]
mNKR-P1B B6	BAC37739	AK079743	Carninci et al., 1999 [17]
mNKR-P1C			

mNKR-P1C	BALB	ABB72027	DQ237929	Carlyle et al., 2006 [15]
mNKR-P1F	B6	NP_694734	NM_153094	Plougastel et al., 2001 [18]
mNKR-P1F	BALB	ABB72029	DQ237931	Carlyle et al., 2006 [15]
mNKR-P1F	NOD	BAE41516	AK170023	Caminci et al., 1999 [17]
mNKR-P1G	B6	ABC71751	DQ336141	Carlyle et al., 2006 [15]
hNKP-P1		NP_002249	NM_002258	Lanier et al., 1994 [19]
cNKR-P1		CAB55571	AJ245903	Kaufman et al., 1999 [20]

Table 2 The assignment of Raman vibrational bands distinguishable in the spectra of rat NKR-P1A protein (see Fig. 4)

Frequency (cm ⁻¹)	Assignment	Notes
509	vS-S in GGG	[44]
539	vS-S in TGT	[44]
572	Trp	[44]
623	Phe	[45, 46]
642	Tyr	[45, 46]
697	vC-S P _N , P _H	[44]
760	Trp W18	[45-47]
828	Tyr doublet, Phe	[44-47]
853	Tyr doublet, Ile	[44-47]
878	Trp W17, Ile, Val	[45-47]
894	Lys, Ala	[45, 46]
937	Lys, Val, Leu	[45, 46]
959	Lys, Leu	[45, 46]
988	Ile	[45, 46]
1005	Phe	[44-46]
1011	Trp W16 shoulder	[44-47]
1033	Phe, Tyr	[44-46]
1062	Lys, Ala, Phe	[45, 46]
1077	Lys, Phe	[45, 46]
1127	Ile, Val, Leu, Trp	[45, 46]
1158	Ile, Val	[45, 46]
1175	Tyr, Phe	[45, 46]
1206	Phe, Tyr, Trp shoulder	[45, 46]
1237	amide III	[45, 46]
1315	amide III, Lys, Val, Ile	[45, 46]
1339	Trp doublet	[44-47]
1358	Trp doublet, Val, Phe	[44-47]
1404	vCO ₂ ⁻ of Asp, Glu	[47]
1425	Trp shoulder	[45, 46]
1449	δCH ₂ , δCH ₃ , Lys, Ile, Leu	[45, 46]
1462	δCH ₂ , Ala, Ile, Val, Leu, Trp, Tyr	[45, 46]
1551	Trp W3	[48]
1579	Trp, Phe	[45, 46]
1607	Phe, Tyr Y8b	[45-47]
1620	Tyr Y8a, Trp	[45-47]
1671	amide I	[44-47]

Table 3 An estimation of secondary structure content of mNKR-P1A/C and rNKR-P1A/B proteins in aqueous solution by pattern recognition least-squares method (LSA) analyzing the amide I band in the case of Raman spectroscopy [49], and amide I and II region in the case of FTIR [50] (using the spectral set from ref. [51]). Secondary structure estimation is compared with modeled structures. Given standard deviations are calculated as standard deviations of the used reference set, therefore they do not reflect the quality of the fits

mouse NKR-P1A			mouse NKR-P1C			
Structure	Raman	FTIR	Model	Raman	FTIR	Model
α -helix	28 \pm 5	22 \pm 10	15	26 \pm 5	20 \pm 10	20
β -sheet	38 \pm 4	26 \pm 9	39	37 \pm 4	24 \pm 9	37
β -turn	17 \pm 2	14 \pm 4	10	18 \pm 2	14 \pm 4	14
β -bend	—	14 \pm 4	18	—	18 \pm 4	10
Unordered	17 \pm 2	24 \pm 6	18	19 \pm 2	24 \pm 6	19
rat NKR-P1A			rat NKR-P1B			
Structure	Raman	FTIR	Structure	Raman	FTIR	Structure
α -helix	15 \pm 5	13 \pm 10	18	19 \pm 5	25 \pm 10	25
β -sheet	50 \pm 4	32 \pm 9	40	48 \pm 4	26 \pm 9	33
β -turn	19 \pm 2	13 \pm 4	15	19 \pm 2	10 \pm 4	12
β -bend	—	16 \pm 4	11	—	14 \pm 4	12
Unordered	15 \pm 2	26 \pm 6	16	14 \pm 2	25 \pm 6	18

Table 4 Amino acids (their primary structure number) that form α -helices or β -strands in every model that was calculated in Procheck. In mouse NKR-P1F, β -strand 4 is divided into two regions. For this reason there are two numbers in the relevant column

	I	II	A	III	B	IV	C	V	VI	VII	VIII
Mouse A	100 – 102	104 – 114	119 – 127	129 – 131	135 – 146	152 – 159		167 – 169	188 – 194	197 – 203	207 – 215
Rat A	97 – 101	103 – 113	114 – 125	128 – 130	134 – 145	150 – 159		164 – 168	188 – 194	196 – 200	206 – 215
Rat B	100 – 102	104 – 114	115 – 126	129 – 131	135 – 149	152 – 155	175 – 181	182, 183	190 – 194	197 – 201	207 – 215
Mouse C	100 – 102	104 – 114	115 – 126	129 – 131	135 – 149	152 – 159		167 – 169	188 – 194	197 – 201	207 – 215
Mouse F	98 – 102	105 – 114	115 – 126	129 – 131	135 – 146	152 – 154/ 156 – 158		167 – 169	190 – 192	199 – 201	207 – 213
Mouse G	98 – 102	104 – 114	115 – 126	129 – 131	135 – 146	152 – 159		167 – 169	188 – 194	197 – 203	207 – 215
Human	98 – 102	105 – 115	116 – 125	129 – 131	135 – 147	153 – 160	175 – 180	165 – 169	189 – 194	197 – 201	205 – 213

Table 5 The number of residues in secondary structure features. In mouse NKR-P1F, β -strand 4 is divided into two regions. For this reason there are two numbers in the relevant column

	1	2	A	3	B	4	C	5	6	7	8
Mouse A	2	11	9	3	12	8		3	7	7	9
Rat A	5	11	12	3	12	10		5	6	5	9
Rat B	3	11	11	3	12	4	7	2	5	4	9
Mouse C	3	11	12	3	15	8		3	6	5	9
Mouse F	5	10	12	3	12	3+3		3	3	3	7
Mouse G	5	11	12	3	12	8		3	7	7	9
Human	5	11	10	3	13	8	6	5	6	5	9

Figure captions

Fig. 1 A comparison of FTIR spectra from mouse NKR-P1A protein (mA) with its mouse NKR-P1C variant (mC) in the region of amide I and amide II bands. The dash-dot lines represents second derivative of the spectrum (smoothed by Savitski-Golay function at 15 points, i.e. ca. 14 cm^{-1}). The peak positions are labeled for NKR-P1A protein. Slight shifts are therefore visible in the second derivative of FTIR NKR-P1C spectrum. Two major bands, observed at 1642 and 1549 cm^{-1} , are assigned to amide I and amide II vibrations of the backbone chain, respectively. The second derivative, which can identify overlapping components, reveals a high content of extended β -sheets by a strong negative band at 1638 cm^{-1} [52]. The second component of β -sheets at 1688 cm^{-1} suggests that β -sheets are antiparallel [53]. The negative band at 1657 cm^{-1} corresponds to α -helices and an unordered structure [52] which points towards an α/β protein fold. The second derivative band at 1688 cm^{-1} , with its shoulder at 1678 cm^{-1} , is connected with high content of loops and turns

