International Journal of Mass Spectrometry 312 (2012) 70–77

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

journal homepage: www.elsevier.com/locate/ijms

Complete sequence determination of hemoglobin from endangered feline species using a combined ESI-MS and X-ray crystallography approach

Jingshu Guo, Saurav Uppal, Lindsey M. Easthon, Timothy C. Mueser, Wendell P. Griffith*

Department of Chemistry, University of Toledo, Toledo, OH 43606, United States

ARTICLE INFO

Article history:
Received 30 March 2011
Received in revised form 25 May 2011
Accepted 1 June 2011
Available online 1 July 2011

Keywords:
Snow leopard hemoglobin
Amur tiger hemoglobin
Amino acid sequence
Mass spectrometry protein sequencing
X-ray crystallography

ABSTRACT

We present the first complete amino acid sequences of Snow leopard (*Uncia uncia*) and Amur tiger (*Panthera tigris altaica*) hemoglobin (Hb) determined using a combined ESI MS and X-ray crystallography approach. Using mass spectrometry alone, we were able to achieve 81%, 90%, 87% and 72% sequence coverage for the α-, and β-globins of Snow leopard Hb and the α-, and β-chain of Amur tiger Hb, respectively. The incomplete mass spectrometry-determined sequences were used to process X-ray diffraction data of Snow leopard and Amur tiger Hb crystals solved to 2.1 Å and 1.8 Å, respectively. Using tandem mass spectrometry in concert with the electron density maps from X-ray structures, we were able to not only achieve 100% coverage of all amino acid sequences, but to also distinguish between isomeric (leucine and isoleucine) and isobaric (glutamine and lysine) residues. These results demonstrate the synergy between mass spectrometry and X-ray crystallography for the determination of full amino acid sequence coverage of abundant proteins.

1. Introduction

The integrity of the functional hemoglobin (Hb) tetramer *in vivo* is dependent on a number of factors including the inter subunit contacts, ligand-binding state, oxidation state of the heme iron, and conformational state (denaturation, hemichrome, etc.). These factors are all affected by the aging of the erythrocyte [1–3]. Human red cells have a life span of approximately 120 days. The inability of anucleate red cells to synthesize new proteins highlights the fact that the events that trigger the removal of old cells from circulation after 120 days must be as a result of alterations in pre-existing proteins, which leads to recognizable changes at the surface of the erythrocyte cell membrane [4]. One such alteration is the clustering of band 3, an integral membrane anion transporter [1,5,6]. It has been shown that this formation of band 3 clusters creates a binding site for autologous anti-band 3 antibodies, which subsequently decorate the cell surface to make the cell recognizable to macrophages for removal by the immune system [2,3,7,8]. Even outside the body, for example when blood is donated for transfusion, age related denaturation of Hb and clustering of band 3 occurs over time of storage, thereby decreasing the quality and effectiveness of the stored blood product [9,10]. Considering the marked differences in erythrocyte life span (for example 40 days in a mouse as compared to 120 days in humans and 1400 days in toads), our research goal is to explore the role of evolution in the governance of the erythrocyte natural aging process and life span focusing on the interaction between the proteins band 3 and Hb. This involves using a combination of mass spectrometric and X-ray crystallographic tools. Achieving this goal necessitates having complete amino acid sequences of Hb and the Hb-binding domain of band 3 from all of the animal species investigated.

Perhaps, the 2 major direct methods for protein sequence determination are by tandem mass spectrometry [11,12] and the Edman degradation reaction [13,14]. It is also possible to determine amino acid sequences indirectly through sequencing of mRNA or cDNA [15], and X-ray crystallography. There are many pros and cons to the use of each of these methods for protein sequence determination. The disadvantages limit the effectiveness and efficiency of these methods. Combinations of 2 or more of these techniques, maximizing on the synergy between them and the advantages that they offer, may facilitate protein-sequencing efforts.

As a tool for protein sequences, though, X-ray crystallography is greatly limited by the quality of the protein crystals and the resolution of diffraction data. It is only at resolutions better than 2.0 Å that the automated fitting of individual atoms is possible. Most amino acid side chains can be assigned even without protein sequence information, but this is extremely difficult to accomplish.

Abbreviations: Hb, hemoglobin; CN-MetHb, cyanomethemoglobin; MS, mass spectrometry; ESI, electrospray ionization; MALDI, matrix-assisted laser desorption/ionization; CID, collision-induced dissociation.

* Corresponding author at: Department of Chemistry, MS 602, University of Toledo, 2801 W. Bancroft Street, Toledo, OH 43606-3390, United States.
Tel.: +1 419 530 7964; fax: +1 419 530 4033.
E-mail address: Wendell.Griffith@utoledo.edu (W.P. Griffith).

1387-3806/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2011.06.001

The majority of crystal structures are refined against data of modest resolution. As a result, the use of crystallographic methods for protein sequence determination is dependent on the availability of some prior sequence or protein structure information. Also, there are a number of residues having similar shape and therefore similar electron density maps, for example glutamine and glutamic acid; asparagine and aspartic acid; and threonine and valine. It is difficult to differentiate between these residues, even at ultrahigh resolution.

Tandem mass spectrometry (MS/MS) is an elegant and efficient approach that has been widely used for many years as a tool to obtain protein primary sequence information. These experiments highlighted a couple of important limitations to the use of MS for protein sequence analysis [16–18]. Unless very high mass accurate and high-resolution MS instrumentation and data are available, it is difficult to differentiate between the isobaric residues glutamine and lysine. Without subsequent and difficult experiments, it is also impossible to differentiate between the isomers leucine and isoleucine by MS.

It is quite clear that every available method for determination of protein sequence can be quite effective. No single method is capable of providing 100% coverage of a protein sequence without a great deal of effort. As a result, combinations of complementary methodologies have been used to facilitate achieving complete sequence coverage. By combining 2 or more of these methods, one can exploit the pros of each method and compensate for their cons. One example of such a combination was the tandem use of X-ray crystallography and MS by Diemer et al. to obtain *ab initio* the complete sequence coverage of HPBP, a human apolipoprotein [19].

