

## Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*

Mann and Jackson

RESEARCH ARTICLE

Open Access

# Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*

Karlheinz Mann<sup>1</sup> and Daniel John Jackson<sup>2\*</sup>

## Abstract

**Background:** With a diversity of pigmented shell morphotypes governed by Mendelian patterns of inheritance, the common grove snail, *Cepaea nemoralis*, has served as a model for evolutionary biologists and population geneticists for decades. Surprisingly, the molecular mechanisms by which *C. nemoralis* generates this pigmented shelled diversity, and the degree of evolutionary conservation present between molluscan shell-forming proteomes, remain unknown.

**Results:** Here, using next generation sequencing and high throughput proteomics, we identify and characterize the major proteinaceous components of the *C. nemoralis* shell, the first shell-proteome for a pulmonate mollusc. The recent availability of several marine molluscan shell-proteomes, and the dataset we report here, allow us to identify 59 evolutionarily conserved and novel shell-forming proteins. While the *C. nemoralis* dataset is dominated by proteins that share little to no similarity with proteins in public databases, almost half of it shares similarity with proteins present in other molluscan shells. In addition, we could not find any indication that a protein (or class of proteins) is directly associated with shell pigmentation in *C. nemoralis*. This is in contrast to the only other partially characterized molluscan-shell pigmentation mechanism employed by the tropical abalone *Haliotis asinina*.

**Conclusions:** The unique pulmonate shell-forming proteome that we report here reveals an abundance of both mollusc-specific and pulmonate-specific proteins, suggesting that novel coding sequences, and/or the extensive divergence of these sequences from ancestral sequences, supported the innovation of new shell types within the Conchifera. In addition, we report here the first evidence that molluscs use independently evolved mechanisms to pigment their shells. This proteome provides a solid foundation from which further studies aimed at the functional characterization of these shell-forming proteins can be conducted.

**Keywords:** Biomineralization, Calcification, Mollusc, Pulmonate, Pigment, Shell, Protein, Evolution, *Cepaea nemoralis*

## Background

The evolutionary origins, mode of construction, patterning, and physical properties of the molluscan shell have held the attention of scientists for centuries. However, the molecular mechanisms by which these structures are constructed are only beginning to be elucidated [1-3]. The molluscan shell is assembled extracellularly and is an ensemble of CaCO<sub>3</sub> and organic macromolecules (proteins, pigments, glycoproteins, lipids and polysaccharides)

which are secreted by an organ known as the mantle. The anterior edge of the mantle underlies the lip of the shell and directs the ordered biomineralization of the different structural layers of the shell and also controls the deposition of pigment features. With advances in nucleic acid sequencing technologies and proteomic methods, the close to complete shell-forming proteomes of several molluscs have now been reported [4-7]. Several proteins from these collections have been more fully characterized [8-11]. However, the vast majority of these previous studies are focused on marine species. While Pavat et al. [12] recently reported the biochemical properties of the shell forming proteome of the pulmonate *Helix aspersa maxima*, the lack of any transcriptome or genome data for this

\* Correspondence: [djackso@uni-goettingen.de](mailto:djackso@uni-goettingen.de)

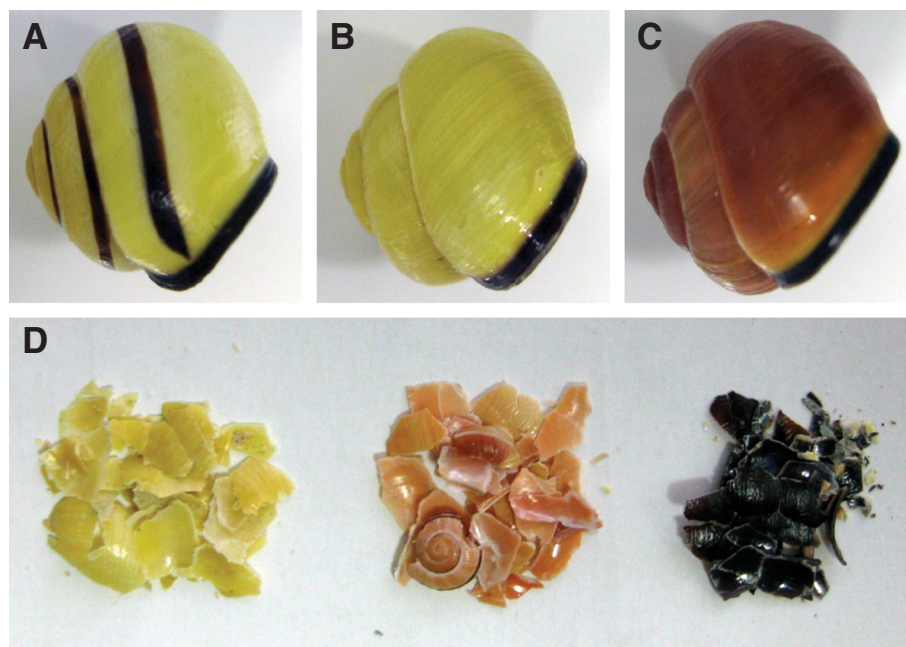
<sup>2</sup>Courant Research Centre Geobiology, Georg-August University of Göttingen, Goldschmidtstrasse 3, 37077 Göttingen, Germany

Full list of author information is available at the end of the article

species limited their proteomic analyses. They were able to characterize nine distinct 2D spots, of which seven returned a total of 14 peptides ranging in length from 4 – 11 residues. A full proteome-scale dataset from a pulmonate gastropod would efficiently highlight the conserved and lineage specific molecular mechanisms of molluscan shell formation, and would provide deep insight into how these proteomes evolve. This is because marine and terrestrial shell-forming molluscs have adapted to significantly different environments that would fundamentally affect both the process of shell formation, and the stability of the secreted composite biomineral; e.g. the abundance and biological availability of calcium, environmental pH, temperature, UV radiation, humidity and so on. While proteins involved in the process of shell formation in different molluscan lineages could be expected to have evolved in response to these different selective pressures, the signature of an ancestral shell-forming program may still be recognizable.

The common grove snail, *Cepaea nemoralis* (Figure 1), has long been studied by ecologists and population geneticists [13-15]. Key insights were gained during the 1950's and 1960's when it was demonstrated that variation in pigmented shell traits of *Cepaea* are inherited in a Mendelian fashion [16,17]. Furthermore, the frequencies of these morphotypes have been suggested to be influenced by two agents of natural selection: predation by birds [18,19], and climatic conditions [20]. Despite

this long history of research concerning the variable pigmentation of the *Cepaea* shell, there is only one study that has aimed to specifically identify the genes that control this morphological variety [21]. Using RAD-Seq (Restriction Site Associated-Sequencing) Richards et al. recently identified 44 anonymous markers putatively linked to loci that control the shell ground color and the presence or absence of dark brown bands on the *Cepaea* shell. Yet despite the association of carotenoids, porphyrins, carbohydrates and polyenes with some molluscan shell pigments [22-25], there exists no example of a complete molecular understanding of any shell pigmentation mechanism in any mollusc. In one case, the molecular basis of a molluscan shell pigment been partially elucidated. One of us previously demonstrated that the protein Sometsuke is directly associated with the blue and red pigmentation of the juvenile shell of the tropical abalone *Haliotis asinina* [4]. This previous finding motivated us to search for proteins associated with the various pigments within the *C. nemoralis* shell using high throughput transcriptomic and proteomic methods. This effort has allowed us to assemble a dataset that is likely to represent the majority of the shell forming proteome of *C. nemoralis*. This in-depth proteome is the first to be reported from a pulmonate gastropod, and also allows us to conduct comparisons between it and several others recently reported shell-forming proteomes from marine species.



**Figure 1** Representative polymorphic shells of *C. nemoralis* surveyed for their protein contents. **A-C**. Examples of the three main shell types we surveyed for both shell-forming proteins and protein-associated pigments **D**. In order to identify pigment-associated proteins *C. nemoralis* shells were crushed and divided into one of three pigmented fractions for subsequent proteomic analyses. With this approach, differentially localized proteins would be visible on LDS-PAGE gels.

