# Chapter 2
# Genetic Diversity in Latvian Spring Barley Association Mapping Population

**Ieva Mezaka, Linda Legzdina, Robbie Waugh, Timothy J. Close, and Nils Rostoks**

**Abstract** Certified organic crop area is continuously increasing in European Union and in Latvia (Eurostat data), despite somewhat lower yield and higher potential for disease damage in organic farming. It is increasingly recognized that breeding varieties for organic farming requires focus on specific traits that may be less important under conventional agriculture. Molecular markers are becoming essential tools for plant breeding allowing reducing time and cost of development of new varieties by early selection of progeny with desired traits. However, there is lack of information on molecular markers for traits that may be important for organic farming, such as plant morphological traits ensuring competitive ability with weeds, yield and yield stability under organic growing conditions, nutrient use efficiency, and resistance to diseases. We have selected 145 Latvian varieties and breeding lines along with 46 foreign accessions for association mapping panel and genotyped those with 1,536 single-nucleotide polymorphism (SNP) markers using Illumina GoldenGate platform and barley oligo pooled array 1. In parallel to genotyping, 154 of the 191 spring barley genotypes contrasting for traits that are important for organic farming are currently in field trials under conventional and organic management. The success of association mapping in structured natural

I. Mezaka • L. Legzdina
State Priekuli Plant Breeding Institute, Zinatnes street 1a, Priekuli 4126, Latvia

R. Waugh
The James Hutton Institute, Invergowrie, Dundee DD2 5DA, Scotland, UK

T.J. Close
University of California, Riverside, CA 92521, USA

N. Rostoks (✉)
Faculty of Biology, University of Latvia, 4 Kronvalda Blvd., Riga 1586, Latvia
e-mail: nils.rostoks@lu.lv

G. Zhang et al. (eds.), *Advance in Barley Sciences: Proceedings of 11th International Barley Genetics Symposium*, DOI 10.1007/978-94-007-4682-4_2,

populations depends on the extent of linkage disequilibrium (LD) and ability to control for the population structure during statistical analyses. Preliminary results based on principal component and phylogenetic analyses of 1,003 SNP markers with average polymorphism information content (PIC) of 0.394 suggested that the set of germplasm is relatively uniform with the exception of a few six-row varieties. STRUCTURE analysis based on the ΔK value suggested that the population could be partitioned into two clusters. The mean LD ($r^2 > 0.1$) extended over 10-cM distance suggesting that the available marker density may be sufficient for association mapping. Plots of pairwise LD along the chromosomes indicated uneven distribution of LD blocks in barley genome.

**Keywords** Barley • Single-nucleotide polymorphism • Linkage disequilibrium • Association mapping • Population structure • Genetic diversity

## 2.1 Introduction

Genetic diversity present in a species is key factor determining its ability to adapt to specific environments. Barley (*Hordeum vulgare* L.) is adapted to a wide range of environments ranging from Mediterranean dry areas to polar circle and to Andean and Himalayan highlands. Barley breeders therefore need to breed not only for yield and quality traits but also for adaptability traits, such as tolerance to biotic and abiotic stresses characteristic to specific target geographic region. Moreover, recent trends toward organic agriculture and novel uses of conventional crops introduce additional challenges, such as necessity for improved disease resistance and nutrient use efficiency or suitability for biofuel production (Sarath et al. 2008; Sticklen 2008; Wolfe et al. 2008). In order to meet these challenges, key traits for the specific breeding purpose need to be identified along with the development of efficient tools to facilitate the breeding (Varshney et al. 2005). Molecular markers can be used for cloning of important genes and QTL as well as to facilitate transfer of useful alleles from exotic germplasm to locally adapted elite varieties (Feuillet et al. 2008; Hajjar and Hodgkin 2007). Characterization of the local barley germplasm and the basis of its adaptability to local environments are critical for successful marker-assisted selection, because of the interactions of the introgressed alleles with the genetic background of adapted germplasm.

Local germplasm has been affected by different factors, such as limited number of founder genotypes, selection by breeders, and introgression of novel alleles from exotic germplasm. Thus, it is important to study not only the overall genetic diversity of a germplasm set but the distribution of genetic diversity in the genome which may allow to identify genome regions under the selection.

Ultimately, for marker-assisted selection (MAS) to be successful, it is important to identify genes that are responsible for agronomically important traits. Once

the genes are known, the molecular mechanisms affecting the trait can be elucidated, and molecular markers based on the specific gene can be designed. Moreover, the allelic diversity in the gene of interest can be studied allowing to estimate the effects of different alleles on the expression of trait. Association mapping in a selection of plant accessions provides alternative to traditionally used biparental mapping populations. The approach takes advantage of linkage disequilibrium (LD) between the marker locus and the trait of interest (Ardlie et al. 2002; Slatkin 2008). The limiting factors in association mapping are number of markers and size of population to be typed with markers. The number of markers depends on the size of LD blocks – the larger the block size, the smaller number of markers needs to be deployed; however, the resolution of the map will also be lower (Kruglyak 2005). The size of LD blocks depends on the material in study. In recently established populations, the extent of LD will be higher because of the limited number of recombination. Cultivated plant varieties represent a special case, as breeding practices have created unique populations with different amount of LD, and the varieties are often related (Ball 2007).

One of the most abundant types of molecular markers is single-nucleotide polymorphisms (SNPs) (Rafalski 2002). Even though majority of them are biallelic, they can be very useful as markers (Kruglyak 1997), in particular, when SNP haplotypes rather than individual SNPs are considered. Utilization of SNP markers is facilitated by development of high-throughput genotyping methods that allow typing of thousands of SNP in large sets of germplasm. Recently, barley EST data was used for resequencing-based SNP discovery (Rostoks et al. 2005; Stein et al. 2007) and for electronic SNP discovery (Close et al. 2009). SNP data was used for development of genotyping platform based on Illumina GoldenGate technology (Rostoks et al. 2006), which also allowed construction of a high-resolution consensus SNP linkage map of barley (Close et al. 2009) and paved the way for whole genome association mapping in barley (Waugh et al. 2009). Association mapping was successfully used to map and identify the *INT-C* gene (Ramsay et al. 2011) as well as to identify candidate gene for anthocyanin pigmentation (Cockram et al. 2010). Latvian barley varieties are mostly bred from Scandinavian, German, and Moravian barley varieties (Gaike 1992). Genome-wide diversity in several of these varieties has been studied using microsatellites (Sjakste et al. 2003) and DArT markers (Kokina and Rostoks 2008). Recently, high-throughput SNP genotyping platform was used for genotyping 95 Latvian barley varieties and breeding lines (Rostoks 2008), which revealed relatively low level of population structure with exception of six-row varieties and linkage disequilibrium extending over the distances suitable for association mapping. Currently, an association mapping project is in progress supported by the European Social Fund with the aim of identifying molecular markers for traits important for low input and organic barley cultivation. As a first step in this project, we report a preliminary characterization of the diversity and the extent of linkage disequilibrium in our association mapping panel.

## 2.2 Materials and Methods

### *2.2.1 Plant Material and DNA Extractions*

The set of germplasm consisted of 25 Latvian barley varieties, 120 Latvian barley breeding lines, and 46 foreign varieties and breeding lines (list of varieties is available upon request). Only spring two-row-type varieties were included, with the exception of six Latvian six-row varieties. DNA for genotyping was extracted from leaves of a single plant using DNeasy Plant Mini Kit (Qiagen, Hilden, Germany).

### *2.2.2 High-Throughput SNP Genotyping and Data Quality Control*

Illumina high-throughput genotyping was done as described (Rostoks et al. 2006). Currently, two barley oligo pooled assay, BOPA1 and BOPA2, each containing 1,536 SNP are available (Close et al. 2009). One set of 95 Latvian varieties and breeding lines was genotyped only with BOPA1, while the second set of 96 Latvian and foreign barley accessions was genotyped with BOPA1 and BOPA2. Current study is based only on BOPA1 data in the total set of 191 accessions. SNP genotyping data was controlled for presence of excessive number of heterozygous loci and missing data points. Barley consensus linkage map based on the SNP markers (Close et al. 2009) was used throughout the study.

### *2.2.3 Data Analyses*

Principal component analysis (PCo) was done using DARwin5 (http://darwin.cirad.fr/darwin). Population structure was studied using Structure 2.2 software using a selection of SNP loci and admixture linkage model (Falush et al. 2003). Forty-one SNPs (five to six per chromosome) were selected based on criteria that they are at least 10 cM apart and have minor allele frequency (MAF) > 0.30. Hypothesis of 1–20 clusters (*K*) was tested using burn-in of 100,000, run length of 200,000, and admixture model in 15 iterations. Accessions were assigned to the most probable number of clusters according to a published method (Evanno et al. 2005). Largest value of an *ad hoc* statistic ΔK, which is based on the rate of change in the log probability of data between successive *K* values, was used as an indicator of the true number of clusters. Polymorphism information content (PIC) was calculated as described (Kota et al. 2008). Linkage disequilibrium between pairs of SNP loci was calculated in TASSEL 3.0 (Bradbury et al. 2007). One thousand and three mapped SNPs with MAF > 0.1

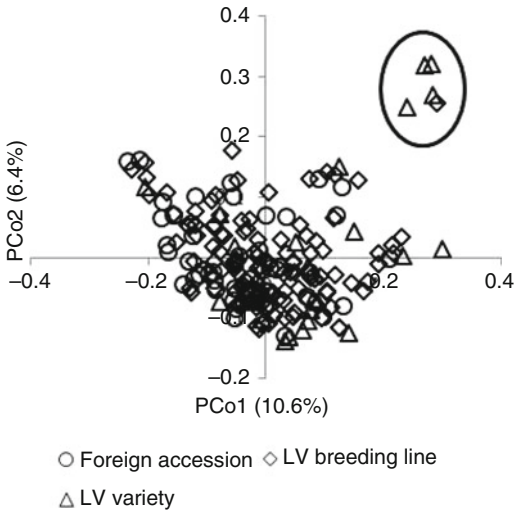



**Fig. 2.1** Principal component analysis of SNP genotype data. Foreign barley accessions and Latvian breeding lines and varieties form a relatively uniform cluster with the exception of several Latvian six-row accessions (*circled*). *X*- and *Y*-axes show percent variation explained by the principal components 1 and 2, respectively

were used for calculations in the whole set of 191 accessions and after removal of the six six-row barley accessions. Once the foreign accessions were removed, MAF was recalculated, and 963 SNPs were retained for subsequent analyses.

## 2.3 Results and Discussion

One hundred and ninety-one Latvian and foreign barley accessions were genotyped with the Illumina oligo pooled array BOPA1 consisting of 1,536 SNPs (Close et al. 2009). SNP genotype data were controlled for the presence of excessive number of heterozygotes and missing data (1,463 SNPs retained), and of those, 1,084 were with minor allele frequency of over 10%. The 1,003 of these SNPs were positioned on the barley consensus linkage map (Close et al. 2009) and were used for all subsequent analyses in this study.

Principal component analysis of the genotype data revealed a relatively compact group of the accessions, while the PCo1 and PCo2 explained only 10.6 and 6.4 % of variation, respectively (Fig. 2.1). The only exception was a group of old Latvian six-row accessions, which were separated from the main cluster. Clustering according to row type is consistent with historical origins of barley (Pourkheirandish and Komatsuda 2007) and with previous studies that analyzed the genetic diversity in populations consisting of both six-row and two-row genotypes (Hamblin et al. 2010; Zhang et al. 2009). Thus, it appears that Latvian barley varieties show little population substructure and high degree of relatedness with foreign barley accessions. This is consistent with the pedigrees of Latvian barley varieties involving early German, Moravian, and Scandinavian varieties and with the recent breeding practices

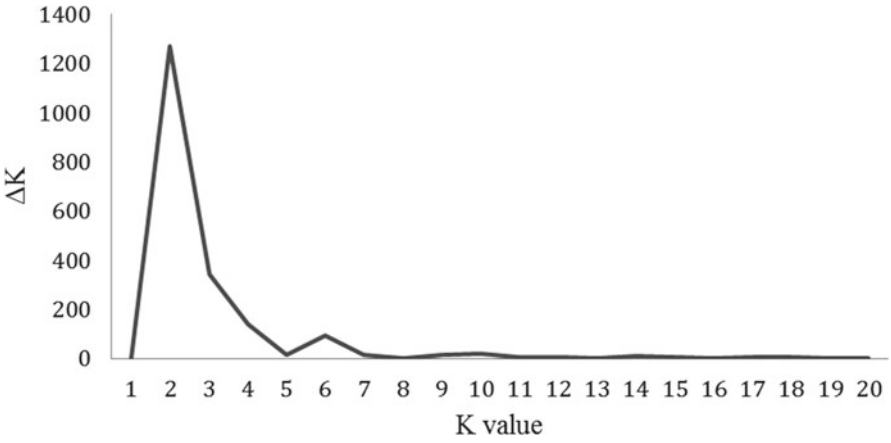



**Fig. 2.2** Detection of number of clusters (*K*) by estimating Δ*K* over 15 iterations for each *K* value with STRUCTURE. Δ*K* was expressed as a mean of the absolute values of ratio of change of the likelihood function with respect to *K*. The highest value of Δ*K* indicates the strength of the signal

that attempt to combine agronomic and quality traits of foreign germplasm with the adaptive traits of the Latvian varieties.

Neighbor-joining dendrogram (results not shown) revealed several major clusters with low bootstrap support and generally good separation of accessions in terminal nodes often consistent with the known pedigrees. No separation according to breeder (Latvian or foreign) was apparent, which again was consistent with the exchange of breeding material between Latvian barley breeders and extensive use of foreign germplasm in Latvian barley breeding programs.

Population structure is known to affect association mapping in various crop species including barley (Hamblin et al. 2010), wheat (Le Couviour et al. 2011), rice (Jin et al. 2010), and oat (Newell et al. 2011). Population structure may result in finding spurious associations between markers and traits; therefore, population structure has to be assessed prior the association mapping (Pritchard et al. 2000). STRUCTURE was used to test highest probability of a number of clusters (*K*) among all barley accessions. The highest value of ΔK was observed with two clusters (Fig. 2.2) suggesting that the accessions most likely separate into two subpopulations, which is in contrast to the ten populations found in the set of 500 UK cultivars (Cockram et al. 2010) and seven populations in the set of 1,816 breeding lines from US breeding programs (Hamblin et al. 2010). This may be explained by more diverse set of genotypes with respect to row number (only six six-row genotypes included in our study) and growth habit (only spring type in our study), because clustering in US and UK genotypes was observed mostly based on row number and growth habit and their combinations. No clustering according to origin of genotypes was observed in our study, unlike the finding that Western European elite wheat varieties cluster according to country of origin and breeding history (Le Couviour et al. 2011).

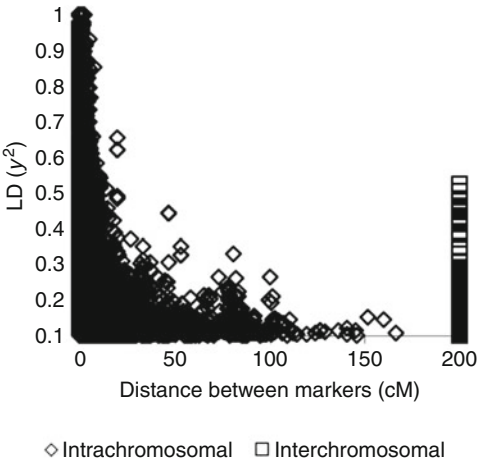



**Fig. 2.3** Extent of the pairwise LD between 1,003 high-quality mapped SNP markers in the whole data set of 191 accessions. LD between pairs of SNP loci on different chromosomes (interchromosomal LD) is shown at a fixed distance of 200 cM

The extent of pairwise LD between all SNP markers was explored in Tassel 3.0 using the genetic distances from Close et al. (2009) (Fig. 2.3). The extent of pairwise LD significantly varied among the barley chromosomes, suggesting that this factor may affect the resolution of association mapping for QTL located on different chromosomes (Table 2.1). Median extent of LD was 2.0 cM on chromosome 6H, while chromosome 2H exhibited much more extensive LD (median 9.3 cM). Moreover, there were some differences in the subsets of germplasm, e.g., more extensive LD was observed on chromosome 1H in a set of Latvian accessions (15 cM) than in the whole set (10 cM). However, on all chromosomes, average distance between markers was smaller than the extent of pairwise LD suggesting that marker coverage could be sufficient for association mapping (Table 2.1).

Distribution of the LD along the chromosomes varied extensively (Fig. 2.4). While the chromosome 2H that showed the highest mean and median pairwise extent of LD also exhibited a relatively uniform distribution along the chromosome, the chromosomes 3H and 6H showed extensive LD mostly in the centromeric regions similar to results by Comadran et al. (2011). Interchromosomal and intrachromosomal variation of LD has been observed in various autogamous crop species, and regions where LD extends up to 50 cM (Hamblin et al. 2010) and 60 cM (Rostoks et al. 2006) in barley, 50 cM in rice (Jin et al. 2010), and 5 cM in oat (Newell et al. 2011) have been detected. Neumann et al. suggested that high LD in elite crop germplasm could be formed during breeding history resulting in assembled blocks of chromosomes containing genes for agronomic fitness (Neumann et al. 2011).

Whole genome association mapping in small and highly structured sets of germplasm has been shown to generate a large number of false-positive associations (Cuesta-Marcos et al. 2010). Recently, we successfully used whole genome association mapping to identify the *NUD* gene as a major determinant of β-glucan content in barley but failed to identify other QTL for β-glucan content in the set of 95 Latvian barley accessions (Mezaka et al. in press). Apparently, simple traits controlled

**Table 2.1** Comparison of polymorphism information content (PIC) and linkage disequilibrium (LD, $r^2>0.1$) in the barley genome

| Chromosome | Number of markers/average distance between markers in cM | PIC | | | Extent of pairwise LD $r^2>0.1$ (average/median in cM) | | |
|---|---|---|---|---|---|---|---|
| | | Whole set | Only two-row | Only LV two-row | Whole set | Only two-row | Only LV two-row |
| 1 H | 111/1.27 | 0.356 | 0.349 | 0.351 | 10.0/5.7 | 10.5/5.9 | 15.0/5.9 |
| 2 H | 157/0.96 | 0.404 | 0.396 | 0.395 | 17.2/9.3 | 14.0/7.1 | 15.8/9.0 |
| 3 H | 153/1.14 | 0.403 | 0.396 | 0.394 | 7.3/3.9 | 6.1/3.5 | 6.7/3.5 |
| 4 H | 117/1.05 | 0.395 | 0.388 | 0.383 | 10.1/4.2 | 7.9/3.6 | 8.7/3.2 |
| 5 H | 202/0.97 | 0.394 | 0.387 | 0.381 | 9.7/5.0 | 7.5/3.9 | 8.8/4.3 |
| 6 H | 147/0.89 | 0.399 | 0.394 | 0.391 | 4.8/2.0 | 4.5/1.7 | 6.8/2.4 |
| 7 H | 117/1.44 | 0.395 | 0.386 | 0.386 | 12.8/5.3 | 10.4/5.3 | 12.0/4.8 |
| Genome-wide | 1,003/1.08 | 0.394 | 0.387 | 0.384 | 10.3/4.7 | 8.5/4.1 | 10.2/4.5 |

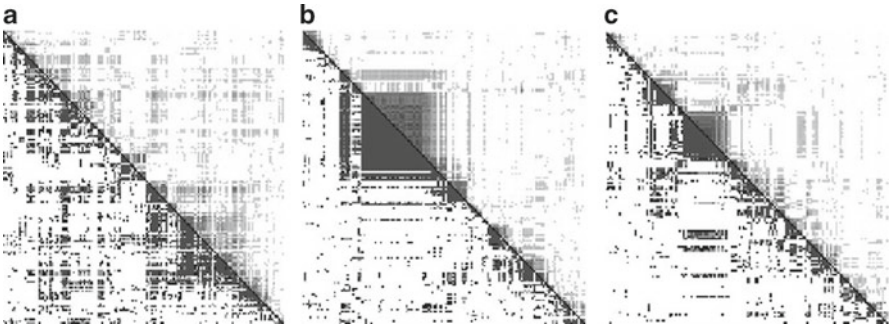



**Fig. 2.4** Extent of the LD along selected chromosomes. (**a**) chromosome 2H; (**b**) chromosome 3H; (**c**) chromosome 6H. *Upper diagonal* in each panel shows LD ($r^2$); *lower diagonal* shows the corresponding *p* values

by a single gene may be successfully mapped in relatively small association mapping panels, provided that population structure is accounted for (Cockram et al. 2010; Comadran et al. 2011). However, other traits controlled by many small-effect QTL may be much more difficult to map. Careful selection of subsets of germplasm for a particular trait and statistical control of population structure may improve association mapping (Cockram et al. 2010), along with other approaches, such as use of SNP haplotypes rather than single SNPs (Lorenz et al. 2010). However, in the summary, the preliminary results reported in this chapter suggest a low degree of population substructure and a suitable extent of LD for association studies in our association mapping population.

**Acknowledgements** The study is funded by the European Social Fund cofinanced project 2009/0218/1DP/1.1.1.2.0/09/APIA/VIAA/099 and the Latvian Council of Science grant Z-956-090.

## References


Ardlie, K. G., Kruglyak, L., & Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nature Reviews. Genetics, 3*, 299–309.

Ball, R. D. (2007). Statistical analysis and experimental design. In N. C. Oraguzie, E. H. A. Rikkerink, S. Gardiner, & H. N. De Silva (Eds.), *Association mapping in plants* (pp. 133–196). New York: Springer.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics, 23*, 2633–2635.

Close, T., Bhat, P., Lonardi, S., Wu, Y., Rostoks, N., Ramsay, L., Druka, A., Stein, N., Svensson, J., Wanamaker, S., Bozdag, S., Roose, M., Moscou, M., Chao, S., Varshney, R., Szucs, P., Sato, K., Hayes, P., Matthews, D., Kleinhofs, A., Muehlbauer, G., DeYoung, J., Marshall, D., Madishetty, K., Fenton, R., Condamine, P., Graner, A., & Waugh, R. (2009). Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics, 10*, 582.

Cockram, J., White, J., Zuluaga, D. L., Smith, D., Comadran, J., Macaulay, M., Luo, Z., Kearsey, M. J., Werner, P., Harrap, D., Tapsell, C., Liu, H., Hedley, P. E., Stein, N., Schulte, D., Steuernagel, B., Marshall, D. F., Thomas, W. T., Ramsay, L., Mackay, I., Balding, D. J., Consortium, T. A., Waugh, R., & O'Sullivan, D. M. (2010). Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 21611–21616.

Comadran, J., Ramsay, L., MacKenzie, K., Hayes, P., Close, T. J., Muehlbauer, G., Stein, N., & Waugh, R. (2011). Patterns of polymorphism and linkage disequilibrium in cultivated barley. *Theoretical and Applied Genetics, 122*, 523–531.

Cuesta-Marcos, A., Szucs, P., Close, T., Filichkin, T., Muehlbauer, G., Smith, K., & Hayes, P. (2010). Genome-wide SNPs and re-sequencing of growth habit and inflorescence genes in barley: implications for association mapping in germplasm arrays varying in size and structure. *BMC Genomics, 11*, 707.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology, 14*, 2611–2620.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics, 164*, 1567–1587.

Feuillet, C., Langridge, P., & Waugh, R. (2008). Cereal breeding takes a walk on the wild side. *Trends in Genetics, 24*, 24–32.

Gaike, M. (1992). Spring barley. In I. Holms (Ed.), *Field crop breeding in Latvia (in Latvian)* (pp. 53–63). Riga: Avots.

Hajjar, R., & Hodgkin, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica, 156*, 1–13.

Hamblin, M. T., Close, T. J., Bhat, P. R., Chao, S., Kling, J. G., Abraham, K. J., Blake, T., Brooks, W. S., Cooper, B., Griffey, C. A., Hayes, P. M., Hole, D. J., Horsley, R. D., Obert, D. E., Smith, K. P., Ullrich, S. E., Muehlbauer, G. J., & Jannink, J. L. (2010). Population structure and linkage disequilibrium in U.S. barley germplasm: implications for association mapping. *Crop Science, 50*, 556–566.

Jin, L., Lu, Y., Xiao, P., Sun, M., Corke, H., & Bao, J. (2010). Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theoretical and Applied Genetics, 121*, 475–487.

Kokina, A., & Rostoks, N. (2008). Genome-wide and Mla locus-specific characterization of Latvian barley varieties. *Proceedings of the Latvian Academy of Sciences, 62*, 103–109.

Kota, R., Varshney, R., Prasad, M., Zhang, H., Stein, N., & Graner, A. (2008). EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. *Functional & Integrative Genomics, 8*, 223–233.

Kruglyak, L. (1997). The use of a genetic map of biallelic markers in linkage studies. *Nature Genetics, 17*, 21–24.

Kruglyak, L. (2005). Power tools for human genetics. *Nature Genetics, 37*, 1299–1300.

Le Couviour, F., Faure, S., Poupard, B., Flodrops, Y., Dubreuil, P., & Praud, S. (2011). Analysis of genetic structure in a panel of elite wheat varieties and relevance for association mapping. *Theoretical and Applied Genetics, 123*, 715–727.

Lorenz, A. J., Hamblin, M. T., & Jannink, J. L. (2010). Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. *PLoS One, 5*, e14079.

Mezaka, I., Bleidere, M., Legzdina, L., & Rostoks, N. (in press). Whole genome association mapping identifies naked grain locus *NUD* as determinant of β-glucan content in barley. *Zemdirbyste – Agriculture*.

Neumann, K., Kobiljski, B., Denčić, S., Varshney, R., & Börner, A. (2011). Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). *Molecular Breeding, 27*, 37–58.

Newell, M. A., Cook, D., Tinker, N. A., & Jannink, J. L. (2011). Population structure and linkage disequilibrium in oat (*Avena sativa* L.): implications for genome-wide association studies. *Theoretical and Applied Genetics, 122*, 623.

Pourkheirandish, M., & Komatsuda, T. (2007). The importance of barley genetics and domestication in a global perspective. *Annals of Botany London, 100*, 999–1008.

Pritchard, J. K., Stephens, M., Rosenberg, N. A., & Donnelly, P. (2000). Association mapping in structured populations. *The American Journal of Human Genetics, 67*, 170–181.

Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology, 5*, 94–100.

Ramsay, L., Comadran, J., Druka, A., Marshall, D. F., Thomas, W. T., Macaulay, M., MacKenzie, K., Simpson, C., Fuller, J., Bonar, N., Hayes, P. M., Lundqvist, U., Franckowiak, J. D., Close, T. J., Muehlbauer, G. J., & Waugh, R. (2011). *INTERMEDIUM-C*, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene *TEOSINTE BRANCHED 1*. *Nature Genetics, 43*, 169–172.

Rostoks, N. (2008, September 9–12) *High throughput genotyping for characterization of barley germplasm*. Proceedings of the EUCARPIA 18th General Congress, Valencia, Spain.

Rostoks, N., Mudie, S., Cardle, L., Russell, J., Ramsay, L., Booth, A., Svensson, J. T., Wanamaker, S. I., Walia, H., Rodriguez, E. M., Hedley, P. E., Liu, H., Morris, J., Close, T. J., Marshall, D. F., & Waugh, R. (2005). Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. *Molecular Genetics and Genomics, 274*, 515–527.

Rostoks, N., Ramsay, L., MacKenzie, K., Cardle, L., Bhat, P. R., Roose, M. L., Svensson, J. T., Stein, N., Varshney, R. K., Marshall, D. F., Graner, A., Close, T. J., & Waugh, R. (2006). Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 18656–18661.

Sarath, G., Mitchell, R. B., Sattler, S. E., Funnell, D., Pedersen, J. F., Graybosch, R. A., & Vogel, K. P. (2008). Opportunities and roadblocks in utilizing forages and small grains for liquid fuels. *Journal of Industrial Microbiology & Biotechnology, 35*, 343–354.

Sjakste, T. G., Rashal, I., & Roder, M. S. (2003). Inheritance of microsatellite alleles in pedigrees of Latvian barley varieties and related European ancestors. *Theoretical and Applied Genetics, 106*, 539–549.

Slatkin, M. (2008). Linkage disequilibrium – Understanding the evolutionary past and mapping the medical future. *Nature Reviews. Genetics, 9*, 477–485.

Stein, N., Prasad, M., Scholz, U., Thiel, T., Zhang, H., Wolf, M., Kota, R., Varshney, R. K., Perovic, D., Grosse, I., & Graner, A. (2007). A 1,000-loci transcript map of the barley genome: New anchoring points for integrative grass genomics. *Theoretical and Applied Genetics, 114*, 823–839.

Sticklen, M. B. (2008). Plant genetic engineering for biofuel production: Towards affordable cellulosic ethanol. *Nature Reviews. Genetics, 9*, 433–443.

Varshney, R. K., Graner, A., & Sorrells, M. E. (2005). Genomics-assisted breeding for crop improvement. *Trends in Plant Science, 10*, 621–630.

Waugh, R., Jannink, J. L., Muehlbauer, G. J., & Ramsay, L. (2009). The emergence of whole genome association scans in barley. *Current Opinion in Plant Biology, 12*, 1–5.

Wolfe, M., Baresel, J., Desclaux, D., Goldringer, I., Hoad, S., Kovacs, G., Loeschenberger, F., Miedaner, T., Ostergard, H., & Lammerts van Bueren, E. (2008). Developments in breeding cereals for organic agriculture. *Euphytica, 163*, 323–346.

Zhang, L. Y., Marchand, S., Tinker, N. A., & Belzile, F. (2009). Population structure and linkage disequilibrium in barley assessed by DArT markers. *Theoretical and Applied Genetics, 119*, 43–52.