In our analysis of hemoglobins from various animals, our lab has recently shown (unpublished data) that the interaction between Hb and Hb-binding proteins has a great dependence on sequence. Unfortunately many animal Hbs have no available sequence, which is necessary for our understanding of the interaction between the two proteins. In addition as the packed red cell samples provided are anucleate and cDNA is unavailable for sequence determination, in this manuscript we present the combination of X-ray crystallography and mass spectrometry (nanoESI-MS/MS and MALDI-TOF/TOF-MS) to determine the complete sequence coverage of hemoglobin from 2 different endangered felines: the Amur tiger and the Snow leopard. One advantage is that Hb crystallization is well established and routine. Hb is also a highly abundant protein and so there is no sample limitation for crystallization trials. One major difference between Diemer's approach and ours are that we only used one enzyme (trypsin) in our analysis. This manuscript also demonstrates the synergy between and the utility of tandem mass spectrometry and X-ray crystallography to determine 100% sequence coverage of proteins that are noncovalent complexes composed of different subunits. In short, mass spectrometry is used to determine >70% coverage of the sequence, which is used to process the crystallography data. The electron density map of the resultant crystal structure is used to fill in the regions of the sequences uncovered by MS as well as to differentiate between isomeric and isobaric residues.

2. Experimental/materials and methods

2.1. Materials

Uncia uncia (Snow leopard) and *Panthera tigris altaica* (Amur tiger) packed red blood cells (pRBC) were generously donated by the Toledo Zoo (Toledo, OH). These samples were opportunistically obtained as they are usually discarded as waste during the collection of serum from the animals for routine hormone testing. Proteomics-grade TPCK-treated trypsin, NaCN, and $K_3[Fe(CN)_6]\cdot 3H_2O$ were purchased from Sigma Chemical Co. (St. Louis, MO). Sephadex G25 resin was obtained from GE Healthcare. The hemoglobin crystal screening kit was prepared as published by Mueser et al. [20]. All other chemicals, buffers, and solvents were of analytical grade or higher. Microcon centrifugal filters with 10 kDa cut-off were obtained from Millipore (Billerica, MA).

2.2. Preparation of CN-MetHb and isolation from pRBCs

Snow leopard and Amur tiger pRBCs were suspended 3 times into a 10× volume of phosphate buffered normal saline (155 mM NaCl, 0.01 M sodium phosphate pH 7.4) and centrifuged at 2000 × *g* for 10 min at 4 °C to rinse the red cells. The buffy coat was removed by aspiration. The rinsed cells were then suspended in a 3× volume of phosphate buffer (0.005 M sodium phosphate pH 7.4) to induce osmotic lysis. The lysed cells were then subjected to high-speed centrifugation at 16,000 × *g* for 15 min at 4 °C to separate the hemolysate, which contains almost pure hemoglobin, from the erythrocyte ghost membrane fraction. Hemoglobin solutions were treated with buffered ferricyanide/cyanide (10 mM $K_3[Fe(CN)_6]\cdot 3H_2O$, 10 mM NaCN, 25 mM phosphate buffer pH 7.4) and then eluted from a 20× volume column of Sephadex G-25 using 10 mM Tris–HCl pH 7.4 to remove excess ligand and hemichrome. CN-MetHb solutions were concentrated using Microcon YM-10 centrifugal concentrators. Concentrations of hemoglobin solutions were determined using UV–vis absorbance and mM extinction coefficient of 11.0 at 540 nm for CN-MetHb [21].

2.3. Liquid chromatography

Globin chain isolation was performed by reversed-phase HPLC on a Shimadzu Biotech Prominence HPLC (Columbia, MD) equipped with a dual wavelength detector. 20 μL of approximately 1 mg/mL of either Hb solution (Snow leopard or Amur tiger Hb) was applied onto a 4.6 mm × 150 mm BDS Hypersil C18 analytical column with 5 μm particle size (Thermo Electron Corp.). Solvent A consisted of 0.1% trifluoroacetic acid, TFA in water; and Solvent B consisted of 0.1% TFA in acetonitrile, ACN. For all separations, the flow rate used was 1 mL/min. Absorbance was measured at 280 nm and 214 nm. For the Snow leopard Hb, globin separation was achieved using an optimized elution gradient profile from 35% B to 60% B over 25 min. For the Amur tiger Hb, globin separation was achieved using an optimized elution gradient profile from 42% B to 52% B over 30 min. Fractions were collected at a rate of 2 fractions min^{-1} and those fractions containing pure single globins were pooled. The identities of the representative peaks for the α- and the β-chains were confirmed by electrospray ionization mass spectrometry (ESI-MS) measurements. The pooled fractions were evaporated to dryness by speed-vac at 30 °C.

2.4. Proteolytic digestion

Pooled fractions containing purified globin chains were reconstituted by dissolution in 100 mM ammonium bicarbonate pH 8.5 and digested with proteomics-grade trypsin using approximately a 1:50 molar ratio of enzyme to globin chain. Incubation was carried out at 37 °C overnight.

2.5. Mass spectrometry

All mass spectrometry data were collected on a Synapt HDMS quadrupole – time of flight ion mobility mass spectrometer equipped with a nanospray source (Waters Corp.). Instrument

parameters were maintained at the following values for each experiment: capillary voltage 3.0 kV, sample cone 40 V, cone gas 0 L/h, and trap collision energy varied from 6 V (for intact globin mass measurement) to 30 V (for peptide MS/MS). Mass spectra were calibrated externally in the range $50 \leq m/z \leq 3000$ using a solution of sodium cesium iodide, and processed using Masslynx 4.1 software (Waters). All mass spectra were averages of 300 scans and presented unprocessed (unsmoothed and without background subtraction). All globin samples (intact or digested) were reconstituted in a solution containing 0.1% formic acid and 50% ACN and diluted to working solution concentrations of approximately 2–10 μM.

For the bottom-up sequencing of the tryptic peptides, after a mass spectrum representative of the digest mixture of tryptic peptides was acquired, each peptide was isolated, fragmented by collision-induced dissociation and tandem mass spectra recorded in the *m/z* range 50–3000. All MS/MS data were smoothed and MaxENT transformed prior to searching against protein database using the online MASCOT utility (freely available at http://www.matrixscience.com). Sequence homology of globin hits from the MASCOT search was used to facilitate compilation of protein sequence from peptide MS and MS/MS data. For the α-globins of both the Snow leopard and Amur tiger, the α-globin sequence from the Northern Persian leopard (HBA_PANPS, P18976) was used. For the Snow leopard and Amur tiger β-globins, the β-globin sequences from the Amur leopard (HBB_PANPO, P04244) and the lion (HBB1_PANLE, P68050) were used, respectively. ESI-MS results were confirmed by MALDI-MS/MS using an ultrafleXtreme MALDI-TOF/TOF mass spectrometer equipped with a smartbeam II laser (Bruker Daltonics) in the positive ionization mode. (Experimental details as well as some representative MALDI mass spectra are included in Supplementary Information.)