## Results and discussion

### General character of the *C. nemoralis* shell-forming proteome

We were able to retrieve a total of 553 proteins/protein groups from the shell proteome of *C. nemoralis* using two different preparative techniques (including or excluding a sodium hypochlorite-plus-sonication pretreatment step). Shell material that had not been washed yielded 418 proteins, while shell material washed with sodium hypochlorite and sonication yielded 525 proteins. 382 proteins were present in both of these datasets. A list of accepted identifications is provided in Additional file 1. The MaxQuant output files containing relevant parameters, such as scores, sequence coverage and number of peptides are provided for proteins/protein groups in Additional file 2 (matrix extracted without hypochlorite pre-treatment) and Additional file 3 (with hypochlorite pre-treatment). Additional file 4 (without hypochlorite treatment) and Additional file 5 (with hypochlorite treatment) contain the MaxQuant output files for peptides. All MaxQuant output files also contain proteins rejected after manual validation of the results, and are therefore not included in the list of accepted identifications in Additional file 1. An iBAQ estimate of protein abundance suggests that 59 proteins/protein groups constitute more than 93% of the total *Cepaea* shell proteome (Table 1). This is a conservative collection of proteins and peptides that met or exceeded stringent bioinformatic and statistical criteria (see Methods). With further work (for example more complete *Cepaea* transcriptomic and/or genomic datasets against which to query the MS data) this estimate will likely increase. It has been reported for corals that in order to reduce the presence of contaminating non-biomineralizing proteins, an extensive sodium hypochlorite cleaning step must be performed on the finely powdered coral biomineral [26]. If this is not done then abundant cytoskeletal proteins such as actins, tubulins and myosins (which are unlikely to be directly involved in biomineralization) will carry through into the final biomineralization dataset [27]. While we observed subtle differences in the LDS-PAGE profiles of proteins derived from *C. nemoralis* shell material prepared with and without a hypochlorite pre-treatment (Figure 2), the proteomic data generated by the two methods were not fundamentally different.

The overall composition of the *C. nemoralis* shell proteome is dominated by uncharacterized and/or novel proteins (Table 1). Indeed the four most abundant proteins of the identifiable *C. nemoralis* proteome (constituting >52% of the total shell proteome) did not share significant similarity with any entries in UniProt (Table 1). Furthermore, 31 out of the 59 distinct proteins (52.5% of all proteins which, by abundance, account for 80% of the identifiable proteome) did not return hits against UniProt. This largely unique proteome reflects the situation for the majority of previously reported molluscan shell forming proteomes,

and highlights the need for the development of reliable and repeatable *in vivo* functional assays in these systems. The most abundant protein in the *C. nemoralis* shell, accounting for more than 25% of the total identifiable proteinaceous material, was isotig\_123 (Table 1). This Gly- and Pro-rich sequence did not share any BLASTp similarity with proteins in UniProt or Refseq, and no domains could be identified by Pfam or HMM searches. The two most abundant *C. nemoralis* shell proteins to share similarity with other proteins were isotig\_2668 and isotig\_821 (Table 1). *C. nemoralis* isotig\_2668 (accounting for 3.87% of the shell proteome) shares significant similarity with a human poly-domain protein named SEL-OB/SVEP1. This human protein is present in osteogenic tissues and plays roles in cell adhesion [28]. Isotig\_821 possesses a chitin-binding peritrophin-A domain, and shares similarity with several other molluscan proteins including BSMP (Blue Mussel Shell Protein) from *P. vulgata* and a large multi-domain containing protein from *C. gigas* (Sushi, von Willebrand, EGF and chitin-binding domains). Conspicuous in their abundance were several other *C. nemoralis* proteins that also possessed other recognizable chitin-interacting domains (14003, 1323, 101824, 84589, 63304, 248122 and 31170), suggesting that chitin is an important organic component of the *C. nemoralis* shell.

Many of the proteins identified in our proteomic analysis possess unusually high proportions of certain amino acids. These types of proteins are often found in molluscan biominerals [5,29,30]. For example, Lustrin (a protein isolated from the nacreous layer of the Californian red abalone *Haliotis rufescens*) contains a 272-residue domain rich in Gly (31%) and Ser (61%) residues. This domain has been suggested to act as an extensor molecule and to impart fracture resistance to the abalone shell [10]. However, it must be pointed out that this function is yet to be experimentally verified. While we cannot assign functions to any of the various *C. nemoralis* proteins that contain domains rich in amino acids such as Gln, Asp, Pro, Gly and Met, their abundance and diversity in this dataset hints at the important role they must play in shell formation. Interestingly, one of these proteins (isotig\_7807), which consists of 13% Ala and 11% Gln, also contains a Whey Acidic Protein (WAP domain), which is also present in the abalone Lustrin proteins [10,31]. WAP domains are thought to possess protease inhibitor activity due to the presence of 4-disulphide core (4-DSC) residues, which serine protease inhibitors also possess [32]. A diversity of protease inhibitor domain-containing proteins has been observed in other molluscan shell forming proteomes [6,29,33]. The presence of protease inhibitors in an external structure such as the molluscan shell may provide the shell with an ability to resist the digestive enzymes secreted by fouling organisms and predators that would dissolve and bore through the shell, for

**Table 1 The major proteins and peptides of the *C. nemoralis* shell: 59 proteins and peptides (with an iBAQ percentage of more than 0.1) constitute 93% of the identifiable *C. nemoralis* shell proteome**

Isotig	Similarity to (best match)	E-value	Identity	Protein features <sup>1</sup>	Fraction	Unique + razor peptides	% of total (iBAQ)
123	None	-	-	(12% G, 18% P)	S > I	15	26.22
1152	None	-	-	(16% G, 15% M, 10% S); SP	S > I	3	15.55
7508	None	-	-	SP	S < I	15	6.44
1265	None	-	-	(12% G)	S > I	9	4.40
2668	Q4LDE5	2.0e-8	32%	Domains: Sushi/SCR/CCP; (10% A, 13% G, 10% S); SP	S > I	13	3.87
	<i>Homo sapiens</i>						
21112	None	-	-	(18% P, 10% N)	S > I	8	3.41
14344	None	-	-	(17% Q, 10% L, 17% P)	S < I	11	2.96
821	J7Q5J6	8.7e-11	29.9%	Similar to BSMP; Domains: CBM_14/ CHIT_BIND_II	S < I	15	2.45
	<i>Patella vulgata</i>						
4164	None	-	-	Similarity to UP2_HALAI (e-value 0.19; 28.1% identity)	S > I	3	2.07
58150	K1QZ49	1.2e-40	35.1%	Similar to adipocyte plasma membrane-associated protein; Domains: TolB_like/ strictosidine synthase; see also contig_221 and contig_16710; SP	S < I	16	1.98
	<i>Crassostrea gigas</i>						
5087	None	-	-	TM	S < I	39	1.60
35852	None	-	-	(11% A, 11% P)	S > I	14	1.46
12941	None	-	-	(18% D, 14% L)	S < I	12	1.37
10584	None	-	-	(16% D, 10% E)	S < I	14	1.35
2108	None	-	-	(10% A, 10% Q)	S < I	8	1.32
1850	None	-	-	(22% P); SP	S > I	15	1.31
201	None	-	-	(11% N, 10% Q); SP	S < I	11	1.20
1504	None	-	-	(16% G, 12% L, 21% M, 10% S in 58AA); SP	S > I	1	1.15
16878	None	-	-	(29% Q, 16% P)	S > I	3	1.08
2744	None	-	-	(10% K, 10% S)	S < I	14	0.90
551	None	-	-	(Fragment of 67AA containing 18% G, 22% M, 12% P)	S > I	3	0.85
7809	None	-	-	(21% D, 16% L)	S < I	17	0.83
25891	J7QJT8	6.3e-25	34.7%	Domains: αCA; SP	S < I	42	0.59
	<i>Patella vulgata</i>						
3938	None	-	-	SP	S > I	2	0.58
1188 + 4282	K1P9P0	1.2e-27	47.0%	Mesenchyme-specific cell surface glycoprotein; Domains:WD40/YVTN repeat-like	S < I	15	0.53
	<i>Crassostrea gigas</i>	4.4e-21	50.9%				
7563	None	-	-	-	S, I	6	0.51
6176	None	-	-	(11% A, 11% G, 12% M)	S > I	7	0.49

**Table 1 The major proteins and peptides of the *C. nemoralis* shell: 59 proteins and peptides (with an iBAQ percentage of more than 0.1) constitute 93% of the identifiable *C. nemoralis* shell proteome (Continued)**