Fig. 2 A comparison of mouse NKR-P1A DCDR spectra (mA) with its mouse NKR-P1C variant (mC); (mA – mC) denotes appropriate spectral difference. In the amide I region at 1669 cm^{-1} , corresponding mostly to stretching $\nu_{\text{C}\alpha=\text{O}}$ skeletal vibrations (for band assignments see Table 2), small negligible differences between mNKR-P1A and C protein variants can be found (mA – mC difference). This is supported by the skeletal NH-vibration region of amide III at 1310 cm^{-1} , where the difference is almost zero between mouse proteins. The band at 508 cm^{-1} corresponds to the lowest free energy *gauche-gauche-gauche* (GGG) conformation of S–S bridges, the most obvious conformation in the protein structures. The band at 535 cm^{-1} reflects unusual conformation that is close to *trans-gauche-trans* – TGT (pure conformation is characterized by vibrations at 540 cm^{-1} , whereas *gauche-gauche-trans* (GGT) has a band at ca. 525 cm^{-1}), which has the highest free energy [44]. The intensity ratio I_{508} / I_{535} clearly indicates that two S–S bridges have GGG conformation and one nearly has TGT conformation. Direct correlation between the Raman frequency of W3 mode and the absolute value of the torsional angle $\chi^{2,1}$ is known [48]. Its position at 1550 cm^{-1} corresponds to a 93° angle $|\chi^{2,1}|$ as an

average angle of six Trp in mNKR-P1A/C proteins. The negative difference band at 1546 cm^{-1} alludes to prevalence on lower angles in the C variant. The Trp W17 mode at 877 cm^{-1} indicates moderate NH-hydrogen bond donation in both A and C variants [55]. The Trp Fermi doublet intensity ratio I_{1358} / I_{1338} ca. 0.88 is a sensitive marker of the amphipathic environment of the aromatic ring [55] and in our case indicates a hydrophilic environment of all Trp in both variants. The two bands at 827 and 852 cm^{-1} are assigned to the Tyr (Y1+Y1a) Fermi resonance doublet. Its intensity ratio I_{852} / I_{827} is used as an Tyr environment indicator [56]. The value 1.0 of mNKR-P1A protein corresponds to moderate Tyr H-bonding, mostly at the surface of the protein. Mouse NKR-P1C protein has only one Tyr with an intensity ratio of 0.75, which may suggest that phenolic OH of this Tyr is a stronger hydrogen bond donor than those placed on the surface of the protein

Fig. 3 A comparison of rat NKR-P1A FTIR spectra (rA) with its rat NKR-P1B variant (rB) in the region of amide I and amide II bands. The dash-dot lines represents second derivative of the spectrum (smoothed by Savitski-Golay function at 15 points, i.e. ca. 21 cm^{-1}). Although the rNKR-P1B spectrum was partially affected by the low stability of the sample which caused partial aggregation (resolved by second derivative bands at 1623 and 1693 cm^{-1}) [52], higher amounts of helical structures (second derivative band at 1654 cm^{-1}) in comparison with antiparallel β -sheets (negative bands at 1638 and shoulder at 1682 cm^{-1}) can be seen

Fig. 4 A comparison of rat NKR-P1A DCDR spectra (rA) with its rat NKR-P1B variant (rB); (rA – rB) denotes appropriate spectral difference. It is clearly visible in the disulfide bridges stretching vibrations region (the bands at 509 and 539 cm^{-1}), that two S–S bridges are in GGG conformation whereas one is in pure TGT conformation. In comparison with mouse protein A/C variants, rat variants show a higher tension on this TGT bridge. The S–S bridges in rNKR-P1B are much more flexible than in the A variant (exhibiting broader and less intense bands) which points to a lower stability of this protein. Only small differences exist between all mouse and rat proteins when comparing the absolute value of the torsional angle $\chi^{2,1}$ that is characterized by the band at 1551 cm^{-1} , which corresponds to an average 95° angle of six Tyr. Nevertheless, there is a positive peak, also visible within the Raman difference spectra, that is reflecting higher flexibility in the rNKR-P1B

protein. In the case of mouse A/C variants, other prominent Tyr bands (878, 1339, 1358 cm^{-1}) reflect the same phenomena as mentioned above whereas each of these vibrations shows higher flexibility in the rNKR-P1B protein. The rat proteins have only one Tyr whose ratio 1.0 of 828 and 853 cm^{-1} bands corresponds to moderate H-bonding which is characteristic for the position on the protein surface.

Fig. 5 Synchronous (left) and asynchronous (right) 2D correlation FTIR spectra of thermal dynamics from 5 °C to 90 °C (with 5 °C increments) of mouse NKR-P1A (mA) and NKR-P1C (mC) proteins. The white and red peaks are positive while the grey and blue peaks are negative. The synchronous second derivative 2DCoS spectra from FTIR spectra of mNKR-P1A/C, under thermal perturbation from 5 °C to 90 °C, exhibit several positive diagonal peaks, so called autopeaks (see left side). The diagonal spectrum of the synchronous 2DCoS mathematically corresponds to autocorrelation function. Thus, any region of the spectrum which changes intensity to a great extent, under the thermal perturbation, will show strong autopeaks, while those remaining almost constant develop no autopeaks. In the case of protein mNKR-P1C, thermal denaturation is connected with changes of aggregate β -sheets (1618 cm^{-1}), native β -sheets (1635 cm^{-1}), α -helices and unordered structures (1646 cm^{-1}), loops (1668 cm^{-1}) and β -turns (1685 cm^{-1}) [52] as demonstrated by its autopeaks. Synchronous 2DCoS describes in-phase intensity variations, i.e. structural changes, so it is questionable if the correlation is positive or negative. This question can be answered by off-diagonal peaks, so called cross peaks. While the autopeaks sign is always positive, the cross peaks sign can be either positive or negative. The synchronous cross peak sign becomes positive if the spectral intensities at the two coordinates are simultaneously increasing or decreasing. On the other hand, a negative sign at the intensity of the cross peak indicates that one of the spectral intensities is increasing while the other is decreasing. Thus, the cross peaks correspond to a simultaneous decrease of α -helices and unordered structures with an increase of β -sheets, loops and β -turns whereas the decrease of β -sheets is then connected with the increase of β -aggregated structures. Similar structural changes can be found in synchronous FTIR 2DCoS in protein mNKR-P1A, however its transition to β -aggregates was not observable because the process was too fast

- Fig. 6** Synchronous (left) and asynchronous (right) 2D correlation Raman spectra of thermal dynamics from 5 °C to 90 °C (with 5 °C increments) of mouse NKR-P1A (mA) and NKR-P1C (mC) proteins. The white and red peaks are positive while the grey and blue peaks are negative. The decrease of α -helix content (the band at 1635 cm^{-1}) is connected with an increase of β -sheet and β -turn structures (1685/1690 cm^{-1} for mNKR-P1C and mNKR-P1A, respectively). This structural change is connected with environmental changes in Trp, Tyr and Phe residues (corresponding bands at 1590, 1600 and 1610/1625 cm^{-1})
- Fig. 7** Heterospectral synchronous 2D correlation analysis Raman spectra of thermal dynamics from 5 °C to 90 °C (with 5 °C increments) of mouse NKR-P1A (mA) and NKR-P1C (mC) proteins. The dynamics of two different regions against the amide I region (1580–1700 cm^{-1}) was followed. The white and red peaks are positive while the grey and blue peaks are negative. An intense search through the entire Raman spectrum did reveal very strong correlations between the amide I region and Trp bands at 760 and 1011 cm^{-1} for both mNKR-P1 variants. It is well known that intensity changes of these Trp Raman bands are connected with amphipathic environment of the indol ring [60]. Thus, 2DCoS reflects changes in hydrophobicity in the Tyr environment by the band shifts from 756 cm^{-1} to 763 cm^{-1} and from 1007 cm^{-1} to 1020 cm^{-1} for mNKR-P1A and similarly for mNKR-P1C. Moreover, the sign for 2DCoS bands is reversed in the amide I region, ca. 1690 cm^{-1} , which corresponds to β -turn structures in proteins
- Fig. 8** Alignments for homology models in the following order; mouse NKR-P1A, rat NKR-P1A, rat NKR-P1B, mouse NKR-P1C, mouse NKR-P1F, mouse NKR-P1G and human NKR-P1
- Fig. 9** Detail view on the rat NKR-P1B model
- Fig. 10** Structures and topologies of resultant models. The regions discussed in the text are depicted here. The same fold like mouse NKR-P1C receptor is also adopted by mouse NKR-P1A/G and rat NKR-P1A. Mouse NKR-P1F, rat NKR-P1B and human NKR-P1 folds are unique