2.6. CN-MetHb nucleation and crystallization

20 mg/mL solutions of Snow leopard and Amur Tiger Hb were used to screen for crystallization conditions using the sitting drop vapor diffusion method. Snow leopard CN-MetHb crystals were obtained at room temperature using a mother liquor containing 0.2 M sodium cacadylate, 0.1 M PIPES pH 6.5, and 14% PEG 4000. Amur Tiger CN-MetHb crystals were obtained at room temperature using a mother liquor containing 0.2 M potassium chloride, 0.1 M PIPES pH 6.5, and 13% PEG 4000. 15% glycerol was used as a crystal cryoprotectant for X-ray diffraction data collection.

2.7. X-ray diffraction data collection and data processing

All X-ray diffraction data were collected on a Rigaku FRE with Saturn 92 CCD at 100 K. Crystal Clear software was used to integrate and scale the data. The space group, unit cell dimension, orientation and other information for all crystals are shown in Table 1. The structure for carbonmonoxy bovine Hb pH 7.2 (PDB ID: 1G08) was used as a model to facilitate solution of the structure. Molecular replacement and refinement of each CN-MetHb data set was then carried out using CCP4 (Suite 6.1.0). Coot (version 0.5.2) and Pymol were used for processing of the final structures.

3. Results

In the chromatographic separation and isolation of Snow leopard (Fig. 1) and Amur tiger (Supplementary Fig. S1) globin chains, baseline separation of the α-globin and the β-globin chains was achieved. In both cases the α-globin eluted before the β-globin. For Snow leopard Hb, the α-chain eluted at 13.5 min while the β-chain eluted at 17.0 min. The over 2 min baseline separation window between the elution of the 2 peaks facilitated fraction collection. Fig. 2 shows the ESI mass spectra of the pooled fractions corresponding to the Snow leopard α-globin. The electrospray deconvolution of the mass spectrum is provided as an inset. (Mass spectra and deconvolutions for the Snow leopard β-globin and the Amur tiger α- and β-globins are provided in Supplementary Information.) The molecular weights of the globin chains were determined to be 15,486, 16,028, 15,456, and 16,027 for the Snow leopard α- and β-globin and the Amur tiger α- and β-globin, respectively. These masses could not be checked against protein sequences due to the unavailability of sequences for these proteins in protein sequence databases.

The ESI mass spectra of the trypsin-digested Snow leopard and Amur tiger α- and β-globins are provided in Supplementary Information. Using mass measurement in concert with peptide mass fingerprint Mascot search, many of the peaks in the mass

Table 1
Summary of crystal structure parameters.

	Snow leopard Hb	Amur tiger Hb
Space group	P1 2_1 1	$P2_12_12_1$
Unit cell dimension (Å)		
a	54.39	55.51
b	74.88	73.04
c	75.64	143.35
Unit cell orientation (°)		
α	90.00	90.00
β	104.66	90.00
γ	90.00	90.00
Resolution range (Å)	43.05–2.10	39.99–1.81
Data redundancy[a]	2.83 (2.37)	13.8 (12.0)
R_{merge} (%)[b]	5.3 (23.8)	6.4 (23.8)
Completeness (%)	98.4 (89.2)	99.9 (99.6)
Mosaicity	1.811	0.26
Model refinement		
R_{factor} (%)[c]	21.13	17.69
R_{free} (%)[d]	26.21	21.43
Rmsd from ideal bond lengths (Å)	0.0069	0.0099
Rmsd from ideal angles (Å)	0.940	1.203

[a] The corresponding value for outer most resolution shell are shown in parentheses.

[b] aaa $R_{merge} = (100 \sum_{h\,k\,l} \sum_j |Ih_{h\,k\,l,j} - Ih_{h\,k\,l}|)/(\sum_{h\,k\,l} \sum_j Ih_{h\,k\,l,j})$, where Ih_{hkl} is the weighted mean intensity of the symmetry-related reflections $Ih_{hkl,j}$.

[c] aaa $R_{factor} = (100 \sum_{h\,k\,l} |F_{obs} - F_{calc}|)/(\sum_{h\,k\,l} F_{obs})$, where F_{obs} and F_{calc} are the observed and calculated structure factors.

[d] R_{free} is calculated from the same equation as R_{factor} using a small subset of randomly selected reflections that are not used in the refinement of the structural model.

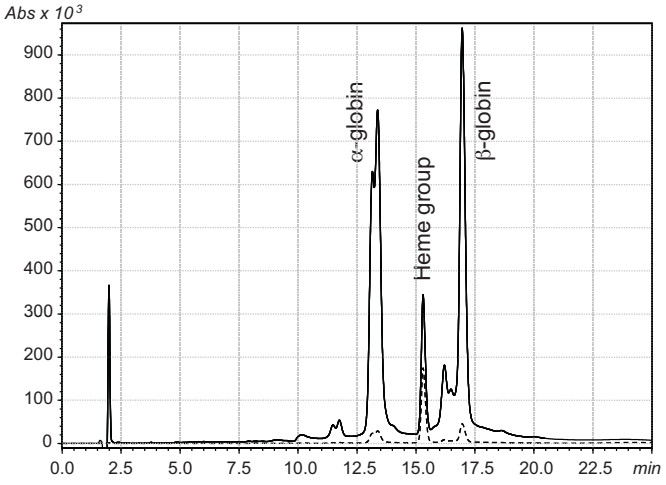

Fig. 1. Reversed-phase HPLC chromatogram showing the separation of the α- and β-globin chains of Snow leopard hemoglobin. The UV absorbance was recorded at 214 nm (solid line) and at 280 nm (dashed line). Peaks corresponding to the α-globin, β-globin, and heme group are labeled.