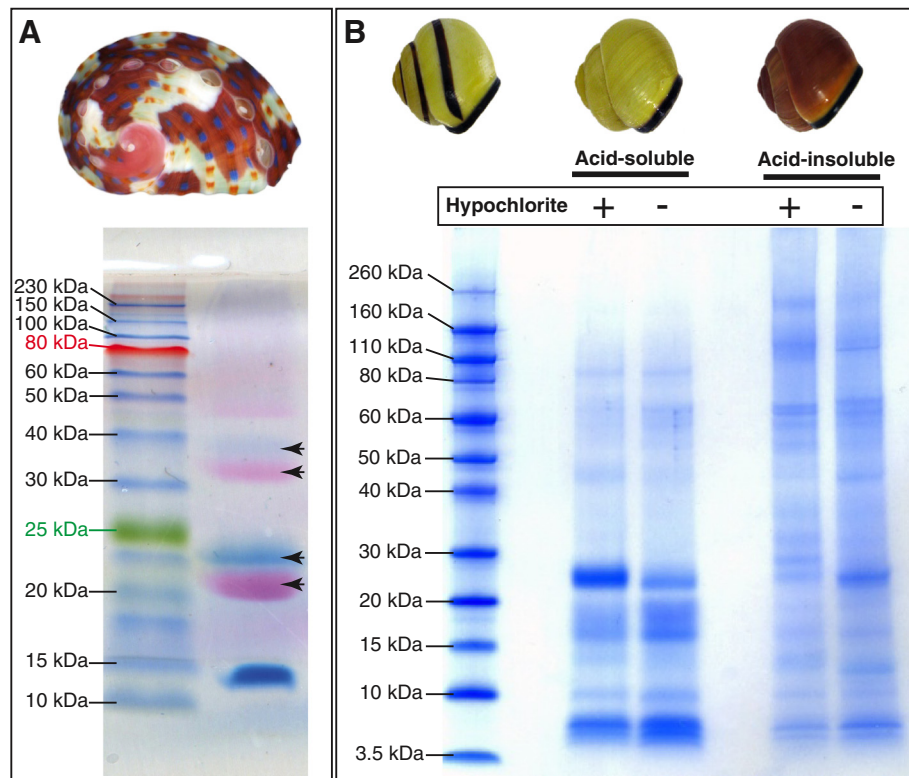
<b>1604</b>	None	-	-	(31% Q, 23% P)	S > I	2	0.48
<b>269</b>	None	-	-	(18% G)	S < I	8	0.44
<b>14003</b>	K1PRD3 <i>Crassostrea gigas</i>	3.6e-22	30.1%	Similar to IgGFc-binding protein; Domains: CBM_14/CHIT_BIND_II; shares peptide with 101824	S, I	18	0.38
<b>1647</b>	None	-	-	Domains:Sushi/SCR/CCP; (12% P, 11% S, 11% T)	S > I	9	0.38
<b>450</b>	68CYM6 <i>Physella acuta</i>	9.9e-67	74.2%	G-type lysozyme;SP	S < I	5	0.36
<b>1323</b>	A7T0W4 <i>Nematostella vectensis</i>	1.3e-26	40.5%	Domains: polysaccharide deacetylase/chitinase	S < I	7	0.32
<b>101824</b>	K1QJK2 <i>Crassostrea gigas</i>	2.1e-22	29.4%	Domains: CBM_14/ CHIT_BIND_II	S < I	29	0.31
<b>84589</b>	K1QIK2 <i>Crassostrea gigas</i>	5.4e-41	29.2%	Domains: CBM_14/ CHIT_BIND_II; also see contig_101824	S < I	29	0.30
<b>132</b>	None	-	-	(12% Q)	S < I	21	0.29
<b>28994*</b>	K1QPM9 <i>Crassostrea gigas</i>	1.5e-22	45.6%	Similar to fatty acid-binding protein, brain	S, I	7	0.24
<b>32297</b>	None	-	-	(14% G, 10% P); SP	S > I	2	0.23
<b>74063</b>	None	-	-	(10% G, 13% L, 20% P); SP	S, I	1	0.22
<b>263</b>	None	-	-	-	S < I	1	0.21
<b>63304</b>	K1Q365 <i>Crassostrea gigas</i>	9.6e-55	29.8%	AA 127-818 similar to lactadherin; Domains: CBM_14/CHIT_BIND_II	S < I	26	0.21
<b>227</b>	D3BGG3 <i>Polysphondylium pallidum</i>	8.1e-5	23.9%	Similar to Zipper-like Domains-containing protein	S < I	14	0.20
<b>691</b>	A5Z1D6 <i>Cernuella virgata</i>	1.4e-89	43.3%	Similar to epiphragmin; Domains: Fibr_C; SP	S < I	31	0.20
<b>2357</b>	None	-	-	AA 395-528 similar to Domains: LDL_recept_a; (15% T); SP	S, I	12	0.19
<b>572*</b>	IFEA <i>Helix aspersa</i>	1.7e-118	96.9%	Non-neuronal cytoplasmic intermediate filament protein	S < I	32	0.19
<b>7323</b>	A7RQD5 <i>Nematostella vectensis</i>	6.2e-41	32.2%	SP	S < I	25	0.19
<b>3883</b>	K1Q9V3 <i>Crassostrea gigas</i>	5.9e-190	86.2%	V-type proton ATPase catalytic subunit A	S < I	27	0.17
<b>943*</b>	G0ZGZ8	5.2e-38	97.8%	Actin	S < I	3	0.16

**Table 1 The major proteins and peptides of the *C. nemoralis* shell: 59 proteins and peptides (with an iBAQ percentage of more than 0.1) constitute 93% of the identifiable *C. nemoralis* shell proteome (Continued)**

<i>Metaphire posthuma</i>							
<b>169764</b>	R7VB66	3.0e-31	54.5%		SP S < I	6	0.16
<i>Capitella telata</i>							
<b>104312</b>	R7TB34	2.4e-5	45.5%	Similar to α-carbonic anhydrase; Domains: α-CA_2	S,I	4	0.14
<i>Capitella telata</i>							
<b>196388</b>	None	-	-	(12% A, 11% G, 11% P, 14% S);	S < I	1	0.14
<b>2858</b>	H2ZUY5	5.5e-41	44.6%	Similar to adipocyte plasma membrane-associated protein; Domains: TolB_like/strictosidine:synthase_related	S < I	11	0.14
<i>Latimeria chalumnae</i>							
<b>248122</b>	C3Z1I6	6.1e-15	38.8%	Similar to chitinase; Domains: glyco_hydro_18/chitinase	S < I	4	0.13
<i>Branchiostoma floridae</i>							
<b>7807</b>	None	-	-	Domains: WAP; (12% A, 11% Q); SP	S < I	15	0.12
<b>1237</b>	K1QI28	1.9e-191	89.8%	V-type proton ATPase subunit B	S < I	18	0.11
<i>Crassostrea gigas</i>							
<b>20308</b>	Q2LZN0	1.6e-9	27.0%	AA 49–467 similar to Dpse/GA10422/alkaline phosphatase; Domains: alkaline phosphatase; SP	S < I	14	0.11
<i>Drosophila pseudoobscura</i>							
<b>20360</b>	H9K6W1	3.9e-37	37.5%	Similar to cadherin; Domains: cadherin; see also contig_75801	S < I	8	0.11
<i>Apis mellifera</i>							
<b>31170</b>	K7S1Q8	3.2e-37	35.0%	Domains: CMB_14/CHIT_BIND_II; SP	S < I	17	0.11
<i>Propionibacterium acidipropionici</i>							

All proteins listed here were found in all pigment fractions (yellow, orange and dark brown) with the exception of isotig\_169764 which was only found in yellow and orange fractions. Proteins and peptides are listed in order of abundance expressed as a percentage of the shell total proteome that we can identify.

<sup>†</sup>The percentages of particularly abundant amino acid residues are given in brackets. "SP" indicates the presence of a signal peptide, "TM" indicates a likely trans-membrane protein.



**Figure 2** A comparison of the protein-associated pigments and PAGE profiles of proteins isolated from the shells of *H. asinina* and *C. nemoralis*. **A.** A representative SDS-PAGE gel of proteins isolated from shells of juvenile *H. asinina*. This gel is unstained (the protein marker is pre-stained) allowing the red and blue pigment-associated proteins to be visualized (arrows). **B.** A representative LDS-PAGE analysis of *C. nemoralis* acid-soluble and -insoluble proteins extracted from shell fragments either treated with or without a sodium hypochlorite-plus-sonication pre-treatment. Protein extractions which displayed such electrophoretic patterns were subjected to FASP (Filter-Aided Sample Preparation) sample preparation and proteomic analysis.

example naticid gastropods [34] polydorid annelids [35] and sponges [36].

We also detected a neurofilament protein (isotig\_572) present in the *C. nemoralis* shell. While this may at first be considered a non-biomineral associated contaminant, a similar protein was reported from the shells of *Helix aspersa* [12], and we also note the presence of such a protein in the shell proteome of *L. gigantea* (see below). The presence of such presumably intra-cellular proteins in extra-cellular structures such as the molluscan shell are difficult to reconcile with our current understanding of how such biominerals are formed, and serve to highlight how far we are from a complete understanding of these processes. While a conventional model of shell formation would account for the presence of such proteins through the non-specific occlusion of cells and cellular debris into a growing face of a biomineral, there are alternative models that should perhaps be considered. The biophysical properties of filament proteins have been well studied, and they are known to be able to reversibly deform to several times their own length [37-39]. The

fracture resistance properties of the molluscan shell, which exceeds that of pure  $\text{CaCO}_3$  by several orders of magnitude, is imparted to the biomineral by the organic components of the shell. A non-canonical secretory pathway for filament proteins, or the specific integration of filament-rich cells into the growing shell may be a mechanism by which the shell acquires such biomechanical properties. However, such hypotheses require further experimental investigation.