Fig. 11 Root mean square fluctuation of amino acid residues during the production phase of the simulation is represented by C_{α} carbon

Fig. 12 Conserved regions among NKR-P1s and their positions in the structural model of rat NKR-P1A. A consensus secondary structure estimation is shown below the alignment in yellow. The red box indicates that a helix is present in human NKR-P1 and rat NKR-P1B only. The color scheme underlying the secondary structure indicates buried (blue) and accessible (white) residues

Fig. 13 A cladogram of all available NKR-P1s based on nucleotide sequences. If in some case different methods show either polytomy (full line) or further branching (dashed line), both are shown, however the polytomy is more relevant, the branching is only a possibility. Probability values are based on the maximum likelihood consensus tree and are displayed as percentages

Figure 1
[Click here to download high resolution image](#)

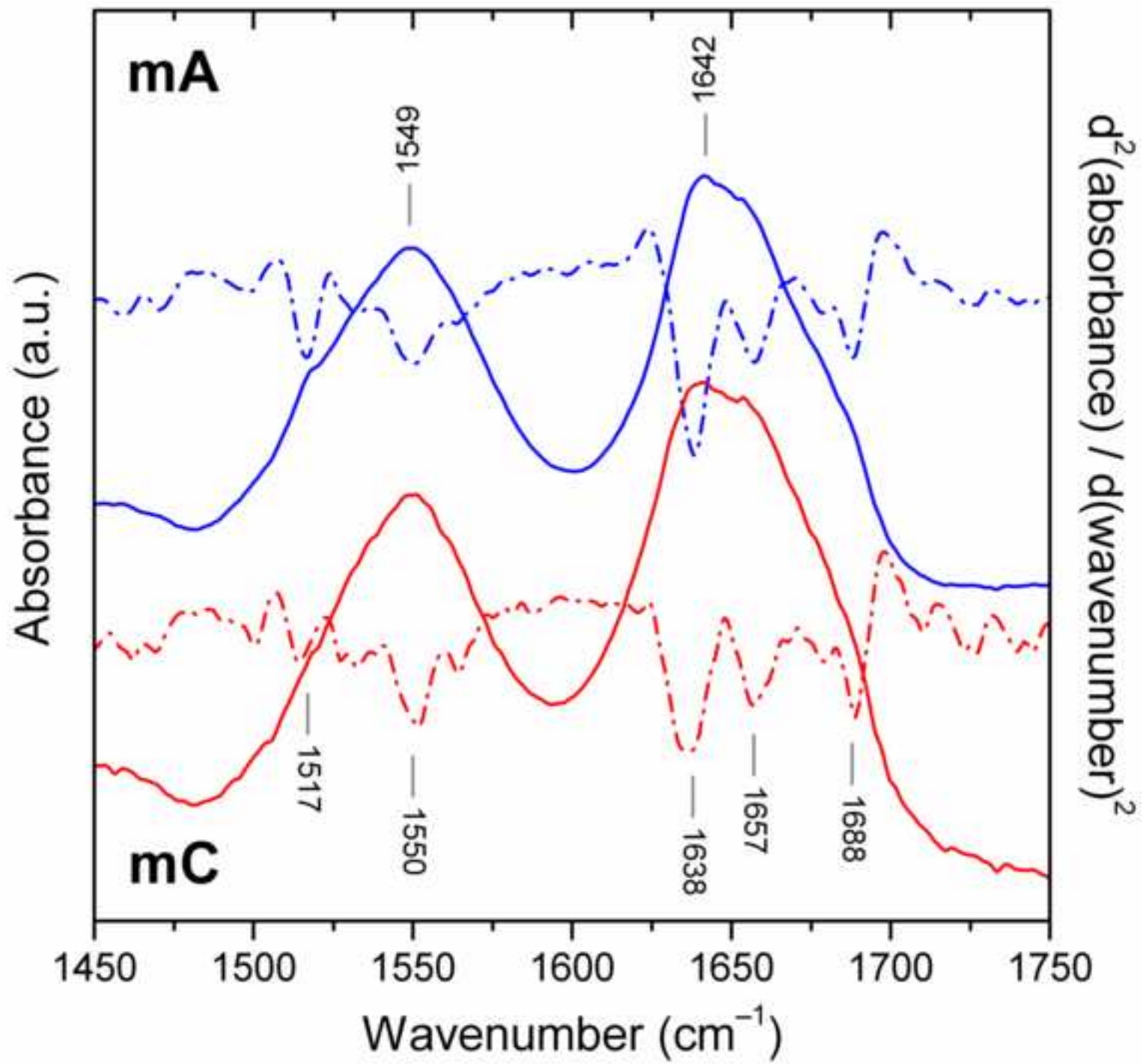


Figure 2
[Click here to download high resolution image](#)