Biological Conservation 153 (2012) 227–238

Contents lists available at SciVerse ScienceDirect

Biological Conservation

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

# Connectedness and connectivity of the Natura 2000 network of protected areas across country borders in the European Union

Otars Opermanis [a,*], Brian MacSharry [a], Ainars Aunins [b], Zelmira Sipkova [a]

[a] European Topic Centre on Biological Diversity, Museum National d'Histoire Naturelle, 57 rue Cuvier, 75231 Paris, France
[b] Department of Zoology and Animal Ecology, Faculty of Biology, University of Latvia, Kronvalda bulv. 4, Riga LV-1010, Latvia

ARTICLE INFO

*Article history:*
Received 20 October 2011
Received in revised form 20 April 2012
Accepted 27 April 2012
Available online 29 June 2012

*Keywords:*
Natura 2000
Connectedness
Connectivity
Political borders
Trans-boundary protected areas
European Union

ABSTRACT

We studied the spatial connectedness of Natura 2000 site boundaries and the functional connectivity of the Natura 2000 network across the 34 terrestrial borders of the European Union. Connectivity was measured by the dispersal success of 192 reptile, amphibian, invertebrate and plant species from Annex II of the European Union Habitats Directive, based on the presence of same species on both sides of the borders. Connectedness and connectivity varied greatly between state borders, with good and bad examples found in all parts of the European Union. Connectedness and connectivity were positively correlated. However, a few outlying examples showed that good connectedness is not necessarily associated with good connectivity and that good connectivity is not always possible because of different habitats and/or different management on both sides of border. In 13 out of 34 borders the connectivity measure was 100% and in 11 other borders it was over 50%. Very few of the geographical and political factors tested to explain variation in connectedness and connectivity were significant. Better connectedness and connectivity, both at the border and site-pair level, was correlated with the rivers forming the border. The time since the designation of the second site in a site-pair was nearly significant, suggesting that connectivity might have always existed in nature but the 'documentation of it' required some time for additional site inventories. It seems that each country border has its own history with a unique subset of factors influencing Natura 2000 site selection and management thus generalising our findings to other international borders would be difficult. We did not find substantial differences in trans-boundary connectivity among taxonomic groups. The application of our method elsewhere in the World would perhaps bring interesting results at a large (continental or sub-continental) scale, provided that similar data are available.



## 1. Introduction

Worldwide it is recognised that protected areas form a cornerstone of efforts to conserve nature and biodiversity (e.g., IUCN, 1994; Maiorano et al., 2007; Chape et al., 2008; Wurzel, 2008). Within the European Union (EU), the Natura 2000 network of protected sites with approximately 26,000 sites covering almost 1,000,000 km$^2$ (EEA, 2010a) across all 27 Member States forms the centrepiece of EU nature and biodiversity policy. The sites are designated for habitats and species of European conservation concern listed in the EU Birds and Habitats Directives which form the legal basis of the Natura 2000 network (European Commission, 1992, 1996).

One of main objectives set out in the Habitats Directive is to achieve a good ecological coherence of the Natura 2000 network. This requirement applies not only to territories of individual Member States, where all protected habitats and species must be sufficiently represented, but also to the linkage between Natura 2000 sites of neighbouring countries to enable movement of species and continuity of habitats where possible. Unlike large federal countries (e.g., United States, Russian Federation, Australia) the territory of the European Union is intersected by many country borders and this obstacle may become an issue in nature conservation planning at the EU level.

Although the term 'ecological coherence' of the Natura 2000 network is often mentioned in the Habitats Directive, there is no explicit definition of ecological coherence and it is often interpreted differently by the European Commission, the Member States and non-governmental organisations (Roth, 2005; Johnson et al., 2008). According to articles 3 and 10 of the directive, 'ecological coherence' is linked with the improvement and management of features of the landscape, which by their linear and continuous structure or their function as stepping stones are essential for the migration, dispersal and genetic exchange of wild species

* Corresponding author. Tel.: +33 (0)1 40 79 38 70; fax: +33 (0)1 40 79 38 67.
*E-mail address:* otars.opermanis@mnhn.fr (O. Opermanis).

0006-3207/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biocon.2012.04.031

(European Commission, 1996; Roth, 2005; Ssymank, 2005). Therefore an important feature of 'ecological coherence' is the connectivity between Natura 2000 sites. This is particularly true in near-border areas where neighbouring sites are designated by different national authorities, which may have used different methodologies in selecting the Natura 2000 sites. This problem is often mentioned as the main reason why the process of building the Natura 2000 network cannot be considered systematic (Gaston et al., 2008).

Cross-boundary coherence can be achieved through cooperation between neighbouring countries and designation of trans-boundary protected areas (Cerovsky, 1996; Fall, 2003). Larger continuous protected areas across borders, which are cooperatively managed, reduce the risk of biodiversity loss. Such sites enable species migration and dispersal, for example, Alpine ibex *Capra ibex* have a summer range in Vanoise National Park in France but winter in the adjacent Gran Paradiso National Park in Italy (Hamilton, 1998).

The number of trans-boundary protected areas has been constantly growing worldwide along with other globalisation processes (Zimmerer et al., 2004). The slogan 'nature knows no boundaries' has been widely exploited in different public campaigns by different international bodies, however, the adequacy of trans-boundary protected areas has rarely been assessed (Parrish et al., 2003; Ardron, 2008). Studies of country borders to date have mainly addressed disputed areas (including disputes over natural resources), border change after major events in history and evolution of borders (Minghi, 1963) but the phenomenon of country borders in nature protection has rarely been addressed although the presence of such impact is often acknowledged (e.g., Cerovsky, 1996; Hanks, 2003).

In the European Union, some attention has been paid to improve trans-boundary cooperation, and thus connectivity, through seminars and bi-lateral exchange of information (Leibenath et al., 2005). To date most studies on efficiency of Natura 2000 network have only vaguely addressed trans-boundary connectivity and have not provided any quantitative assessment of it (e.g., Gaston et al., 2008; Rosati et al., 2008; Ilies et al., 2010; Jantke et al., 2011), but see Sundblad et al. (2011). Our aim was to make an EU-wide assessment of trans-boundary connectivity involving quantitative measures enabling comparisons between European Union Member States and regions. We also aimed to analyse the effects of environmental and political factors upon the connectivity.

In ecology, connectivity has two components: the first component refers to the physical links between elements of the spatial structure of a landscape (often referred to as 'connectedness') and can be described from mapped elements. The second component is functional connectivity, a parameter of landscape function, which measures the processes by which sub-populations of organisms are interconnected into a functional demographic unit (Baudry and Merriam, 1988; Jongman et al., 2004; Boitani et al., 2007). Functional connectivity, depending on species and research opportunities, has been measured as the distance between sites, structure of landscape, dispersal success between sites and search time travelling from one to another site (Tischendorf and Fahrig, 2000; Jongman et al., 2004). In this paper we use the term 'connectedness' to describe the spatial relationships between Natura 2000 site boundaries and 'connectivity' to describe the functional connectivity. In our study of connectivity we focused on species with limited dispersal ability (reptiles, amphibians, invertebrates and plants) where connectivity issues play a more important role than for good dispersers.

Landscape connectedness and biological connectivity are often assumed to be correlated as connectedness is assumed to be a key factor contributing to connectivity (Bennett, 2003; Campagne et al., 2009). In this study we tested this relationship by obtaining connectedness and connectivity measures separately using different methods.

## 2. Methods

### 2.1. Trans-boundary protected areas

Trans-boundary Protected Areas are defined as protected area complexes that in different countries adjoin across international boundaries. They may be comprised of separate units that are connected or may be joined into a single site (IUCN and WCPA, 2001). In our study we extended this definition of trans-boundary protected areas to Natura 2000 sites that are not physically connected but close enough to each other on either sides of the boundary to enable dispersal of species.

To study connectedness we included both types of Natura 2000 sites, i.e., Special Protection Areas (SPAs) designated under the Birds Directive and Sites of Community Interest (SCIs) designated by the Habitats Directive. Although SPAs are primarily designated for the bird species listed in the Annex I of the Birds Directive, most of them directly or indirectly contribute to the conservation of species protected under the Habitats Directive. In our calculations of connectivity at the site-pair level (see below) we used only SCIs as only this category had systematic records of species protected under the Habitats Directive. If SPAs were used as well, we would obtain a significant underestimation of connectivity for the selected species.

In our study we focus only on Natura 2000 sites which form a cornerstone of EU nature conservation policy (Bonnin et al., 2007). However, many if not all European Union Member States have additional protected areas of national importance (e.g., Araujo et al., 2007; Maiorano et al., 2007) that may be situated in the border regions but these are usually areas with weaker conservation status where nature conservation is not a primary function, e.g., biosphere reserves, although such protected areas may be very large. Given that comparable data on species and habitats present in these sites were often lacking or difficult to obtain, we did not use these sites in our study.

We analysed all 34 terrestrial political borders where both adjoining countries are members of the European Union. We used spatial and descriptive data from the official consolidated European Natura 2000 dataset, December 2009, prepared by the European Environment Agency (EEA, 2010a).

### 2.2. Measuring connectedness

Connectedness describes the structural links between elements of the spatial structure of a landscape and can be described using mapped elements (Baudry and Merriam, 1988), in this case, boundaries of Natura 2000 sites. We chose to measure the connectedness by a quantified proportion of Natura 2000 sites on the both sides of a border. The country borders used in this analysis were derived from the European Boundary Map at a scale of 1:100,000 obtained from the European Environment Agency. Unfortunately Member States use country borders at different scales and projections during the process of digitising Natura 2000 site boundaries. Where two countries share a border, this border is represented slightly differently in the cartographic system of each country due to factors such as the national projection and the scale at which the border has been mapped. This means that where in reality a Natura 2000 site shares a common border when this site is digitised, the 'common border' is displayed differently and in effect creates a cartographical distortion of reality. In order to overcome these obstacles, and examine this issue at the European scale, we degraded all datasets to 1 km grids (the 1 km ETRS 89 LAEA 5210

grid was used). The initial step was to intersect the borders with the 1 km grid. As a second step, all Natura 2000 sites that intersected the gridded border were extracted and separated by country. These sites, split by country, were then intersected with the same 1 km grid. Following on from this, all grid cells where sites from different countries occurred in the same cell were considered to represent areas where the border consisted of Natura 2000 sites on both sides. Those cells where only sites from one country occurred in were considered to be areas where the border was composed of Natura 2000 site but only on one side (Fig. 1). Thus we counted (1) the number of cells which included Natura 2000 site only on one border side and (2) the number of cells which included Natura 2000 sites on both sides of border. We quantified connectedness as a proportion of value 2 against the sum of values 1 and 2.

In order to test possible variation of connectedness according to political and geographical factors, we classified the 34 boundaries as to whether they represented old or new European Union Member States according to the recent enlargement process (the 12 Member States that joined since 2004 were classified as 'new') and whether the boundary is a division of nations representing the same or different major language families (e.g., Romance, Germanic, Balto-Slavic, Finno-Ugric, Greek (National Geographic Society, 2005)). The last variable was introduced to test whether language barrier was an obstacle in communication between authorities of neighbouring countries thus making the selection of optimal trans-border protected areas more difficult.

As only a small part of the former Iron Curtain line (i.e., division between East and West earlier in 20th century; Cerovsky, 1998) matches with today's terrestrial borders between the EU Member States (only Austrian/Czech, Austrian/Hungarian, Austrian/Slovak, Bulgarian/Greek), so we could not address this potential effect due to too small sample size in one level of this factor (4 versus 30).

We tested if connectedness is related to the proportion of Natura 2000 sites cover (i.e., density) within 1 km border buffer and to the physical features of the borders, namely the proportion of mountainous areas and rivers forming the political border. To assess the proportion of 'mountainous areas' present in the borders we used the 'mountains' dataset created by the European Environment Agency during the work on a report on mountains in Europe (EEA, 2010b). This 'mountains' dataset used 16 combinations of criteria to designate 'mountainous' areas in Europe. River data were obtained from the European Catchments and Rivers Network System database (ECRINS) version 1.0 of the European Environment Agency (ECRINS, 2012). We considered all rivers in this database (without size limitation) that were forming the border, but not rivers that were crossing the border. Both mountain and river datasets were plotted against the country borders as derived from the European Boundary Map to obtain a percentage of grids with mountains and rivers.

### 2.3. Measuring connectivity

Connectivity according to Baudry and Merriam (1998) is a parameter of landscape function, which measures the processes by which sub-populations of organisms are interconnected to a functional demographic unit. Species survival is dependent on

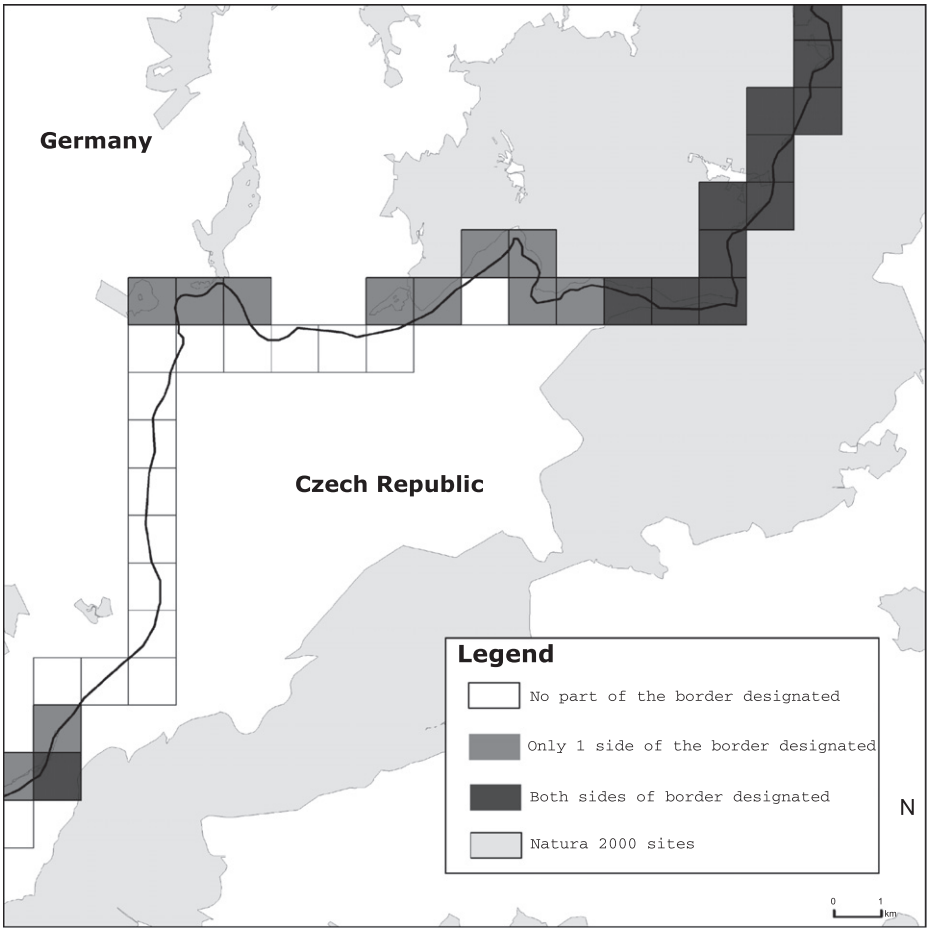



**Fig. 1.** Gridding of country borders and method of calculation of connectedness. Connectedness was expressed as number of dark grey grids/number of light grey grids * 100.

habitat quality, resource availability and, for most species, the ability move through landscape. Movement is needed daily for foraging, rest and shelter, and as periodic migrations for reproduction and to avoid unfavourable environments and for dispersal (e.g., Hansson, 1991; Caughley and Sinclair, 1994).

Plants and animals both disperse by wind, water, with the help of other species or by their own movements. Dispersal can only function if there are sites to disperse from and to and a means of dispersal (Jongman et al., 2004). Dispersal is often associated with emigration and immigration, i.e., movement from one population or meta-population to another, very often from one protected area to another, especially in an urbanised landscape matrix as found in large parts of the European Union. We measured connectivity by a proportion of pairs of Natura 2000 sites on both sides of political borders that host the same species, which indirectly indicates a presence of successful dispersal (emigration or immigration) between sites in a historical sense. For each border, we first identified site-pairs and then checked if the paired sites host the same species. A 'pair' consisted of two sites, each from the other side of a country border, with a maximum distance of 1 km between their external borders. To start off all Natura 2000 sites within 1 km of a border were selected. Each site was buffered by 1 km and all sites from adjacent countries selected by this buffer were deemed to be potential pairs for the site (Fig. 2). Accordingly, one site could participate in more than one pair. The procedure was run from one side of the border, then the other, and then duplicates were eliminated.

We restricted our analyses to taxonomic groups with limited dispersal ability and where dispersal ability may be a key factor influencing survival: amphibians, reptiles, invertebrates and plants listed in the Annex II of the Habitats Directive (European Commission, 1996). Species with limited dispersal ability are more dependent on continuous habitats (sites) and therefore can be a good measure of biological connectivity. Annex II of the Habitats Directive includes 24 reptile, 25 amphibian, 135 invertebrate and 461 plant species. However, many (59.8%) of these species occurred only in one Member State, and thus they could not be included in this study. Finally, 21 reptiles and amphibians, 78 invertebrates and 93 plants were included in this study, which constitutes 21% of the taxa listed in the Annex II of the Habitats Directive.

Knowledge of the biology of dispersing organisms is currently inadequate, particularly regarding the distances that organisms can move through a landscape from one habitat patch to another (Ruckelhaus et al., 1997; Johnson et al., 2008). Distance limitation to a maximum of 1 km between sites in a pair seems fair to study possible border effects, as well as it reflects well the possibilities of most taxa to travel between sites given their maximum dispersal capacities (Table 1).

Data on the presence/absence of species in paired Natura 2000 sites were taken from the respective Standard Data Forms submitted by the Member States (European Commission, 1997). It is acknowledged that Natura 2000 data have been collected primarily for nature conservation administration purposes, rather than scientific, and involve different methods and effort by different

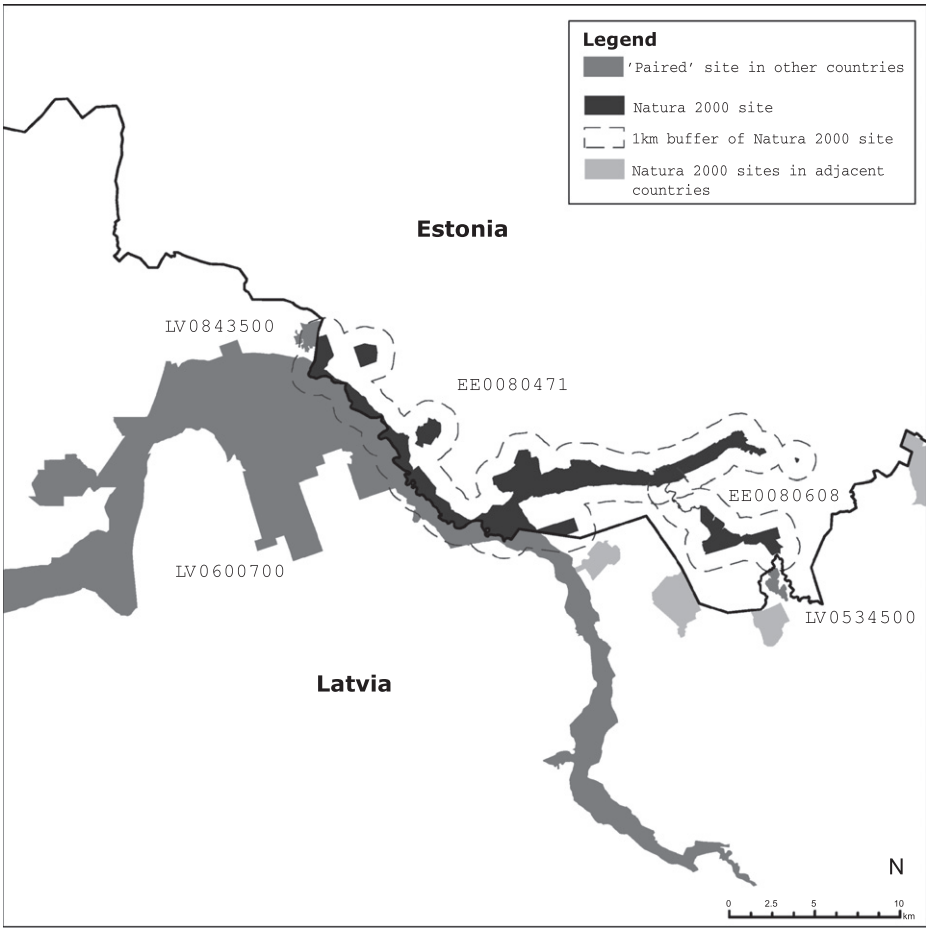



**Fig. 2.** Identification of the site pairs to measure connectivity. An example shows a pairing of Estonian site EE0080471 with 2 Latvian sites LV0843500 and LV0600700 and Estonian site EE0080608 with the Latvian site LV0534500.

**Table 1**
Observed dispersal capacity of some species listed in the Annex II of the EU Habitats Directive.

| Species or group | Distance (m) | Comments on distance | References |
|---|---|---|---|
| **Plants** | | | |
| Lady's-slipper Orchid *Cypripedium calceolus* | >100 | Plants with identical multi-locus genotypes were found | Brzosko et al. (2002) |
| 17 Grassland plant species (not Annex II) | 140–780 | Potential dispersal | Kiviniemi and Eriksson (1999) |
| **Invertebrates** | | | |
| Ramshorn snail *Anisus vorticulus* | 179 | Median between the two occupied suitable habitats | Niggebrugge et al. (2007) |
| Ramshorn snail *Anisus vorticulus* | 679 | Median between the two occupied and unoccupied suitable habitats | Niggebrugge et al. (2007) |
| Violet copper *Lycaena helle* | 560 | Maximum travel | Fischer et al. (1999) |
| Marsh fritillary *Euphydias aurinia* | 400 | Maximum travel | Baguette and Schtickzelle (2006), Junker and Schmitt (2010) |
| False ringlet *Coenonympha oedipus* | 399 | Maximum travel | Celik et al. (2009) |
| *Maculinea* butterflies | 100 | Approximate travel | Nowicki et al. (2005) |
| Southern damselfly *Coenagrium mercuriale* | 1060 | Maximum dispersal capacity | Purse et al. (2003) |
| Stag beetle *Lucanus cervus* | 2065 | Maximum travel | Rink and Sinsch (2007) |
| Hermit beetle *Osmoderma eremita* | 190 | Maximum travel | Ranius and Hedin (2001) |
| Long-horned beetle *Rosalia alpina* | 1000 | Maximum dispersal range | Gatter (1997), Russo et al. (2011) |
| **Amphibians** | | | |
| Crested newt *Triturus cristatus* | 1290 | Maximum travel | Kupfer (1998) |
| Crested newt *Triturus cristatus* | 1000 | Average dispersal rate per year | Arntzen and Wallis (1991) |
| Italian crested newt *Triturus carnifex* | 299 | Maximum travel | Schabetsberger et al. (2004) |
| Alpine salamander *Salamandra aurorae* | 30 | Maximum travel | Bonato and Francasso (2003) |

Member States. Nevertheless, we consider this information reliable at the level of presence or absence of species, applied in this study. The presence or absence of the feature is very important and with legal consequences for the Member States due to the fact that the listing of the feature in the Standard Data Form of a site is not only a statement of its presence but also initiates a responsibility by the Member State to ensure its conservation at the site in a favourable condition (European Commission, 1996). Therefore Member States must run national site surveillance and monitoring programmes and regularly update species information in the Standard Data Forms.

### 2.4. Statistics used

All statistics were performed using R software (R Development Core Team, 2011). As the percentages which we used to quantify connectedness and connectivity as well as variables, such as river presence and the portion of the border region consisting of Natura 2000 sites, had significant departures from the normal distribution we used non-parametric statistics. For comparison of two samples we used the Wilcoxon ranks test and for comparison of three samples (only where Member States were classified as old, new and different) we used the Kruskal–Wallis rank sum test. The Spearman Rank Correlation ($r_s$) was used to detect possible correlation between connectedness measures, as well as between connectedness and border length, between connectedness and number of sites within a 1 km buffer of the border and between connectedness and river presence on the borders. Exceptions were the connectivity and mountains where we could not use rank correlation because of too many tied ranks. In these cases we used Pearson correlation ($r$) and square root transformation in order to best fit variables to the normal distribution.

To model the factors that affect connectivity we created a data matrix where the entry unit was site-pair. Due to the hierarchical structure of the dataset, we used Generalised Linear Mixed Models (GLMM) from the R statistical software with the "lme4" package (R Development Core Team, 2011; Bates et al., 2011) to explore possible impacts of predictor variables upon connectivity (Table 2). We fitted a binomial family model using the logit-link function with a binary response variable that indicated whether or not common features exist in the site-pair (further Binary response model) and a Poisson family model using a log-link function with a discrete numerical response variable that indicated the number of common features in the site-pair (further Numerical response model). To control for the effects of individual borders between countries, we included each country pair as a random effects factor in the model. We explored all possible combinations of uncorrelated variables. From the competing models we chose one binary response and one numerical response model consisting of significant and near-significant predictors ($P < 0.1$) having the smallest Akaike's Information Criterion (AIC) and highest log likelihood (there was no contradiction between these two).

The covariates ARETOT, AREGAP, FEATOT and FEAGAP (Table 2) were strongly inter-correlated (Spearman rank correlation: $r_s$ = 0.34–0.90, $P < 0.001$) and so were LASTTIM and GAPTIM (Spearman rank correlation: $r_s = -0.24$, $P < 0.001$). Inter-correlated variables were not allowed to occur simultaneously in the same model. As variables LASTTIM and GAPTIM held missing values (Table 2), models having any of these parameters were estimated from 422 complete cases and only 30 country-pairs were used in MSPAIR as all site pairs on Hungarian–Romanian, Bulgarian–Romanian, Belgian–German and Belgian–Luxembourg borders had only missing values, i.e., no data.

## 3. Results

### 3.1. General observations

The Natura 2000 database from the end of 2009 included information on 25,828 sites in 27 EU Member States covering 922,271 km$^2$. Of these, 1226 sites were within a 1 km buffer of EU internal terrestrial boundaries. The total terrestrial border length between

**Table 2**
Predictors used to explain variation in connectivity using Generalised Linear Mixed Models and their parameters.

| Variable | Description | Variable type | Count/ basic statistics | | |
|---|---|---|---|---|---|
| ENLARG | Indicates whether site-pair was located on the border between 'old', 'new' or 'old and new' EU member states. | Fixed 3-level factor | Old: 252, new: 142, different: 120 | | |
| LANGUA | Indicates whether site-pair was located on the border between different or same or different major language families | Fixed 2-level factor | Same: 283, different: 231 | | |
| MOUNT | Presence of mountains in the site-pair area | Fixed 2-level factor | Yes: 340, no: 174 | | |
| RIVER | Presence of river forming the border in the site-pair | Fixed 2-level factor | Yes: 180, no: 334 | | |
| MSPAIR | Country-pair to which the site-pair belongs | Random 34-level factor | 514 | | |
| | | | N | Mean | Range |
| ARETOT | Total area (ha) of the sites in a pair | Covariate | 514 | 30377.7 | 71.6 – 587187.0 |
| AREGAP | Difference in the area (ha) between the sites in a pair | Covariate | 514 | 21270.6 | 1.2 – 317723.0 |
| FEATOT | Total number of species in the sites in a pair | Covariate | 514 | 9.2 | 0 - 38 |
| FEAGAP | Difference in the number of species in the sites in a pair | Covariate | 514 | 4 | 0 - 31 |
| LASTTIM | Time (months) since designation of the last site in a pair | Covariate | 422 | 70.7 | 6.1 – 171.5 |
| GAPTIM | Time gap (months) between designation of the sites participating in a pair | Covariate | 422 | 40.5 | 0 – 168.4 |
| ACTDIST | Actual distance (m) between the sites in a pair | Covariate | 514 | 74.0 | 0 - 1000 |

EU Member States was 14,714 km. There was a high variation in Natura 2000 site densities within 1 km of the border, similarly in the ratio between the border region and the interior of the state (Fig. 3). There were no statistically significant differences between Natura 2000 site densities (% cover), between the border regions and interior of the state (Wilcoxon ranks test, $P = 0.17$).

### 3.2. Connectedness

Connectedness scores varied significantly among borders (Table 3) with a mean 46.2 ± 3.9SE%, range 5–98%. The best connectedness was along the Swedish–Finnish border in the North. The worst connectedness was recorded for the Austrian–Slovenian, French–Luxembourg and Polish–Lithuanian borders, which were also amongst the shortest. In general, good and bad examples of connectedness can be found in all parts of European Union (Fig. 4A). Relationships between border length versus connectedness ($r_s = 0.30$, $P = 0.07$), number of sites at the border versus connectedness ($r_s = -0.17$, $P = 0.31$) and between average density of Natura 2000 sites in the 1 km border region of two bordering EU Member States versus connectedness ($r_s = 0.24$, $P = 0.17$) were not statistically significant.

Connectedness did not depend on political factors. We compared connectedness between boundaries where both countries were new EU Member States (joined in 2004, or 2007), where both were old Member States and where one was new and one was an old Member State (Kruskal–Wallis rank sum test, $P = 0.63$). Likewise, we compared connectedness of boundaries where both Member States represented the same versus different major language families (Wilcoxon ranks test, $P = 0.97$).

There were no significant correlation between the proportion of mountains at border and connectedness ($r = 0.10$, $P = 0.57$), but more rivers as borders significantly increased connectedness ($r_s = 0.61$, $P < 0.001$).