A previously described molluscan shell matrix protein, dermatopontin, was suggested to be the major proteinaceous component of the shell of the freshwater snail *Biomphalaria glabrata* [40,41]. Dermatopontin was also reported from the shells of other gastropods [42], and bivalves, where it is thought to play a role in nacre formation [43], and can also be found in taxa ranging from bacteria to humans. Interestingly, we did not detect Dermatopontin in the shell of *C. nemoralis*. To investigate this further we constructed an HMM profile of molluscan Dermatopontin proteins and used a local installation of HMMsearch [44] to query this against our translated



*C. nemoralis* transcriptome. This search returned two significant hits (contig\_46837 e-value 6.8e-27; contig\_162693 e-value 5e-06). These contigs contain clear Dermato-pontin domains, with contig\_46837 also possessing a signal sequence. This discrepancy between the presence of Dermato-pontin transcripts in the *C. nemoralis* mantle transcriptome, and the absence of Dermato-pontin proteins in the shell-proteome may be a technical artifact, or a biological reality. While the diverse technical challenges of working with samples such as biominerals make the first possibility likely, the second scenario should also be considered in light of the apparent evolvability of molluscan shell forming secretomes [4,29,33]. Shell-forming genes under recent negative selection pressures could conceivably still be transcribed, but not translated or actively involved in shell formation. However, such a scenario requires additional investigation.

While a high proportion of the 59 proteins identified in the shell of *C. nemoralis* did not share any BLAST similarity with proteins in UniProt, some of them did contain domains that could be recognized by HMM searches. These are indicated in column 5 of Table 1. Many of these were hits were against “uncharacterized” domains or proteins of unknown function. In some cases trans-membrane (TM) regions could be identified. Such a finding is interesting as it leads to the question of how could such membrane-embedded proteins be located within the mature biomineral. Such a finding was also recently reported in a dataset of coral biomineralizing proteins [45] and several other studies [46,47]. In the coral study of Ramos-Silva et al. the majority of the MS peptides identified in the biomineral could only be matched against the extra-cellular regions of trans-membrane proteins, suggesting that these extra-cellular domains are specifically cleaved from the trans-membrane portions of TM proteins. We observe a similar phenomenon in our *C. nemoralis* data. Seventy-two peptides were observed in the MS data for isotig\_5087, of these 70 were located in the putative extra-cellular domain of the protein (Additional file 6). The role that trans-membrane proteins play in molluscan shell-formation has thus far received little attention.

#### **Pigmentation of the *C. nemoralis* shell is not directly associated with a proteinaceous component**

Previously, one of us described the Sometsuke protein from the shell of *Haliotis asinina* [4]. This protein is most likely coupled to a chromophore, which is involved in imparting both the red, and blue colors to the juvenile abalone shell (Figure 2A), and is perhaps the currently best understood molluscan shell-pigmentation mechanism at the genetic level. One of our primary motivations for the current work was to determine whether *C. nemoralis* also uses a protein-associated pigmentation mechanism to pattern its shell, and if so, to identify those proteins.

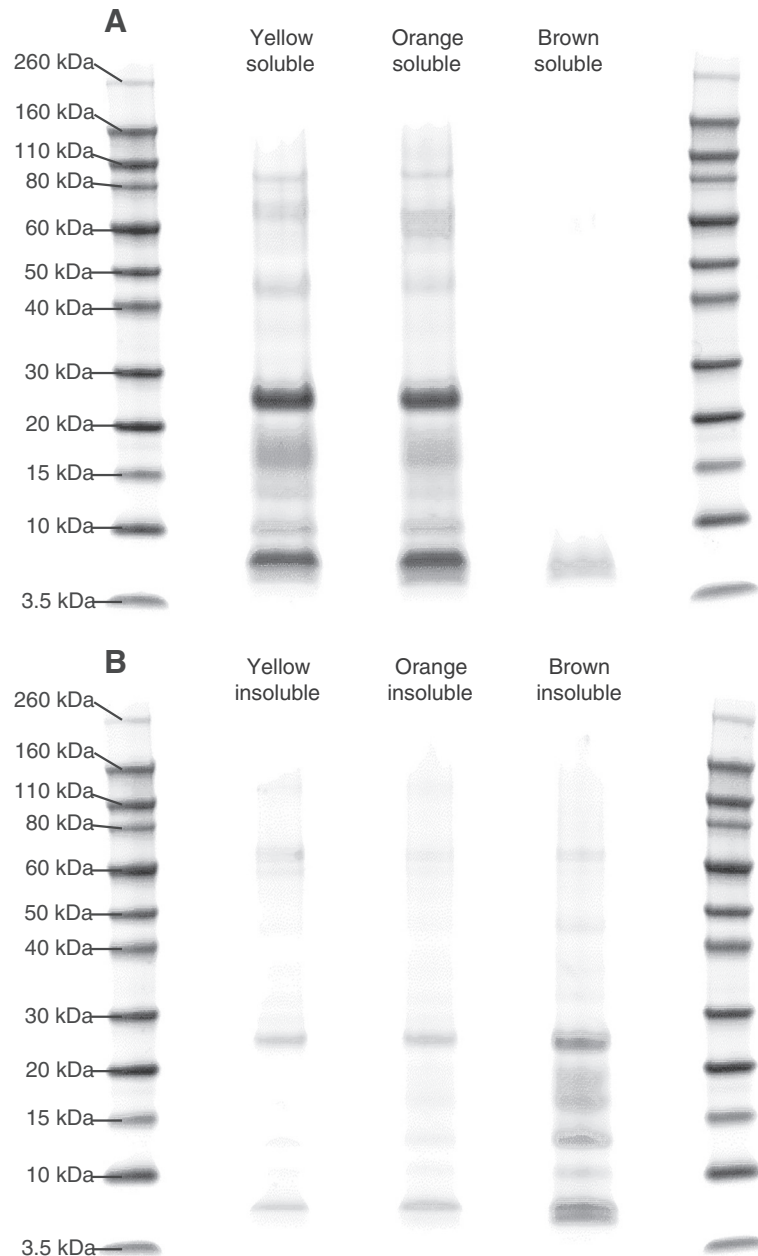
Multiple protein extractions from a variety of *C. nemoralis* shells, including protocols without the potentially destructive washing with hypochlorite, suggested that this was not the case. Pigmented LDS-PAGE bands (as per Sometsuke from *H. asinina*) were never observed (Figure 2B). A dark brown material, which consistently accompanied extractions from dark brown shell fractions, remained predominantly in the PAGE sample buffer-insoluble material that was removed by centrifugation before electrophoresis in order to obtain clear and comparable LDS-PAGE electropherograms. While it could be argued that an insoluble *C. nemoralis* pigment-associated protein would need to be rendered soluble in order for it to be visualized on a gel, we point out that the denaturation treatments applied to the *C. nemoralis* samples (70°C for 10 minutes in detergent-containing loading buffer with mercaptoethanol) were more than adequate to solubilize the water-insoluble Sometsuke protein (Figure 2A).

Unfortunately we were unable to relatively quantify the proteins associated with each of the three pigment classes using a MaxQuant-implementation of label-free quantitation (LFQ) due to the significantly different solubility behavior of the protein extracts (see below). However, a second line of evidence suggests that *C. nemoralis* shell pigments are not associated with proteins. A qualitative assessment of the 59 proteins extracted from the shell (derived from three different pigment fractions) reveals that 58 were present in all three pigment fractions. These 58 proteins (which passed our stringent quality filters) account for >93% of the identifiable proteome. If we assume that a *C. nemoralis* pigment-associated protein would be at least moderately differentially abundant between the three pigment fractions (see Figure 1D, and as is certainly the case for Sometsuke in *H. asinina*, see [4], such a protein should either be easily observable on SDS/LDS-PAGE gels, or qualitatively differentially distributed across the three pigment fractions. The only protein to be qualitatively differentially distributed across the three pigment fractions was isotig\_169764. This sequence was only detected in yellow and orange fractions, and is apparently highly conserved as it shares significant similarity with proteins in organisms ranging from bacteria (10e-23) and green algae (2e-05) to hemichordates (7e-34) and segmented worms (6e-46). Despite this conservation, there are no recognizable functional domains in the *C. nemoralis* orthologue of this protein. It was also relatively rare at just 0.16% of the total identifiable proteome, suggesting that it is unlikely to be directly involved in pigmenting the shell. Considering all of these points, our favored interpretation is that *C. nemoralis* shell pigments are not associated with proteins, and most likely have no homology with the Sometsuke pigmentation mechanism employed by *H. asinina*. If correct, this indicates that molluscan shell

pigmentation mechanisms may have multiple evolutionary origins. However this scenario requires further investigation and experimental evidence.

The LDS-PAGE profiles of acid-soluble vs. acid-insoluble proteins isolated from the three different *C. nemoralis* shell-pigment fractions differed significantly (Figure 3). Essentially, the profiles of soluble proteins derived from Yellow and Orange fractions were similar to each other, and were in general more abundant and heterogeneous

than the soluble proteins isolated from the Dark Brown fraction. In contrast, the acid-insoluble proteins isolated from the Dark Brown fraction were more abundant than those isolated from the Yellow or Orange fractions. Despite this difference, the prominent bands present in yellow-soluble, orange-soluble and brown-insoluble appear to be largely similar (Figure 3). While we cannot explain this observation, it is clear that the biochemical properties of the proteins present in the brown fraction



**Figure 3** Representative LDS-PAGE analyses of soluble and insoluble proteins derived from three different *C. nemoralis* shell-pigment fractions (yellow, orange and brown). **A.** Soluble proteins. **B.** Insoluble proteins. The most striking difference between these fractions was the abundance of the proteins present in the yellow-soluble and orange-soluble fractions relative to the brown-soluble fraction. This pattern is reversed in the acid-insoluble fractions.

are qualitatively different from those in the yellow and orange fractions. This may be the result of post-translational modifications that affect the solubility of each protein fraction, such as different degrees of cross-linking. However, this idea requires further investigation.