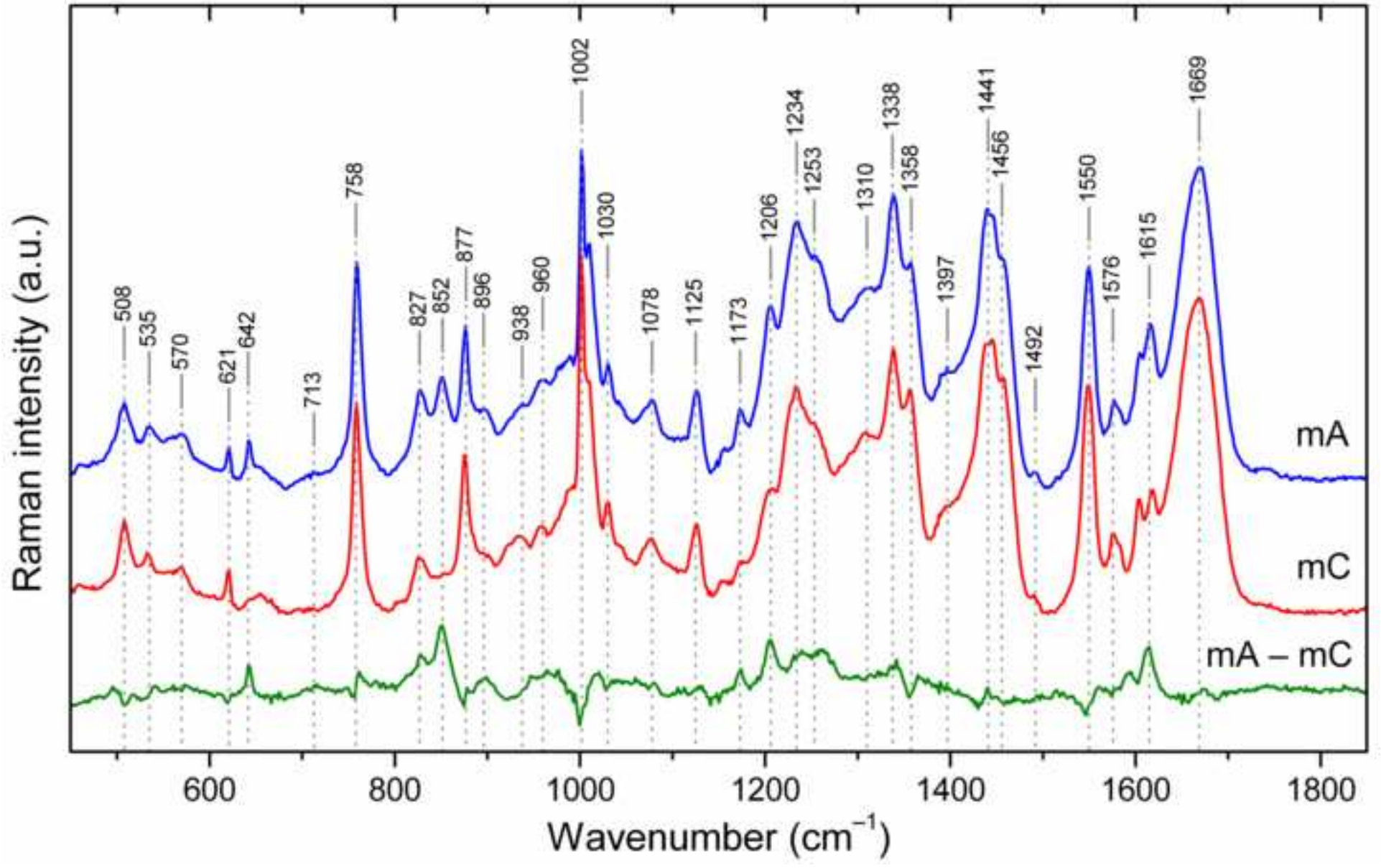


Figure 3
[Click here to download high resolution image](#)

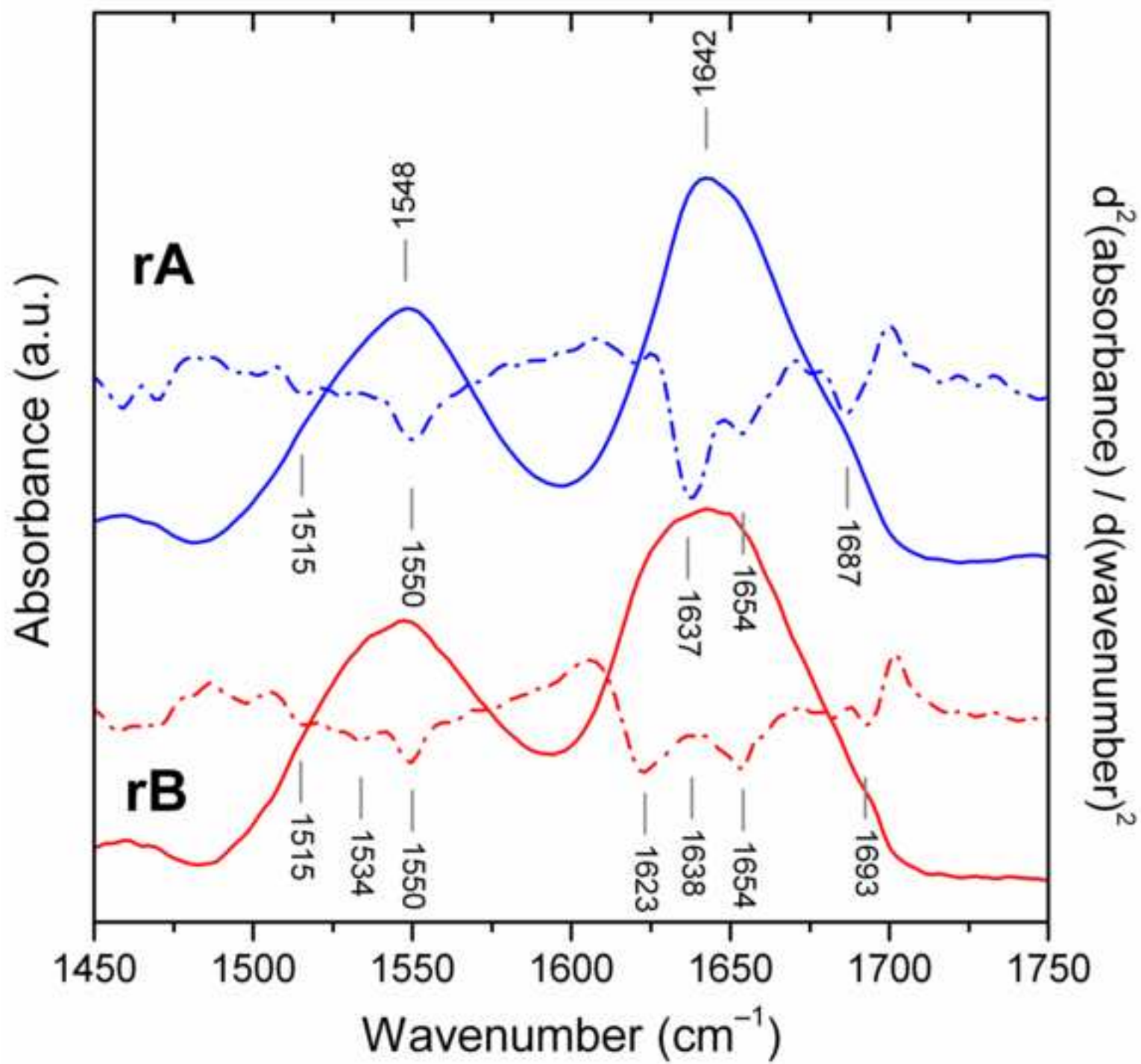


Figure 4
[Click here to download high resolution image](#)