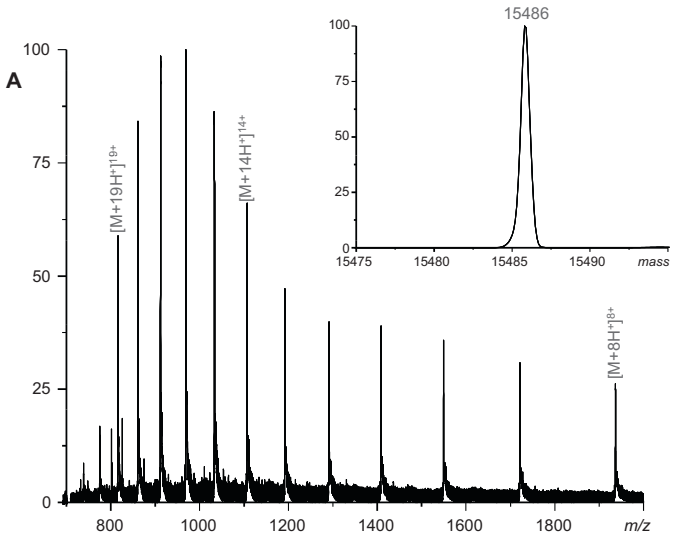

Fig. 2. ESI mass spectrum showing the isolation and mass measurement of the pooled HPLC fractions containing the Snow leopard α-globin. Inset shows the deconvolution of the mass spectrum to represent neutral mass.

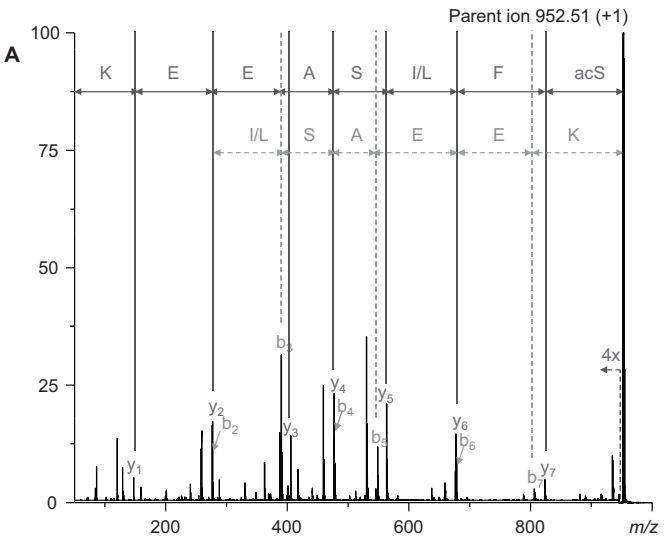

Fig. 3. Tandem mass spectrum for the tryptic peptide from digest of Snow leopard β-globin. b- and y-type fragment ions used to confirm the sequence identity for the peptides are labeled along with the determined sequence for residues 1–8, which contains N-terminal acetylserine.

spectra could have potential sequences assigned based on sequence homology with globin hits from other animal species. This potential sequence information greatly facilitated and expedited analysis of MS/MS data. All protein sequences were confirmed by tandem MS. A number of the peaks though could not be matched by mass to any homologous protein sequence. De novo sequencing of MS/MS data was used to generate sequence information for these peptides. Sequences for all peptides, their positions in the protein sequences as well as their experimental and calculated masses can be found in Table 2. (It is important to note that the assignment of leucine and isoleucine in these sequences were confirmed from electron density maps from the crystal structures.) Although, impossible with isomeric residues, the mass accuracy of the instrument used in this research was adequate to allow the differentiation between the isobaric residues glutamine and lysine (residue monoisotopic masses 128.05858 and 128.09496, respectively).

Collision-induced dissociation (CID) fragmentation mass spectra for selected peptides from the Snow leopard β-globin are provided in Figs. 3 and 4. (See Table 2 for sequence information and Supplementary Information for all other tandem mass spectra.) The tandem mass spectrum of a +1 charge state ion at *m*/*z* 952.51 (Fig. 3) was identified as the N-terminal peptide of the Snow leopard β-globin. Data was consistent with the N-terminus of this β-globin being acetylated. In fact, the β-globins from both Snow leopard and Amur tiger were found to be acetylated. This posttranslational modification has been previously demonstrated for a number of hemoglobin proteins from other feline species. Fig. 4 shows the tandem mass spectrum of the peptide at *m*/*z* 657.83 (+2 charge state), which corresponds to the Snow leopard β-globin peptide 18–30 with sequence VNVDEVGGEALGR. The resolution and mass accuracy of the data was adequate enough to allow the differentiation between aspartic acid and asparagine residues; as well as glutamic acid and glutamine residues. Residue 28, which could be either leucine or isoleucine could not be assigned with this data and was later assigned using electron density maps from the X-ray diffraction data.

Diffraction quality crystals from crystal growth optimized conditions for 20 mg/mL solutions of Snow leopard and Amur tiger Hb (Supplementary Information) resulted in resolutions of 2.1 Å (Snow leopard Hb) and 1.8 Å (Amur Tiger Hb). Even at these resolutions and using the CO-ligated bovine Hb structure as a model, complete solution of the structures would have proven difficult. Structure solution from the diffraction data was facilitated and expedited by the incorporation of MS-determined protein sequences. Structures of the Snow leopard and Amur tiger Hb (Supplementary Information) show the expected $\alpha_2\beta_2$ tetrameric structure of all mammalian hemoglobins as well as the expected CN ligation of the heme group. Electron density maps, as discussed later, were invaluable in facilitating the differentiation between isomeric leucine and isoleucine.

Peptides sequences that were not detected/covered by mass spectrometry include Snow leopard α: residues 62–68, Amur tiger α: residues 12–16, 32–40, and Amur tiger β: residues 88–95, and 105–132. Examples of the use of electron density maps were used to determine the sequences of these peptides are provided in Fig. 6 for the amur tiger α-globin peptide 12–16 (A) and the tiger β-globin peptide 88–95 (B). The tiger β-globin peptide 88–95 contains 2

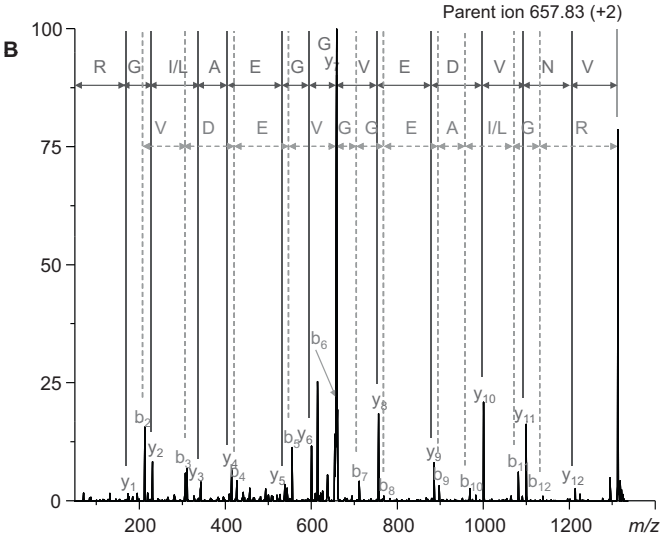

Fig. 4. Tandem mass spectrum for Snow leopard β-globin tryptic peptide residues 18–30. b- and y-type fragment ions used to confirm the sequence identity for the peptides are labeled along with the determined sequence. This peptide contains 1 Ile/Leu residue indistinguishable in this MS experiment. It also contains 1 Asn, 1 Asp, and 2 Glu residues which were confidently assigned by mass.