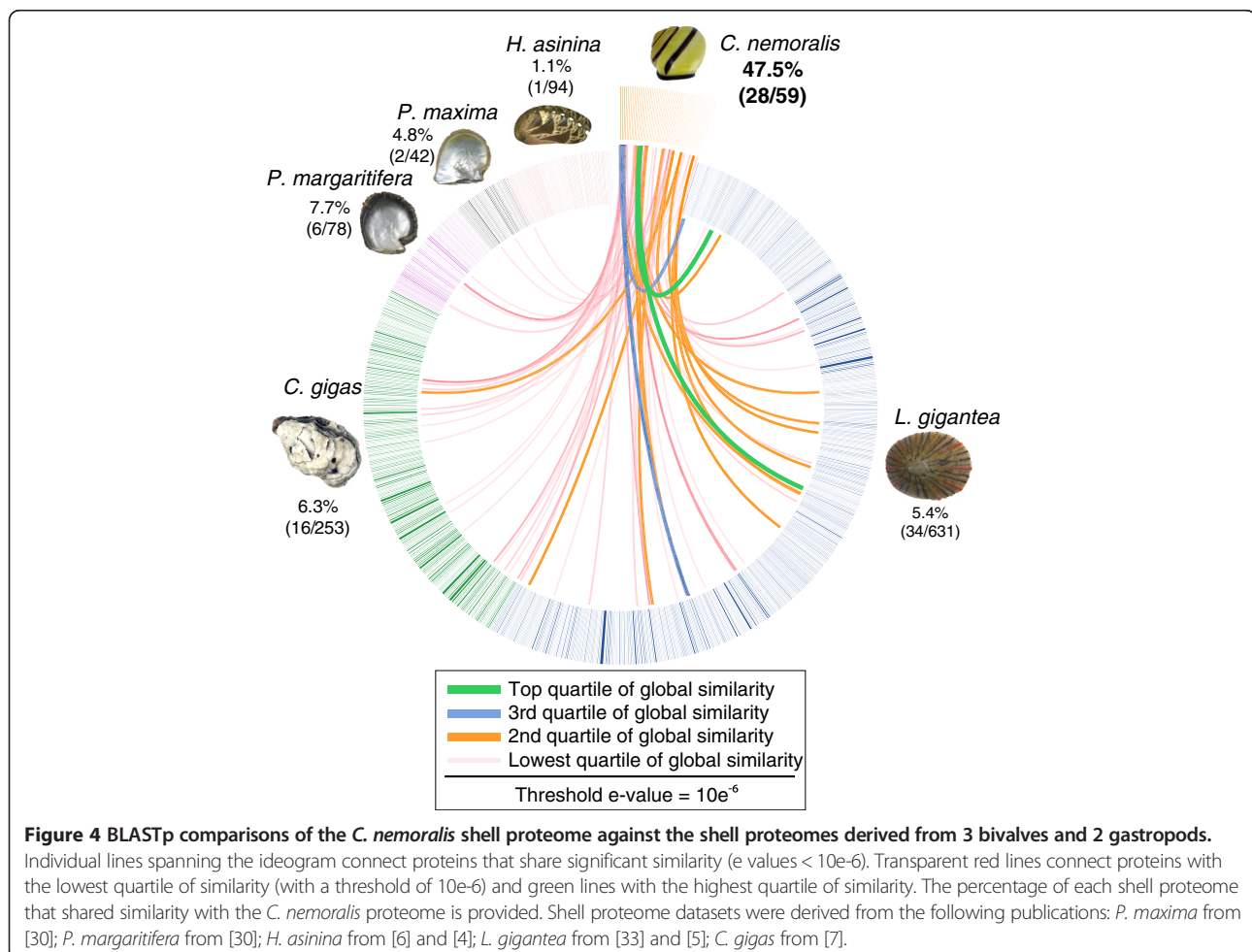
#### Comparisons of molluscan shell forming proteomes

With the *C. nemoralis* shell proteome in hand we were able to conduct broad level comparisons against five other molluscan shell forming proteomes. Importantly, all six of these datasets are not transcriptome (RNA)-based datasets but are primarily composed of proteins that have been isolated from the shells of the respective species (mapped back to either RNASeq scale mantle transcriptomes or genome assemblies), and therefore are likely to be somehow directly involved in shell formation. To our knowledge, this is the first time such a proteome level comparison of molluscan shell forming proteins has been made.

Of the 59 proteins we isolated from the *C. nemoralis* shell, 28 (47.5%) shared similarity (at an e-value threshold of  $10e^{-6}$ ) with one or more proteins derived from

the five other molluscan shell proteomes we investigated here (Figure 4). Interestingly only one *C. nemoralis* protein shared similarity with any of the 94 *H. asinina* shell forming proteins. This single protein shares no significant similarity with any proteins in public databases and contains no identifiable conserved domains. It was previously reported that the shell forming proteome of *H. asinina* is highly divergent from other such molluscan proteomes, and that this could be interpreted as evidence of a rapidly evolving shell-forming secretome [4]. Given that *C. nemoralis* and *H. asinina* share more recent common ancestry than *C. nemoralis* does with any of the three bivalves investigated here (all three of which include more proteins with similarity to the *C. nemoralis* proteome) the result we report here appears to support that hypothesis.

Several *C. nemoralis* shell proteins displayed extremely high similarity with other molluscan shell forming proteins. The four *C. nemoralis* proteins to share the highest similarity with any other species were all shared with *L. gigantea*. In order of similarity these were: isotig\_572 (a filament protein – see discussion above) at an e-value of  $4e^{-143}$



(green link in Figure 4); isotig\_84589 (a chitin-binding domain-containing protein – see below) at an e-value of  $3e-93$  (green link in Figure 4); isotig\_63304 (another chitin-binding domain-containing protein) at an e-value of  $2e-90$  (blue link in Figure 4); and isotig\_25891 (carbonic anhydrase) at an e-value of  $6e-74$  (blue link in Figure 4). In order to identify deeply conserved molluscan shell forming proteins, we ordered all *C. nemoralis* proteins that were found in any other shell proteome according to the number of databases they were found in (Additional file 7). This revealed two proteins that were found in four of the five molluscan shell proteomes: Cnem821 and Cnem248122. Cnem821 possesses a chitin-binding Peritrophin-A domain (Carbohydrate Binding Module 14: Pfam PF01607). The second protein found in four of the five molluscan shell proteomes (Cnem248122) shared significant similarity with chitinase proteins in Swissprot. These results emphasize the prominent role that chitin is likely to play in the construction of disparate molluscan shells and indeed in many metazoan biominerals [48-52].

Other conserved *C. nemoralis* shell forming proteins of interest include two carbonic anhydrase domain-containing proteins (Cnem25891 and Cnem104312) and two V-ATPase subunits (Cnem3883 and Cnem1237). V-ATPases have not previously been shown to play a role in molluscan shell formation (beyond their presence in mantle EST or RNASeq datasets), however it could be expected that their ability to transport  $H^+$  across membranes would afford them a central role in the regulation of shell formation. Indeed such proton pumps are known to play roles in the calcification of a variety of metazoan biominerals [53-55]. While one of the carbonic anhydrase domain-containing isotigs (Cnem104312) is clearly incomplete, the other (Cnem25891) is potentially complete and encodes a protein of 1,028 amino acids (the corresponding isotig contains 3,859 bp). This protein contains a signal sequence, a carbonic anhydrase domain with phylogenetic affinity to the secreted and membrane bound  $\alpha$ -CAs (see Additional file 8 for a phylogenetic analysis) and a carboxyl region of relatively low complexity (Additional file 9). The shell-forming Nacrein proteins also contain carbonic anhydrase domains and have been previously described from gastropod and bivalve shells [56,57]. Interestingly the CA-domain in the *C. nemoralis* protein is not interrupted by the low-complexity region as it is in the Nacreins, and the *C. nemoralis* low complexity domain is composed of Gln residues rather than Gly and Asn residues (Additional file 9). The hydropathy profile of all of these proteins display similar characteristics (Additional file 10), and suggests that the low complexity domains interact with the water-soluble phases of the biomineralization processes. Miyamoto et al. [58] reported that a recombinant Nacrein protein inhibited the precipitation of  $CaCO_3$  in *in vitro* calcification assays, and that removal of the repetitive

low-complexity domain attenuated this inhibitory activity. While the results of such *in vitro* calcification assays should always be interpreted with caution, this result indicates that the low complexity domains of molluscan shell forming  $\alpha$ -CAs have a significant impact on the activity of the enzyme to which they are fused or inserted. The fact that the  $\alpha$ -CA we have identified here has a significantly different domain of low-complexity from previously described shell-associated  $\alpha$ -CAs suggests that this gene may have a different evolutionary history.