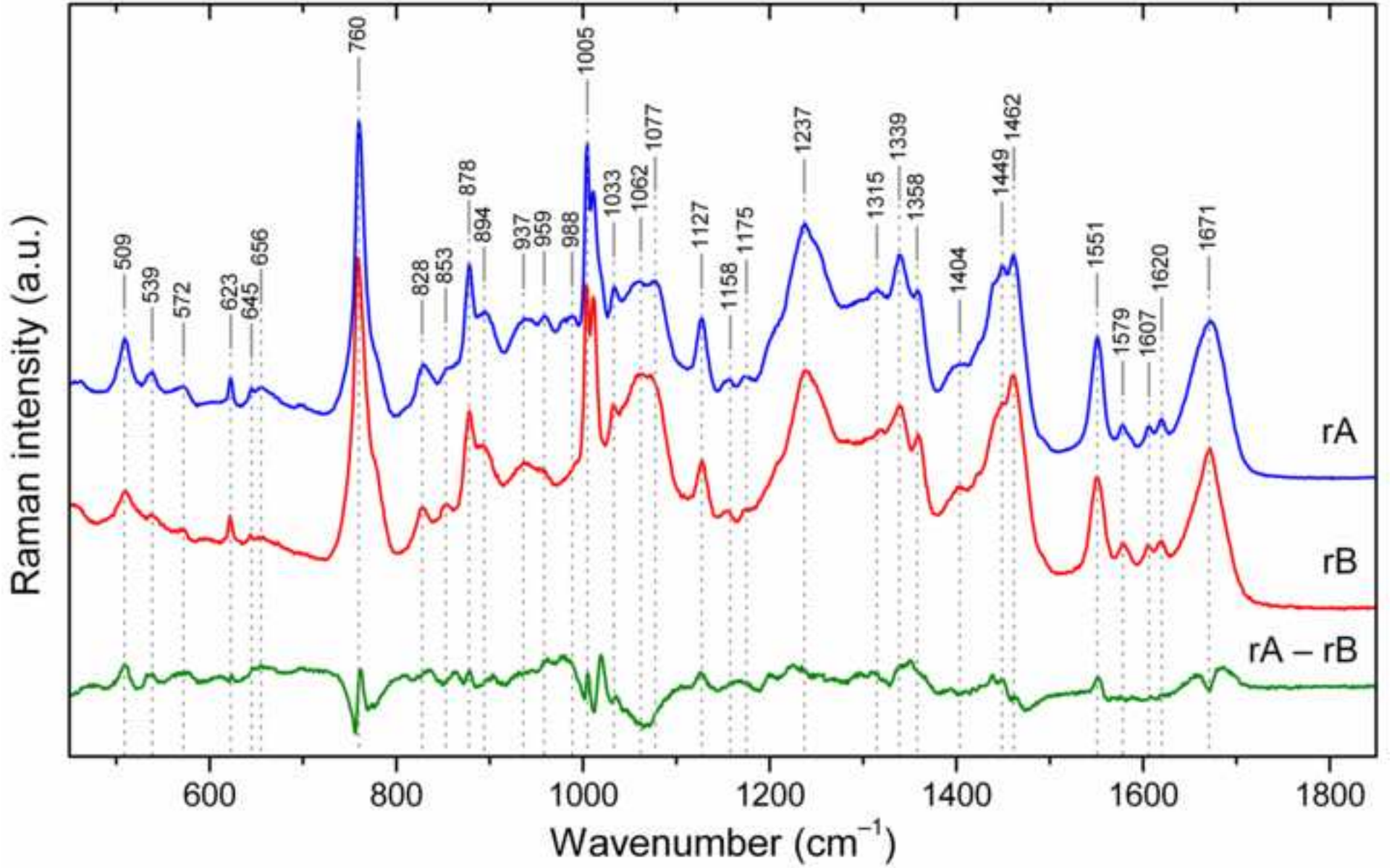


Figure 5
[Click here to download high resolution image](#)

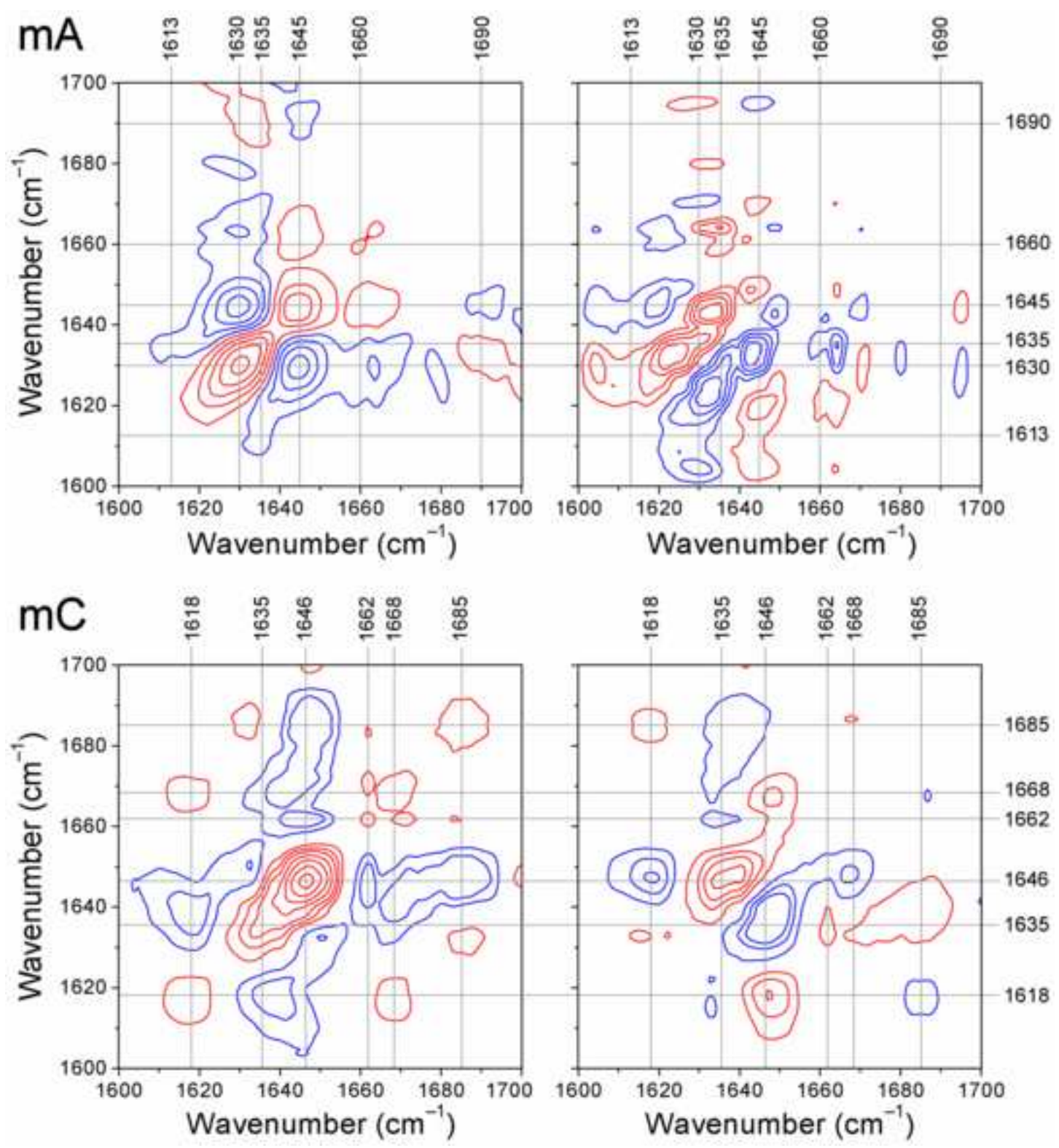


Figure 6
[Click here to download high resolution image](#)