Table 2
Summary of peptides identified by mass spectrometry.

No.	Globin chain	Sequence position	Parent ion (monoisotopic)	$[M+H]^+_{meas}$ (monoisotopic)	$[M+H]^+_{calc}$ (monoisotopic)	Δm	Peptide sequence
1	Snl-α	1–7	719.42(+1)	719.42	719.39	0.03	VLSSADK
2		8–11	474.29(+1)	474.29	474.27	0.02	NNVK
3		17–31	780.37(+2)	1159.74	1559.74	0.00	IGSHAGEYGAEALER
4[a]		32–40	1030.51(+2)	2060.02	2060.98	−0.96	TFCSFPTTK
5		41–61	1193.10(+2)	2385.20	2385.14	0.06	TYFPHFDLSHGSAQVQTHGQK
6		69–90	793.42(+3)	2378.26	2378.20	0.06	AVAHIDDLPNALSDLSDLHAYK
7		93–99	818.45(+1)	818.45	818.44	0.01	VDPVNFK
8[a]		100–127	1048.53(+3)	3143.59	3143.56	0.03	FLSHCLLVTLACHHPEEFTPTVHASLDK
9		128–139	643.80(+2)	1286.60	1286.70	0.10	FFSAVSTVLTSK
10	Snl-β	1–8	952.51(+1)	952.51	952.46	0.05	acSFLSAEEK
11		9–17	974.58(+1)	974.58	974.53	0.05	GLVDGLWSK
12		18–30	657.83(+2)	1314.66	1314.66	0.00	VNVDEVGGEALGR
13		31–40	637.87(+2)	1274.74	1274.73	0.01	LLVVYPWTQR
14		41–59	1018.48(+2)	2035.96	2035.94	0.02	FFQSFGDLSSADAIMSNAK
15		62–65	412.25(+1)	412.25	412.23	0.02	AHGK
16		67–76	1079.59(+1)	1079.59	1079.57	0.02	VLNSFSDGLK
17		77–82	717.40(+1)	717.40	717.38	0.02	NIDDLK
18		83–87	493.30(+1)	493.30	493.28	0.02	GAFAK
19		96–104	563.78(+2)	1126.56	1126.56	0.00	LHVDPENFR
20[a]		88–95& 105–132	1353.41(+3)	4058.22	4059.08	−0.85	LSELHCDK LLGNVLVCVLAHHFGHEFNPQVQAAFQK
21		133–144	575.84(+2)	1150.68	1150.67	0.01	VVAGVASALAHR
22	Tig-α	1–7	719.40(+1)	719.40	719.39	0.01	VLSSADK
23		8–11	474.26(+1)	474.26	474.27	−0.01	NNVK
24		17–31	780.37(+2)	1559.74	1559.74	0.00	IGSHAGEYGAEALER
25		41–61	795.70(+3)	2385.10	2385.14	−0.04	TYFPHFDLSHGSAQVQTHGQK
26		62–68	717.42(+1)	717.42	717.41	0.01	VADALTK
27		69–90	793.40(+3)	2378.20	2377.21	−0.01	AVAHIDDLPNALSDLSDLHAYK
28		93–99	818.44(+1)	818.44	818.44	0.00	VDPVNFK
29[a]		100–127	1038.37(+3)	3113.11	3113.55	−0.44	FLSHCLLVTLACHHPEEFTPAVHASLDK
30		128–139	643.85(+2)	1286.70	1286.70	0.00	FFSAVSTVLTSK
31	Tig-β	1–8	952.35(+1)	952.35	952.46	−0.11	acSFLSAEEK
32		9–17	974.42(+1)	974.42	973.55	−0.13	GLVDGLWSK
33		18–30	657.75(+2)	1314.50	1314.66	−0.16	VNVDEVGGEALGR
34		31–40	637.78(+2)	1274.56	1274.73	−0.17	LLVVYPWTQR
35		41–59	1018.35(+2)	2035.70	2035.94	−0.16	FFQSFGDLSSADAIMSNAK
36		62–65	412.24(+1)	412.24	412.23	0.01	AHGK
37		67–76	540.72(+2)	1080.44	1080.56	−0.12	VLDSFSDGLK
38		77–82	717.29(+1)	717.29	717.38	−0.09	NIDDLK
39		83–87	493.22(+1)	493.22	493.28	−0.06	GAFAK
40		96–104	563.71(+2)	1126.42	1126.56	−0.14	LHVDPENFR
41		133–144	575.76(+2)	1150.52	1150.67	−0.15	VVAGVASALAHR

[a] Cysteine-containing peptides that were found to be disulfide cross-linked.

leucine residues, which could not have been assigned our mass spectrometry methodology. This peptide also contains one glutamic acid and one aspartic acid residue. Although these cannot be differentiated using the electron density maps, exploiting the high sequence homology between hemoglobins and using sequence comparisons with sequences from other feline species assigned them.

4. Discussion

Hemoglobin has been a topic of research for many decades, primarily focused on the analysis of normal and clinical variants of human hemoglobin. Due to its ease of crystallization and availability in large quantities, hemoglobin was extremely popular for crystallographers including the very first crystal structures by Perutz and coworkers [22]. In recent years, mass spectrometry has become a preferred method used in the sequence and variant/mutation analysis of human hemoglobin. Also, a fair number of vertebrate hemoglobin sequences are known but is limited to representative animal species across the evolutionary Tree of Life. Rapid analysis using mass spectrometry combined with structural verification provides an avenue for hemoglobin sequences from a large variety of endangered species. Snow leopard and Amur tiger hemoglobin sequences presented here are currently unavailable in any protein sequence repository or database.