### 3.3. Connectivity

We identified 514 pairs of sites in the 1 km buffer of the 34 political borders of the European Union. These pairs included 3720 presences of selected species and 870 'matches', i.e., when the species was present in the sites in both sites of a pair. Such matches were present in 60% of site-pairs. On average there were 1.69 matches per site-pair; maximum number of matches per one site-pair was 15.

We further examined potential factors that might affect connectivity at the site-pair level using multi-factor modelling. Statistically significant predictors in the GLMM model with a binary response variable were the total number of features in the site-pair (more features, better connectivity), Member State pair classification according to EU enlargement (best connectivity on the borders of two old and worst connectivity on the borders of two new Member States), presence of river forming the border (better connectivity if a river is present) and time since designation of the last site in a pair (better connectivity if more time has passed since designation; Table 4).

The best numerical response model was similar but with the exception that Member State pair classification according to EU enlargement did not appear as a significant factor. Instead, the effect of potential language barrier was near-significant suggesting better connectivity on borders between nations belonging to the same language group. During the model selection procedure it became clear that both binary response and numerical response models have a good agreement as the effect of fixed variables had the same direction and similar significant predictors (Table 4).

Examining the time difference between the designation dates of sites participating in one site-pair, we found no significant differences between different border categories with relation to a date of membership in the European Union (means ± SE: old/new 41 ± 1.9 months, old/old 40 ± 1.7, new/new 38 ± 1.6) with an EU mean of 40.4 months, or 3.4 years.

Connectivity scores per border varied significantly across the European Union (Table 3; Fig. 4B) with a mean of 67.9%. Thirteen political borders had 100% connectivity, i.e. all site pairs on these borders had at least one common species. Connectedness positively correlated with connectivity ($r = 0.40$, $P = 0.017$). There were a few cases with better connectivity but relatively poor connectedness (e.g., Austrian/Hungarian, Austrian/Slovenian, Belgian/Luxembourg, French/Luxembourg, Hungarian/Romanian, Irish/United Kingdom and Italian/Slovenian) but very few examples where connectedness was good but connectivity poor (e.g., Czech/German and German/Danish; Table 3).

### 3.4. Species and taxonomic groups

The total of 870 matches was unevenly distributed among 192 species used in this study. Of these species, 85 had less than 5 presences in selected site-pairs, and of those 65 species had no matches. Only one species (stag beetle *Lucanus cervus*) was present in more than 200 site pairs (242) and only 11 species between 100 and 200 site-pairs. Individual taxonomic group connectivity averaged from 18.7% to 30.1% (Table 5). Each group had similar proportions of species with high and low individual connectivity scores. Notably, species with the highest individual connectivity tended

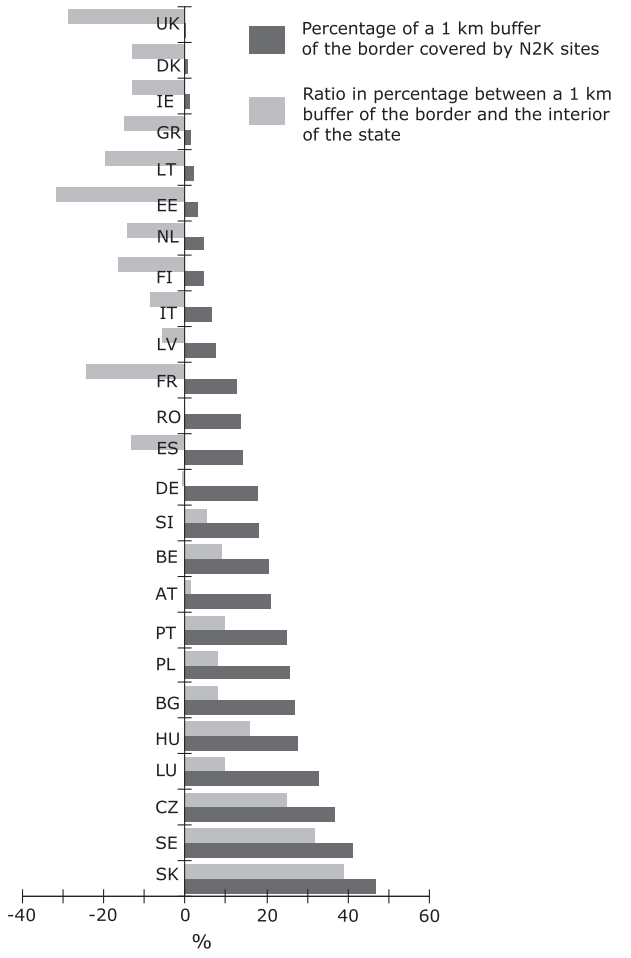



**Fig. 3.** Natura 2000 site cover in the 1 km buffer and the ratio between the cover in 1 km buffer versus the cover in the interior of EU Member States. Member States are arranged according to greatest proportions in selected buffer of borderland. Names of the European Union Member states follow the ISO 2-digit codes.

to be geographically localised, i.e., occurring in 2–3 Member States. The most common species at the border, those with presence in at least 50 site pairs, had individual connectivity values typically between 15% and 35%.

## 4. Discussion

Looking at the level of the European Union, both connectedness and connectivity varied greatly amongst borders (Table 3; Fig. 4), but were positively correlated. However, in few cases good connectedness was not associated with good connectivity. Moreover, very few political and environmental factors used to explain variation in trans-boundary connectedness and connectivity proved to be statistically significant in this study. Each of the 34 country borders has its own history with a unique subset of factors influencing Natura 2000 site selection and management with few common patterns and any generalisations should be made with care.

To interpret the above findings in a broader inter-disciplinary aspect, it is worthwhile to examine the historical factors that determined the presence of protected areas in the border areas which could influence connectedness and further connectivity. In Europe, until the 20th century, the most common motivations to preserve nature were religion, hunting interests and protection of the entire landscape for aesthetic reasons (Welzholz and Johann, 2007). Even the first protected areas, whose declared aim was biodiversity protection, were mostly created in areas which were too remote or unproductive to be important economically (Pressey et al., 1996), for example at high altitudes (Nilsson and Gotmark, 1992; Rosati et al., 2008). However, areas historically 'remote and unproductive' are not necessarily on the political borders as they are today. Many major cities and industry centres are not necessarily situated in the central geographical positions within Member States which would determine the locations of protected areas being more along country borders. Countless political events in 20th century Europe have resulted in changed political borders. Thus locations of many capitals and other cities, and even of early protected areas (many of which later became Natura 2000 sites), are actually older than the country borders themselves.

With the start of the building of the Natura 2000 network, conservation efforts at the European level became more systematic and concerted by ensuring that all species listed in Annex II of the EU Habitats Directive are adequately protected given their abundance, distribution and ecological and genetic variation (Evans, 2012). Thus the entire territory of each country had to be assessed, irrespective of economic or social interests. Fig. 3 shows that there was a high variation in the ratio between Natura 2000 site cover (%) at the border and in the interior. Ignoring the island states of United Kingdom and Ireland, and a number of other states with a proportionally large marine border versus terrestrial (e.g., Denmark, Greece, Finland and Sweden), Slovakia and the Czech Republic with their top positive ratio pose no surprise, as a large proportion of their boundaries are 'most natural', i.e., situated along natural barriers, such as mountains, where Natura 2000 sites would be expected to be especially numerous. In contrast in the Baltic region (Estonia, Latvia, and Lithuania) the ratio between the Natura 2000 cover in the border region versus the interior was negative. A possible explanation for the smaller number of sites on the borders of the Baltic region is that these countries still host a high level of biological diversity throughout the region. Some EU priority habitats according to Annex I of the Directive, for example active raised bogs, western taiga and bog woodland, are very widely distributed in this region either in the interior or at the border, therefore there was no specific reason to pay special attention to the border region. Moreover, in the Baltics many of the most important wetlands are not on political borders (Kimmel et al., 2010) and a relatively low proportion of the borders are formed by rivers. In addition, during the relatively recent Soviet period, Baltic terrestrial borders were not 'active borders' with an absence of border security and limitations in agricultural or industrial development.

It may seem logical that a greater proportion of border region covered by the Natura 2000 sites would lead to better connectedness and thus to high connectivity. This was true in many cases, but with certain exceptions. The large proportion of Natura 2000 sites along the Czech Republic's borders yielded only moderately good connectedness and bad connectivity at the Czech–German border (Table 3). The reason for this can be due to different habitat management on both sides of the border (e.g., Roth, 2005; Konvicka et al., 2008).

Many country borders follow rivers and mountain ridges. Therefore sites on different border sides can have different biotic and abiotic conditions (exposition, micro-climate, abrasion/accumulation, vegetation etc.) which can lead to different species occurring on each border side. In such cases neighbouring Member States would have different reasons for determining boundaries of sites on their respective side of the border. For example, the Arctic marsh grass *Arctophila fulva* is present in the Finnish site (FI1302105 Kainuunkylän saaret) and absent from the adjacent Swedish site (SE0820430 Torne och Kalix älvsystem) though both paired sites are situated by the Tornio River which forms the border between these countries. However, upon closer examination, it

**Table 3**
Connectedness and connectivity of 34 European Union internal terrestrial borders. Connectivity was positive for the site-pair if at least one common feature was found. Names of the EU Member states follow the ISO 2-digit codes.

| Border | Length (number of grids) | Number of sites within 1 km of border buffer | Connectedness; all Natura 2000 sites (%) | Connectivity; SCI only (%) |
|---|---|---|---|---|
| AT/CZ | 443 | 21 | 58.1 | 57.1 |
| AT/DE | 786 | 48 | 58.1 | 72.7 |
| AT/HU | 352 | 26 | 33.2 | 100.0 |
| AT/IT | 415 | 51 | 43.8 | 80.0 |
| AT/SI | 312 | 23 | 5.8 | 100.0 |
| AT/SK | 111 | 61 | 87.9 | 92.9 |
| BE/DE | 152 | 31 | 39.7 | 20.0 |
| BE/FR | 632 | 36 | 49.1 | 41.7 |
| BE/LU | 148 | 31 | 15.6 | 60.0 |
| BE/NL | 449 | 96 | 27.7 | 46.2 |
| BG/GR | 489 | 11 | 42.8 | 95.5 |
| BG/RO | 609 | 15 | 57.3 | 100.0 |
| CZ/DE | 743 | 11 | 47.7 | 18.9 |
| CZ/PL | 762 | 77 | 39.8 | 20.0 |
| CZ/SK | 241 | 71 | 24.0 | 63.6 |
| DE/DK | 68 | 64 | 32.3 | 0.0 |
| DE/FR | 444 | 48 | 61.5 | 95.2 |
| DE/LU | 134 | 19 | 72.2 | 77.8 |
| DE/NL | 569 | 47 | 22.3 | 46.7 |
| DE/PL | 431 | 25 | 72.7 | 60.0 |
| EE/LV | 338 | 48 | 44.6 | 8.3 |
| ES/FR | 791 | 52 | 63.3 | 52.6 |
| ES/PT | 1198 | 61 | 65.9 | 100.0 |
| FI/SE | 541 | 39 | 98.4 | 100.0 |
| FR/IT | 508 | 13 | 49.0 | 100.0 |
| FR/LU | 76 | 13 | 8.0 | 100.0 |
| HU/RO | 446 | 13 | 38.8 | 100.0 |
| HU/SI | 85 | 13 | 57.3 | 100.0 |
| HU/SK | 624 | 9 | 68.6 | 100.0 |
| IE/UK | 455 | 19 | 69.6 | 100.0 |
| IT/SI | 226 | 23 | 49.5 | 100.0 |
| LT/LV | 533 | 70 | 9.0 | 0.0 |
| LT/PL | 97 | 35 | 4.7 | 0.0 |
| PL/SK | 505 | 6 | 63.1 | 100.0 |

can be seen that the Swedish site includes only freshwater habitats, i.e., water column, while the Finnish site includes a variety of other habitats including marshes which are the habitat for the Arctic marsh grass.

Therefore connectedness and connectivity as such across country borders cannot be considered 'bad' or 'good' (although we use such expressions before and below in this paper, meaning low or high scores), and good connectivity is not always achievable. Moreover, depending on the biology of the species, connectivity can be both positively and negatively related to species conservation (Taylor et al., 2006). Thus we refrained from setting thresholds satisfactory connectedness and connectivity, and the figures obtained from our calculations serve rather as a comparison between the individual European Union terrestrial borders and between regions.

Using GLMM to study connectivity at the site-pair level we found, not surprisingly, that if more protected species were in a site-pair there was a higher chance of having at least one or more of them in both sites. Presence of a river as the border positively affected connectivity. However, rivers can be both obstacles and agents in species dispersal. Factors that can affect dispersal, i.e., river width and depth, water velocity (Riis and Sand-Jensen, 2006) spatial arrangement and overall proportion of dead water zones (Bond et al., 2000), dispersal phenology and weather (Meritt and Wohl, 2002), presence of animals, such as birds, as agents of dispersal (Clausen et al., 2002), interplay with different dispersal strategies of different organisms included in this study. In general most streams are dynamic ecosystems with changing water levels and disturbances which are temporarily unpredictable (Riis and Sand-Jensen, 2006). Positive effect of rivers on connectivity in this study can be explained by the fact that the used river database (ECRINS) contained also relatively small streams (down to few metres wide) and thus many such streams could not act as serious barriers even for poor dispersers. On the other hand, the river presence itself would ensure higher diversity of habitats and likely also protected species richness in site-pairs and thus connectivity. Nevertheless, it is uncertain whether this pattern would repeat elsewhere if similar study was repeated.

'Political' factors had a less pronounced effects on connectivity. Better connectivity was achieved when a site-pair was represented by the old EU Member States in the binary response model and the language barrier seemed to be a problem in the numerical response model (Table 4). Although each of these variables appears only in one of the final models, this pattern was present in all tested models that included these variables. Surprisingly, the time gap between designations of paired sites did not play a considerable role on connectivity. If there is a large time gap between designations, the Member State designating the second site has an opportunity to examine the Standard Data Forms of the already designated sites from the other side of the border and to take this information into account in its own survey planning and decision-making. On the contrary, if both Member States are working in parallel with similar time schedules, it is possible to co-operate in field investigations, information exchange and decision-making. Our study showed that neither of these assumptions worked. In fact, of 'political' factors, a better predictor of connectivity was the time since the second site of a pair was designated. This can be explained by the fact that very often inventories and other studies in Natura 2000 sites continue after the designation, thus improving knowledge of the species that are present. Neighbouring Member

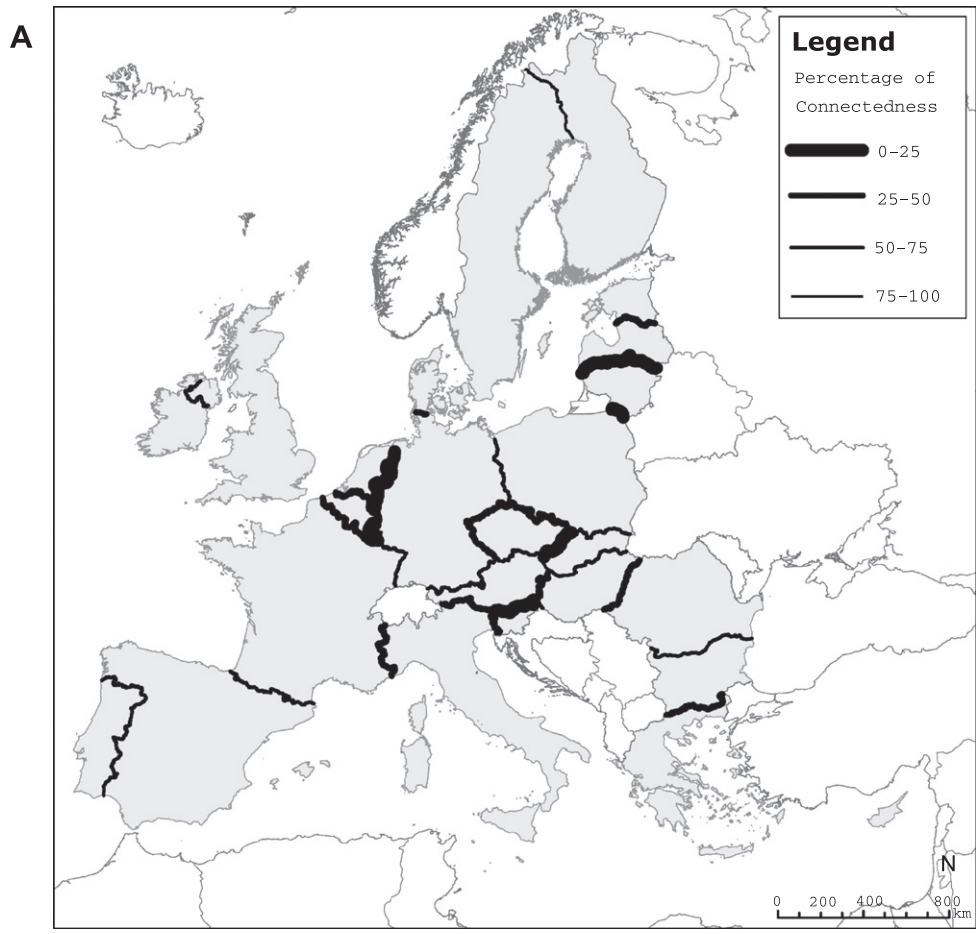



**Fig. 4.** Connectedness (A) and connectivity (B) of Natura 2000 sites across country borders in different parts of the European Union. Worse connectedness and connectivity scores are marked by thicker lines.

States might have had an opportunity to organise a search for 'missing species'. Thus, the longer the time period since both sites 'were in place', apparently, more complete were these inventories. This suggests that the presence of 'documented connectivity' is more linked with available scientific knowledge about sites rather than the right choice of Natura 2000 site boundaries at the time of site designation.

How far connectedness and connectivity can be influenced by an intended co-operation between neighbouring countries? According to the few published self-assessments, there was no cooperation between Czech Republic and Austria (Roth, 2005) and it brought 'below average' values of connectedness and connectivity (Table 3). There was very good cooperation between the Czech Republic and Germany resulting in average connectedness but not good connectivity. On the contrary, in the case of several good trans-border projects between Hungary and Romania (Ilies et al., 2010) there was poor connectedness but 100% connectivity (Table 3). The cooperation between Germany and Poland was difficult because the countries were at different stages of the Natura 2000 designation process, nevertheless the result was good connectedness and connectivity. Cooperation between Poland and Czech Republic was not possible because of different approaches in site designation process in these states, even though both countries represented the same European Union enlargement group in 2004 (Makomaska-Juchiewicz, 2005). Consequently the Czech–Polish border had low connectedness and connectivity (Table 3, Fig. 4).

Overall, there is an opinion that, in general, cross-border cooperation for establishing ecological networks in Europe is not well developed: most plans are being or have been developed only at the regional or sub-national level (Jongman et al., 2004). It seems that often cooperation is mainly 'at the Park level', i.e., it deals with large protected areas, like National Parks, but less regarding the smaller Natura 2000 sites which may not have recreational value. However, our data seem to show that, even though cooperation may have been lacking, especially at the official level, the Habitats Directive requirement that Natura 2000 sites must be selected on purely scientific criteria within Member States must have brought results that probably seem to be adequate. In 13 out of 34 borders connectivity measure was 100% and in 11 another borders over 50% and this seems to be a good result. Clearly the bio-geographical seminars (Evans, 2012) have had a positive impact on connectedness and connectivity. These seminars were held once most Member States in a bio-geographical region had proposed a significant share of potential SCIs. At the bio-geographical seminars, Member States had an opportunity to look at the trans-boundary issues and fill in the gaps through designation of additional sites in the following years.

This study did not reveal substantial differences in trans-border connectivity between different taxonomic groups (Table 5), while great variation existed within taxonomic groups, obviously depending on the dispersal strategy and ability of individual species. It is remarkable that the distribution of the best-performing species was limited to a few Member States. This suggests that the

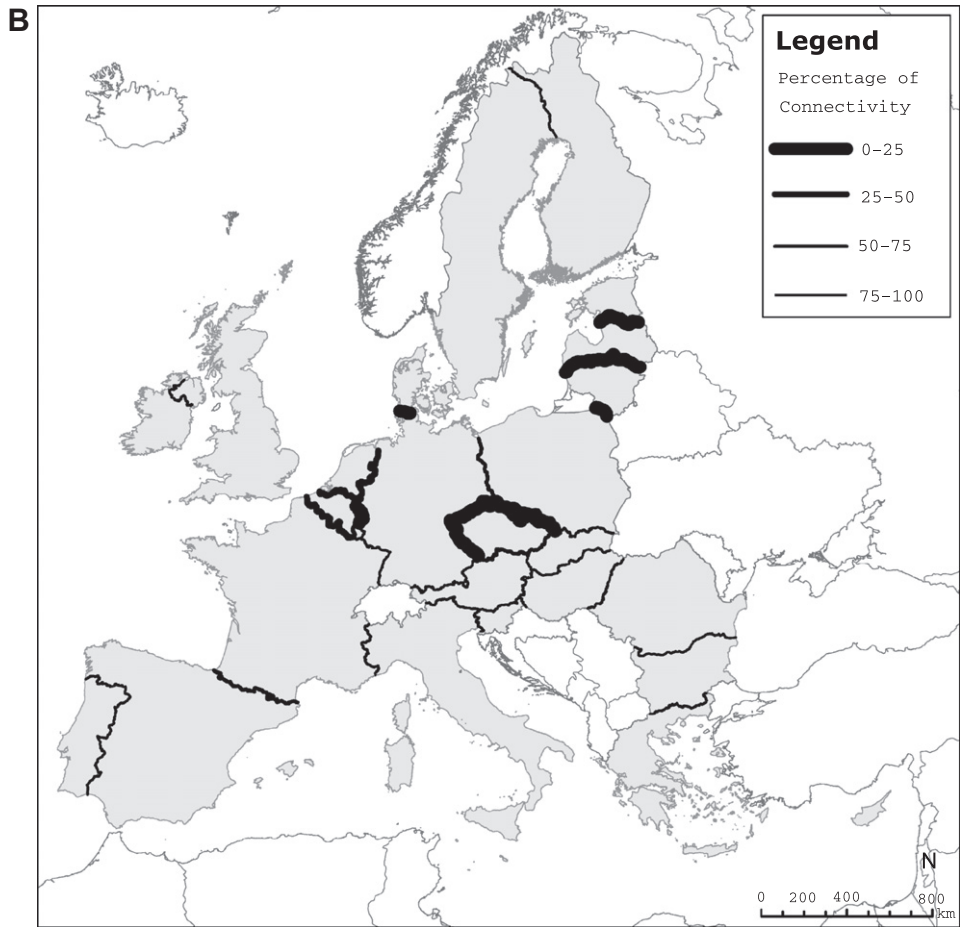



**Fig. 4** (*continued*)

**Table 4**
Factors affecting connectivity in cross-border site-pairs.

| Model parameter | Binary response model | | | Numerical response model | | |
|---|---|---|---|---|---|---|
| | Estimate | SD | *P* value | Estimate | SD | *P* value |
| Intercept | 82.7536 | 0.7082 | 0.0001 | −1.2562 | 0.1974 | 1.98e-10 |
| FEATOT | 0.4443 | 0.0510 | <2e-16 | 0.0977 | 0.0152 | <2e-16 |
| LANGUA (same) | | | | 0.2980 | 0.1686 | 0.08 |
| ENLARG (different) | −1.6157 | 0.7870 | 0.04 | | | |
| ENLARG (new) | −1.9709 | 0.7565 | 0.01 | | | |
| RIVER (yes) | 0.9133 | 0.3579 | 0.01 | 0.2650 | 0.0882 | 0.003 |
| LASTTIM | 0.0105 | 0.0061 | 0.09 | 0.0034 | 0.0019 | 0.08 |
| Intercept variance (MSPAIR) | 1.3168 | 1.1475 | | 0.1052 | 0.3243 | |
| Log likelihood | −167.6 | | | −235.9 | | |
| AIC | 349.1 | | | 483.7 | | |

responsible Member States were aware about the high priority of detecting and protecting these species, and this is very important for their global survival probability (Schmeller et al., 2008).

A few limitations of this study should be highlighted. In the European Union, the Natura 2000 network makes a crucial contribution to the protection of the core areas in the Pan-European Ecological Network (Bonnin et al., 2007), thus it is important to be aware that this study looked only at a part of the European ecological network. Other existing elements, such as nationally protected areas, corridors and stepping stones that are outside Natura 2000 network, possibly ensure higher connectivity than observed in this study.

Protected area networks, and other sites of key importance to biodiversity are often viewed only as large islands in a mixed landscape matrix (Donald and Evans, 2006). We investigated connectedness and connectivity of such 'islands' across country borders which are 'artificial lines' in an ecological sense, because these lines could be bottlenecks for connectivity at the European Union level. We assume that within-country connectedness and connectivity must be ensured relying on the understanding of this necessity by the nature conservation administrations in each European Union Member State. This, however, needs confirmation through further studies.

**Table 5**
Connectivity in different taxonomic groups. Panel A: Average connectivity value between groups. Panel B: Best performing species and their distribution pattern. Names of the European Union Member States follow the ISO 2-digit codes.

| Taxonomic group | Total site-pairs | Matches | | % Connectivity | |
|---|---|---|---|---|---|
| *Panel A* | | | | | |
| Amphibians and reptiles | 883 | 266 | | 30.1 | |
| Invertebrates | 2131 | 472 | | 22.1 | |
| Plants | 693 | 130 | | 18.7 | |
| Species | Tax. group | Total | Matches | % Connectivity | Distribution |
| *Panel B* | | | | | |
| *Carabus menetriesi pacholei* | Invertebrate | 6 | 4 | 66.7 | AT, CZ, DE |
| *Ophiogomphus serpentinus* | Invertebrate | 2 | 1 | 50.0 | DE, FR |
| *Gentiana ligustica* | Plant | 26 | 12 | 46.2 | FR, IT |
| *Hydromantes strinatii* | Amphibian | 26 | 12 | 46.2 | FR, IT |
| *Tozzia carpatica* | Plant | 5 | 2 | 40.0 | PL, SK |
| *Bombina bombina* | Amphibian | 197 | 78 | 39.6 | Many |
| *Androsace pyrenaica* | Plant | 18 | 7 | 38.9 | ES, FR |
| *Stenobothrus eurasius* | Invertebrate | 8 | 3 | 37.5 | CZ, HU, SK |
| *Cucujus cinnaberinus* | Invertebrate | 52 | 19 | 36.5 | Many |
| *Mauremys leprosa* | Reptile | 55 | 19 | 34.5 | ES, FR, PT |
| *Chioglossa lusitanica* | Invertebrate | 12 | 4 | 33.3 | ES, PT |
| *Erebia calcaria* | Invertebrate | 6 | 2 | 33.3 | AT, IT, SI |
| *Lacerta schreiberi* | Reptile | 33 | 11 | 33.3 | ES, PT |
| *Saxifraga florulenta* | Plant | 6 | 2 | 33.3 | FR, IT |
| *Lacerta bonnali* | Reptile | 34 | 11 | 32.4 | ES, FR |

One of the main problems in the evaluation of protected area efficiency is the absence of systematically collected data about the subjects of site designation (e.g., Gaston et al., 2006; Ioja et al., 2010). Also in this case, the only way to study connectivity of the Natura 2000 network given the limitations of our data source (Natura 2000 database) was to choose dispersal success, based on evidence that the same species is present in two neighbouring sites. Ideally, the study should have been implemented by looking at how Natura 2000 sites match with the detailed mapping of species distributions or species' habitats. Unfortunately, such information is not available at the EU-wide scale for all the species involved. In our opinion, our method was a good way forward in the absence of such ideal data. However, there is also temporal uncertainty about our findings on connectivity: the fact, that same species occur in both sites of a trans-border pair, does not necessarily mean that connectivity exists today as species may have travelled between sites a long time ago. So our work relates to both existing and existed connectivity but the relationships between the two remain unknown.

In this study we relied on the data recorded in the Standard Data Forms and provided by the EU Member States. It is known that not all SCIs in different countries have been surveyed systematically with the same effort and precision. Therefore trans-boundary connectedness in this study may have been overestimated (due to historical data which was used and the fact that in some cases the species may no longer occur in some Natura 2000 sites) or underestimated (as some sites may not have been surveyed with appropriate method, timing or effort to detect all possible species of the Annex II of the Habitats Directive). We, however, have good reason to believe in the underestimation of connectivity as most EU Member States are running Natura 2000 survey and monitoring schemes and this will likely lead to new discoveries of species and consequently to an improvement of documented connectivity.

The application of our method to study connectedness and connectivity across country borders elsewhere in the World would perhaps bring interesting results at a large (continental or sub-continental) scale, but provided that similar (or better) data collection pattern and recording, as in case of Natura 2000, exists. Within the European Union, similar assessment could be done across borders of provinces of large federal states (e.g., Germany, Spain), since it is possible that these boundaries may pose similar threat to species dispersal as the country borders (Ssymank, 2005).

## Acknowledgements

Data collection and analysis was performed as a part of the work of the European Topic Centre on Biological Diversity for the European Environment Agency report on the Protected Areas in Europe. The views expressed are those of the authors and should not be taken as views of the European Topic Centre on Biological Diversity. We thank Doug Evans, Carlos Romao and Michael Hosek for useful comments on earlier drafts of this paper.