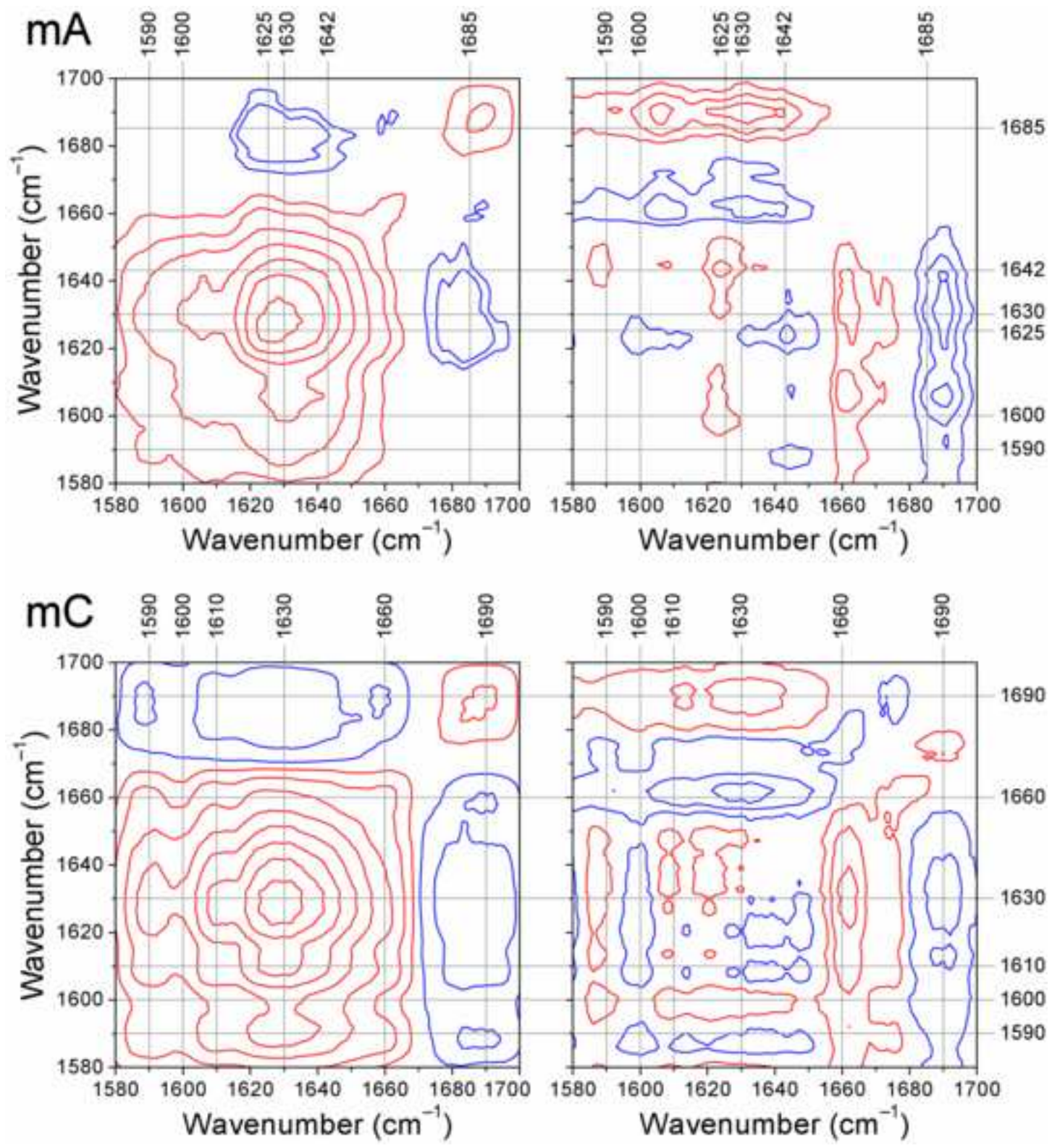


Figure 7
[Click here to download high resolution image](#)

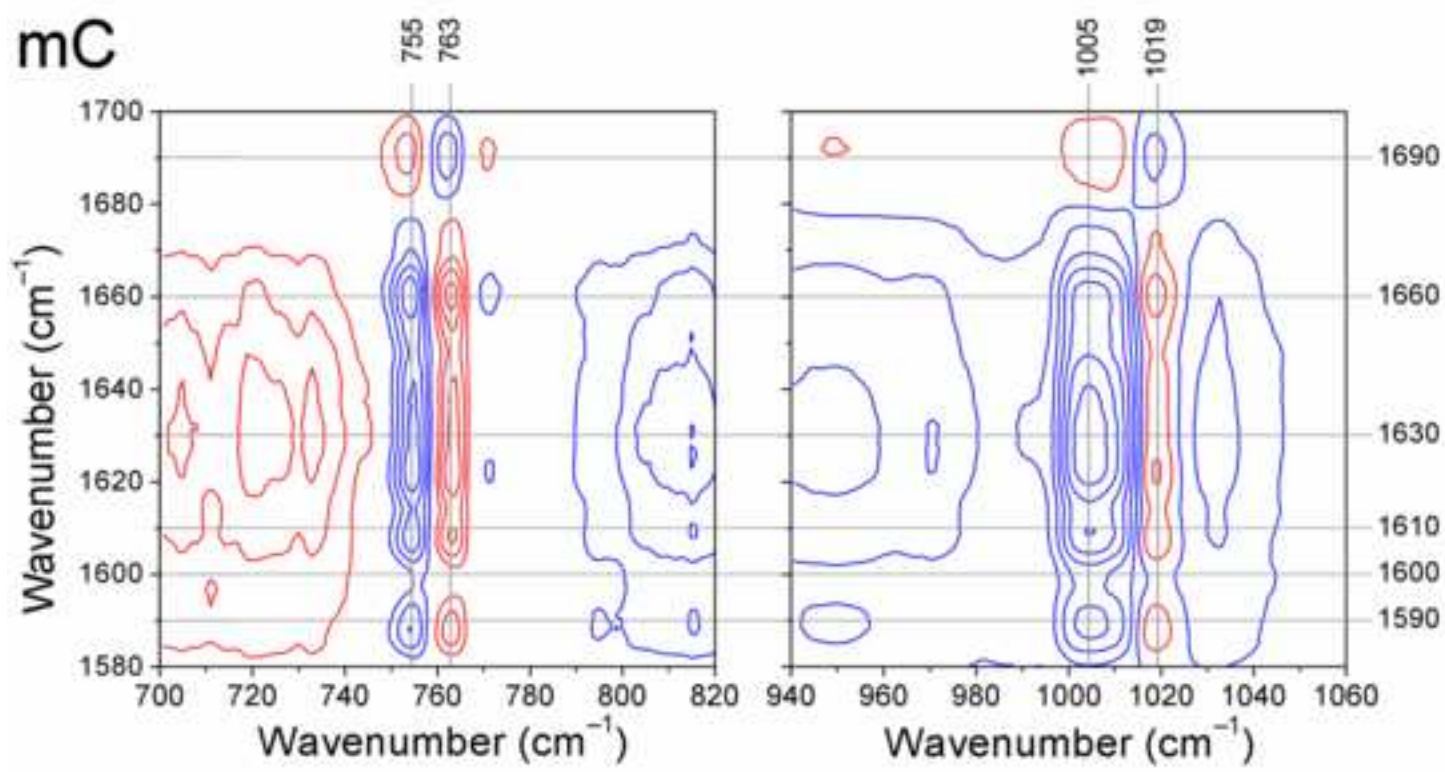
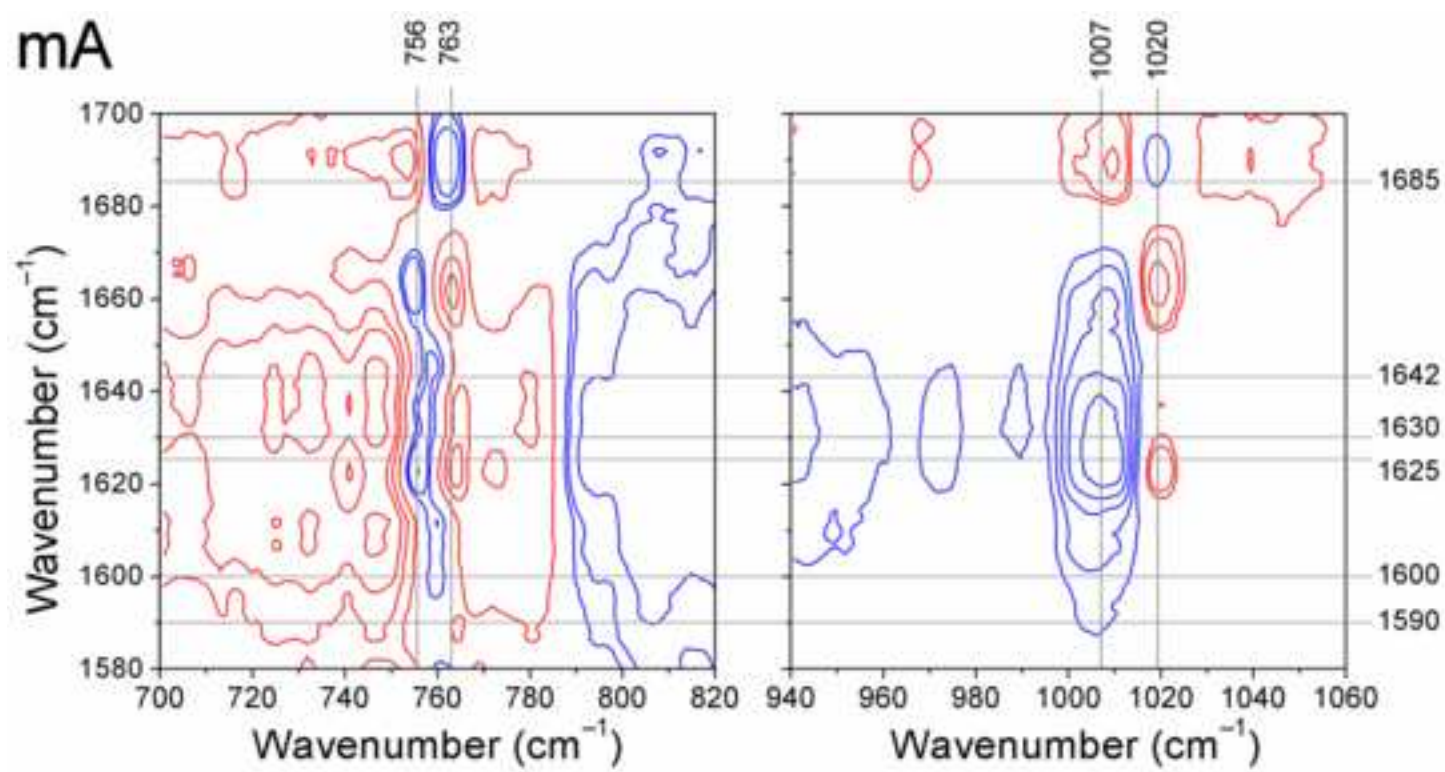