Our interest is to better understand the interactions between the erythrocyte Band 3 anion channel and hemoglobin, the principle signal for cellular senescence in these anuclear red blood cells. We require both structural and sequence analysis to understand the evolution of the differential interactions between these 2 proteins. An understanding of how these protein–protein interactions are controlled by sequence and structure may aid in the comprehension of how these interactions affect/effect the aging process of the red blood cell across the evolutionary Tree of Life. Our lab has recently shown (manuscript in preparation) that although the Hb-binding domain of human band 3 binds specifically to human Hb, it does not bind to bovine Hb, which has an extremely high degree of sequence homology and structural identity. These results highlight the importance of sequence information to protein interaction studies.

Molecular weight information on the HPLC purified isolated globins (Fig. 2 and Supplementary Information) were derived from deconvolution of the ESI mass spectra of their pooled fractions. Experimentally determined molecular weights for the globin chains were determined to be 15,486, 16,028, 15,456, and 16,027 for the Snow leopard α- and β-globin and the Amur tiger α-

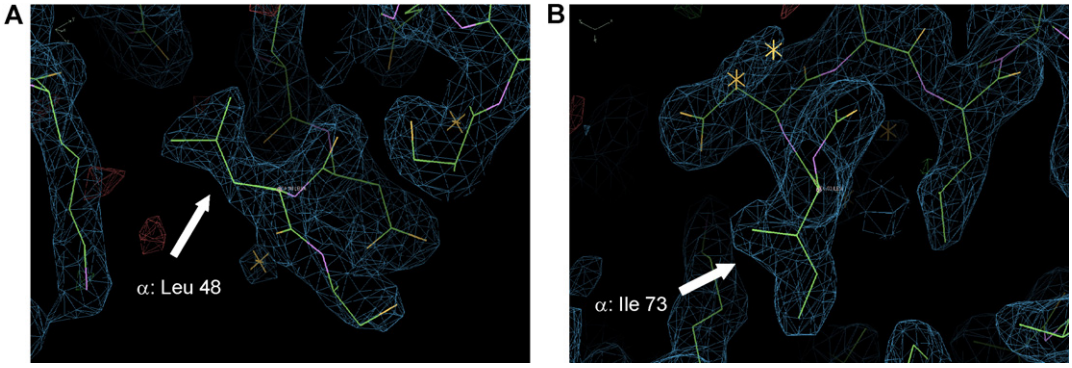

Fig. 5. Electron density maps from the Snow leopard CN-MetHb crystal structure showing the side chains of leucine 48 (A) and isoleucine 73 (B) and the facile use of these maps to differentiate between these isomeric residues.

and β-globin, respectively. Although they contain equivalent numbers of residues, the α- and β-globins of Snow leopard and Amur tiger hemoglobin had molecular weights larger than analogous hemoglobins from mammalian sources. This however is consistent with known feline sequences where substitutions with larger amino acids have occurred.

The traditional bottom-up approach to protein sequencing used in this work was achieved by trypsin proteolysis in solution followed by tandem mass spectrometry. The ESI mass spectrum of the trypsin digest of Snow leopard β-globin (Supplementary Information) shows a number of peptides whose masses (after MaxENT transformation of the entire TIC) were used in a peptide mass fingerprint Mascot search. This Mascot search facilitated both MS and MS/MS data analysis as it provided a number of sequences of potentially highly homologous globins that would result in tryptic peptides of the same mass. Tandem mass spectra of each of these peptides would in most cases confirm the sequences of these peptides. On the other hand, the de novo sequence data analysis for unique peptides, whose masses were not used in Mascot search (due to no match being found) was facilitated by exploiting the high homology of hemoglobin sequences. The representative tandem mass spectra for tryptic peptides of the Snow leopard β-globin (Figs. 3 and 4 and Supplementary Information) show series of b-type and y-type fragment ions and their associated amino acid sequence. As can be seen, in each case the y-ion series provided full coverage of the peptide sequence while not all b-ions could be identified in the tandem mass spectra. A summary of all peptides identified by mass spectrometry for each globin chain (Table 2) shows the 82%, 90%, 87% and 72% sequence coverage for the Snow leopard α-globin, β-globin and the Amur tiger α-globin, and β-globin, respectively. It is important to note that with our experimental platform, in all of the MS-derived sequence information it is impossible to differentiate between isomeric leucine and isoleucine residues. Assignments of Ile or Leu in the sequences in Table 2 are confirmed from the X-ray structure electron density maps.

The tandem mass spectra for the N-terminal peptides for the β-globin chains from both Snow leopard (Fig. 3) and Amur tiger Hb (Supplementary Information) showed the presence of N-terminal α-amino acetylation. In a recent publication, Ashiuchi et al. showed using engineered adult human hemoglobin that the presence of a serine residue at the N-terminus of the β-globin led to complete acetylation during expression and a large decrease in the strength of the tetramer-dimer allosteric interface [23]. N-terminal acetylation has been previously demonstrated in other feline β-globins [24,25]. There are very few published attempts to ascertain the physiological relevance of N-terminal acetylation in feline hemoglobin.

One of the drawbacks of using sequencing by MS is that in most cases it does not provide complete coverage of protein sequences. In the experiments presented here, peptides sequences that were not covered by mass spectrometry include Snow leopard α: residues 62–68, Amur tiger α: residues 12–16, and 32–40, and Amur tiger β: residues 88–95, and 105–132. Fig. 6A and B demonstrates the utility of using the electron density maps to identify the amino acid residue sequences in these regions. Electron density maps for these proteins allowed the assignment of all leucine and isoleucine residues. Within these regions not sequenced by MS though, due to their similarity in electron density maps it was not possible to differentiate between a small number of Asp/Asn and Glu/Gln residues: Snow leopard α 62–68 (one Asp); tiger β 88–95 (1 Glu and 1 Asp), and 105–132 (3 Gln, 1 Glu, and 2 Asn). Considering the extremely high degree of sequence homology in the hemoglobin family, these residues though could be assigned by sequence comparison with other feline hemoglobin sequences.

The use of MS sequencing and X-ray crystallography proved to be complementary. The X-ray crystal structures of Snow leopard and Amur tiger CN-MetHb were modeled using the sequence information derived from our MS experiments and molecular replacement model (with PDB ID 1G08). The electron density maps for these structures allowed the differentiation between leucine and isoleucine residues, which were indistinguishable in our MS experiments. Examples of the electron density maps for one leucine residue (Snow leopard a Leu-48) and one isoleucine residue (Snow leopard a Ile-73) are provided in Fig. 5. Based on the shapes of these electron density maps, it is quite a simple task to tell these 2 residues apart. Electron density maps allowing the assignment of 2 other leucine residues in the Amur tiger β-peptide 88–95 are shown in Fig. 6B. For crystallographic analysis, the presence of amino acid residues that have similar shapes, for example Asp/Asn and Glu/Gln, are easily distinguished in the MS sequencing allowing accurate tracing of the electron density maps. Also, in the X-ray crystal structures, the N- and C-termini of some chains exhibited high B-factors. This disorder and its consequent weak (or lack of) density at the termini resulted in the inability of X-ray crystallography to provide a sequence for these regions of the globin chains. Mass spectrometry, on the other hand, was able to provide the necessary sequence information for these disordered regions. These disordered regions were limited to the N-terminal acetylserine in both Snow leopard and Amur tiger β-globins and the 3 C-terminal residues (KYR) in the Snow leopard α-globins.