## References

Araujo, M.B., Lobo, J.M., Moreno, J.C., 2007. The effectiveness of iberian protected areas in conserving terrestrial biodiversity. Conserv. Biol. 21, 1423–1432.
Ardron, J.A., 2008. The challenge of assuring whether the OSPAR network of marine protected areas is ecologically coherent. Hydrobiologia 606, 45–53.
Arntzen, J.W., Wallis, G.P., 1991. Restricted gene flow in a moving hybrid zone of the newts Triturus cristatus and Triturus marmoratus in western France. Evolution 45, 805–826.
Baguette, M., Schtickzelle, N., 2006. Negative relationship between dispersal distance and demography in butterfly metapopulations. Ecology 87, 648–654.
Bates, D., Maechler, M., Bolker, M., 2011. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-41. <http://CRAN.R-project.org/package=lme4>.
Baudry, J., Merriam, H.G., 1988. Connectivity and connectedness: functional versus structural patterns in landscapes. In: Schreiber, K.-F. (Ed.), Connectivity in Landscape Ecology. Proceedings of the Second International Seminar of the International association for Landscape Ecology, Munster, 1987, vol. 29. Munsterusche Geographische Arbeiten, pp. 23–28.
Bennett, F.A., 2003. Linkages in the landscape. The role of corridors and connectivity in wildlife conservation. In: IUCN, IUCN Forest Conservation Programme. Conserving Forest Ecosystems Series No. 1, Australia.
Boitani, L., Falcucci, A., Mariorano, L., Rondinini, C., 2007. Ecological networks as conceptual frameworks or operational tools in conservation. Conserv. Biol. 21, 1414–1422.
Bonato, L., Francasso, G., 2003. Movements, distribution pattern and diversity in a population of Salamandra atra aurorae (Caudata: Salamandridae). Amph.-Reptilia 24, 251–260.
Bond, N., Perry, G.L.W., Downes, B.J., 2000. Dispersal of organisms in a patchy stream environment under different scenarios. J. Anim. Ecol. 69, 608–619.
Bonnin, M., Bruszik, A., Lethier, H., Richard, D., Terry, A., 2007. The Pan-European Ecological Network: taking stock. Council of Europe Publishing, Nature and Environment, No. 146.
Brzosko, E., Wroblewska, A., Ratkiewicz, M., 2002. Spatial genetic structure and clonal diversity of island populations of lady's slipper (Cypripedium calceolus) from the Biebrza National Park (northeast Poland). Mol. Ecol. 11, 2499–2509.
Campagne, P., Affre, L., Baumel, A., Roche, P., Tatoni, T., 2009. Fine-scale response to landscape structure in Primula vulgaris Huds.: does hedgerow network connectedness ensure connectivity though gene flow? Popul. Ecol. 51, 209–219.
Caughley, G., Sinclair, A.R.E., 1994. Wildlife Ecology and Management. Blackwell Scientific Publications, Boston, MA.

Celik, T., Vres, B., Seliskar, A., 2009. Determinants of within-patch microdistribution and movements of endangered butterfly Coenonympha oedippus (Fabricius, 1787) (Nymphalidae: Satyrinae). Hacquetia 8, 115–128.
Cerovsky, J., 1996. Biodiversity conservation in Transboundary protected areas in Europe. Praha.
Chape, S., Spalding, M., Jenkins, M.D., 2008. The World's protected areas. Prepared by the UNEP World Conservation Monitoring Centre. University of California Press, Berkeley.
Clausen, P., Nolet, B.A., Fox, A.D., Klaasen, M., 2002. Long-distance endozoochorus of submerged macrophyte seeds by migratory waterbirds in Northern Europe – a critical review of possibilities and limitations. Acta Oecol. 23, 191–203.
Donald, P.F., Evans, A.D., 2006. Habitat connectivity and matrix restoration: the wider implications of agri-environment schemes. J. Appl. Ecol. 43, 209–218.
EEA, 2010b. Europe's ecological backbone: recognising a true value of mountains. European Environment Agency. Copenhagen.
European Commission, 1992. Council Directive 79/409/EEC of April 2 1979, on the conservation of wild birds. European Community Environment Legislation, vol. 4. pp. 2–49.
European Commission, 1996. Council Directive 92/43/EEC of 21 May 1992, on the conservation of natural habitats of wild fauna and flora. European Community Environment Legislation, vol. 4. pp. 81–158.
European Commission, 1997. Commission decision of 18 December 1996 concerning a site information for proposed SCIs. Official Journal of the European Communities. (24.04.97).
Evans, D., 2012. Building the European Union's Natura 2000 network. Nat. Conserv. 1, 11–26.
Fall, J.J., 2003. Planning protected areas across boundaries. New paradigms and old ghosts. J. Sustain. Forest. 17, 81–102.
Fischer, K., Beinlich., B., Plachter, H., 1999. Population structure, mobility and habitat preferences of the violet copper Lycaena helle (Lepidoptera: Lycaenidae) in Western Germany: implications for conservation. J. Insect Conserv. 3, 43–52.
Gaston, K., Charman, K., Jackson, S.F., Armsworth, P.R., Bonn, A., Briers, R.A., Callaghan, C.S.Q., Catchpole, R., Hopkins, J., Kunin, W.E., Latham, J., Opdam, P., Stoneman, R., Stroud, D.A., Tratt, R., 2006. The ecological effectiveness of protected areas: the United Kingdom. Biol. Conserv. 132, 76–87.
Gaston, K.J., Jackson, S.F., Nagy, A., Cantu-Salazar, L., Johnson, M., 2008. Protected areas in Europe. Principle and practice. Ann. N.Y. Acad. Sci. 1134, 97–119.
Gatter, W., 1997. Förderungsmöglichkeiten für den Alpenbock. Allg Forst Z Waldwirtsch Umweltvorsorge 213, 1305–1306.
Hamilton, L.S., 1998. Guidelines for effective trans-boundary cooperation: philosophy and best practices. In: Parks for Peace. Proceedings of the International Conference on Transboundary Protected Areas as a Vehicle for International Cooperation, 1997.
Hanks, J., 2003. Transfrontier Conservation Areas (TFCAs) in southern Africa: their role in conserving biodiversity, socioeconomic development and promoting a culture of peace. J. Sustain. Forest. 17, 127–148.
Hansson, L., 1991. Dispersal and connectivity in metapopulations. Biol. J. Linn. Soc. 42, 89–103.
Ilies, D.C., Blaga, L., Josan, I., Baias, S., Morar, C., Herman, G., 2010. Cross border natural parks, support for regional development, case study of the Northern and Western Romanian border. Revista Romana de Geografie Politica 12, 126–139.
Ioja, C.I., Patroescu, M., Rozylowicz, L., Popescu, V.D., Verghelet, M., Zotta, M.I., Felciuc, M., 2010. The efficacy of Romania's protected areas network in conserving biodiversity. Biol. Conserv. 143, 2468–2476.
IUCN, 1994. United Nations List of National Parks and Protected Areas. Prepared by WCMC and CNPPA. IUCN, Gland, Switzerland and Cambridge, UK.
Jantke, K., Schleupner, C., Schneider, U.A., 2011. Gap analysis of European wetland species: priority regions for expanding the Natura 2000 network. Biodivers. Conserv. 20, 581–605.
Johnson, M.P., Crowe, T.P., McAllen, R., Allcock, A.L., 2008. Characterising the marine Natura 2000 network for the Atlantic region. Aquatic Conserv.: Mar. Fresh. Ecosyst. 18, 86–97.
Jongman, R.H.G., Kulvik, M., Kristiansen, I., 2004. European ecological networks and green ways. Landscape and Urban Plan. 68, 305–319.
Junker, M., Schmitt, T., 2010. Demography, dispersal and movement pattern of Euphydryas aurinia (Lepidoptera: Nymphalidae) at the Iberian Penninsula: an alarming example in an increasingly fragmented landscape? J. Insect Conser. 14, 237–246.
Kimmel, K., Kull, A., Salm, J.-O., Mander, U., 2010. The status, conservation and sustainable use of Estonian wetlands. Wetlands Ecol. Manage. 18, 375–395.
Kiviniemi, K., Eriksson, O., 1999. Dispersal, recruitment and site occupancy of grassland plants in fragmented habitats. Oikos 86, 241–253.
Konvicka, M., Benes, J., Cizek, O., Kopecek, F., Konvicka, O., Vitaz, L., 2008. How too much care kills species: grassland reserves, agri-environmental schemes and extinction of Colias myrmidone (Lepidoptera: Pieridae) from its former stronghold. J. Insect Conserv. 12, 519–525.
Kupfer, A., 1998. Wanderstrecken einzelner kammolche (Triturus cristatus) in einem agrarlebensraum. Z. Feldherpethol. 5, 238–242.
Leibenath, M., Rientjes, S., Lintz, G., Kolbe-Weber, C., Walz, U. (Eds.), 2005. Crossing Borders: Natura 2000 in the Light of EU Enlargement. European Centre for Nature Conservation. The Netherlands.
Makomaska-Juchiewicz, M., 2005. Cross-border implementation and coherence of Natura 2000 in Poland. In: Leibenath, M., et al. (Eds.), Crossing borders: Natura 2000 in the Light of EU enlargement. Proceedings of an International Workshop. Dresden, pp. 13–19.
Maiorano, L., Falcucci, A., Garton, E.O., Boitani, L., 2007. Contribution of the Natura 2000 network to biodiversity conservation in Italy. Conserv. Biol. 21, 1433–1444.
Meritt, D.M., Wohl, E.E., 2002. Processes governing hydrochory along rivers: hydraulics, hydrology, and dispersal phenology. Ecol. Appl. 12, 1071–1087.
Minghi, J.V., 1963. Boundary studies in political geography. Ann. Assoc. Am. Geogr. 53, 407–428.
National Geographic Society, 2005. Europe: mother tongues (map). National Geographic: April supplement.
Nowicki, P., Witek, M., Skorka, P., Settele, J., Woyciechowski, M., 2005. Population ecology of the endangered butterflies Maculinea teleius and M. Nausithous and the implications for conservation. Popul. Ecol. 47, 193–202.
Niggebrugge, K., Durance, I., Watson, A.M., Leuven, R.S.E.W., Ormerod, S.J., 2007. Applying landscape ecology to conservation biology: spatially explicit analysis reveals dispersal limits of threatened wetland gastropods. Biol. Conserv. 139, 286–296.
Nilsson, C., Gotmark, F., 1992. Protected areas in Sweden: is natural variety conservation adequately represented. Conserv. Biol. 6, 232–242.
Parrish, J.D., Braun, D.P., Unnasch, R.S., 2003. Are we conserving what we say we are? Measuring ecological integrity within protected areas. Bioscience 53, 851–860.
Pressey, R.L., Ferrier, S., Hager, T.C., Woods, C.A., Tully, S.L., Weinman, K.M., 1996. How well protected are the forests of north-eastern New South Wales? — Analyses of forest environments in relation to formal protection measures, land tenure, and vulnerability to clearing. For. Ecol. Manage. 85, 311–333.
Purse, B.V., Hopkins, G.W., Day, K.J., Thompson, D.J., 2003. Dispersal characteristics and management of a rare damsefly. J. Appl. Ecol. 40, 716–728.
R Development Core Team, 2011. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
Ranius, T., Hedin, J., 2001. The dispersal rate of a beetle, Osmoderma eremita, living in tree hollows. Oecologia 126, 363–370.
Riis, T., Sand-Jensen, K., 2006. Dispersal of plant fragments in small streams. Freshw. Biol. 51, 274–286.
Rink, M., Sinsch, U., 2007. Radio-telemetric monitoring of dispersing stag beetles: implications for conservation. J. Zool. 272, 235–243.
Rosati, L., Margignani, M., Blasi, C., 2008. A gap analysis comparing Natura 2000 vs National Protected Area network with potential natural vegetation. Commun. Ecol. 9, 147–154.
Roth, P., 2005. Cross-border implementation and coherence of Natura 2000 in the Czech Republic. In: Leibenath, M., et al. (Eds.), Crossing borders: Natura 2000 in the Light of EU enlargement. Proceedings of an International Workshop. Dresden, pp. 21–24.
Ruckelhaus, M., Hartway, C., Kareiva, P.M., 1997. Assessing the data requirements of spatially explicit dispersal models. Conserv. Biol. 11, 1298–1306.
Russo, D., Cistrone, L., Garonne, A.P., 2011. Habitat selection by the highly endangered long-horned beetle Rosalia alpina in Southern Europe: a multiple scale assessment. J. Insect Conserv. 15, 685–693.
Schabetsberger, R., Jehle, R., Maletzky, A., Pesta, J., Sztatecsny, M., 2004. Delineation of terrestrial reserves for amphibians: post-breeding migrations of Italian crested newts (Triturus c. carnifex) at high altitude. Biol. Conserv. 117, 95–104.
Schmeller, D.S., Gruber, B., Budrys, E., Framsted, E., Lengyel, S., Henle, K., 2008. National responsibilities in European Species Conservation: a methodological review. Conserv. Biol. 22, 593–601.
Ssymank, A., 2005. Cross-border implementation and coherence of Natura 2000 in Germany. In: Leibenath, M., et al. (Eds.), Crossing borders: Natura 2000 in the Light of EU enlargement. Proceedings of an International Workshop. Dresden, pp. 25–39.
Sundblad, G., Bergstrom, U., Sandstrom, A., 2011. Ecological coherence of marine protected area networks: a spatial assessment using species distribution models. J. Appl. Ecol. 48, 112–120.
Taylor, P.D., Fahrig, L., With, K.A., 2006. Landscape connectivity: a return to the basics. In: Crooks, K.R., Sanjayan, M. (Eds.), Connectivity Conservation. Cambridge University Press, Cambridge, United Kingdom, pp. 29–43.
Tischendorf, L., Fahrig, L., 2000. How should we measure landscape connectivity? Landscape Ecol. 15, 633–641.
Welzholz, J.C., Johann, E., 2007. History of Protected Forest Areas in Europe. In: Frank, G., Parviainen, J., Vandekerhove, K., Latham, J., Schuck, A., Little, D. (Eds.), Protected Forest Areas in Europe: Analysis and Harmonization (PROFOR): Results, Conclusions and Recommendations. pp. 17–39.
Wurzel, R.K.W., 2008. European Union environmental policy and Natura 2000. In: Keulartz, J., Leistra, G., (Eds.), Legitimacy in European Nature Conservation Policy: Case Studies in Multilevel Governance. pp. 259–282.
Zimmerer, K.S., Galt, R.E., Buck, M.V., 2004. Globalization and multi-spatial terends in the coverage of protected-area conservation (1980–2000). Ambio 33, 520–529.

## *Web references*

ECRINS, 2012. Information: <http://sia.eionet.europa.eu/Thematic_Assessment/ECRINS>, data: <http://projects.eionet.europa.eu/ecrins> (accessed 12.03.12.).
EEA, 2010a. http://www.eea.europa.eu/data-and-maps/data/natura-2000 (accessed 12.03.12.).
IUCN and WCPA, 2001. Transboundary Protected Areas for Peace and Cooperation. <http://www.iucn.org> (accessed 13.03.12).

BASIC SCIENCE

Nanomedicine: Nanotechnology, Biology, and Medicine
9 (2013) 65–73

Research Article

nanomedjournal.com

# Fibronectin-binding nanoparticles for intracellular targeting addressed by *B. burgdorferi* BBK32 protein fragments

Renate Ranka, PhD[a,b,*], Ivars Petrovskis, PhD[a], Irina Sominskaya, PhD[a], Janis Bogans, MSc[a], Ruta Bruvere, PhD[a], Inara Akopjana, BSc[a], Velta Ose, PhD[a], Irena Timofejeva, MSc[a], Kalvis Brangulis, MSc[a], Pauls Pumpens, PhD[a], Viesturs Baumanis, PhD[a,b]

[a]*Latvian Biomedical Research and Study Centre, Riga, Latvia*
[b]*University of Latvia, Riga, Latvia*

Received 22 December 2011; accepted 2 May 2012

**Abstract**

Virus-like particles (VLPs) are created by the self-assembly of multiple copies of envelope and/or capsid proteins from many viruses, mimicking the conformation of a native virus. Such noninfectious nanostructures are mainly used as antigen-presenting platforms, especially in vaccine research; however, some of them recently were used as scaffolds in biotechnology to produce targeted nanoparticles for intracellular delivery. This study demonstrates the creation of fusion VLPs using hepatitis B core protein-based system maintaining a fibronectin-binding property from *B. burgdorferi* BBK32 protein, including the evidence of particles' transmission to BHK-21 target cells via caveolae/rafts endocythosis. These results make this construct to be an attractive model in development of HBc-based nanoparticles for cellular targeting applications and highlights the fragment of *B. burgdorferi* BBK32 as a novel cellular uptake-promoting peptide.

***From the Clinical Editor:*** This paper discusses the nanotechnology-based application of self-assembling viral-like peptides (VLP-s) for targeted delivery using a hepatitis B core protein based system. Creating fusion VLPs may be an attractive model for cellular targeting applications.


*Key words:* Fibronectin; Nanoparticles; *B. burgdorferi*

Various strategies in the design of nanoparticles (NPs) — particles in the size range 1 – 1000 nm — aim to create new generations of drug-delivery vehicles, contrast agents, and diagnostic devices.[1] One of the relevant potentials of the nanomedicines is their intracellular targeting possibility. Effective intracellular drug delivery is important for therapeutic agents that have specific molecular targets inside a cell as well as for drugs that undergo extensive efflux from the cell by the efflux transporters. Thus, the ability to penetrate inside cells bypassing lysosomal degradation is one of the key problems in the rational design of pharmaceutical nanocarriers.[2] For this purpose, the surface of nanocarriers could be modified by certain internalizable ligands (e.g., folate, transferrin) or by cell-penetrating peptides, such as a trans-activating transcriptional activator (TaT), an integrin-binding peptide (RGD peptide) or polyArginine.[3] This approach is receiving increasing attention over the last years because it is efficient for a range of cell types, and various endocytotic mechanisms can be engaged to facilitate the internalization of a carrier.[2]

Virus-like particles (VLPs) is a broad group of nanocarrier systems that exhibit great potential in biomedicine research, including targeted drug delivery.[4] VLPs are self-assembling noninfectious supramolecular structures that have been produced from structural proteins of a wide variety of virus families. The unique features of VLPs are proper dimensions for nanoscale applications, size homogeneity, a large surface area-to-mass ratio, a symmetric macromolecular organization, biodegradability, biocompatibility and ease of production/purification.[5] In addition, a number of targeting molecules and peptides can be displayed in a biologically functional form on their capsid surface using chemical or genetic means.[1] However, the issues of potential toxicity, biodistribution, clearance rates, stability, and immunogenicity are not fully addressed for these novel molecular devices for in vivo biomedical applications.

No conflict of interest was reported by the authors of this article.

This work was supported by grant of Latvian Council of Science, Nr. 10.0029.3 and by ESF Project No. 1DP/1.1.2.0/09/APIA/VIAA/150.

*Corresponding author: Latvian Biomedical Research and Study Centre, Riga, LV-1067, Latvia.

*E-mail address:* renate_r@biomed.lu.lv (R. Ranka).

1549-9634/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.nano.2012.05.003

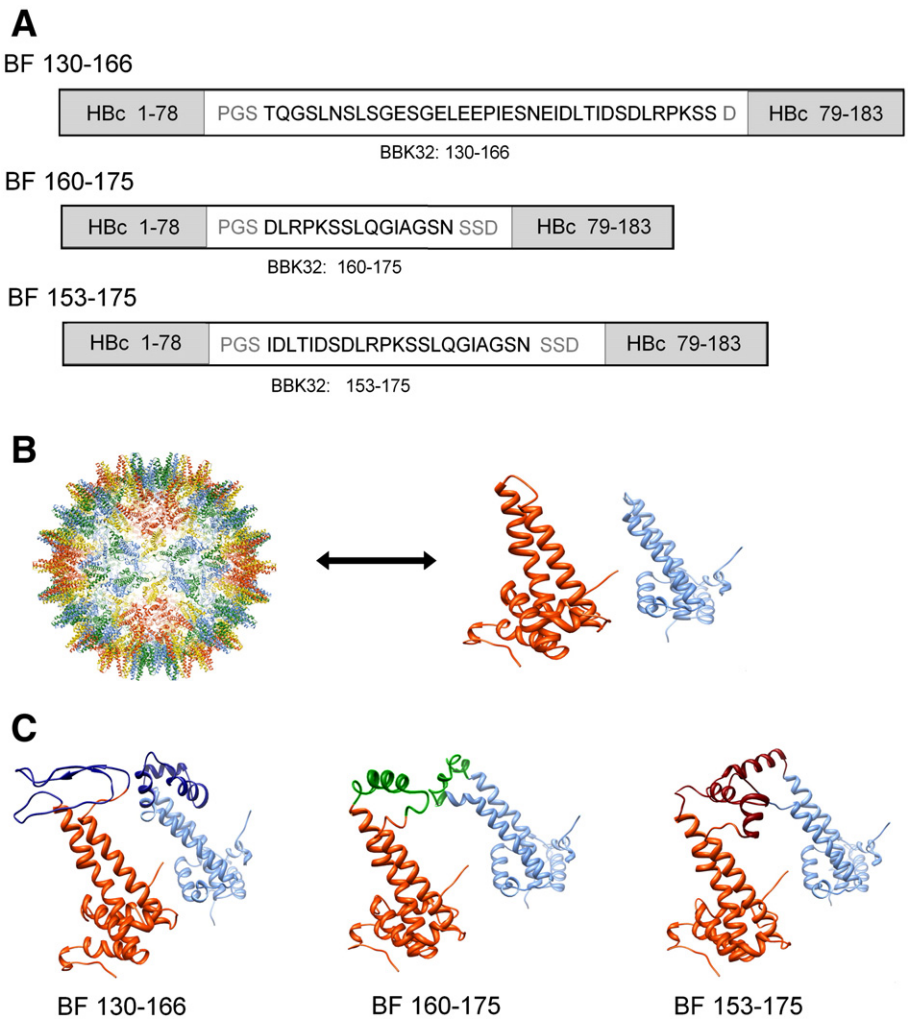



Figure 1. **(A)** Schematic presentation of the fusion HBc/BBK32 constructs used in this study. BBK32 fragments were inserted in MIR of HBc protein. *Gray boxes*: HBc fragments. *White boxes*: BBK32 fragments. Abbreviations of grey color represents linkers (amino acids GS and SS), and amino acids, added due to introduction of the *Bam*HI restriction site (amino acids P and D). **(B)** Three-dimensional prediction of the HBc VLP and individual HBc protein chains. **(C)** Three-dimensional prediction of the individual HBc protein chains carrying different fragments of the BBK32 protein. Predictions were performed on the basis of the x-ray structure of the HBc, genotype A,[7] by a comparative modeling program 3D-JIGSAW (http://bmm.cancerresearchuk.org/3djigsaw/)[30] and presented by Chimera software.[31] Chains A (*orange red*) and D (*cornflower blue*) of the HBc asymmetric tetramer unit are presented, but B and C chains are omitted for visual clarity. Three BBK32 protein insertions are shown by blue, green and red.

Hepatitis B virus core (HBc) is a 183-aa, 21 kDa protein which self-assembles to form ~35-nm particles that comprise the virion nucleocapsid. Dimer clustering of two HBc monomers produces spikes on the surface of the capsid, and the loop region located on top of these spikes (residues 78 – 83) is the major target of the humoral response, i.e., major immunodominant region (MIR).[6,7] Generally HBc-based VLPs have been used as an epitope presentation system in vaccine design and antigen presentation studies.[8-11] On the other hand, the ability of HBc to incorporate peptides of different origin may be useful in the development of NPs for different application.[12] For instance, packaging of enzymatically active nuclease, green fluorescent protein, RNA, and siRNA into the interior of recombinant hepatitis B virus capsids was demonstrated.[13-16] Moreover, it was proposed that the inner VLP lumen may be modulated by C terminal domain modifications; the authentic arginine-rich C terminal domain packages RNA but may be replaced, within limits by heterologous moieties that could act as cargo or provide specific functions.[17] However, transport of such particles in extracellular space and their uptake by different cells is of crucial significance for expression of their desired biological activity, so that it is necessary to design the appropriate targeting structures on the surface of VLPs to enhance cellular internalization.

Fibronectin (FN) is a common constituent of the extracellular matrix with important functions such as structural support and signaling for cell survival, migration, contractility, and differentiation.[18] FN is a large glycosylated mosaic protein composed of multiple copies of three types of modules: FNI, FNII, and FNIII. These modules comprise several functional domains that mediate interactions with cell-surface receptors (integrin-binding tripeptide, Arg-Gly-Asp [RGD] peptide), other ECM components (e.g., two heparin-binding domains, the

Table 1
Oligonucleotide primers designed for amplification of BBK*32* gene fragments

| Construct | Forward primer | Reverse primer |
|---|---|---|
| BF130-166 | 5′-CCGGATCCTGGATCAACTCAAGGAAGTTTA | 5′-TTGGATCCGAACTCTTTGGCCT |
| BF160-175 | 5′-ACCGGATCCTGGATCAGATTTACGTCCAAAGAG | 5′-TTGGATCCGAAGAGTTTGATCCTGC |
| BF153-175 | 5′-ACCGGATCCTGGATCAATTGATCTTACTATAG | 5′-TTGGATCCGAAGAGTTTGATCCTGC |

collagen-binding domain [gelatin-binding domain, GBD], and two fibrin-binding sites), and FN itself. FN is secreted by different cell types as a disulfide-bonded, soluble inactive dimer that is subsequently assembled into an insoluble fibrillar network.[19] FN acts as a ligand for bacterial and viral adherence to host cells and plays a role in modulating the cellular entry of retroviral vector particles.[20,21] In addition, a number of studies suggested an active role of FN adsorption in cellular uptake of nonviral nanocarriers (for review, see Adler et al.[22]

A large amount of bacterial FN-binding protein has been identified so far; among them is BBK32, a 47 kDa surface-exposed lipoprotein and FN-binding adhesin of *Borrelia burgdorferi* (*B. burgdorferi*).[23,24] BBK32 contains multiple FN-binding motifs located to an extended intrinsically disordered segment.[25] It was shown that the region of BBK32 comprising amino acids 147 – 205 binds to the N-terminal FNI and FNIII modules, but residues 120 – 147 of BBK32 protein are considered similar to the FN-binding region of SfbI protein from *S. pyogenes* that had been shown to bind to the GBD of FN.[26-28]

Here we present the construction of HBc-based chimeric VLPs with FN-binding property by insertion of *B. burgdorferi* BBK32 fragments into the MIR region of capsid. The evidence of intracellular internalization of these VLPs into nonphagocytic cells through caveolae-mediated endocytosis makes this construct an attractive model in development of HBc-based NPs for medically relevant cellular targeting applications and highlights the 130 – 166 aa fragment of *B. burgdorferi* BBK32 as a novel cellular uptake-promoting peptide.

## Methods

### *Construction of plasmids and prediction of structure of chimeric HBc/BBK32 VLPs*

The construction design of fusion VLPs in this study involved the insertion of a foreign sequence into the MIR of HBc protein (Figure 1, *A*). The cloning vector pHBc encoding the full-length (amino acids 1–183) of HBc protein, subtype ayw, with inserted *BamH*I restriction site at MIR was constructed by M. Mihailova like those described in another study.[29]

Three-dimensional predictions of the reconstructed HBc structures were performed on the basis of the x-ray structure of the HBc, genotype A,[7] by a comparative modeling program 3D-JIGSAW (http://bmm.cancerresearchuk.org/3djigsaw/)[30] and presented by Chimera software (Figure 1, *B, C*).[31]

The fragments of BBK*32* gene-encoding amino acids 130 to 166, 160 to 175, and 153 to 175 (BF130-166, BF160-175, and BF153-175 constructs, respectively) were amplified by PCR from *B. burgdorferi* sensu stricto isolate B31 DNA sample. The primers used to amplify these segments of BBK*32* are listed in Table 1. A linker of two amino acids (GS) was added to the N-terminal end of the insert, and a serine (S) was added to the C-terminal end of insert. (The exception was BF130-166 construct, where both C-end serines were of *BBK32* origin). *BamH*I restriction site was incorporated into both ends of inserts; the final peptide sequences are shown in Figure 1, *A*. The amplification products were digested with *BamH*I restriction endonuclease (Fermentas, Vilnius, Lithuania), purified using QIAEX II Gel Extraction Kit (Qiagen, Hilden, Germany), ligated into the linearized cloning vector, and competent *E. coli* RR1 ($F^-$ *leuB6 proA2 thi-1 araC14 lacY1 galK2 xyl-5 mtl-1 rpsL20* ($Str^r$) *glnV44* Δ*(mcrC-mrr)* cells were transformed with plasmid. The recombinant plasmids were isolated from positive clones by standard techniques, and correct constructs were verified by sequencing analysis.

### *Expression and purification of recombinant VLPs*

Recombinant proteins were expressed in *E. coli* strain K802 by standard techniques and purified by the combination of ion exchange chromatography (IEC) on Q-HP Sepharose (GE Healthcare, Helsinki, Finland) and size-exclusion chromatography (SEC) on Sepharose 4FF (GE Healthcare) column methods (Supplementary Materials Sections S1 and S2). Protein samples were analyzed on a SDS-PAGE gel and by western blotting with mouse monoclonal anti-HBc 13C9 antibodies,[32] as well as by double radial immunodiffusion according to Ouchterlony using polyclonal rabbit anti-HBc antibodies (Dako, Glostrup, Denmark).

### *Electron microscopy*

The purified fusion VLPs in suspension were adsorbed on carbon-formvar-coated copper grids and negatively stained with 1% uranyl acetate aqueous solution. The grids were examined with a JEM-100C electron microscope (JEOL Ltd., Tokyo, Japan) at an accelerating voltage of 100 kV.

### *Evaluation of the FN-binding activity by ligand affinity blot*

For the qualitative dot blot assay chimeric HBc/BBK32 VLPs were spotted onto the nitrocellulose membrane and probed with FN from human plasma followed by incubation with anti-FN antibodies (Supplementary Materials Section S3). As a control, recombinant full-length (HBc) core particles expressed in *E. coli* were used.