## Conclusions

The 59 shell-associated proteins we report here represent the largest collection of proteins from a pulmonate shell to date. The abundant proteins in this dataset either display no similarity to known proteins, or similarity to uncharacterized proteins. Comparisons of this dataset with other molluscan shell-forming proteomes indicate that almost half of the *C. nemoralis* shell-forming proteome we describe here shares similarity with the shell-forming components of other molluscs. Two lines of experimental evidence failed to identify the presence of pigment-associated proteins in the *C. nemoralis* shell. Considering the clear association between a protein and a pigment in the juvenile *Haliothis* shell, this finding indicates that molluscan shell pigmentation mechanisms may have diverse evolutionary origins. This dataset will serve as an important platform from which further studies aimed at the characterization of pulmonate shell forming and pigmentation genes can be performed.

## Methods

### Generation of a reference *C. nemoralis* mantle transcriptome for proteomic surveys

Seven total RNA extractions derived from the distal-most edge (i.e. the shell forming region) of the mantle tissue of 4 juvenile *C. nemoralis* individuals (two extractions from each of three individuals for Illumina sequencing and one extraction from a fourth individual for 454 sequencing) was extracted using Trizol according to the manufacturer's instructions. These total RNA extractions were sequenced on the Illumina HiSeq2000 and Roche 454 platforms and assembled using the CLC Genomics workbench (version 5.5.2). This assembly generated a total of 676,358 contigs >100 bp, summing to a total of 193,892,905 bp (max contig size = 14,765 bp, median contig size = 180 bp). In order to reduce the redundancy of this assembly for proteomic interrogation (see below) the following steps were applied. First, contigs shorter than 500 bp were removed. All remaining contigs were then clustered into isotigs using "usearch" [59]. The longest possible open reading frame (which was required to be >50 amino acids) was then extracted from each isotig using standard Perl scripts. These translated ORFs were then

clustered again using “usearch” to produce a total of 55,623 putative coding translated fragments. This dataset was used to conduct the proteomic surveys (see below). All assembled nucleic acid sequences discussed in this work are available in Additional file 11.

#### Preparation of shell matrix and peptides

Shells from approximately 100 freshly collected snails were crushed into approximately 5 mm<sup>2</sup> fragments and carefully sorted into three populations: dark brown (20.5 g derived from shells with a yellow background); yellow (23.7 g); and orange (12.3 g; see Figure 1). The shell pieces were either briefly washed with deionized water, or with sodium hypochlorite (6-14% active chlorine; Merck, Darmstadt, Germany) for 2 h at room temperature with 5 min sonication and change of hypochlorite solution every 30 min. Hypochlorite-treated shells were then washed with deionized water and dried. Washed shell pieces were demineralized in 50% acetic acid (20 mL/g of shell) for 16 h at 4°C. The resulting suspensions were dialyzed successively against 3 × 10vol 10% acetic acid and 2 × 10vol 5% acetic acid at 4-6°C (Spectra/Por 6 dialysis membrane, mw cut-off 2000; Spectrum Europe, Breda, The Netherlands). The dialyzed suspensions were centrifuged for 1 h at 4°C and 12000g<sub>av</sub>. Pellets and supernatants were lyophilized and analyzed separately.

LDS-PAGE was performed with pre-cast 4-12% gradient Novex Bis-Tris gels in MES buffer using reagents and protocols supplied by the manufacturer (Invitrogen, Carlsbad, CA). The sample buffer was complemented with β-mercaptoethanol to a final concentration of 1% and samples were heated to 70°C for 10 min. Sample buffer-insoluble material was removed by centrifugation at 16000 g for 5 min. Gels were loaded with the soluble fraction at 200 µg matrix/lane.

Reduction, carbamidomethylation and enzymatic cleavage of matrix proteins were performed using a modification of the FASP (Filter-aided sample preparation) method [60]. Aliquots of 200 µg of acid-soluble or acid-insoluble shell matrix were suspended in 200 µl of 0.1 M Tris, pH8, containing 6 M guanidine hydrochloride and 0.01 M dithiothreitol (DTT). This mixture was heated to 56°C for 60 min, cooled to room temperature, and centrifuged at 13000 rpm in an Eppendorf bench top centrifuge 5415D for 15 min. The supernatant was loaded into an Amicon Ultra 0.5 ml 30 K filter device (Millipore; Tullagreen, Ireland). DTT was removed by centrifugation at 13000 rpm for 15 min and washing with 2 × 1vol of the same buffer. Carbamidomethylation was done in the device using 0.1 M Tris buffer, pH8, containing 6 M-guanidine hydrochloride and 0.05 mM iodoacetamide and incubation for 45 min in the dark. Carbamidomethylated proteins were washed with 0.05 M ammonium hydrogen carbonate buffer, pH8, containing 6 M urea, and centrifugation as before. Each sample

was then incubated with 2 µg of Lysyl endopeptidase (WAKO Chemicals GmbH, Neuss, Germany) in 40 µl of 0.05 mM ammonium hydrogen carbonate buffer containing 6 M urea for 6 h at 37°C. This was followed by addition of 4 µg of trypsin (Sequencing grade, modified; Promega, Madison, USA) in 80 µl of 0.05 M ammonium hydrogen carbonate buffer and further incubation at 37°C for 16 h. Peptides were collected by centrifugation and the filters were washed twice with 40 µl of 0.05 M ammonium hydrogen carbonate buffer. The peptide solutions were acidified to pH ~ 1 with trifluoroacetic acid and desalted for mass spectrometric analysis with C18 Stage Tips [61].

#### LC-MS analysis

Peptide mixtures were analyzed by on-line nanoflow liquid chromatography using the EASY-nLC 1000 system (Proxeon Biosystems, Odense, Denmark, now part of Thermo Fisher Scientific) with 20 cm capillary columns of an internal diameter of 75 µm filled with 1.8 µm Reprosil-Pur C18-AQ resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). Peptides were eluted with a linear gradient from 5-30% buffer B (80% acetonitrile in 0.1% formic acid) for 100 min, 30-60% B for 12 min and 80-95% B for 8 min at a flow rate of 250 nl/min. The eluate was electrosprayed into an Orbitrap Elite (Thermo Fisher Scientific, Bremen, Germany) using a Proxeon nanoelectrospray ion source. The Orbitrap Elite was operated essentially as previously described [62] in a HCD top 10 mode with dynamic selection of the 10 most intense peaks of each survey scan (300-1750Th) for fragmentation. The resolution was 120,000 for full scans and 15,000 for fragments (both specified at m/z 400). Ion target values were 1e6 and 5e4ms, respectively. Dynamic exclusion time was 30 sec.

#### Analysis of proteomic data

Raw files were processed using the Andromeda search engine-based version 1.3.9.21 of MaxQuant (<http://www.maxquant.org/>) with “second peptide” enabled, iBAQ (intensity-based absolute quantitation; [63] as implemented in recent versions of MaxQuant, and “match between runs” options (match time window 0.5 min; alignment time window 20 min) [64-66]. For protein identification the *C. nemoralis* mantle transcriptome database (see above) was converted into a FASTA-formatted protein sequence database, downloaded into MaxQuant, and automatically combined with the reversed sequences and sequences of common contaminants, such as human keratins. Carbamidomethylation was set as fixed modification. Variable modifications were oxidation (M), N-acetyl (protein), pyro-Glu/Gln (N-term) and phospho (STY). The initial mass tolerance for full scans was 7 ppm and 20 ppm for MS/MS. Two missed cleavages were allowed and the minimal length required for a peptide was seven amino acids. The peptide and protein false discovery rates (FDR) were

set to 0.01. The maximal posterior error probability (PEP), which is the individual probability of each peptide to be a false hit considering identification score and peptide length, was set to 0.01. The minimal peptide score was 50. Two sequence-unique peptides were required to occur at least three times in two different replicates for high-confidence protein identifications. In exceptional cases single-sequence-unique identifications were accepted if the peptide occurred in at least three different replicates and was identified in both hypochlorite washed samples and water washed samples.

Identifications with one or two sequence-unique peptides were routinely validated using the MaxQuant Expert System software [67] considering the assignment of major peaks, occurrence of uninterrupted  $\gamma$ - or  $\beta$ -ion series of at least four consecutive amino acids, preferred cleavages N-terminal to proline bonds, the possible presence of  $a_2/b_2$  ion pairs and immonium ions, mass accuracy and score. Based on the sum of peak intensities, the iBAQ [63] option of MaxQuant was used to calculate the approximate proportion of each protein in the total identifiable proteome.

Sequence similarity searches were performed with FASTA (<http://www.ebi.ac.uk/Tools/sss/fasta/>) [68] against current releases of the Uniprot Knowledgebase (UniProt KB). Other bioinformatics tools used were Clustal Omega for sequence alignments (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) [69], InterProScan (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>) [70] for domain predictions, and SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>) [71] for signal sequence prediction.