Figure 8
Click here to download high resolution image

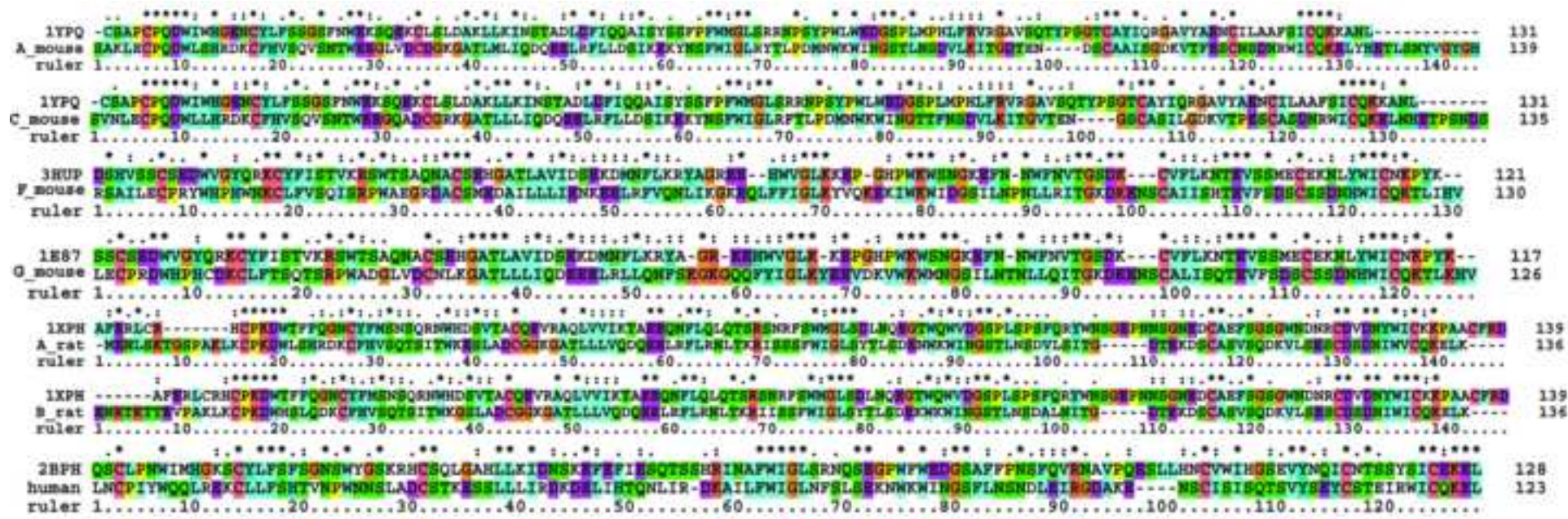


Figure 9
[Click here to download high resolution image](#)