### *Quantitative FN-binding assays on microtiter wells*

Microtiter wells were coated with purified recombinant BF130-166 or HBc proteins and probed with increasing concentrations (0-10 μg/mL) of FN from human plasma followed by incubation with anti-FN antibodies. For experiments analyzing the role of gelatin- and heparin- binding domains of

FN in the interactions between BF130-166 and FN, competitive binding analyses were performed by incubating 0.5 μg/mL of FN to immobilized BF130-166 and HBc proteins in the presence of gelatin or heparin at four competitor/FN molar ratios (0.0/1, 0.25/1, 0.5/1 and 1/1) (Supplementary Materials Section S4).

*Cellular entry assay*

For the cellular entry assays, BF130-166 VLPs were added to the medium of baby hamster kidney BHK-21 cells to final concentration 5 μg/mL, and cells were incubated under different conditions, each experiment was repeated three times. (i) To investigate time-dependent uptake, cells were incubated with VLPs for 5, 15, 30, 45 and 60 minutes at 37°C. (ii) To study the uptake in the presence of gelatin, cells were pretreated with gelatin (5 μg/mL or 20 μg/mL) for 15 minutes and incubated with VLPs in the presence of gelatin for 45 minutes at 37°C. (iii) To study the effect of sodium azide treatment, cells were incubated with VLPs in the presence of 100 mM sodium azide in the medium for 45 minutes at 37°C. (iii) To study the effect of chlorpromazine (5 μg/mL) and filipin (5 μg/mL) treatment, cells were pretreated with drug for 30 minutes[33] and incubated with VLPs in the presence of drug for 45 minutes at 37°C.

After incubation supernatants were removed and the cells were washed once with ice-cold PBS, twice with an ice-cold mild acidic wash buffer to remove external particles. The localization of VLPs and FN in cellular space was evaluated by fluorescent double immunostaining method (Supplementary Materials Section S5). For comparison, HBc particles were used. Samples were examined with a laser confocal fluorescence microscope Leica TCS SP2SE. As a negative control cells incubated without VLPs were used.

## Results

*Synthesis and purification of fusion VLPs*

Three recombinant plasmids encoding different fusion HBc/BBK32 proteins were successfully constructed according to the sequencing analysis of plasmid DNA (data not shown). Figure 1, *C* presents three-dimensional predictive reconstructions of the chimeric HBc protein chains, which are exposing the appropriate BBK32 fragments on the tips of the HBc spikes. The BBK32 structure 130-166 seems to be the most flexible, because predictions for the four HBc chains vary from the mostly α-helical (chain D) to the β-sheet-forming (chains A, B, and C). The 153-175 and 160-175 insertions are predicted as mostly α-helice-forming for all four HBc chain variants.

Obtained plasmids were transformed into *E. coli* and recombinant proteins were expressed. SDS-PAGE and western blot analysis indicated the expression of soluble recombinant proteins of expected molecular weight, the levels of expression were sufficient for the further purification procedures (Figure 2, *A* and 2, *B*). Expressed recombinant HBc/BBK32 proteins were purified from *E. coli*; the purity of the isolated proteins according to Coomassie blue staining of the SDS-PAGE gel was 95% (Figure 2, *A*). The final yields of the recombinant proteins varied between 5 mg and 8 mg per 1 g of *E. coli* cells (data not shown). Further examination of these preparations by electron microscopy revealed the presence of well-assembled particles of similar size (approximately 30 nm-35 nm) and shape for all three constructs (Figure 2, *C, D, E*).

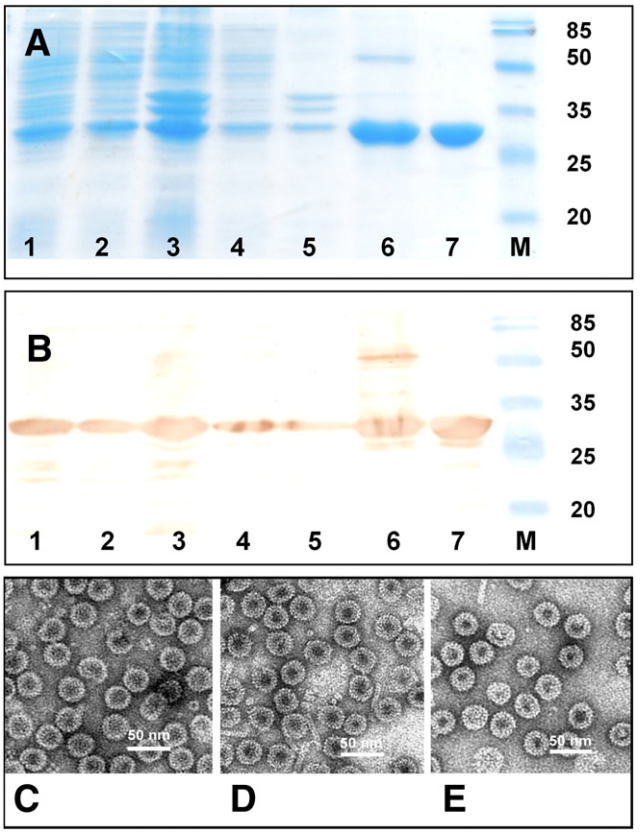



Figure 2. Expression, purification and morphologic analysis of fusion HBc/BBK32 VLPs. Fusion HBc/BBK32 proteins were expressed in *E. coli* cells and purified by the combination of techniques. The expression levels of the proteins of interest were similar; the purity of the isolated proteins according to Coomassie blue staining of the SDS-PAGE gel was 95%. Spherical particles of uniform size were obtained from all three constructs. Construct BF130-166 is shown as an example. **(A)** SDS-PAGE of protein samples, Coomassie brilliant blue staining, **(B)** Western blot with mouse monoclonal anti-HBc 13C9 antibodies. Lane 1: lysate of *E. coli* cells, lane 2: 0.5 M Urea supernatant, lane 3: 0.5 M Urea cellular debris, lane 4: supernatant after precipitation with ammonium sulfate at 10% saturation, lane 5: precipitate after precipitation with ammonium sulfate at 10% saturation, lane 6: protein fraction purified by IEC on Q-HP Sepharose, lane 7: protein fraction purified by SEC on Sepharose 4FF column, lane M: prestained protein molecular weight marker in kDa. **(C, D, E)**: Negative-staining electron microscopy images of the purified BF130-166 **(C)**, BF160-175 **(D)** and BF 153-175 **(E)** protein. Scale bars (*in white*) are 50 nm as indicated.

*FN-binding property of chimeric VLPs*

It was proposed that incorporation of *B. burgdorferi* BBK32 protein peptides harboring a FN-binding property into the hepatitis B core virus-like particles would allow us to create VLPs with enhanced FN-binding activity. The FN-binding property of obtained VLPs was evaluated by dot blot analysis by

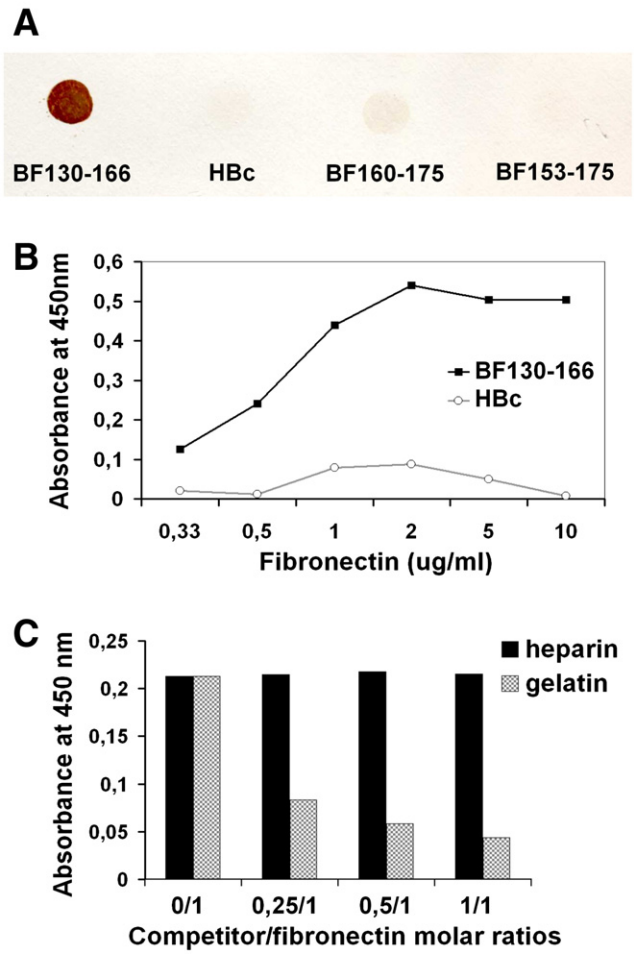



Figure 3. FN-binding properties of fusion HBc/BBK32 VLPs. **(A)** Dot blot analysis. Here, 1 μg of purified VLPs was dotted on nitrocellulose membrane for ligand blotting with FN from human plasma. The VLPs designation is as follows: BF130-166, fusion HBc VLPs including amino acids 130 to 166 of *B. burgdorferi* BBK32 protein; HBc, recombinant VLPs containing unmodified hepatitis B core protein; BF160-175 and BF153-175, fusion HBc VLPs including amino acids 160 to 175 or 153 to 175 of *B. burgdorferi* BBK32 protein, respectively. Only the BF130-166 VLPs bind human FN efficiently. **(B)** Dose-dependent, saturable binding of human FN to BF130-166 protein. Binding of FN to immobilized BF130-166 protein and HBc protein was analyzed by ELISA, with bound FN detected by anti FN antibodies. **(C)** Effects of gelatin and heparin on BF130-166 – FN interactions. Binding of human FN to immobilized BF130-166 protein in the presence of gelatin (*gray bars*) and heparin (*black bars*) was analyzed by ELISA, at gelatin or heparin/FN molar ratios of 0/1, 0.25/1, 0.5/1 and 1/1. Bound FN was detected by anti-FN antibodies. Dose-dependent inhibition of FN binding by gelatin was observed. As a negative control of the ELISA assays, corresponding amounts of FN were added to 5% milk-blocked wells, which did not contain any other immobilized protein. Values represent FN binding minus background readings for negative control wells.

use of a human FN as possible ligand, and recombinant full-length HBc VLPs were used as a control. The results show that only BF130-166 VLPs have pronounced FN-binding activity (Figure 3, *A*). Very slight FN-binding activity was detected for the construct BF160-175. No FN-binding activity was detected for BF153-175 construct and HBc VLPs. Reverse ligand affinity blot, in which FN, collagen, gelatin, and BSA were spotted on the membrane and probed with recombinant VLPs followed by incubation with polyclonal rabbit anti-HBcAg antibody, demonstrated similar results. No binding was observed for collagen/gelatin or BSA (data not shown).

The ability of BF130-166 to bind FN was further evaluated by quantitative ELISA. Saturable, dose-dependent binding of FN to BF130-166 was observed (Figure 3, *B*). The binding of FN was specific for BF130-166 as the results for the recombinant insert-free HBc VLPs could not be considered as positive.

### *Binding of BF130-166 to GBD of FN*

FN is a multifunctional molecule that interacts with a number of ligands via several distinct structural and functional domains. The BBK32 fragment of the BF130-166 construct contains putative GBD-binding motif[28]; therefore we next examined the potential role of this FN domain in BF130-166 interactions with FN. For this reason, FN binding by immobilized BF130-166 was examined by ELISA in the presence of increasing concentrations of gelatin, a known ligand for GBD. In a parallel assay, the adhesion of FN to immobilized BF130-166 was examined in the presence of increasing concentrations of heparin targeted to block the heparin-binding domains of FN. As shown in Figure 3, *C*, dose-dependent inhibition of FN binding by gelatin was observed. In contrast, the presence of heparin had no effect on FN – BF130-166 binding activity. These results indicated that recombinant BF130-166 VLPs interact with GBD of the FN molecule.

### *Cell-permeability of fusion BF130-166 VLPs*

Because clear and strong FN-binding properties were observed only for the construct BF130-166, the possible internalization of these VLPs into BHK-21 cells was further studied by confocal laser scanning microscopy. Culture of BHK-21 cells was inoculated by BF130-166 particles and the time course of internalization of VLPs was studied over 60 minutes' incubation time at 37°C. For comparison, HBc particles were used. The localization of VLPs and FN in cellular space was evaluated by fluorescent double immunostaining method. Only green fluorescence (FN) was observed in mock treated cells, as expected (Figure 4, Control).

In BF130-166 NP-treated cells, the VLPs appeared as nano-sized fluorescent dots (red) mostly associated with cellular membranes after 5 minutes' incubation time. These VLPs were not removed from the cellular surface during the acid wash; thus we can suggest that this may reflect the initial cellular internalization stage. The staining pattern of VLPs was associated with the FN fluorescence (green); therefore, these results suggested co-localization of VLPs with FN in the cell line studied (Figure 4). Further analysis revealed enhancement of intracellular red signal in cell samples at 15 minutes' and 30 minutes' incubation time. In these images VLP- positive staining mostly accumulated into discrete foci resembled a typical endosomal or lysosomal punctuate pattern (Figure 4, 15`, and 30`). Subsequently the TRITC fluorescence signal became reduced and more dispersed as the time of incubation increased; only a few red fluorescent dots were observed in cells after 60

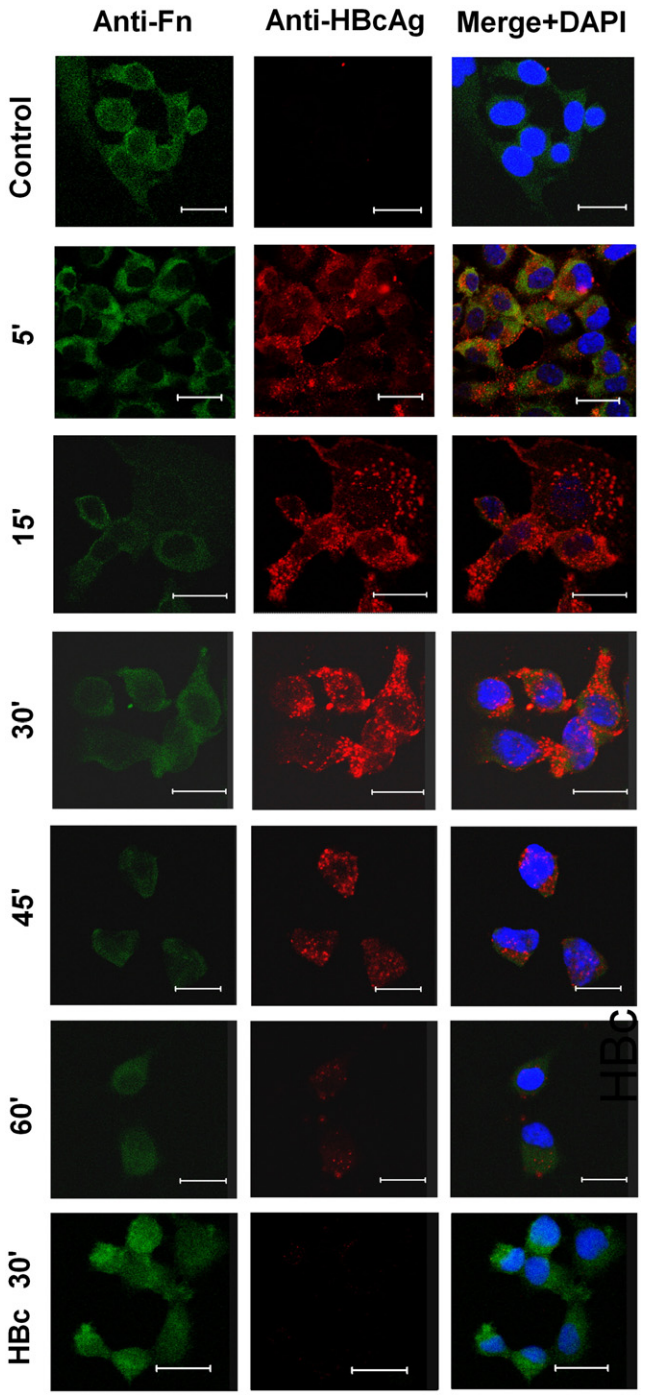



Figure 4. Fusion BF130-166 VLPs transmission to BHK-21 cells assessed by confocal microscopy. Culture of BHK-21 cells were inoculated by VLPs and the uptake of VLPs is depicted after 5, 15, 30, 45, and 60 min incubation time at 37°C following acid wash to remove external particles. Cells were stained with polyclonal rabbit anti-HBc antibody followed by TRITC-conjugated anti-rabbit IgG secondary antibody (red), and monoclonal anti-FN antibody followed by FITC-conjugated antimouse secondary antibody (green) in double immunostaining experiment. Cells' nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI, blue). No signal was observed in mock-treated cells (Control). No intracellular VLP-positive signal was observed for HBc VLPs. Scale bars (*in white*) are 20 μm as indicated.

minutes' incubation time (Figure 4, Anti-HBcAg, 60`). These results suggest a process of intracellular degradation of the NPs. In contrast, we did not observe pronounced changes in the green fluorescence pattern in cells over time (Figure 4, Anti-FN, images 5` - 60`). The intracellular localization of HBc VLPs was not observed (Figure 4, HBc 30`).

### *Cellular uptake of BF130-166 particles is inhibited by gelatin*

Given the separable nature of particle attachment and cell entry, we wanted to explore the possible role of binding to FN in the cellular uptake. For this purpose increasing concentrations of gelatin (5 and 20 μg/mL) were added to BHK-21 cells 15 minutes prior to experiments and remained present during the incubation with VLPs. The result shows a decrease of VLPs-positive signal in cells pretreated with 5 μg/mL gelatin and an absence of this staining in cells pretreated with 20 μg/mL gelatin (Figure 5, Anti-HBcAg, +gel. 5μg/mL and +gel. 20μg/mL). This observation confirms that BF130-166 particles internalized the cells after binding to GBD of FN and verifies the specificity of binding and uptake.

### *Mode of uptake of BF130-166 particles*

Next, we carried out a series of investigations on the uptake mechanism. Endocytosis can be inhibited with the depletion of cellular energy resources. The BHK-21 cells were incubated with BF130-166 in the presence of sodium azide, which inhibits active transport processes. Results show that the level of fluorescent intensity in the cytosol was reduced dramatically relative to cells cultured in standard conditions (Figure 5, Anti-HBcAg, $+NaN_3$). These results suggest that recombinant BF130-166 particles were internalized by endocytosis.

Further to inhibit clathrin-mediated and caveolin-mediated endocytosis, BHK-21 cells were treated with either 5 μg/mL chlorpromazine or 5 μg/mL filipin for 30 minutes, respectively. The results show that chlorpromazine treatment did not inhibit the entry of BF130-166; this observation suggests the clathrin-independent internalization. In contrast, filipin, a sterol-binding agent that disrupts caveolae and caveolae-like structures, strongly inhibit the entry of BF130-166, suggesting a filipin-sensitive pathway (Figure 5, Anti-HBcAg, +Chlorprom. and +Filipin). These observations provide evidence that FN-binding BF130-166 VLPs enter the cells via caveolae/raft-dependent endocytosis.

## Discussion

VLPs, including HBc-based NPs, are emerging nanocarrier platforms that can be tailored at the genetic level. The surface modification of such particles with an appropriate peptide molecule is one of the possible strategies to influence the ability of the engineered VLPs to interact with cells and tissues at a molecular level and mediate the uptake by mammalian cells. To date, several studies have been performed aimed at developing HBc-based chimeric VLPs that could serve as potential nanovehicles to target various cells.[16,34,35] In this study, we tested the possibility of exploiting the *B. burgdorferi* FN-binding

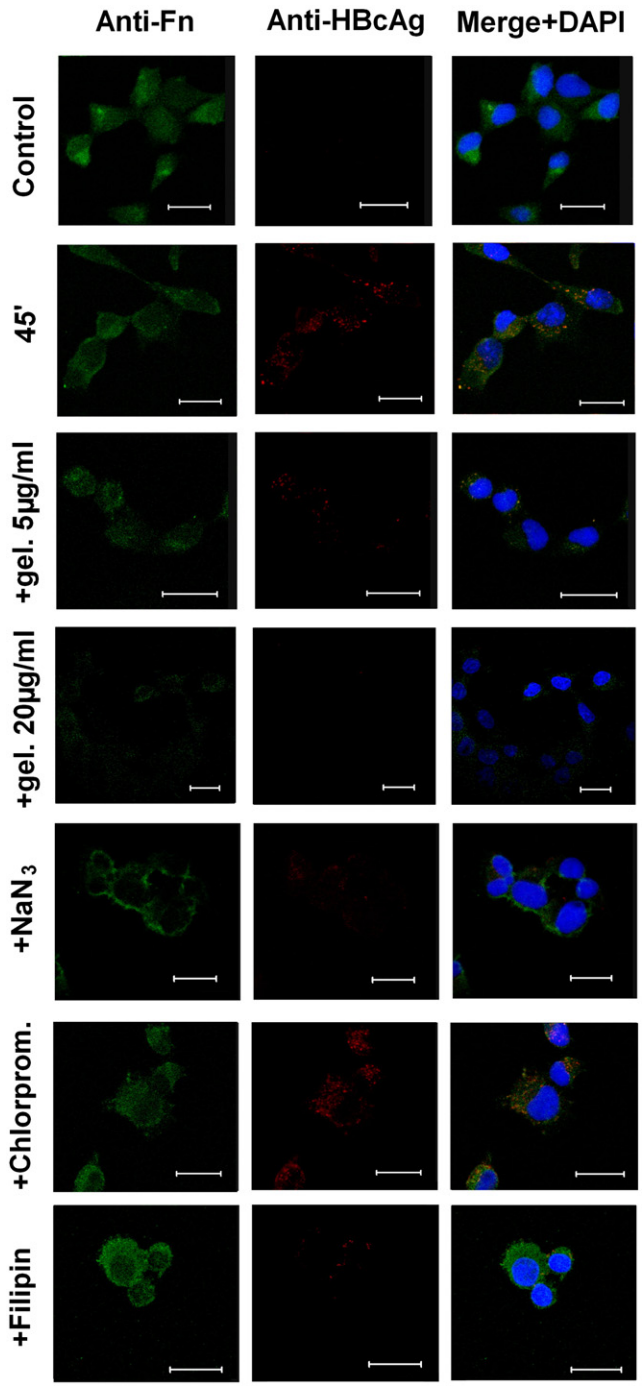



Figure 5. Fusion BF130-166 VLPs are endocytosed following attachment to FN. Culture of BHK-21 cells were inoculated by VLPs and the uptake of VLPs is depicted after 45 min incubation time at 37°C following acid wash to remove external particles. Cells were stained with polyclonal rabbit anti-HBc antibody followed by TRITC-conjugated anti-rabbit IgG secondary antibody (*red*), and monoclonal anti-FN antibody followed by FITC-conjugated antimouse secondary antibody (*green*) in double immunostaining experiment. Cells' nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI, blue). No signal was observed in mock-treated cells (*Control*). Pronounced red signal was observed in cells incubated with BF130-166 VLPs (45'). Similar signal was detected in cells pretreated with 5 µg/mL chlorpromazine and incubated with BF130-166 VLPs in the presence of chlorpromazine (+Chlorprom.) Intracellular VLP-positive signal was decreased in cells pretreated with 5 µg/mL gelatin and incubated with BF130-166 VLPs in the presence of gelatin for 45 min at 37°C (+gel.5 µg/mL). Intracellular VLP-positive signal was absent in cells pretreated with 20 µg/mL gelatin and incubated with BF130-166 VLPs in the presence of gelatin for 45 min at 37°C (+gel.20 µg/mL). The red signal was markedly reduced in cells pretreated and incubated in the presence of 100 mM sodium azide (+$NaN_3$). The red signal was markedly reduced in cells pretreated and incubated in the presence of 5 µg/mL filipin (+filipin). Scale bars (*in white*) are 20 µm as indicated.

protein BBK32 for modification of HBc particles to create an FN-targeting system with subsequent transduction of target cells.

For this purpose, three fragments of *B. burgdorferi* FN-binding protein BBK32 were inserted in MIR of the HBc protein: amino acids 130 to 166, 160 to 175, and 153 to 175 (BF130-166, BF160-175, and BF153-175 constructs, respectively), and fusion proteins as VLPs were expressed in *E. coli* cells. HBc fusion proteins often maintain the self-assembling ability and display inserted foreign epitopes on the tips of particle spikes if they are inserted into the MIR.[6] However, folding and capsid formations of the chimeric proteins are not always achieved easily, and the insert size and composition often are limitation factors in this process due to the conformational stress. Earlier it was shown that failure of some chimeric proteins to assemble into VLPs could be explained by the length of the insert; 120 amino acids are usually accepted as a maximum. On the other hand, Kratz et al[36] reported successful insertion of the GFP protein (238 amino acids) in the HBc; thus the structural importance of proper and independent folding of foreign sequences was clearly demonstrated. The potential role of different characteristics in the efficient assembly of VLPs was addressed in numerous studies, such as β-sheet forming properties, the distance between the N and the C-terminus, and volume and hydrophobicity of the amino acids of the insert. However, the nature of the insert makes it often difficult to manipulate, and the design of successful constructs is still largely empirical. One of the most useful strategies to influence the flexibility of construct conformation is the addition of linkers. The linker size and composition was shown to have a strong effect on the VLPs' assembling property.[6,37] To exploit these possibilities to overtake the assembly process, 41, 21, and 28 amino acids were inserted in HBc (BF130-166, BF160-175, and BF153-175 constructs, respectively) including linkers. In addition, three-dimensional predictions of the reconstructed HBc structures were performed. The expression of all constructs in *E. coli* resulted in stable, soluble proteins that were effectively purified. Moreover, the electron microscopy data confirmed that all three HBc/BBK32 fusion proteins assembled properly into symmetric particles indicating the successful design strategy (Figure 3). The average size of the particles calculated from the electron microscopy images was similar to the size of HBc VLPs, indicating that small-sized inserts did not have a great influence on the final dimensions of chimeric VLPs. Interestingly, recently the strategy of splitting the HBc protein inside the insertion loop (SplitCore) was offered to overcome almost accidental assembly process of chimeric VLPs.[17]

Another task of this study was to evaluate the FN-binding properties of obtained fusion VLPs. Different BBK32 peptides

used in this study have been shown to bind different FN regions.[25-28] Our results show that only BF130-166 VLPs demonstrated clear, strong FN-binding ability. Moreover, we were able to show that this binding occurs via GBD of FN. Unfortunately, the nature of inserts in BF160-175 and BF153-175 constructs seems to be insufficient to ensure stable ligand binding. It could be explained by the manner of binding of this BBK32 region to the N-terminal modules of FN that require an antiparallel orientation of the binding partners with the peptide forming additional β-strands at the edge of the triple-stranded β-sheets of the two FNI modules (i.e., a "tandem-zipper" mechanism).[26] The 153-175 and 160-175 insertions were predicted as α-helice-forming for all four possible chain variants; therefore we could assume the lack of conformational compatibility in FN binding for these VLPs (Figure 1). However, induction of ordered aggregation of soluble FN and inhibition of endothelial cell proliferation similar to that of anastellin were shown for these BBK32 fragments, so there may be further interest in these constructs in the future.[27]

Recently, retroviral binding to FN was associated with increased infection of target cells and efficient gene transfer.[21,38] Therefore, we tested the possible intracellular entry of the BF130-166 VLPs using BHK-21 as target cells. Indeed, the results of confocal microscopy analysis clearly demonstrated time-dependent particle uptake in BHK-21 cells. It was observed that entry process started with the attachment of VLPs to the cell surface followed by rapid internalization by energy dependent endocytosis. Initial particle co-localization with FN supported the idea that recombinant BF130-166 particles interact with FN molecules in the cellular microenvironment. Furthermore we were able to demonstrate the dose-dependant inhibition of the intracellular uptake by gelatin pretreatment. These results indicate that BF130-166 VLPs binding to the GBD of FN has an active role in the internalization process.

The relatively fast uptake was similar to the process observed for nonenveloped hepatitis C virus capsids in study by Katsarou et al.[39] Clathrin-mediated endocytosis was reported as an entry pathway for these capsids as well as for HBc particles.[39,40] In contrast, binding to FN has been previously reported to result in caveolae-dependent endocytosis of γ - retrovirus vectors.[20] Noteworthy, α5β1 integrin, a cellular receptor required for FN internalization, can be internalized by both clathrin-dependent and caveolar endocytosis.[41] In the present study we have tested BF130-166 particles' entry in the presence of chlorpromazine and filipin, agents that block clathrin- and caveolae- dependent uptake, respectively. The results show that inhibition of clathrin-mediated endocytosis had no effect on the uptake, but the disruption of membrane microdomains by the cholesterol-chelating agent filipin completely opposes the internalization of fusion VLPs. Accordingly, these data suggest that BF130-166 NP endocytosis in BHK 21 cells occurs via a non-clathrin-mediated mechanism and likely involves caveolae/lipid rafts. These findings may have significant implications because certain forms of caveolae-mediated endocytosis are thought to avoid the degradative lysosomal compartment.[22]

In conclusion, we were able to obtain FN-binding HBc-based VLPs with intracellular targeting properties. Although the potential applications of these NPs in humans may be limited due to possible immunogenicity and the abundance of FN in plasma, the evidence of particles' fast and efficient transmission to target cells opens the possibility for using this system as an intracellular delivery vector in vitro, as well as a model platform for investigation of multifunctional NPs with targeted features.

## Acknowledgments

We thank Dr. Kaspars Tars for the suggestion of linkers used in recombinant constructs, Dr. Marija Mihailova for kindly providing the cloning vector, and Dr. Tatjana Kozlovska for helpful discussions.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.nano.2012.05.003.