### Comparisons of molluscan shell forming proteomes

BLASTp based comparisons of the *C. nemoralis* shell proteome were made against five previously published molluscan shell proteomes. These included 42 *Pinctada maxima* proteins reported by Marie et al. [30], 78 *Pinctada margaritifera* proteins reported by Marie et al. [30], 253 *Crassostrea gigas* proteins reported by Zhang et al. [7], a combined set of 94 *Haliotis asinina* proteins reported by Marie et al. [6] and Jackson et al. [4], and a combined set of 631 *Lottia gigantea* proteins reported by Marie et al. [33] and Mann et al. [5]. The e-value threshold was set to  $1e-06$ . These comparisons were made using a modified version of Circoletto [72] which uses an implementation of the legacy BLAST package. The following command was issued to the circoletto.pl script: circoletto.pl -query XX -database XX -untangling\_off -e\_value  $10e-6$  -best\_hit\_type local -score2colour eval. The \*.conf files generated by circoletto.pl were modified using custom Perl scripts and then passed to Circos [73] in order to generate an ideogram. The Circoletto \*.blasted file which details all of the BLASTp results is provided as Additional file 12.

### Additional files

**Additional file 1:** A list of all accepted identifications after manual validation of all protein/protein group identifications provided by MaxQuant.

**Additional file 2:** A complete list of protein identifications in matrices extracted from shells not treated with hypochlorite plus ultra-sonication.

**Additional file 3:** A complete list of protein identifications in matrices extracted from shells treated with hypochlorite plus ultra-sonication before demineralization in acid.

**Additional file 4:** The peptide data complementing protein data of Additional file 2.

**Additional file 5:** The peptide data complementing protein data of Additional file 3.

**Additional file 6:** A schematic representation of a *C. nemoralis* putative trans-membrane protein (derived from isotig\_5087), onto which the spatial distribution of the 72 LC-MS peptides are mapped.

**Additional file 7:** Top BLASTp hits returned against *C. nemoralis* queries from five molluscan shell proteomes and Swissprot.

**Additional file 8:** A 50% majority rule consensus tree generated by Bayesian methods representing the phylogenetic relationships of metazoan CAs.

**Additional file 9:** The Nucleotide and derived protein sequence of Cnem25891.

**Additional file 10:** Hydropathy profiles of Cnem25891 and two previously reported molluscan shell-forming proteins which also possess CA domains.

**Additional file 11:** Contains the 59 isotigs (nucleic acid sequences) that were recovered by MaxQuant.

**Additional file 12:** The Circoletto generated "\*.blasted" file used to generate the Circos figure (Figure 4).

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

DJJ carried out the RNA extractions for RNA-Seq and collected and sorted *C. nemoralis* shell material for proteome work. KM carried out the protein extractions and conducted the LC-MS analyses. Both authors conducted the bioinformatic analyses, and drafted the manuscript. Both authors read and approved the final manuscript.

### Acknowledgements

We thank Matthias Mann for supporting this project, Korbinian Mayr and Igor Paron for keeping the mass spectrometers in excellent condition, and Gaby Sowa for preparing the capillary columns. Illumina sequencing was performed by Markus B Schilhabel and his team at the Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel. Juvenile *H. asinina* shells were kindly provided by Carmel McDougall and Bernie Degnan. This work was supported by Deutsche Forschungsgemeinschaft (DFG) funding to DJJ through the CRC for Geobiology and the German Excellence Initiative, and DFG project #JA2108/2-1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Author details

<sup>1</sup>Max Planck Institute for Biochemistry, Department of Proteomics and Signal Transduction, Am Klopferspitz 18, D-82152 Martinsried, Munich, Germany.

<sup>2</sup>Courant Research Centre Geobiology, Georg-August University of Göttingen, Goldschmidtstrasse 3, 37077 Göttingen, Germany.

Received: 27 January 2014 Accepted: 25 March 2014

Published: 31 March 2014

## References

- Marin F, Luquet G: Molluscan shell proteins. *Comptes Rendus Palevol* 2004, **3**:469–492.
- Wilt FH: Developmental biology meets materials science: Morphogenesis of biomineralized structures. *Dev Biol* 2005, **280**:15–25.
- Wilt FH, Killian CE, Livingston BT: Development of calcareous skeletal elements in invertebrates. *Differentiation* 2003, **71**:237–250.
- Jackson DJ, McDougall C, Green KM, Simpson F, Wörheide G, Degnan BM: A rapidly evolving secretome builds and patterns a sea shell. *BMC Biol* 2006, **4**:40.
- Mann K, Edsinger-Gonzales E, Mann M: In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*. *Proc Natl Acad Sci U S A* 2012, **109**:20986–20991.
- Marie B, Marie A, Jackson DJ, DuBost L, Degnan BM, Milet C, Marin F: Proteomic analysis of the calcified organic matrix of the tropical abalone *Haliotis asinina* shell. *Proc Natl Acad Sci U S A* 2010, **107**:854.
- Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H, Xie Y, Holland PWH, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng C, Meng J, Yang L, Liu J, Wen B, Zhang N, Huang Z, Zhu Q, Feng Y, Mount A, Hedgecock D *et al*: The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 2012, **490**:49–54.
- Amos FF, Ponce CB, Evans JS: Formation of framework nacre polypeptide supramolecular assemblies that nucleate polymorphs. *Biomacromolecules* 2011, **12**:1883–1890.
- Montagnani C, Marie B, Marin F, Belliard C, Riquet F, Tayalé A, Zanella-Cléon I, Fleury E, Gueguen Y, Piquemal D, Cochenne-Laureau N: Pmarg-pearlin is a matrix protein involved in nacre framework formation in the pearl oyster *Pinctada margaritifera*. *Chembiochem* 2011. doi:10.1002/cbic.201100216.
- Shen X, Belcher AMB, Hansma PK, Stucky GD, Morse DEM: Molecular cloning and characterization of Lustrin A, a matrix protein from shell and pearl nacre of *Haliotis rufescens*. *J Biol Chem* 1997, **272**:32472–32481.
- Suzuki M, Saruwatari K, Kogure T, Yamamoto Y, Nishimura T, Kato T, Nagasawa H: An acidic matrix protein, Pif, is a key macromolecule for nacre formation. *Science* 2009, **325**:781–788.
- Pavat C, Zanella-Cléon I, Becchi M, Medakovic D, Luquet G, Guichard N, Alcaraz G, Dommergues J-L, Serpentin A, Lebel J-M: The shell matrix of the pulmonate land snail *Helix aspersa maxima*. *Comp Biochem Physiol B* 2012, **161**:303–314.
- Fisher R, Diver C: Crossing-over in the Land Snail *Cepaea nemoralis*. *Nature* 1934, **133**:834–835.
- Lang A: Fortgesetzte Vererbungsstudien. *Mol Gen Genet* 1911, **5**:97–138.
- Lang A: Vererbungswissenschaftliche Miscellen. *Mol Gen Genet* 1912, **8**:233–283.
- Cain AJ, King J, Sheppard P: New data on the genetics of polymorphism in the snail *Cepaea nemoralis* L. *Genetics* 1960, **45**:393–411.
- Cook L: The genetics of *Cepaea nemoralis*. *Heredity* 1967, **22**:397–410.
- Cain AJ, Sheppard PM: Natural selection in *Cepaea*. *Genetics* 1954, **39**:89–116.
- Surmacki A, Ożarowska-Nowicka A, Rosin ZM: Color polymorphism in a land snail *Cepaea nemoralis* (Pulmonata: Helicidae) as viewed by potential avian predators. *Naturwissenschaften* 2013, **100**:1–8.
- Jones JS, Leith B, Rawlings P: Polymorphism in *Cepaea*: a problem with too many solutions? *Annu Rev Ecol Syst* 1977, **8**:109–143.
- Richards PM, Liu MM, Lowe N, Davey JW, Blaxter ML, Davison A: RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Mol Ecol* 2013, **22**:3077–3089.
- Comfort AC: The pigmentation of molluscan shells. *Biol Rev* 1951, **26**:285–301.
- Gysi JR, Chapman DJ: Comparative biochemistry of *Haliotis* pigmentation: unusual bilipeptides of *Haliotis cracherodii*. *Comp Biochem Physiol B* 1979, **63B**:355–361.
- Hedegaard C, Bardeau JF, Chateigner D: Molluscan shell pigments: an *in situ* resonance raman study. *J Molluscan Stud* 2006, **72**:157–162.
- Koizumi C, Nonaka J: Yellow pigments of pearl - I. carotenoid pigment in yellow nacre. *Bull Jap Soc Sci Fish* 1970, **36**:1054–1058.
- Ramos-Silva P, Marin F, Kaandorp J, Marie B: Biomineralization toolkit: the importance of sample cleaning prior to the characterization of biomineral proteomes. *Proc Natl Acad Sci U S A* 2013, **110**:E2144–E2146.
- Drake JL, Mass T, Haramaty L, Zelzion E, Bhattacharya D, Falkowski PG: Proteomic analysis of skeletal organic matrix from the stony coral *Stylophora pistillata*. *Proc Natl Acad Sci U S A* 2013, **110**:3788–3793.
- Shur I, Socher R, Hameiri M, Fried A, Benayahu D: Molecular and cellular characterization of SEL OB/SVEP1 in osteogenic cells in vivo and in vitro. *J Cell Physiol* 2006, **206**:420–427.
- Jackson DJ, McDougall C, Woodcroft B, Moase P, Rose RA, Kube M, Reinhardt R, Rokhsar DS, Montagnani C, Joubert C, Piquemal D, Degnan BM: Parallel evolution of nacre building gene sets in molluscs. *Mol Biol Evol* 2010, **27**:591–608.
- Marie B, Joubert C, Tayale A, Zanella-Cleon I, Belliard C, Piquemal D, Cochenne-Laureau N, Marin F, Gueguen Y, Montagnani C: Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. *Proc Natl Acad Sci U S A* 2012, **109**:20986–20991.
- Gaume B, Denis F, Van Wormhoudt A, Huchette S, Jackson DJ, Avignon S, Auzoux-Bordenave S: Characterisation and expression of the biomineralising gene Lustrin A during shell formation of the European abalone *Haliotis tuberculata*. *Comp Biochem Physiol B* 2014, **169**:1–8.
- Ranganathan S, Simpson KJ, Shaw DC, Nicholas KR: The whey acidic protein family: a new signature motif and three-dimensional structure by comparative modeling. *J Mol Graph Model* 1999, **17**:106–113. 134.
- Marie B, Jackson DJ, Zanella-Cléon I, Ramos-Silva P, Guichard N, Marin F: The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage specific novelties. *FEBS J* 2013, **280**:214–232.
- Kabat AR: Predatory ecology of naticid gastropods with a review of shell boring predation. *Mal Int J Malac* 1990, **32**:155–193.
- Caceres-Martinez J, Macias-Montes de Oca P, Vasquez-Yeomans R: *Polydora* sp. infestation and health of the Pacific oyster *Crassostrea gigas* cultured in Baja California, NW Mexico. *J Shell Res* 1998, **17**:259–264.
- Le Cam S, Viard F: Infestation of the invasive mollusc *Crepidula fornicata* by the native shell borer *Cliona celata*: a case of high parasite load without detrimental effects. *Biol Inv* 2011, **13**:1087–1098.
- Herrmann H, Bär H, Kreplak L, Strelkov SV, Aebi U: Intermediate filaments: from cell architecture to nanomechanics. *Nat Rev Mol Cell Biol* 2007, **8**:562–573.
- Qin Z, Kreplak L, Buehler MJ: Hierarchical structure controls nanomechanical properties of vimentin intermediate filaments. *PLoS One* 2009, **4**:e7294.
- Kreplak L, Fudge D: Biomechanical properties of intermediate filaments: from tissues to single filaments and back. *Bioessays* 2007, **29**:26–35.
- Marxen JC, Nimtz M, Becker W, Mann K: The major soluble 19.6 kDa protein of the organic shell matrix of the freshwater snail *Biomphalaria glabrata* is an N-glycosylated dermatopontin. *Biochim Biophys Acta* 2003, **1650**:92–98.
- Marxen JC, Becker W: The organic shell matrix of the freshwater snail *Biomphalaria glabrata*. *Comp Biochem Physiol B* 1997, **118**:23–33.
- Sarashina I, Yamaguchi H, Haga T, Iijima M, Chiba S, Endo K: Molecular evolution and functionally important structures of molluscan Dermatopontin: implications for the origins of molluscan shell matrix proteins. *J Mol Evol* 2006, **62**:307–318.
- Jiao Y, Wang H, Du X, Zhao X, Wang Q, Huang R, Deng Y: Dermatopontin, a shell matrix protein gene from Pearl Oyster *Pinctada Martensii*, participates in Nacre Formation. *Biochem Biophys Res Commun* 2012, **425**:679–683.
- Finn RD, Clements J, Eddy SR: HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 2011, **39**:W29–W37.
- Ramos-Silva P, Kaandorp J, Huisman L, Marie B, Zanella-Cléon I, Guichard N, Miller DJ, Marin F: The skeletal proteome of the coral *Acropora millepora*: the evolution of calcification by cooption and domain shuffling. *Mol Biol Evol* 2013.
- Mann K, Poustka AJ, Mann M: In-depth, high-accuracy proteomics of sea urchin tooth matrix. *Proc Natl Acad Sci U S A* 2008, **6**:33.
- Mann K, Poustka AJ, Mann M: The sea urchin (*Strongylocentrotus purpuratus*) test and spine proteomes. *Proc Natl Acad Sci U S A* 2008, **6**:22.
- Peters W: Occurrence of chitin in Mollusca. *Comp Biochem Physiol B* 1972, **41**:541–550.
- Weiss I, Schönitzer V, Eichner N, Sumper M: The chitin synthase involved in marine bivalve mollusk shell formation contains a myosin domain. *FEBS Lett* 2006, **580**:1846–1852.
- Ehrlich H, Kaluzhnaya OV, Tsurkan MV, Ereskovsky A, Tabachnick KR, Ilan M, Stelling A, Galli R, Petrova OV, Nekipelov SV: First report on chitinous holdfast in sponges (Porifera). *Proc R Soc B* 2013, **280**:20130339.
- Brunet PCJ, Carlisle DB: Chitin in Pogonophora. *Nature* 1958, **182**:1689.
- Falini G, Fermani S: Chitin mineralization. *Tissue Eng* 2004, **10**:1–6.