The complete sequences of the globin chains of Snow leopard and Amur tiger Hb are provided in Fig. 7. Molecular weights (average masses) calculated for the Snow leopard α-, Snow leopard β-, Amur tiger α-, and Amur tiger β-globin chains were 15,486, 16,029, 15,457, and 16,030, respectively. These are in good agreement with the molecular weights of 15,486, 16,028, 15,456, and 16,027 (in the same order) measured.

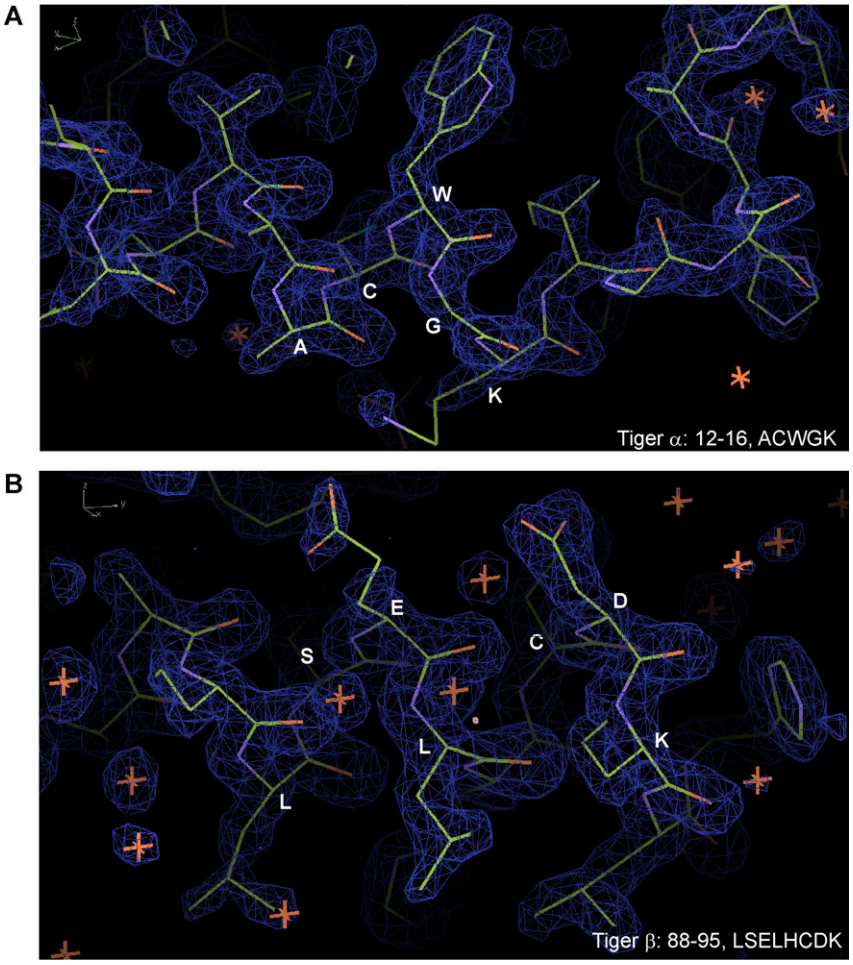

Fig. 6. Electron density maps from the Amur tiger CN-MetHb crystal structure including the peptides α: 12–16 with sequence ACWGK (A) and β: 88–95 with sequence LSELHCDK (B). These demonstrate the facile use of these maps to identify amino acid residues in the sequence. The tiger β:88–95 peptide also contains an Asp and a Glu residue which could not be distinguished from Asn and Gln, respectively using these maps.

```
Snow leopard Alpha
  1 VLSSADKNNV KACWGKIGSH AGEYGAEALE RTFCSFPTTK TYFPHFDLSH GSAQVQTHGQ
 61 KVADALTKAV AHIDDLPNAL SDLSDLHAYK LRVDPVNFKF LSHCLLVTLA CHHPEEFTPT
121 VHASLDKFFS AVSTVLTSKY R

Snow leopard Beta (N-terminus acetylated)
  1 SFLSAEEKGL VDGLWSKVNV DEVGGEALGR LLVVYPWTQR FFQSFGDLSS ADAIMSNAKV
 61 KAHGKKVLNS FSDGLKNIDD LKGAFAKLSE LHCDKLHVDP ENFRLLGNVL VCVLAHHFGH
121 EFNPQVQAAF QKVVAGVASA LAHRYH

Amur Tiger Alpha
  1 VLSSADKNNV KACWGKIGSH AGEYGAEALE RTFCSFPTTK TYFPHFDLSH GSAQVQTHGQ
 61 KVADALTKAV AHIDDLPNAL SDLSDLHAYK LRVDPVNFKF LSHCLLVTLA CHHPEEFTPA
121 VHASLDKFFS AVSTVLTSKY R

Amur Tiger Beta (N-terminus acetylated)
  1 SFLSAEEKGL VDGLWSKVNV DEVGGEALGR LLVVYPWTQR FFQSFGDLSS ADAIMSNAKV
 61 KAHGKKVLDS FSDGLKNIDD LKGAFAKLSE LHCDKLHVDP ENFRLLGNVL VCVLAHHFGH
121 EFNPQVQAAF QKVVAGVASA LAHRYH
```

Fig. 7. Complete sequences of Snow leopard and Amur tiger hemoglobin α-globin and β-globin chains.