## References

1. Cho K, Wang X, Nie S, Chen Z, Shin DM. Therapeutic nanoparticles for drug delivery in cancer. *Clin Cancer Res* 2008;**14**:1310-6.
2. Petros RA, DeSimone JM. Strategies in the design of nanoparticles for therapeutic applications. *Nat Rev Drug Discov* 2010;**9**:615-27.
3. Torchilin VP. Targeted pharmaceutical nanocarriers for cancer therapy and imaging. *AAPS J* 2007;**9**(2):E128-47.
4. Ma Y, Nolte R, Cornelissen J. Virus-based nanocarriers for drug delivery. *Adv Drug Deliv Rev* 2012;**64**:811-25.
5. Grasso S, Santi L. Viral nanoparticles as macromolecular devices for new therapeutic and pharmaceutical approaches. *Int J Physiol Pathophysiol Pharmacol* 2010;**2**:161-78.
6. Pumpens P, Grens E. HBV core particles as a carrier for B cell/T cell epitopes. *Intervirology* 2001;**44**:98-114.
7. Wynne SA, Crowther RA, Leslie AG. The crystal structure of the human hepatitis B virus capsid. *Mol Cell* 1999;**3**:771-80.
8. Ding FX, Xian X, Guo YJ, Liu Y, Wang Y, Yang F, et al. A preliminary study on the activation and antigen presentation of hepatitis B virus core protein virus-like particle-pulsed bone marrow-derived dendritic cells. *Mol Bio Syst* 2010;**6**:2192-9.
9. Lee BO, Tucker A, Frelin L, Sallberg M, Jones J, Peters C, et al. Interaction of the hepatitis B core antigen and the innate immune system. *J Immunol* 2009;**182**:6670-81.
10. Sun C, Ding FX, Wang F, He XW, He Y, Li ZS, et al. Screen of multifunctional monoclonal antibodies against hepatitis B core virus-like particles. *Microbiol Immunol* 2009;**53**:340-8.
11. Watts NR, Cardone G, Vethanayagam JG, Cheng N, Hultgren C, Stahl SJ, et al. Non-canonical binding of an antibody resembling a naïve B cell receptor immunoglobulin to hepatitis B virus capsids. *J Mol Biol* 2008;**379**:1119-29.
12. Garcea RL, Gissmann L. Virus-like particles as vaccines and vessels for the delivery of small molecules. *Curr Opin Biotechnol* 2004;**15**(6):513-7.
13. Beterams G, Böttcher B, Nassal M. Packaging of up to 240 subunits of a 17 kDa nuclease into the interior of recombinant hepatitis B virus capsids. *FEBS Lett* 2000;**481**:169-76.
14. Lee KW, Tan WS. Recombinant hepatitis B virus core particles: association, dissociation and encapsidation of green fluorescent protein. *J Virol Methods* 2008;**151**(2):172-80.
15. Porterfield JZ, Dhason MS, Loeb DD, Nassal M, Stray SJ, Zlotnick A. Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. *J Virol* 2010;**84**(14):7174-84.

16. Choi KM, Choi SH, Jeon H, Kim IS, Ahn HJ. Chimeric capsid protein as a nanocarrier for siRNA delivery: stability and cellular uptake of encapsulated siRNA. *ACS Nano* 2011;**5**(11):8690-9.
17. Walker A, Skamel C, Nassal M. SplitCore: An exceptionally versatile viral nanoparticle for native whole protein display regardless of 3D structure. *Sci Rep* 2011;**1** Article 5.
18. Pankov R, Yamada KM. Fibronectin at a glance. *J Cell Sci* 2002;**115**: 3861-3.
19. Mao Y, Schwarzbauer JE. Fibronectin fibrillogenesis, a cell-mediated matrix assembly process. *Matrix Biol* 2005;**24**:389-99.
20. Beer C, Pedersen L. Matrix fibronectin binds gammaretrovirus and assists in entry: new light on viral infections. *J Virol* 2007;**81**: 8247-57.
21. O'Neill LS, Skinner AM, Woodward JA, Kurre P. Entry kinetics and cell-cell transmission of surface-bound retroviral vector particles. *J Gene Med* 2010;**12**:463-76.
22. Adler AF, Leong KW. Emerging links between surface nanotechnology and endocytosis: impact on nonviral gene delivery. *Nano Today* 2010; **5**(6):553-69.
23. Henderson B, Nair S, Pallas J, Williams MA. Fibronectin: a multidomain host adhesin targeted by bacterial fibronectin-binding proteins. *FEMS Microbiol Rev* 2011;**35**:147-200.
24. Probert WS, Johnson BJ. Identification of a 47 kDa fibronectin-binding protein expressed by *Borrelia burgdorferi* isolate B31. *Mol Microbiol* 1998;**30**:1003-15.
25. Kim JH, Singvall J, Schwarz-Linek U, Johnson BJ, Potts JR, Höök M. BBK32, a fibronectin binding MSCRAMM from *Borrelia burgdorferi*, contains a disordered region that undergoes a conformational change on ligand binding. *J Biol Chem* 2004;**279**:41706-14.
26. Raibaud S, Schwarz-Linek U, Kim JH, Jenkins HT, Baines ER, Gurusiddappa S, et al. *Borrelia burgdorferi* binds fibronectin through a tandem beta-zipper, a common mechanism of fibronectin binding in staphylococci, streptococci, and spirochetes. *J Biol Chem* 2005;**280**: 18803-9.
27. Prabhakaran S, Liang X, Skare JT, Potts JR, Höök M. A novel fibronectin binding motif in MSCRAMMs targets F3 modules. *PLoS One* 2009;**4**:e5412.
28. Probert WS, Kim JH, Höök M, Johnson BJ. Mapping the ligand-binding region of *Borrelia burgdorferi* fibronectin-binding protein BBK32. *Infect Immun* 2001;**69**:4129-33.
29. Mihailova M, Boos M, Petrovskis I, Ose V, Skrastina D, Fiedler M, et al. Recombinant virus-like particles as a carrier of B- and T-cell epitopes of hepatitis C virus (HCV). *Vaccine* 2006;**24**:4369-77.
30. Bates PA, Kelley LA, MacCallum RM, Sternberg MJE. Enhancement of protein modelling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins* 2001(Suppl 5):39-46.
31. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF chimera - a visualization system for exploratory research and analysis. *J Comput Chem* 2004;**25**:1605-12.
32. Bichko V, Schodel F, Nassal M, Gren E, Berzinsh I, Borisova G, et al. Epitopes recognized by antibodies to denatured core protein of hepatitis B virus. *Mol Immunol* 1993;**30**:221-31.
33. Payne CK, Jones SA, Chen C, Zhuang X. Internalization and trafficking of cell surface proteoglycans and proteoglycan-binding ligands. *Traffic* 2007;**8**(4):389-401.
34. Brandenburg B, Stockl L, Gutzeit C, Roos M, Lupberger J, Schwartlander R, et al. A novel system for efficient gene transfer into primary human hepatocytes via cell-permeable hepatitis B virus-like particle. *Hepatology* 2005;**42**:1300-9.
35. Lee KW, Tey BT, Ho KL, Tan WS. Delivery of chimeric hepatitis B core particles into liver cells. *J Appl Microbiol* 2012;**112**:119-31.
36. Kratz PA, Bottcher B, Nassal M. Native display of complete foreign protein domains on the surface of hepatitis B virus capsids. *Proc Natl Acad Sci USA* 1999;**96**:1915-20.
37. Janssens ME, Geysen D, Broos K, De Goeyse I, Robbens J, Van Petegem F, et al. Folding properties of the hepatitis B core as a carrier protein for vaccination research. *Amino Acids* 2010;**38**:1617-26.
38. Sondergaard CS, Haldrup C, Beer C, Andersen B, Kohn DB, Pedersen L. Preloading potential of retroviral vectors is packaging cell clone dependent and centrifugation onto CH-296 ensures highest transduction efficiency. *Hum Gene Ther* 2009;**20**:337-49.
39. Katsarou K, Lavdas AA, Tsitoura P, Serti E, Markoulatos P, Mavromara P, et al. Endocytosis of hepatitis C virus non-enveloped capsid-like particles induces MAPK-ERK1/2 signaling events. *Cell Mol Life Sci* 2010;**67**:2491-506.
40. Cooper A, Shaul Y. Clathrin-mediated endocytosis and lysosomal cleavage of hepatitis B virus capsid-like core particles. *J Biol Chem* 2006;**281**:16563-9.
41. Caswell PT, Vadrevu S, Norman JC. Integrins: masters and slaves of endocytic transport. *Nat Rev Mol Cell Biol* 2009;**10**:843-53.

doi: 10.1111/j.1365-3083.2012.02720.x

# The Relationship Between Maximal Exercise-Induced Increases in Serum IL-6, MPO and MMP-9 Concentrations

D. Reihmane*†, A. Jurka† & P. Tretjakovs†‡

*Xlab, Center for Healthy Aging, Department of Biomedical Sciences, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark; †Institute of Clinical and Experimental Medicine, University of Latvia, Riga, Latvia; and ‡Pauls Stradins Clinical University Hospital, Riga, Latvia

*Received 16 March 2012; Accepted in revised form 3 May 2012*

*Correspondence to*: D. Reihmane, Department of Biomedical Sciences, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen N, Denmark and Institute of Clinical and Experimental Medicine, University of Latvia, Ojara Vaciesa str. 4, LV-1004, Riga, Latvia. E-mail: dace.reihmane@sund.ku.dk; dace.reihmane@lu.lv

## Abstract

The aim of this study was to test the hypothesis that exercise would induce inflammatory response characterized by increased pro-inflammatory cytokines – interleukin-6 (IL-6) and tumour necrosis factor-α (TNF-α), adhesion molecule, matrix metalloprotease-9 (MMP-9) and myeloperoxidase (MPO) levels. Additional aim was to elucidate the possible source of maximal exercise-induced increase in MMP-9 concentration. To examine our hypothesis, 26 professional male ice hockey players [age 25 ± 1 (mean ± SEM) years; BMI 25.8 ± 0.4 kg/m$^2$] performed an incremental bicycle test until exhaustion, when maximal oxygen consumption was recorded. Venous blood samples were collected 30 min before and 2 min after exercise. There was an increase in the count of leucocytes (8.7 ± 1.8 versus 5.7 ± 1.3 × 10$^9$ cells per l) and IL-6 (1.24 ± 0.17 versus 0.69 ± 0.13 pg/ml), MPO (72 ± 7 versus 50 ± 4 ng/ml) and MPP-9 (139 ± 9 versus 110 ± 6 ng/ml) concentrations ($P < 0.05$) comparing post- and pre-exercise levels. Maximal exercise-induced increase in MPO correlated with the increases in IL-6 ($P < 0.05$, $R = 0.54$) and MMP-9 ($P < 0.01$, $R = 0.62$) concentrations. Furthermore, increase in IL-6 correlated with the increase in MMP-9 concentrations ($P < 0.05$, $R = 0.60$). Maximal exercise induces an inflammatory response characterized by leucocytosis and increased IL-6, MPO and MMP-9 concentrations. Correlations between increased MPO (marker of neutrophils degranulation) and both increased IL-6 and MMP-9 concentrations may suggest that neutrophils could be the main source of these inflammatory biomarkers during maximal exercise. Furthermore, correlation between increases in serum IL-6 and MMP-9 concentrations may suggest that IL-6 could exert modulatory effects on MMP-9 release during maximal exercise.

## Introduction

Interleukin-6 (IL-6) and tumour necrosis factor-α (TNF-α) are pro-inflammatory cytokines that have been considered as a key orchestrator of the acute phase inflammatory response [1, 2]. Moreover, increases in IL-6 [3, 4] and TNF-α [5] plasma and serum levels play an important role in the host response to acute exercise [6], for example, promoting leucocyte adherence and chemotaxis capacities, by increased expression of adhesion molecules on the surface of leucocytes and endothelial cells [7, 8]. Further on, leucocytes participate in aseptic muscle inflammation associated with muscle fibre injury caused by intensive exercise, and neutrophils reinforce muscle tissue injury assisting in its clearance [9]. Exercise-induced activation of the secretory function of neutrophils results in their degranulation that leads to an increase in plasma concentration of marker neutrophil proteins, including myeloperoxidase (MPO) [10]. Recent studies suggest that also MMP-9 plays a critical role in cleaving muscle-specific proteins and contributing to extracellular matrix formation, remodelling and regeneration in skeletal muscle [11], but it remains unclear which tissues are a basic source of the increased MMP-9 concentration [12] induced by physical exercise. It has been shown that a single bout of exercise can induce MMP-9 expression in skeletal muscle [13] which is stimulated by





SCANDINAVIAN JOURNAL OF Immunology AN INTERNATIONAL MONTHLY JOURNAL

increased levels of TNF-α [14]. Moreover, increased level of IL-6 during inflammation augments MMP-9 expression in leucocytes [1].

The aim of this study was to test the hypothesis that exercise would induce inflammatory response characterized by increased pro-inflammatory cytokines – interleukin-6 (IL-6) and tumour necrosis factor-α (TNF-α), adhesion molecule, matrix metalloprotease-9 (MMP-9) and MPO levels. Additional aim was to elucidate the source of maximal exercise-induced increase in MMP-9.

## Methods

*Subjects.* After approval of experimental procedures by the Ethical Committee of the Institute of Experimental and Clinical Medicine, University of Latvia, informed consent was obtained from 26 professional male ice hockey players (Table 1). All subjects were asked to fast 3 h before the maximal exercise test, to refrain from caffeine for 12 h, and from alcohol, nicotine and any medication for 24 h, as well as to avoid physical overload or other stressors. Body composition was assessed in all subjects using bioelectrical impedance analyser T Scan (Jawon Medical, Kyungsan, Korea).

*Maximal exercise protocol.* We measured maximal oxygen consumption on a cycle ergometer Monark Ergomedic 839E (Monark, Sweden) because the muscle groups used in ice skating are quite similar to those used in cycling. Cortex Metalyzer 3B system (Cranlea & Company, Birmingham, UK) was used to evaluate cardio-respiratory functions (ventilation, respiratory exchange ratio, electrocardiogram, heart rate, blood pressure). Athletes achieved their maximal oxygen consumption, when three of four commonly accepted criteria were met: (1) volitional exhaustion; (2) maximal heart rate measured at exhaustion was superior to 90% of the age-predicted maximal heart rate; (3) respiratory exchange ratio was above 1.10; and (4) capillary blood lactate concentration was >8 mM.

*Blood sampling and analysis.* Venous blood samples for IL-6, TNF-α, sE-selectin, sICAM-1, sVCAM-1, MMP-9, MPO and standard blood tests (leucocyte formula etc.) were taken 30 min before and 2 min after maximal exercise. As these professional athletes had an active training season, time limitation was the main reason for the lack of data during recovery phase. Blood samples for the determination of cytokines, adhesion molecules and other bioactive molecules were collected without anticoagulant and were allowed to coagulate for 20–30 min at room temperature. This was followed by centrifugation for 10 min, at $1600 \times g$ in room temperature. All specimens were immediately aliquoted and put in the freezer at −80 °C. Commercially available multiplex immunoassay kits (MILLIPLEX MAP kit Cat. No.: HCVD1-67 AK; MILLIPLEX MAP kit Cat. No.: HADK2-61K-B) were used for quantitative determination of IL-6, TNF-a, sE-Selectin, sVCAM-1, sICAM-1, MMP-9 and MPO by Luminex 200 analyzer (Luminex Corp., Austin, TX, USA). Cortisol and high sensitivity C-reactive protein (hsCRP) were measured by Immulite 2500 analyzer (Siemens Medical Solutions, USA). Capillary blood lactate concentration was determined using lactate analyzer Biosen 5030 AutoCal (EKF Diagnosticb GmbH, Magdeburg, Germany). Other blood tests were performed in clinical laboratory 'E.Gulbja laboratorija', Riga, Latvia.

Haemoglobin and haematocrit were also measured (Table 2) to take into account the possible influence of exercise-induced plasma volume change. Correction of plasma concentrations for blood cells and inflammatory molecules was made according to the method described previously [15].

*Statistical analysis.* The levels of measured mediators below the detection limit were arbitrarily assumed to be one-half of the detection limit value [16]. Data were analysed by SigmaPlot 11.0 software (Systat Software Inc., San Jose, CA, USA). After testing normality (Shapiro–Wilk test), data with normal distribution were analysed using paired *t*-test. Wilcoxon signed-rank test was used as nonparametric method (data marked as †). Data were expressed as mean ± standard error of the mean (SEM). Data for correlation analysis (Pearson correlation or Spearman rank *R* test) were expressed as difference between absolute values of measured parameters before and after exercise and further labelled in the article as delta (Δ). A value of $P < 0.05$ was considered to be significant.

**Table 1** Subject characteristics.

| | |
|---|---|
| Age (years) | 25 ± 1 |
| Body mass index (BMI) (kg/m$^2$) | 25.8 ± 0.4 |
| Fat mass (%) | 22 ± 1 |
| Muscle mass (%) | 74 ± 1 |
| LDL cholesterol (mM) | 2.7 ± 0.2 |
| HDL cholesterol (mM) | 1.5 ± 0.1 |
| Triglyceride (mM) | 1.2 ± 0.1 |
| $VO_{2max}$ (ml/kg/min) | 50.3 ± 1.0 |
| Maximal heart rate (bpm) | 188 ± 2 |
| Maximal respiratory exchange ratio | 1.12 ± 0.01 |
| Maximal lactate concentration (mM) | 10.3 ± 0.4 |

Values are means ± SEM of 26 male subjects.

## Results

Athletes' maximal oxygen consumption at incremental bicycle test was 50 ± 1 ml/kg/min (Table 1). There was a significant increase in haemoglobin concentration and haematocrit (Table 2), probably, because of the reinforced sweating during maximal exercise. Maximal exercise induced significant changes in absolute counts of all leucocyte subsets. Athletes' total leucocyte count was significantly increased after maximal exercise (Table 2).

**Table 2** Pre-exercise and post-exercise blood mediators' measures and haematological parameters.

| | Pre-exercise | Post-exercise | *P* value | *n* |
|---|---|---|---|---|
| IL-6 (pg/ml)† | 0.69 ± 0.13 | 1.24 ± 0.17 | <0.001 | 20 |
| TNF-α (pg/ml) | 6.31 ± 0.49 | 6.68 ± 0.56 | NS | 25 |
| sE-selectin (ng/ml) | 29 ± 2 | 29 ± 2 | NS | 25 |
| sICAM-1 (ng/ml)† | 120 ± 13 | 111 ± 15 | NS | 25 |
| sVCAM-1 (ng/ml) | 1103± 45 | 1105 ± 42 | NS | 25 |
| MMP-9 (ng/ml)† | 110 ± 6 | 139 ± 9 | <0.001 | 25 |
| MPO (ng/ml) | 50 ± 4 | 72 ± 7 | <0.001 | 26 |
| hsCRP (mg/l) | 0.55 ± 0.08 | 0.51 ± 0.08 | <0.05 | 26 |
| Cortisol (nM) | 446 ± 27 | 420 ± 29 | NS | 26 |
| Insulin (μU/ml)† | 7.3 ± 0.9 | 7.4 ± 0.8 | NS | 26 |
| Glucose (mM) | 5.5 ± 0.1 | 6.4 ± 0.2 | <0.001 | 26 |
| Haemoglobin (mM) | 9.4 ± 0.1 | 10.0 ± 0.1 | <0.001 | 22 |
| Haematocrit (%) | 46 ± 1 | 50 ± 1 | <0.001 | 22 |
| Leucocytes × $10^9$ cells per l | 5.7 ± 0.3 | 8.7 ± 0.4 | <0.001 | 22 |
| Neutrophils × $10^9$ cells per l† | 2.86 ± 0.2 | 3.91 ± 0.4 | <0.001 | 22 |
| Eosinophils × $10^9$ cells per l | 0.17 ± 0.01 | 0.20 ± 0.02 | <0.05 | 22 |
| Basophils × $10^9$ cells per l | 0.05 ± 0.00 | 0.07 ± 0.00 | <0.001 | 22 |
| Lymphocytes × $10^9$ cells per l | 2.1 ± 0.1 | 3.8 ± 0.1 | <0.001 | 22 |
| Monocytes × $10^9$ cells per l | 0.50 ± 0.03 | 0.73 ± 0.03 | <0.001 | 22 |

Values are means ± SEM. NS, non-significant.
†Data nonparametric methods were used.

Elevation of circulating lymphocytes and neutrophils was the main reason for the change in total white cell count. There was a significant increase in glucose ($P < 0.001$), decrease in high sensitivity C-reactive protein ($P < 0.001$), but there was no change in cortisol or insulin concentrations after maximal exercise (Table 2).

Sera IL-6 concentration increased significantly between the rest and the end of maximal exercise ($P < 0.001$, see Table 2). Likewise, there were significant increases in MMP-9 and MPO concentrations ($P < 0.001$, see Table 2). Sera concentrations of TNF-α, sE-selectin, s-ICAM-1 and sVCAM-1 did not differ from pre-exercise levels ($P > 0.05$, see Table 2).

There were a number of correlations between maximal exercise-induced increases in determined variables. Maximal exercise-induced increase in MPO concentration correlated with Δ neutrophils abs count ($P < 0.05$, $R = 0.52$, $n = 22$) and Δ IL-6 concentration ($P < 0.01$, $R = 0.64$, $n = 19$, see Fig. 1). Furthermore, Δ MMP-9 correlated with Δ MPO ($P < 0.01$, $R = 0.62$, $n = 24$, see Fig. 2) and Δ IL-6 concentrations ($P < 0.05$, $R = 0.60$, $n = 17$, see Fig. 3).

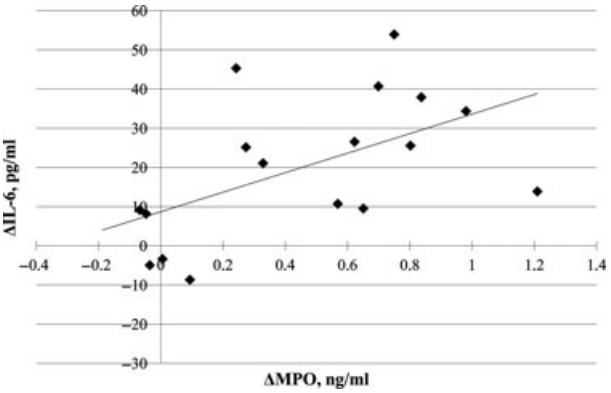



**Figure 1** The correlation between maximal exercise-induced increases in MPO and IL-6 concentrations. $P < 0.05$, $R = 0.54$, regression equation: $y = 8.6949 + 24.91*x$.

## Discussion

Exercise is associated with temporary changes in the immune system, for example, count of immune cells [17] and concentrations of cytokines [18], adhesion molecules [19], MMPs [12] and MPO [9]. The present study shows that maximal exercise induced significant changes in all subpopulations of leucocytes, especially neutrophils and lymphocytes ($P < 0.05$). Moreover, our data are in agreement with other studies that have shown exercise-induced degranulation of neutrophils and subsequent release of MPO ($P < 0.05$), confirmed by correlation between these two biomarkers ($P < 0.05$) [10].

Muscle damage often caused by eccentric exercise [20] and low plasma glucose or muscle glycogen levels [21] increase IL-6 concentrations during exercise. In fact, we observed an increase in serum glucose concentrations, most likely due to exercise-induced sympatho-adrenergic hepatic stimulation and increase in IL-6 concentration after maximal exercise ($P < 0.05$). It appears that the increase in IL-6 concentration in this study is less

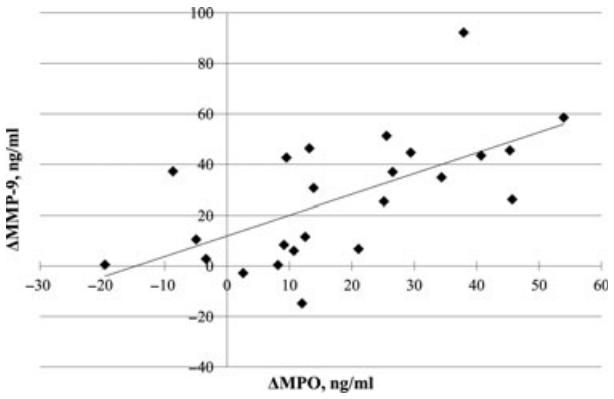



**Figure 2** The correlation between maximal exercise-induced increases in MPO and MMP-9 concentrations. $P < 0.01$, $R = 0.62$, regression equation: $y = 11.8963 + 0.8147*x$.

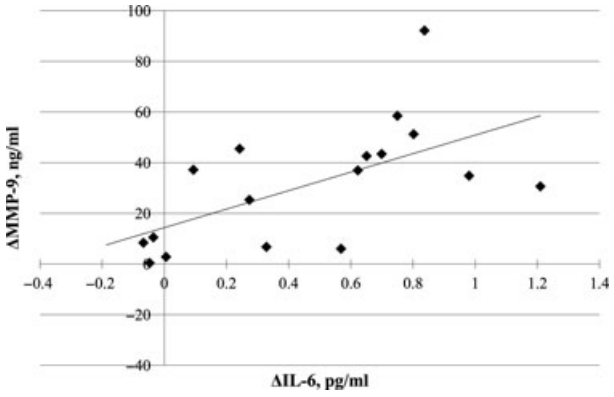



**Figure 3** The correlation between maximal exercise-induced increases in IL-6 and MMP-9 concentrations. $P < 0.05$, $R = 0.60$, regression equation: $y = 14.3641 + 36.5221*x$.

connected with glucose homoeostasis. Moreover, our data showed that maximal exercise-induced MPO expression correlated with the increase in IL-6 concentration ($P < 0.05$). With the release of MPO, marker of neutrophil degranulation, it is plausible that also IL-6 at least to some extent is released from the neutrophils during maximal exercise.

There are data that show pro-inflammatory cytokine (e.g. IL-6, TNF-α) incentive effect on the adhesion molecule up-regulation on the endothelial cells [7]. However, we did not observe sICAM-1, sVCAM-1 or sE-selectin changes during exercise despite the maximal exercise-induced increase in IL-6 concentration [22]. Our data are in agreement with other studies and show no change in TNF-α levels in healthy male athletes following maximal exercise [2]. Gokhale *et al.* [23] also found that majority of the athletes and non-athletes demonstrated a rise in IL-6 and a fall in TNF-α levels. This relation is in agreement with the opinion that IL-6 also exerts anti-inflammatory effects by inhibiting production of TNF-α, possibly also during maximal exercise.

### MMP-9

Exercise can cause a damage of skeletal muscles and connective tissue, which leads to activation of tissue MMPs. MMPs, including MMP-9, are the major components of neutrophilic tertiary granules and are also expressed by other types of leucocytes including monocytes and lymphocytes [1]. Furthermore, a single bout of exercise as cycling can induce an increment in concentration of total MMP-9 protein and its mRNA in human skeletal muscle [24]. In addition, the studies on the effect of exercise on plasma concentrations of MMP-9 are equivocal, showing no change [25] or increase after maximal exercise [26]. It has been shown that there is no correlation between exercise-induced increase in MMP-9 and creatine kinase activity, suggesting that the rapid and transient increase in the serum MMP-9 concentration may reflect accelerated release of MMP-9 to circulation because of exercise-induced changes in leucocyte number rather than an increased extracellular matrix breakdown [27]. Our data showed that maximal exercise-induced MPO expression correlated with the increase in MMP-9 concentration ($P < 0.05$), supporting the previous mentioned hypothesis, that neutrophils could be the major contributors for increased MMP-9 levels. Moreover, our data showed that maximal exercise-induced MMP-9 expression correlated with the increase in IL-6 concentration ($P < 0.05$), suggesting a close interplay between inflammatory and proteolytic processes. There is no substantial evidence in scientific literature showing IL-6 induced MMP-9 release from neutrophils; however, it is known that IL-6 stimulate MMP-9 expression in different cell types, for example, fibroblasts [28]. The correlations found in this study possibly could point out IL-6 modulatory effects on MMP-9, which needs to be clarified in further studies. It has been suggested that IL-6 might also regulate mobilization of neutrophils into circulatory system [29], which could be another way of IL-6 contribution to increased MMP-9 levels after maximal exercise.

In conclusion, maximal exercise induces an inflammatory response characterized by greater count of all subpopulations of leucocytes and increased IL-6, MPO and MMP-9 concentrations. Although study design does not allow estimating precise source of increased IL-6 and MMP-9, their correlations with increased MPO (maker of neutrophils degranulation) levels may suggest that neutrophils could be the main source of these inflammatory biomarkers during maximal exercise. Furthermore, correlation between increases in serum IL-6 and MMP-9 concentrations may suggest that IL-6 could exert modulatory effects on MMP-9 release during maximal exercise.

## Acknowledgment

This work has been supported by the European Social Fund within the project 'Support for Doctoral Studies at University of Latvia' and in part by grant No. 2010.10-4/VPP-4/5 of the framework of the Latvian National Program.