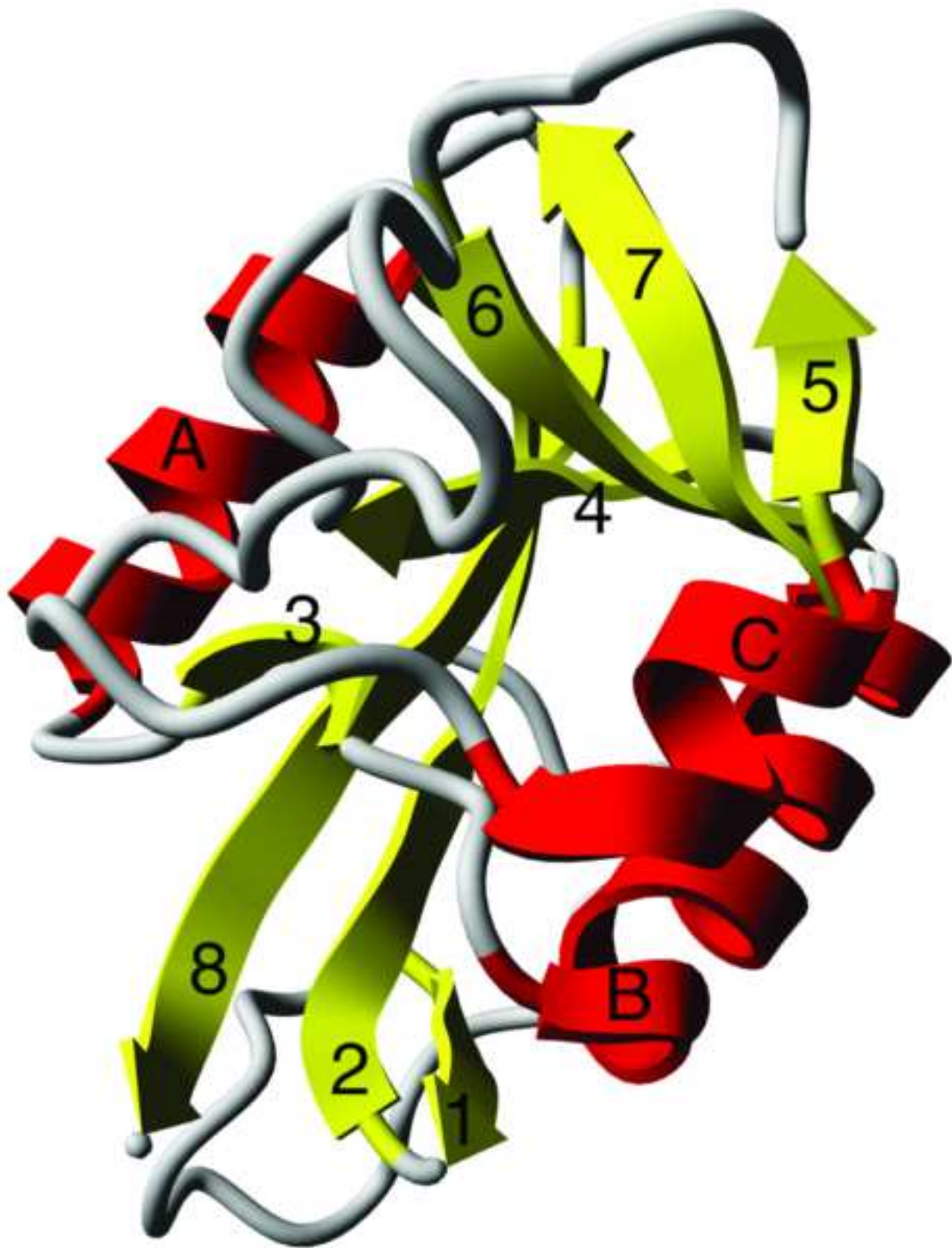


Figure 10
[Click here to download high resolution image](#)

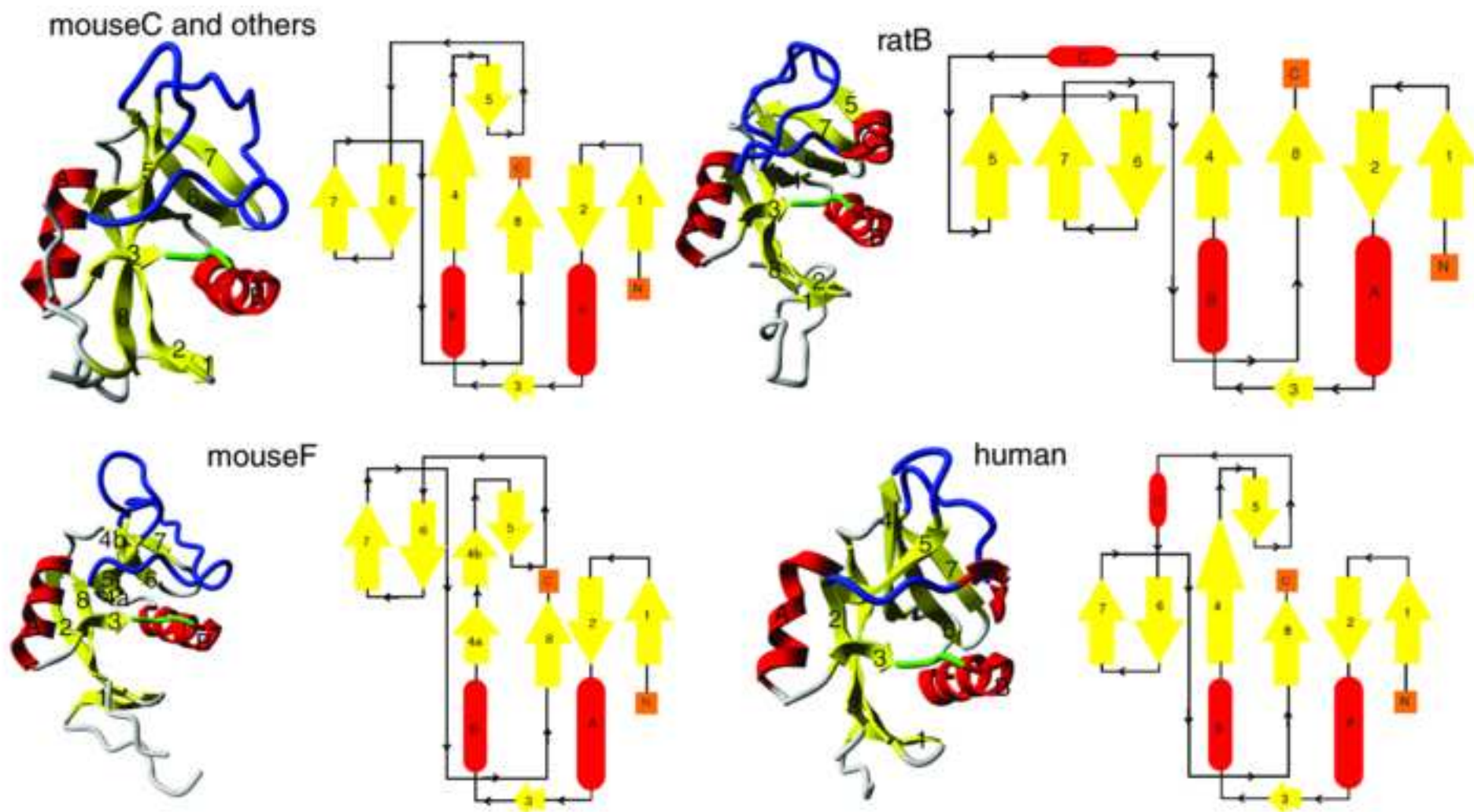


Figure 11
[Click here to download high resolution image](#)

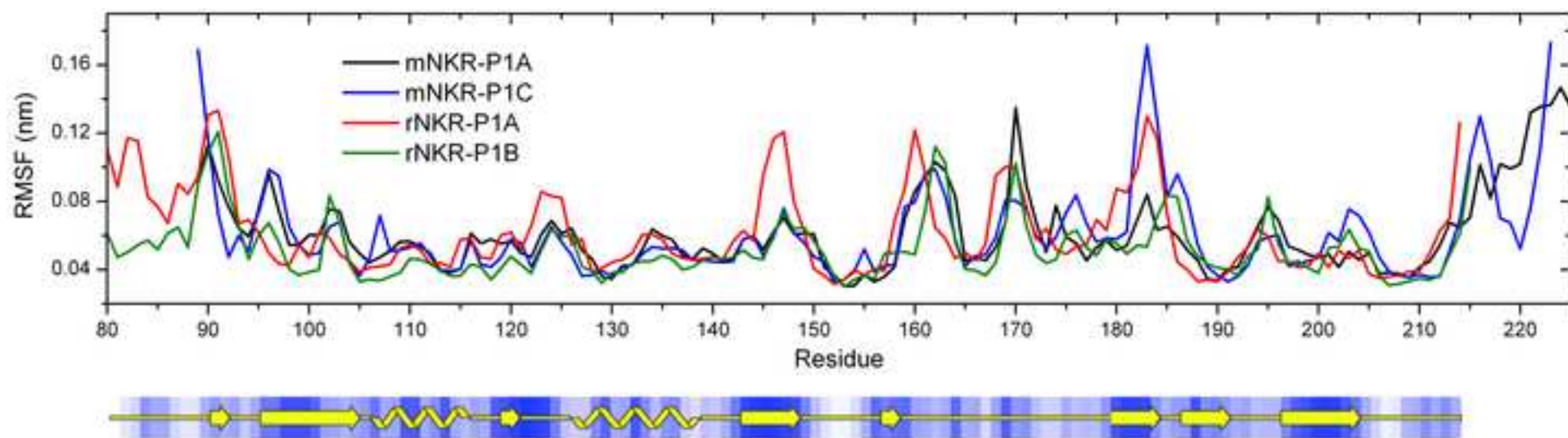


Figure 12

[Click here to download high resolution image](#)