5. Conclusions

We present here the complete sequences of hemoglobin from 2 felines: the Snow leopard and the Amur tiger. This complete coverage of the protein sequences was only possible through exploitation of the synergy between tandem mass spectrometry and X-ray crystallography. Using mass spectrometry and sequence homology we were able to determine 82%, 90%, 87% and 72% for the Snow leopard α-globin, β-globin and the Amur tiger α-globin, and β-globin, respectively. Though incomplete, this sequence information was necessary to the solution of the protein structures from X-ray diffraction patterns and invaluable to differentiating between

residues that have similar structures in crystallography (for example Glu and Gln, Asp and Asn, and Thr and Val). Blanks in the protein sequences, which resulted from peptides that were too short, too long, or undetected in MS were all filled using electron density maps in the crystal structure. Only through the combination of these two very different analytical methods were we able to determine the complete primary sequences of these proteins. This strategy can be applied to many other proteins and noncovalent protein complexes.

Acknowledgements

This work was supported by start-up funds to Wendell P. Griffith from the University of Toledo, and by a grant from the deArce Memorial Endowment Fund. *U. uncia* (Snow leopard) and *P. tigris altaica* (Amur tiger) packed red blood cells were generously donated by the Toledo Zoo. We acknowledge the NSF MRI Program (Award #0923184, PI: WPG) for funding the acquisition of the MALDI-TOF/TOF mass spectrometer. We would also like to thank Dr. Bryant Leif Hanson for assistance with the use of the X-ray crystallography equipment as well as useful advice and discussions. We are grateful to the University of Toledo Arts and Sciences Instrumentation Center, where all X-ray diffraction and MALDI-TOF/TOF-MS data were collected.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ijms.2011.06.001.

References

[1] P.S. Low, Interation of native and denatured hemoglobins with band 3: consequences for erythrocyte structure and function , in: P. Agre, J.C. Parker (Eds.), Red Blood Cell Membranes: Structure, Function, Clinical Implications, Marcel Dekker Inc., NY, 1989, pp. 237–260.
[2] M.R. Clark, Senescence of red blood cells: progress and problems , Physiol. Rev. 68 (2) (1988) 503–554.
[3] M.R. Clark, S.B. Shohet, Red cell senescence , Clin. Haematol. 14 (1) (1985) 223–257.
[4] M.P. Rettig, P.S. Low, J.A. Gimm, N. Mohandas, J. Wang, J.A. Christian, Evaluation of biochemical changes during in vivo erythrocyte senescence in the dog , Blood 93 (1) (1999) 376–384.
[5] P.S. Low, Role of hemoglobin denaturation and band 3 clustering in initiating red cell removal , Adv. Exp. Med. Biol. 307 (1991) 173–183.
[6] P.S. Low, S.M. Waugh, K. Zinke, D. Drenckhahn, The role of hemoglobin denaturation and band-3 clustering in red blood-cell aging , Science 227 (4686) (1985) 531–533.
[7] M.M.B. Kay, Generation of senescent cell antigen on old cells initiates igg binding to a neoantigen , Cell. Mol. Biol. 39 (2) (1993) 131–153.
[8] H.U. Lutz, Innate immune and non-immune mediators of erythrocyte clearance , Cell. Mol. Biol. 50 (2) (2004) 107–116.
[9] C.G. Koch, L. Li, D.I. Sessler, P. Figueroa, G.A. Hoeltge, T. Mihaljevic, E.H. Blackstone, Duration of red-cell storage and complications after cardiac surgery , N. Engl. J. Med. 358 (12) (2008) 1229–1239.
[10] S.S. Wang, What's the shelf life of blood? Focus on whether older donations impair recovery of transfusion recipients , Wall Street J. (2009).
[11] H. Steen, M. Mann, The ABC's (and XYZ's) of peptide sequencing , Nat. Rev. Mol. Cell Biol. 5 (9) (2004) 699–711.
[12] D.F. Hunt, J.R. Yates 3rd, J. Shabanowitz, S. Winston, C.R. Hauer, Protein sequencing by tandem mass-spectrometry , Proc. Natl. Acad. Sci. U.S.A. 83 (17) (1986) 6233–6237.
[13] P. Edman, Method for determination of the amino acid sequence in peptides , Acta Chem. Scand. 4 (1950) 283–293.
[14] H.D. Niall, J.W. Jacobs, J. van Rietschoten, G.W. Tregear, Protected edman degradation – new approach to microsequence analysis of proteins , FEBS Lett. 41 (1) (1974) 62–64.
[15] K.A. Walsh, L.H. Ericsson, D.C. Parmelee, K. Titani, Advances in protein sequencing , Ann. Rev. Biochem. 50 (1981) 261–284.
[16] Y. Wada, Advanced analytical methods for hemoglobin variants , J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 781 (1–2) (2002) 291–301.
[17] W.P. Griffith, A. I.Kaltashov, Mass spectrometry in the study of hemoglobin: from covalent structure to higher order assembly , Curr. Org. Chem. 10 (2006) 535–553.
[18] W.P. Griffith, I.A. Kaltashov, Phenotypic analysis of Atlantic cod hemoglobin chains using a combination of top-down and bottom-up mass spectrometric approaches , Int. J. Mass Spectrom. 278 (2008) 114–121.
[19] H. Diemer, M. Elias, F. Renault, D. Rochu, C. Contreras-Martel, C. Schaeffer, A. Van Dorsselaer, E. Chabriere, Tandem use of X-ray crystallography and mass spectrometry to obtain ab initio the complete and exact amino acids sequence of HPBP, a human 3B-kDa apolipoprotein , Proteins 71 (4) (2008) 1708–1720.
[20] T.C. Mueser, P.H. Rogers, A. Arnone, Interface sliding as illustrated by the multiple quaternary structures of liganded hemoglobin , Biochemistry 39 (50) (2000) 15353–15364.
[21] A. Riggs, Preparation of blood hemoglobins of vertebrates , Methods Enzymol. 76 (1981) 5–29.
[22] J. Boyes-Watson, E. Davidson, M.F.R. Perutz, An X-ray study of horse methaemoglobin , Proc. R. Soc. Med. 191 (1024) (1947) 83–132.
[23] M. Ashiuchi, T. Yagami, R.J. Willey, J.C. Padovan, B.T. Chait, A. Popowicz, L.R. Manning, J.M. Manning, N-terminal acetylation and protonation of individual hemoglobin subunits: Position-dependent effects on tetramer strength and cooperativity , Protein Sci. 14 (6) (2005) 1458–1471.
[24] N. Spencer, Genetics of cat hemoglobins – quantitative polymorphism , Biochem. Genet. 17 (7–8) (1979) 747–756.
[25] H.P. Driessen, W.W. de Jong, G.I. Tesser, H. Bloemendal, The mechanism of n-terminal acetylation of proteins , CRC Crit. Rev. Biochem. 18 (4) (1985) 281–325.