## References

1 Chen Y, Fan Y, Poon KY *et al.* MMP-9 expression is associated with leukocytic but not endothelial markers in brain arteriovenous malformations. *Front Biosci* 2006;11:3121–8.
2 Limongelli G, Calabro P, Maddaloni V *et al.* Cardiotrophin-1 and TNF-alpha circulating levels at rest and during cardiopulmonary exercise test in athletes and healthy individuals. *Cytokine* 2010; 50:245–7.
3 Edwards KM, Burns VE, Ring C, Carroll D. Individual differences in the interleukin-6 response to maximal and submaximal exercise tasks. *J Sports Sci* 2006;24:855–62.
4 Chaar V, Romana M, Tripette J *et al.* Effect of strenuous physical exercise on circulating cell-derived microparticles. *Clin Hemorheol Microcirc* 2011;47:15–25.
5 Kinugawa T, Kato M, Ogino K *et al.* Interleukin-6 and tumor necrosis factor-alpha levels increase in response to maximal exercise in patients with chronic heart failure. *Int J Cardiol* 2003;87:83–90.
6 Rahman ZA, Abdullah N, Singh R, Sosroseno W. Effect of acute exercise on the levels of salivary cortisol, tumor necrosis factor-alpha and nitric oxide. *J Oral Sci* 2010;52:133–6.
7 Monchanin G, Serpero LD, Connes P *et al.* Effects of progressive and maximal exercise on plasma levels of adhesion molecules in athletes with sickle cell trait with or without alpha-thalassemia. *J Appl Physiol* 2007;102:169–73.
8 Karatzis EN. The role of inflammatory agents in endothelial function and their contribution to atherosclerosis. *Hellenic J Cardiol* 2005;46:232–9.
9 Morozov VI, Tsyplenkov PV, Golberg ND, Kalinski MI. The effects of high-intensity exercise on skeletal muscle neutrophil myeloperoxidase in untrained and trained rats. *Eur J Appl Physiol* 2006; 97:716–22.
10 Morozov VI, Pryatkin SA, Kalinski MI, Rogozkin VA. Effect of exercise to exhaustion on myeloperoxidase and lysozyme release from blood neutrophils. *Eur J Appl Physiol* 2003;89:257–62.
11 Urso ML, Pierce JR, Alemany JA, Harman EA, Nindl BC. Effects of exercise training on the matrix metalloprotease response to acute exercise. *Eur J Appl Physiol* 2009;106:655–63.
12 Madden MC, Byrnes WC, Lebin JA, Batliner ME, Allen DL. Plasma matrix metalloproteinase-9 response to eccentric exercise of the elbow flexors. *Eur J Appl Physiol* 2011;111:1795–1805.
13 Rullman E, Norrbom J, Stromberg A *et al.* Endurance exercise activates matrix metalloproteinases in human skeletal muscle. *J Appl Physiol* 2009;106:804–12.
14 Srivastava AK, Qin X, Wedhas N *et al.* Tumor necrosis factor-alpha augments matrix metalloproteinase-9 production in skeletal muscle cells through the activation of transforming growth factor-beta-activated kinase 1 (TAK1)-dependent signaling pathway. *J Biol Chem* 2007;282:35113–24.
15 Johansen LB, Videbaek R, Hammerum M, Norsk P. Underestimation of plasma volume changes in humans by hematocrit/hemoglobin method. *Am J Physiol* 1998;274:R126–30.
16 Corhay JL, Henket M, Nguyen D, Duysinx B, Sele J, Louis R. Leukotriene B4 contributes to exhaled breath condensate and sputum neutrophil chemotaxis in COPD. *Chest* 2009;136:1047–54.
17 Lippi G, Banfi G, Montagnana M, Salvagno GL, Schena F, Guidi GC. Acute variation of leucocytes counts following a half-marathon run. *Int J Lab Hematol* 2010;32:117–21.
18 Suzuki K, Nakaji S, Yamada M *et al.* Impact of a competitive marathon race on systemic cytokine and neutrophil responses. *Med Sci Sports Exerc* 2003;35:348–55.
19 Nielsen HG, Lyberg T. Long-distance running modulates the expression of leucocyte and endothelial adhesion molecules. *Scand J Immunol* 2004;60:356–62.
20 Toft AD, Jensen LB, Bruunsgaard H *et al.* Cytokine response to eccentric exercise in young and elderly humans. *Am J Physiol Cell Physiol* 2002;283:C289–95.
21 Helge JW, Stallknecht B, Pedersen BK, Galbo H, Kiens B, Richter EA. The effect of graded exercise on IL-6 release and glucose uptake in human skeletal muscle. *J.Physiol* 2003;546:299–305.
22 Tripette J, Connes P, Hedreville M *et al.* Patterns of exercise-related inflammatory response in sickle cell trait carriers. *Br J Sports Med* 2010;44:232–7.
23 Gokhale R, Chandrashekara S, Vasanthakumar KC. Cytokine response to strenuous exercise in athletes and non-athletes – an adaptive response. *Cytokine* 2007;40:123–7.
24 Rullman E, Rundqvist H, Wagsater D *et al.* A single bout of exercise activates matrix metalloproteinase in human skeletal muscle. *J Appl Physiol* 2007;102:2346–51.
25 Tayebjee MH, Lip GY, Blann AD, Macfadyen RJ. Effects of age, gender, ethnicity, diurnal variation and exercise on circulating levels of matrix metalloproteinases (MMP)-2 and -9, and their inhibitors, tissue inhibitors of matrix metalloproteinases (TIMP)-1 and -2. *Thromb Res* 2005;115:205–10.
26 Danzig V, Mikova B, Kuchynka P *et al.* Levels of circulating biomarkers at rest and after exercise in coronary artery disease patients. *Physiol Res* 2010;59:385–92.
27 Koskinen SO, Hoyhtya M, Turpeenniemi-Hujanen T *et al.* Serum concentrations of collagen degrading enzymes and their inhibitors after downhill running. *Scand J Med Sci Sports* 2001;11:9–15.
28 Dasu MR, Barrow RE, Spies M, Herndon DN. Matrix metalloproteinase expression in cytokine stimulated human dermal fibroblasts. *Burns* 2003;29:527–31.
29 Yamada M, Suzuki K, Kudo S, Totsuka M, Nakaji S, Sugawara K. Raised plasma G-CSF and IL-6 after exercise may play a role in neutrophil mobilization into the circulation. *J Appl Physiol* 2002; 92:1789–94.

doi:10.1016/j.jmb.2012.06.032 J. Mol. Biol. (2012) 423, 63–78

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

# Structures of Hepatitis B Virus Cores Presenting a Model Epitope and Their Complexes with Antibodies

A. M. Roseman[1,2*], O. Borschukova[3], J. A. Berriman[1], S. A. Wynne[1], P. Pumpens[3] and R. A. Crowther[1]

[1]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK
[2]Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK
[3]Latvian Biomedical Research and Study Centre, 1 Ratsupites Street, LV-1067 Riga, Latvia

*Received 24 February 2012;*
*received in revised form*
*16 May 2012;*
*accepted 20 June 2012*
Available online
28 June 2012

***Edited by J. Johnson***

***Keywords:***
cryomicroscopy;
image processing;
vaccine carriers;
virus-like particles;
improved immunogenicity

The core shell of hepatitis B virus is a potent immune stimulator, giving a strong neutralizing immune response to foreign epitopes inserted at the immunodominant region, located at the tips of spikes on the exterior of the shell. Here, we analyze structures of core shells with a model epitope inserted at two alternative positions in the immunodominant region. Recombinantly expressed core protein assembles into $T=3$ and $T=4$ icosahedral shells, and atomic coordinates are available for the $T=4$ shell. Since the modified protein assembles predominantly into $T=3$ shells, a quasi-atomic model of the native $T=3$ shell was made. The spikes in this $T=3$ structure resemble those in $T=4$ shells crystallized from expressed protein. However, the spikes in the modified shells exhibit an altered conformation, similar to the DNA containing shells in virions. Both constructs allow full access of antibodies to the foreign epitope, DPAFR from the preS1 region of hepatitis B virus surface antigen. However, one induces a 10-fold weaker immune response when injected into mice. In this construct, the epitope is less constrained by the flanking linker regions and is positioned so that the symmetry of the shell causes pairs of epitopes to come close enough to interfere with one another. In the other construct, the epitope mimics the native epitope conformation and position. The interaction of native core shells with an antibody specific to the immunodominant epitope is compared to the constructs with an antibody against the foreign epitope. Our findings have implications for the design of vaccines based on virus-like particles.



## Introduction

The core protein of hepatitis B virus (HBV) provides an appropriate framework for display of foreign epitopes and has been suggested to provide a basis for novel vaccines.[1–3] Core shells expressing part of the foot and mouth disease virus capsid gave protective immunity to virus challenge in guinea pigs.[4] Clinical trials are in progress for a malaria vaccine based on a core protein displaying B- and T-cell epitopes from the circumsporozoite surface protein.[5–7] A segment of the influenza virus M2 protein by display on core was rendered strongly

*Corresponding author. Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK. E-mail address: Alan.Roseman@manchester.ac.uk.

Present addresses: O. Borschukova, 30 Gardner Road, Apt. 3H, Brookline, MA 02445, USA; J. A. Berriman, 51 Oakhill Road, Putney, London SW15 2QJ, UK.

Abbreviations used: EM, electron microscopy; Fab, fragment of antibody; Fv, fragment of antibody variable region; HBV, hepatitis B virus; L1, linker 1; L2, linker 2; MIR, major immunodominant region; PBS, phosphate-buffered saline; 3D, three-dimensional.

immunogenic and protective in mice.[8] It is even possible to display whole domains, for example, the 28-kDa ectodomain of the OspA (outer surface protein A) of the Lyme disease agent, where the resulting construct conferred immunity to the agent in mice.[9] The utility of the core shell arises in part from the special properties it has in stimulating different aspects of the immune system.[10]

The recombinant core protein can be readily expressed in *Escherichia coli*, where it assembles into icosahedral shells of two sizes, containing 180 ($T=3$) or 240 ($T=4$) subunits, respectively.[11] Hepatitis B virions contain the $T=4$ form of the core shell.[12] The core protein has a basic C-terminal tail that interacts with and packages nucleic acid but that is dispensable for shell formation, provided the truncation is not before amino acid 140.[13–15] The position of the truncation influences the ratio of $T=3$ to $T=4$ particles.[16] The fold of the core protein was determined by electron cryomicroscopy[17] and an atomic structure of the $T=4$ shell was solved by X-ray crystallography.[18] The shells are built from dimers of core protein molecules, with each dimer producing a spike on the surface consisting of a four-helix bundle formed by apposed α-helical hairpins (Fig. 1).[17–19] During assembly of the virion, the core undergoes a maturation step involving a relative scissor-like movement of the α-helical hairpins, which enables envelopment of the core and secretion of the virus.[12]

The core protein can remain assembly competent while accommodating foreign inserts at the N- and C-termini and more importantly at the tips of the spikes, where the immunodominant epitope of the native core protein is located. The tip itself has a large insertion capacity, accepting, for example, a 120-amino-acid hantavirus nucleocapsid fragment,[20] the entire green fluorescent protein (238 aa),[21] or a domain of the Lyme disease OspA outer protein (255 aa).[9] For short peptides, display on the core protein greatly increases the immunogenicity of the foreign protein segment.[4,22] Moreover, insertion in the context of the helical hairpin forming the spike is likely to constrain the conformation of an inserted peptide.[23]

The epitopes of the native core structure have been extensively investigated by using electron microscopy (EM) to examine complexes of core shells with various antibodies.[24–27] However, no detailed structural analysis has been made of the display of foreign epitopes and their complexes with antibodies. We here describe the structures determined by electron cryomicroscopy of engineered core shells displaying at the spike tip a well-characterized epitope DPAFR from the preS1 region of the surface protein of HBV,[28,29] against which a specific monoclonal antibody MA18/7 is available.[30] More precise mapping revealed a DPxF sequence as a necessary and sufficient epitope for the antibody recognition.[31] This antibody inhibits attachment of HBV to human hepatocytes[32] and blocks model infection of primary *Tupaia* hepatocytes.[33] Moreover, the DPAFR epitope is structurally similar to the native immunodominant DPASR/DPISR stretch at the tip of the core spike. We examined two constructs, which position the foreign epitope differently. The insertions result in an increased ratio of $T=3$ to $T=4$ particles. Besides showing local differences at the spike tips when compared with the native protein, the construct maps show bodily rotation of the core protein molecules, similar to those seen in maturation of the core shell,[12] and not unexpected as the shell is known to react to conformational stress.[34] We compared antibody complexes of the two constructs with the native structure in complex with an antibody against the immunodominant native epitope. The results demonstrate that the different insertions result in a different geometry of antibody interactions with varying degrees of steric hindrance between antibody fragments on neighboring epitopes. Understanding the changes in shell structure caused by insertions and the overall geometry of the antibody complexes may help to define the most efficient ways of displaying foreign epitopes to produce effective vaccines.

## Results and Discussion

### The native $T=3$ core shell structure

The structure of the $T=4$ form of the core shell is shown in Fig. 1a, with a detailed view of the α-helical hairpins forming the spike in Fig. 1b and c.[17,18] The nomenclature[16] for the subunits in the $T=4$ and $T=3$ shells is shown in Fig. 1d and e. In the $T=4$ shell, there are A,B and C,D dimers formed by the four computationally nonequivalent subunits. In the $T=3$ shell, the dimers are of the form A,B and C, C, with the C,C dimers lying on strict 2-fold axes.

The assembly of the core protein into $T=3$ or $T=4$ sized shells is influenced by truncation of the core protein at different points near the C-terminus and by modifications elsewhere.[16,34] Since the proportion of $T=3$ particles increased markedly with insertions into the major immunodominant region (MIR) at the tip of the spike (Table 1), we needed a detailed model of the native $T=3$ form of the shell, as a basis to understand the changes in the shells with insertions. Using a construct truncated at position 140, which forms mainly $T=3$ core shells (75%), we computed a map of the native $T=3$ shell at 8 Å resolution (Fig. 1f). We made a quasi-atomic model by docking in dimers from the $T=4$ crystal structure,[18] which fit well in the $T=3$ shell (Fig. 1g).

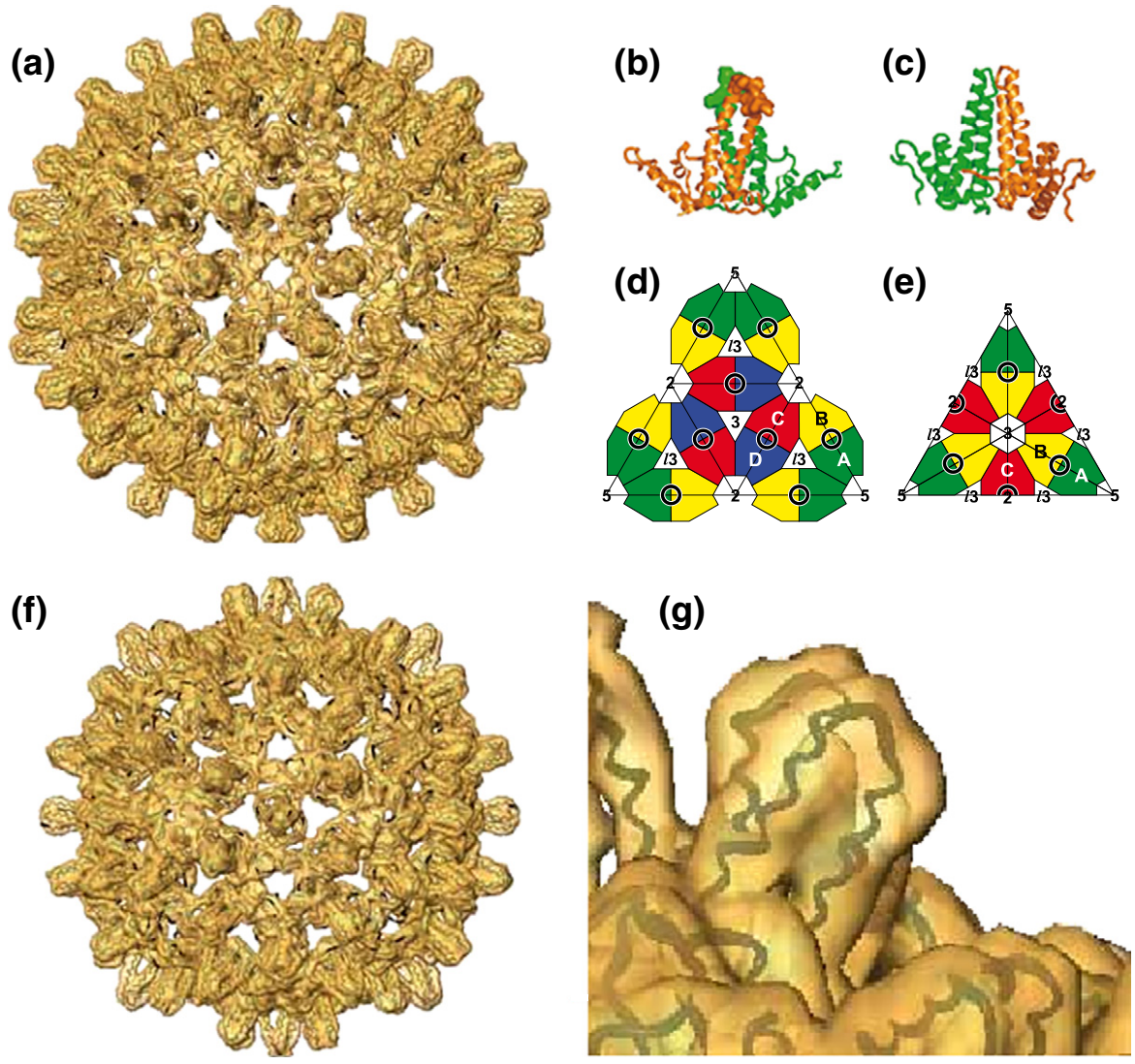



**Fig. 1.** Structures of native HBV core shells. (a) $T=4$ map from EM[12] with the crystal structure superimposed.[18] (b and c) Enlarged views from two directions of the dimer structure forming the spike. This is the A,B dimer (see below), which forms rings around the 5-fold axes. The A chains are closest to the 5-fold axis while the B chains face outwards from it and are closer to the 2-fold axis. The C,D spikes are positioned around the 2-fold axes. In (b), the MIR is shown in space-filling representation, as defined in Fig. 2. (d and e) Nomenclature for the structurally independent subunits A, B, C, and D in the $T=4$ shell and A, B, and C in the $T=3$ shell, respectively.[16] The rings represent the positions of the dimer spikes; 2, 3, and 5 indicate the positions of the icosahedral 2-fold, 3-fold, and 5-fold axes, respectively, and *l*3 denotes the local 3-fold axes. (f) Electron microscope map of the $T=3$ shell with dimers from the crystal structure fitted (see the text). This shell is 32 nm in diameter. (g) Detailed view of the A,B dimer of the $T=3$ shell, showing good fit of chain from the crystal structure.

The native $T=3$ core particle matches the $T=4$ particle in terms of the shape of the spikes and the shape of density around the 3-fold and 5-fold axes. The two types of $T=3$ spikes closely resemble the $T=4$ spikes, clearly exhibiting the bulbous knob-shaped profile created by the crossing of the helices at the dimer interface (see Fig. 1), as opposed to the narrower conformation observed in mature $T=4$

**Table 1.** Percentage of shells ($T=3$ or $T=4$) for various constructs

| | CW full length | CW trunc140 | Riga full length[a] | Riga trunc144[a] | S2-16 trunc144 | S1-8 trunc144 |
|---|---|---|---|---|---|---|
| $T=3$ | 10 | 75 | 10 | 20 | 60 | 80 |
| $T=4$ | 90 | 25 | 90 | 80 | 40 | 20 |

[a] From Borisova *et al.*[29]

viral cores.[12] The A,B dimer from the crystal structure of the $T=4$ shell (in the immature conformation) docked precisely into the density of the spike at the local 2-fold A,B position in the $T=3$ map. When the chains in the dimer were allowed to dock independently, there was no significant adjustment of the position or orientation of the individual subunits. At the strict 2-fold C,C position, the C-chain from the $T=4$ crystal structure fitted best. The level of differences between the conformations of the $T=3$ subunits themselves, and with the four different $T=4$ solved chains, is within the degree of variability of the four chains in the $T=4$ crystal structure.[18]

### Constructs displaying the model epitope

Two different insertion constructs are studied here, each containing the DPAFR epitope from the preS1 region of hepatitis B surface protein. These constructs, with a C-terminal truncation at position 144, were selected from a larger panel of 12 constructs that were characterized biochemically and immunologically.[29] They are named S2-16 and S1-8. These two were chosen for structural analysis because both were expressed and formed shells relatively well, and competitive ELISA showed that the inserted epitope was fully accessible to antibody. The constructs differed in that when tested in mice, though both were immunogenic against the DPAFR epitope, the strength of the immune response induced to this epitope by the S2-16 construct was 10-fold higher than that of S1-8. The titer of the antibody response was measured by direct ELISA against preS1 peptide 21–47 containing the DPAFR epitope, which showed that dilutions of sera of $10^3$ for S1-8 but $10^4$ for S2-16 took the optical absorbance values down to background level. Background level was determined as the mean of three measurements of sera from non-immunized mice (see Table 2 of Borisova *et al.*[29]).

In the native spike, the helical hairpin consists of an outward-going helix 3 and a returning kinked helix 4, split into helix 4a and helix 4b at the kink (Fig. 2a). The immunodominant native epitope (DPASR or DPISR depending on viral strain) stands exposed on the corner of the spike at the start of helix 4a. In these two constructs, the native epitope, DPISR, has been deleted and replaced by the DPAFR epitope, inserted between two short linker sequences (Fig. 2b). On the N-terminal side, the loop following helix 3 is extended by a 5-aa linker sequence [linker 1 (L1), DHDHV]. The DPAFR epitope follows and is joined to helix 4 by a 4-aa linker sequence [linker 2 (L2), YVDR or YVDH]. The difference between S1-8 and S2-16 is that in S2-16, the 6 aa of helix 4a following the native epitope DPASR/DPISR have also been deleted. Thus, in

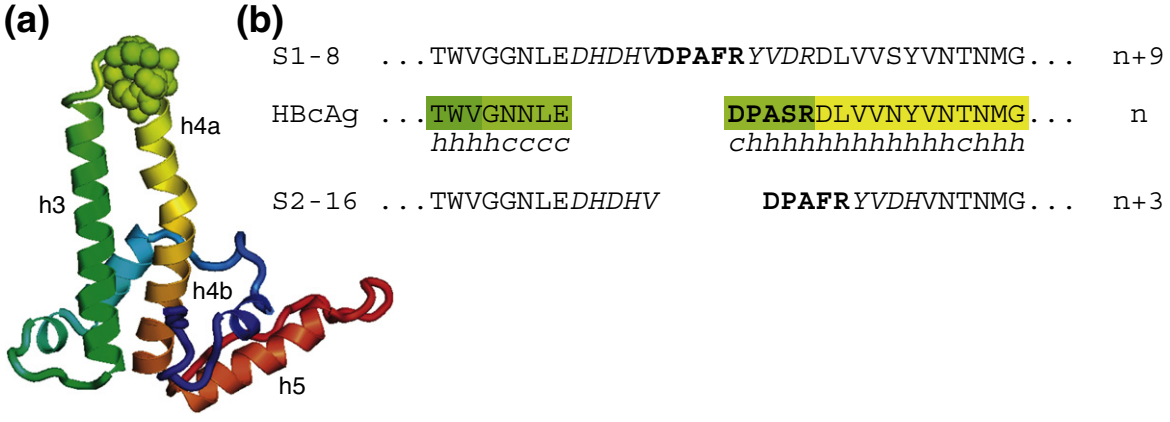



**Fig. 2.** (a) Detailed structure of the subunit, shown rainbow color coded, from N-terminus (blue) to C-terminus (red). The fold that contributes to the spike consists of an outward-going helix 3 and a returning helix 4 forming a hairpin. Helix 4 is kinked midway, giving subhelices 4a and 4b. A short loop region of 5 aa joins helix 3 to helix 4 and, together with the first few residues of helix 4a, forms the MIR (shown as space filling). This region at the tip of the spikes can be replaced with other epitopes. (b) Sequences in the region of the spike tip of the native and of the constructs with inserted epitopes. The conformation of each residue (*h* for helical, *c* for coil) in the native crystal structure[18] is indicated beneath the native sequence. The two constructs S1-8 and S2-16 used in this study were selected from a larger panel of 12 constructs, which were characterized biochemically and immunologically.[29] In these constructs, the native epitope (DPISR or DPASR depending on strain) has been deleted and replaced by the epitope DPAFR from the preS1 region of HBV surface protein. In the S2-16 construct, an additional 6 aa C-terminal to the native epitope were deleted. In both constructs, the loop region in the native protein following helix 3 (GGNLE) has been extended on the C-terminal side with a linker module, L1, sequence DHDHV. The DPAFR epitope follows and joins onto another linker module, L2, sequence YVDR in S1-8 and YVDH in S2-16. Hence, in both cases, the DPAFR epitope is inserted between the two linker modules, with the result that S2-16 has three and S1-8 has nine more amino acids than the native sequence.

these constructs, S2-16 contains three more amino acids and S1-8 contains nine more amino acids than the native sequence. In the S1-8 construct, the foreign epitope and linker L2 might be expected to extend from helix 4a in a helical conformation, whereas in S2-16, the deletion would cause the epitope to be embedded in helix 4a.

### Comparison of the structures of the core particle constructs

Images of the shells formed by the constructs showed that the particles were predominantly of the $T=3$ size, and there was a significant number of malformed or incomplete shells (Fig. S1). Good particles were selected from the micrographs and a map of the $T=3$ shell for each of the constructs was computed to 8 Å (Fig. 3a and b). The maps of the modified cores displaying the model epitopes look similar in overall character to the native $T=3$ core, with well-defined helices forming the spikes, but also show some clear differences, as outlined below.

The S1-8 map has longer spikes than the native structure, while in the S2-16 map, the spikes are shorter. When sections of the maps are viewed as densities, each spike tip has a fuzzy weak density around it, not observed in the native map, which indicates that parts of the inserted sequences may be disordered (see Fig. S2). Also, the spikes in the modified cores are slender and narrow, resembling

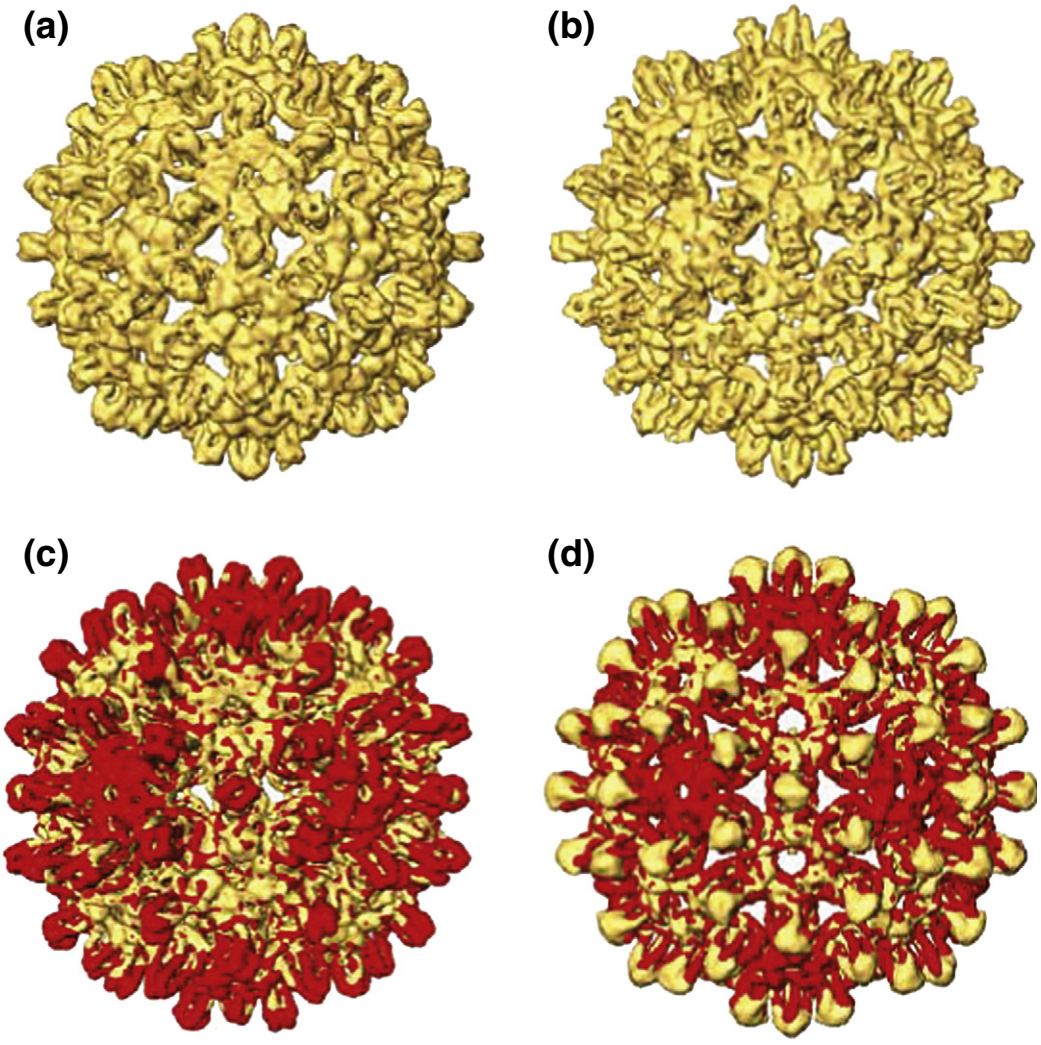



**Fig. 3.** Maps of $T=3$ core shells with inserted epitope. (a) S2-16. (b) S1-8. (c and d) The same maps in yellow with the native $T=3$ map superimposed in red. In each case, the region of the shell surface around the 3-folds (local 6-folds) lies at the same radius as in the native construct. Around the 5-fold axes, the shell surface in the constructs has dropped, and the native shell extends to higher radius, as indicated by the red color at this position. In (c), the spike tips appear red, indicating that the spike in the S2-16 construct is shorter than that in the native, whereas in (d), the tips appear yellow, indicating that the spike in the S1-8 construct is longer than that in the native.

more closely those observed in the mature $T=4$ viral cores than those in the $T=3$ or $T=4$ cores of the recombinantly expressed protein.[12] The density in the shell region (excluding the spikes) agrees closely in most places, except that around the 5-fold axes, the shell region of the construct maps has

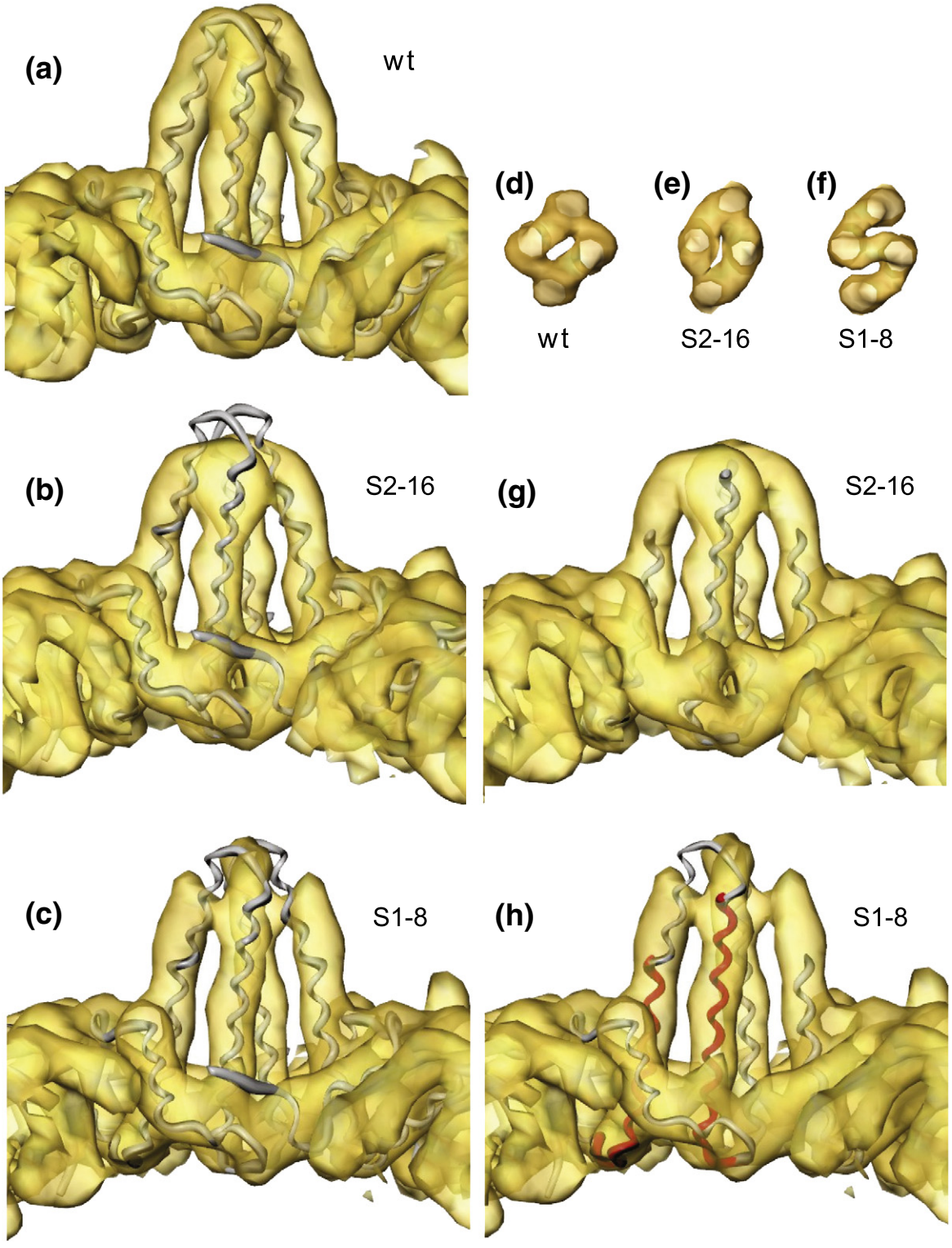



**Fig. 4** (*legend on next page*)

dropped to a lower radius (Fig. 3c and d). The changes in conformation can be accounted for largely by a bodily rotation of the subunit, but some internal changes are also needed to fully explain the differences.