53. Takano Y, Ozawa H, Crenshaw MA: **Ca-ATPase and ALPase activities at the initial calcification sites of dentin and enamel in the rat incisor.** *Cell Tissue Res* 1986, **243**:91–99.
54. Tambutti E, Allemand D, Mueller E, Jaubert J: **A compartmental approach to the mechanism of calcification in hermatypic corals.** *J Exp Biol* 1996, **199**:1029–1041.
55. Roer RD: **Mechanisms of resorption and deposition of calcium in the Carapace of the Crab *Carcinus Maenas*.** *J Exp Biol* 1980, **88**:205–218.
56. Miyamoto H, Yano M, Miyashita T: **Similarities in the structure of nacrein, the shell-matrix protein, in a bivalve and a gastropod.** *J Molluscan Stud* 2003, **69**:87–89.
57. Miyamoto H, Miyashita T, Okushima M, Nakano S, Morita T, Matsushiro A: **A carbonic anhydrase from the nacreous layer in oyster pearls.** *Proc Natl Acad Sci U S A* 1996, **93**:9657–9660.
58. Miyamoto H, Miyoshi F, Kohno J: **The carbonic anhydrase domain protein nacrein is expressed in the epithelial cells of the mantle and acts as a negative regulator in calcification in the mollusc *Pinctada fucata*.** *Zool Sci* 2005, **22**:311–315.
59. Edgar RC: **Search and clustering orders of magnitude faster than BLAST.** *Bioinformatics* 2010, **26**:2460–2461.
60. Wisniewski JR, Zougman A, Nagaraj N, Mann M: **Universal sample preparation method for proteome analysis.** *Nat Meth* 2009, **6**:359–362.
61. Rappsilber J, Mann M, Ishihama Y: **Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips.** *Nat Protoc* 2007, **2**:1896–1906.
62. Michalski A, Damoc E, Lange O, Denisov E, Nolting D, Muller M, Viner R, Schwartz J, Remes P, Belford M, Duniyach JJ, Cox J, Horning S, Mann M, Makarov A: **Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes.** *Mol Cell Proteomics* 2012, **11**:O111.013698.
63. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M: **Global quantification of mammalian gene expression control.** *Nature* 2011, **473**:337–342.
64. Cox J, Mann M: **MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.** *Nat Biotechnol* 2008, **26**:1367–1372.
65. Cox J, Matic I, Hilger M, Nagaraj N, Selbach M, Olsen JV, Mann M: **A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics.** *Nat Protoc* 2009, **4**:698–705.
66. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M: **Andromeda: a peptide search engine integrated into the MaxQuant environment.** *J Proteom Res* 2011, **10**:1794–1805.
67. Neuhauser N, Michalski A, Cox J, Mann M: **Expert system for computer-assisted annotation of MS/MS spectra.** *Mol Cell Proteomics* 2012, **11**:1500–1509.
68. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R: **A new bioinformatics analysis tools framework at EMBL-EBI.** *Nucleic Acids Res* 2010, **38**:W695–W699.
69. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG: **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.** *Mol Syst Biol* 2011, **7**:539.
70. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, de Castro E, Coggill P, Corbett M, Das U, Daugherty L, Duquenne L, Finn RD, Fraser M, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J et al: **InterPro in 2011: new developments in the family and domain prediction database.** *Nucleic Acids Res* 2012, **40**:D306–D312.
71. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Meth* 2011, **8**(10):785–786.
72. Darzentas N: **Circoletto: visualizing sequence similarity with Circos.** *Bioinformatics* 2010, **26**:2620–2621.
73. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome Res* 2009, **19**:1639–1645.

doi:10.1186/1471-2164-15-249

**Cite this article as:** Mann and Jackson: Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*. *BMC Genomics* 2014 **15**:249.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