When the atomic model of the $T=3$ core, obtained by fitting the subunits to the native $T=3$ map (Fig. 4a), is overlaid on the construct maps (Fig. 4b and c), the hairpin helices forming the spike do not lie correctly in the density of either construct map. It is apparent that the subunits need to be tilted to fit the helices into the map density (Fig. 4g and h), which can be achieved quantitatively by docking of the native core protein chains as individual subunits into the construct maps. The change in conformation in the shell can be accounted for mostly by a rigid-body rotation of the subunits by a few degrees about an axis roughly perpendicular to the plane of the interface between the subunits of the dimer. The effect is to reduce the crossover of the helical hairpins from apposing dimers, resulting in a narrower profile of the spikes, much like that observed in the $T=4$ cores obtained from mature virions.[12] The rotation (of ~6°) is largest for the A chains, which are located in rings around each 5-fold axis. These C-terminal helix 5s (Fig. 1b, c, e, and f) associate in a ring around the 5-fold axis, forming the shell layer at this position. The rigid-body rotation of the subunit, bringing the spike helices into a more radial orientation, also has the effect of pivoting the helix 5s about a point near the base of the helices forming the hairpin, causing the shell layer around the 5-fold axis to drop to a lower radius (Fig. 3c and d). The subunit at the B position rotates less and pivots about the C-terminal tip of helix 5, leaving the shell radius unchanged here. The subunits forming the spike at the C positions also show a small rotation, also pivoting about the C-terminal tips of helix 5, and achieving the more slender spike profile.

It is clear that though the fit of the subunits into the spike density of the constructs has improved through rigid-body reorientation of the subunits, smaller additional changes are also needed to explain the observed differences. In particular, the outer segment of kinked helix 4 (helix 4a) needs to be reoriented into density, predominantly as a rigid body, but the exact position of the tip of helix 4a cannot be certain from interpreting maps at this resolution, though the strong density associated with the spike structure is consistent with a helical structure. The degree of tilt of helix 4a needed to fit it into the density in the outer region of the spike is ~20° for S2-16 and ~10° for S1-8. Since α-helices are resolved as strong tubular densities in these maps at 8 Å resolution, the helices in the constructs can be compared with those in the native core, and their lengths can be estimated with an accuracy of one turn of the α-helix (~3 or 4 aa). It is clear that helix 3 and helix 4a in the constructs show differences in length compared with those in the native core. This affects the position of the foreign epitope in the constructs, as will now be described.

## Positioning of the foreign epitope

The S2-16 spike appears to be a shortened version of the native one (Fig. 4b and g). Helix 3 looks the same length but helix 4a is shorter by ~1 turn (Fig. 4b). (In the diagrams shown in Fig. 4, helix 4 of the core protein subunit appears on the left and helix 3 appears on the right in the hairpin nearer to the observer.) This is consistent with the fact that 2 aa have been removed from within this helix and suggests that the inserted linker sequence L2 must in this case be mostly helical. The connection between the unchanged helix 3 and the shortened helix 4 appears the same as in the native; hence, the 5-aa linker L1 added to the loop region does not form a defined density and is not seen. It probably gives rise to the fuzzy weaker density apparent around the spikes. In summary, in S2-16, L1 is disordered, and L2 has been incorporated within helix 4 to replace the deleted helical part of the native structure. The DPAFR epitope is likely to adopt a helical configuration to replace the native DPASR/DPISR epitope on the end of the shortened helix 4a.

In the map of the S1-8 construct, helix 4a also appears about one turn shorter at its N-terminal end compared with helix 4a in the native structure

**Fig. 4.** Spike conformations in the modified cores compared with the native type (wt). (a) Spike at the strict 2-fold symmetry axis showing the native $T=3$ map and atomic model. (b) Map of the S2-16 construct overlaid on the native $T=3$ model spike, same view as in (a). (c) Map of the S1-8 construct overlaid on the native $T=3$ model spike, same view as in (a) and (b). (d–f) Radial views of the tip of the spike, showing native (d), S2-16 (e), and S1-8 (f). In each case, the C,C dimer is shown, but the computationally independent A,B dimers in each construct display very similar features. In particular, the additional bridging density seen in S1-8 (f) compared with wt (d) or S2-16 (e) is thus unlikely to have arisen from the averaging of noise. (g) S2-16 spike, same view as in (a) to (c), with molecular fragment of spike, as shown, reoriented to fit in the density. (h) S1-8 spike, same view as in (a) to (c), and (g), with fragment of spike reoriented to fit in the density. The fitted fragment is shown in red and is the same fragment that was fitted in (g). The larger fragment in gray has been aligned with the red fragment to show the improved fitting in the extended part of helix 4a as well as the loop region. The fragment fitted included helix 3 and helix 4b, but not helix 4a or the loop region between helix 3 and helix 4a. One turn of α-helix before helix 3 and beyond helix 4b was also included. (b) and (c) show the mismatch of the native structure to the constructs. (g) and (h) show how the change can be modeled largely as a rigid-body movement of the subunits. (f) shows the bridge of density joining the subunits in S1-8. (e) shows a distortion compared to (d), but not the bridge shown in (f).

(Fig. 4c and h), indicating that in this construct, the inserted linker sequence L2 does not continue the α-helix. Helix 3 appears to be extended by one turn. The DPAFR epitope is positioned between the two linkers L1 and L2. Since the S1-8 spike is longer than the native spike, the linkers must extend outwards from their constrained end points and present the epitope at the tip of the extended spike. This is a quite different configuration to the native epitope or to the foreign epitope in S2-16. The strong additional feature in S1-8 at the tip of the spike lying across the 2-fold axis (Fig. 4f *versus* Fig. 4d and e) indicates extra material and a new interaction between the subunits at the dimer interface. This compact density may in part correspond to the inserted DPAFR epitope forming a short helix. In the solution structure of the preS1 protein from the HBV surface antigen, a short region containing part of the DPAFR epitope was shown to be helical.[35] The fuzzy density surrounding the spike indicates that there is some disordered chain in this structure too (Fig. S2).

Thus, the sequence modifications at the tips of the spikes cause entire subunits to undergo similar rigid-body rotations, giving rise to the similar observed changes in the shell region of the structure. The overall rigid-body rotation of the subunits is

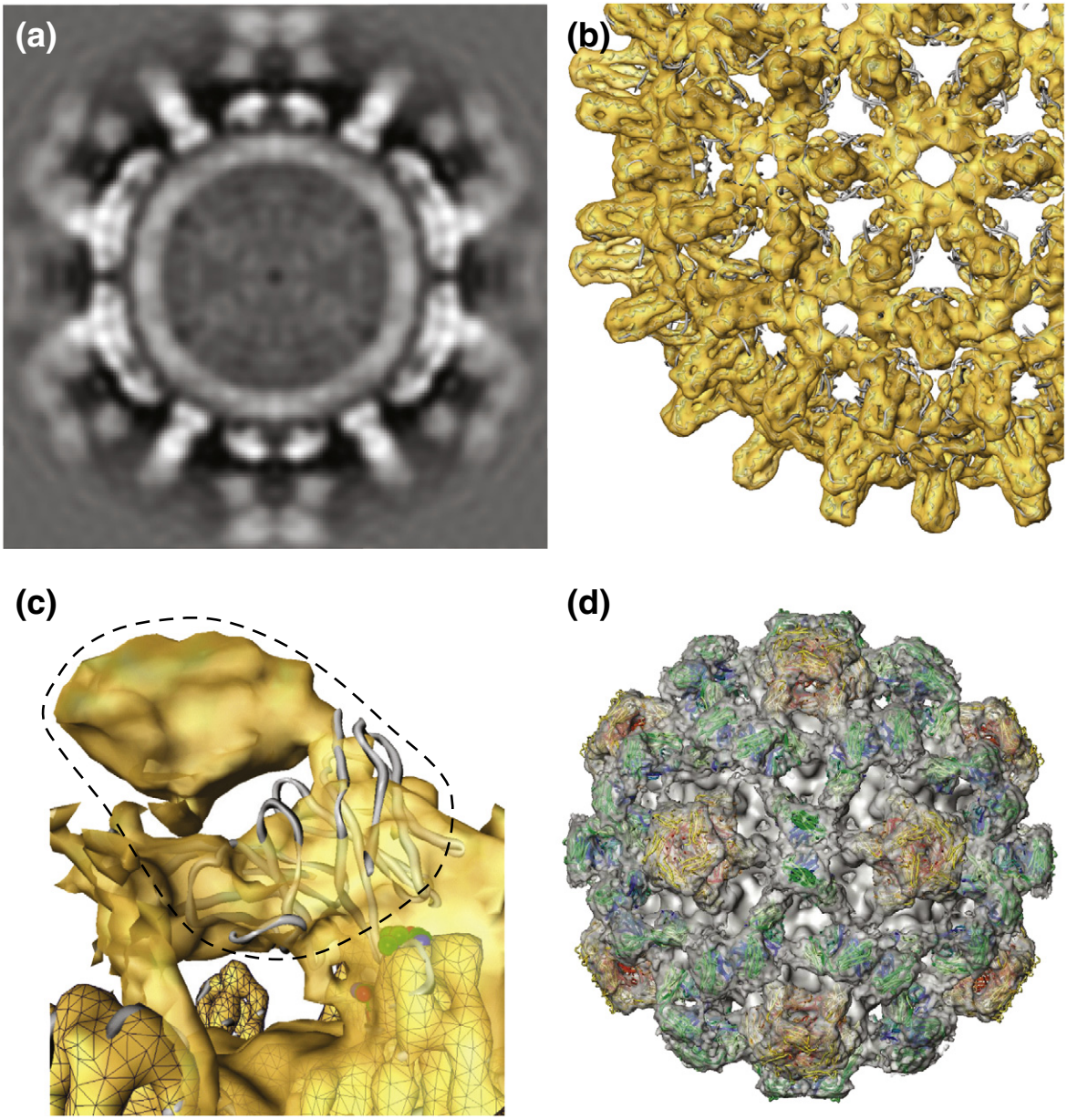



**Fig. 5** (*legend on next page*)

also responsible for the changes in the profile of the spikes, making them more slender than in the native shell. In addition, helix 4a reorients, causing or caused by a reconfiguration of the tip structure, which is related to the overall subunit reorientation. In both constructs, some of the inserted linker sequence is disordered. The structure of S1-8 was not as expected, since the inserted linker sequences caused helix 3 to be extended rather than helix 4a. Together with the additional partially ordered material, this makes the S1-8 spikes longer than those in the native structure. In the S2-16 construct, the shortening of helix 4a and the disordering of the loop at the tip make the spikes look shorter than in the native structure.

### Complexes of Fabs with the native core shell

An initial study of a complex of the well-characterized C1-5 antibody,[36–38] which recognizes the immunodominant DPISR epitope (more precisely 78-DPIxxD-83)[38] at the tips of the spikes of the native core shell, was made to allow comparison with the interactions of the MA18/7 antibody with the engineered core shells. Particles were mainly (90%) of the larger $T=4$ form, and a map of the $T=4$ complex computed to 8 Å resolution gives an indication of the level of detail that can be obtained in a favorable case. As this complex has already been studied,[27] only brief details are given here to allow comparison with the constructs with a foreign epitope.

The map of the complex (Fig. 5a) shows that the diameter of the particle has increased from 350 Å for the unlabeled core to 470 Å in the complex, illustrating the large size of the antibody compared to the spikes on core shell (see also Fig. S3). Because the density for the fragment of antibodies (Fabs) is weaker than that for the core (see below), the core itself in the complex can be visualized by choosing a density threshold intermediate between Fab and core (Fig. 5b). The structure thus displayed closely matches the uncomplexed core (Fig. 1a), as shown by the good fit of the crystal structure into the density (Fig. 5b). Comparison of the map of the complex displayed at a lower density level (Fig. 5d) with the view of the core shell (Fig. 5b) shows that the Fabs form an additional outer shell of density with two main motifs, a turret-like nearly cylindrical density on the 5-fold axis and a rectangular bridge of density across the 2-fold (local 6-fold) axis. These features represent superpositions of antibody fragments bound to a subset of the icosahedrally related sites. The computed map is icosahedrally averaged, and at each position, the density represents the average of the occupancy over all the symmetrically related sites.

Examination of the additional density due to the Fabs indicates that the four independent binding positions on the A, B, C, and D subunits have differing occupancies, demonstrating that their environments are distinct from each other. Antibody binding at one site may block binding at another site, as has been reported previously, for example.[39] In the core protein, the spike, being a dimer, contains two closely positioned quasi-equivalent antibody binding sites. The footprint of binding of the Fab to one site overlaps with the other; thus, only one of these can be occupied at once, giving a maximum averaged occupancy level of 50%, for all potential sites (Fig. S4). The Fabs project off the epitope at the corner of the spike at an angle of about 45° to the radius (Fig. 5c), as observed previously.[27] Because of this orientation, the Fab fragments extend out sideways over a relatively long distance and block access to binding sites on neighboring spikes. The A chains are arranged around the 5-fold axis, and their bound Fabs face inwards towards the 5-fold axis, forming the turret-like feature. The average occupancy for the A site is 0.4, indicating that, on average, two of the possible five A sites are filled. In the icosahedrally averaged map, this is represented as a partial occupancy of 0.4 at all the sites, and the density at the turret-like feature is an average of 2 Fabs per 5-fold motif, bound at random positions. Molecular modeling (Fig. 5c and d and Fig. S4) indicates that a maximum of two of these sites (of five potential A sites) can be filled at once, in good agreement with the measured occupancy.

**Fig. 5.** Map of the complex of C1-5 Fab with the native $T=4$ shell. (a) Central slice through Fab complex map. (b) Core protein region of antibody complex. The Fab density is weaker than the core [see (a)], because of low occupancy of sites due to blocking, and has been suppressed by setting the display surface at a high level. The core density is sharply defined and the coordinates from the crystal structure (overlaid gray tubes) fit precisely. (c) Close-up view of one spike and its associated Fab density (outlined with the broken line) with the modeled Fv chain (drawn as a gray tube) docked into the density adjacent to the epitope (shown as colored space filling). The density more distant from the epitope corresponds to the constant domain of the Fab. The surface of the core protein [as in (b)] is indicated by mesh work. (d) Map of the whole complex with Fabs modeled in to show how the overall density arises, although not all sites can be occupied because of blocking (see the text; Fig. S4). For clarity, the model represents 100% occupancy of A and C sites, with no labeling on B or D sites. Fabs on A sites have their heavy and light chains colored red or yellow, respectively. Those on C sites are colored green (heavy chain) and blue (light chain). Since the occupancy of the B sites is very low, and there is a large overlap of the Fabs on the C and D sites, the modeled Fabs fill most of the observed EM density. Figure S3 shows a section similar to (a), with annotation delimiting the shell region from bound Fabs.

Similar effects cause the partial occupancy around the 2-fold (local 6-fold) axes, where the three different types of quasi-symmetry-related sites (on the B, C, or D chains), six in total, can clash, giving rise to the bridge-like features. The occupancy at the B site, B being the outward facing chain of the dimers arranged around the 5-fold axis, was measured as less than 0.1. This low occupancy can be explained by the blocking effect of Fabs at the C and D sites. The occupancy on the C and D chains positioned around the local 2-fold axis was 0.4 and 0.2, respectively. The difference in occupancy at these two similar sites could be due to their different structures or environments—which is a consequence of the quasi-symmetry. The orientation of Fab binding and separation of spikes on the core appear to be compatible with bivalent binding of an antibody.

### Fab complexes with the modified core shells containing inserted epitopes

The core shell constructs were labeled with Fabs derived from the monoclonal antibody MA18/7 to the DPAFR epitope. The maps of the labeled S2-16 complex are similar in character to the C1-5-labeled native complex. Figure 6a and b show equivalent views of $T=4$ maps of the native and S2-16 Fab complexes, at the same nominal resolution of 15 Å. The Fabs project out from the spikes at a similar angle to form similar motifs at the 5-fold or 2-fold symmetry axes. Slight orientation differences could be due to the different nature of the interaction in the two complexes, as well as the epitope being presented differently. In addition, it is possible that the two different antibodies (C1-5 and MA 18/7) exhibit slightly different elbow angles.[40] However, the major axis of corresponding Fabs lies in a very similar direction and thus gives rise to very similar maps. Since they are so similar, it is very likely that the position and orientation of presentation of the epitopes are very similar and that the model epitope has essentially replaced the native epitope at the top of helix 4.

Cross-reactivity of the antibodies has been tested in Western blot, direct and competitive ELISA, and immunogold EM (P.P. *et al.*, unpublished results). The anti-DPAFR MA18/7 antibody does not cross-react with the native DPASR or DPISR core epitope in Western blot, ELISA, or immunogold EM. No sign of even weak reaction was ever seen. Vice versa, the C1-5 antibody does not recognize S1-8, S2-16, other HBc (core protein from HBV) VLP carriers bearing the DPAFR epitope and lacking native DPISRD sequence, or RNA phage coat protein-preS1 fusions, which contain DPAFR epitope in different amino acid neighborhood (P.P. *et al.*, unpublished results).

Comparison of the $T=4$ and $T=3$ S2-16 maps (Fig. 6b and c), which have identical Fab binding at the molecular level, illustrates how the same features manifest differently according to the different $T=3$ or $T=4$ packing. The turret-like motifs around the 5-fold axes correspond closely, but in the $T=3$ structure, the bridge across the local 6-fold axis has strict 3-fold symmetry, as opposed to having strict 2-fold symmetry in the $T=4$ structure. Apart from the symmetry difference, these bridges are similar in character.

Since the majority of the S1-8 particles were of the $T=3$ type, only a $T=3$ map of the S1-8 Fab complex was calculated. In Fig. 6c and d, the S2-16 $T=3$ map is compared with the S1-8 $T=3$ map. The additional density due to the Fabs in the S1-8 map extends radially outwards from the spikes on the shell, rather than at an angle (~45°) as for the S2-16:MA18/7 Fab or native core:C1-5 Fab complexes. The S1-8:MA18/7 Fab map does not show the same motifs as the native core:C1-5 Fab or S2-16:MA18/7 Fab maps on the 5-fold or 3-fold axes. Because of the radial disposition of the Fabs in the S1-8 complex, the density due to Fabs is less distinct than for the S2-16 complex, as it does not form the distinctive "bridges" or interacting motifs across the symmetry axes. Thus, blocking between sites on adjacent spikes does not occur, and the blocking effect is restricted to the overlap of the two closely spaced sites on the tip of a given spike.

The Fab density on S1-8 is strongest near the tips of the spikes on the core shell. It becomes weaker with increasing distance from the surface, indicating a level of flexibility or disorder. The density lying close to the dimer interface (2-fold axis for C,C spike, or local 2-fold for A,B spike) is enhanced by 2-fold averaging (computational 2-fold averaging at the C, C spike, or superposition of quasi equivalent densities on the A,B spike). This combination of disorder and 2-fold averaging makes it difficult to interpret the outer density in detailed terms of the shape of a Fab.

An additional density feature observed at the tip of the spike in the S1-8 unlabeled core (Fig. 4f) indicates that the DPAFR epitope lies close to the 2-fold or local 2-fold axes at the tip of the spike of the dimer in the S1-8 construct. This explains why the DPAFR epitope on S1-8 is accessible, though in a very different environment to S2-16. It also raises the possibility of interference of the two epitopes on the dimer with one another, since they are in close proximity. The location of the epitope in an unconstrained position in a loop at the tip of the spike may also explain why in this complex the Fab fragments do not appear well ordered. Since S1-8 and S2-16 contain exactly the same epitope, the different levels of immunogenicity observed in mice must be due to the way it is presented.

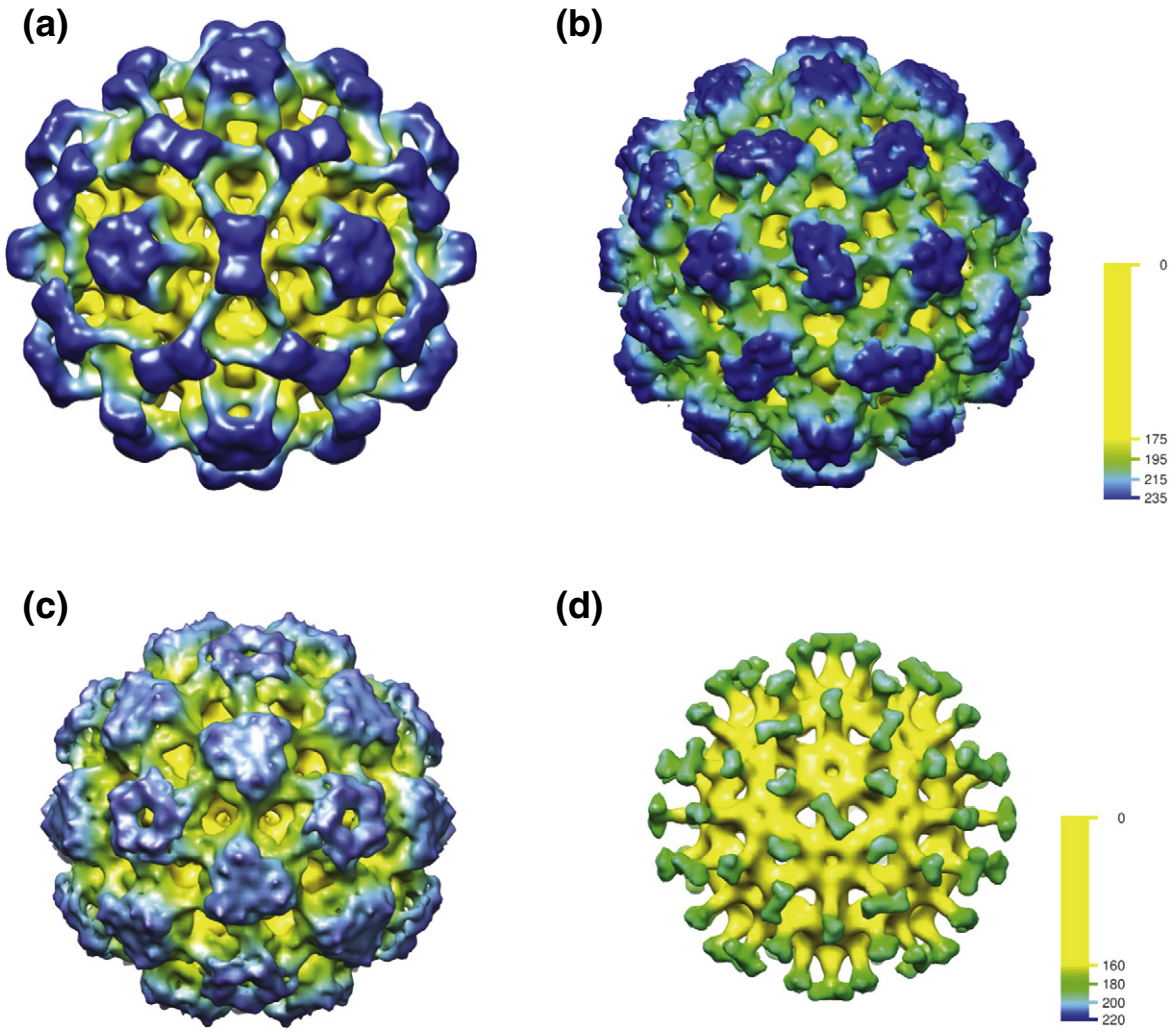



**Fig. 6.** Maps of various complexes of Fabs and core shells. (a) $T=4$ native shell with C1-5 Fab (as Fig. 5a and d, but at lower resolution to compare with construct maps). (b) $T=4$ S2-16 construct with MA18/7 Fab. (c) $T=3$ S2-16 construct with MA18/7 Fab. (d) $T=3$ S1-8 construct with MA18/7 Fab. Note that the maps in (a), (b), and (c) have a similar character, with the Fabs coming together to make strong features around the symmetry positions [5-fold and 2-fold in (a) and (b); 5-fold and 3-fold in (c)], whereas in (d), the density arising from the Fab is radial and does not create features around the symmetry axes. The maps were filtered to 15 Å resolution. The radial color scales shown are calibrated in angstroms (Å). The core shell is colored yellow, and then the color graduates in 20-Å bands from green, to light blue, to dark blue as the radius increases. The outer limit of the shell radius is 160 Å or 175 Å for the $T=3$ or $T=4$ shells, respectively.

In summary, in S2-16, the model epitope is in a position and conformation closely resembling the immunodominant epitope in the native shell. Bound Fabs project from the "corner" of the spike, at an angle of ~45°, causing blocking effects between sites on different spikes. In S1-8, the DPAFR epitope is located more centrally on the tip of the spikes, near the dimer interface and the 2-fold axis. The density of bound Fabs projects radially outwards from the center of the spike. No blocking between sites on different spikes will occur with this geometry. The footprint of antibody binding will only allow one antibody or Fab to bind per spike on either of the constructs or to the MIR on native cores.

## Conclusions

HBV core protein has an inherent and complex immune stimulating capability,[10] which is both T-cell independent and T-cell dependent.[41] The native core protein assembled into particles can present many copies of an epitope with high density and appropriate spacing to cross-link B-cell receptors

optimally.[42] For the native particle, this produces a strong antibody response against the major immunodominant epitope positioned at the tips of the spikes on the core shells. By removing the native immunodominant epitope and inserting a foreign sequence in its place, it is possible to elicit a strong immune response to the inserted sequence. The inserted sequence may be a large protein domain or just a small peptide. In the latter case, as we have studied here, the exact positioning of the foreign sequence can have large effects on the immunogenicity of the construct.

For the modified core construct to induce an effective immune response, the displayed foreign epitope must be able to bind strongly to receptors on B-cells in a way that promotes the cross-linking of clustered receptors. The structure of the MIR at the tips of the HBV core protein spikes potentially allows the inserted epitope to be either displayed as a loop constrained at both ends by fixed α-helices in the hairpin or incorporated in a helical conformation at the end of one of the radially extending helices. The two constructs studied here appear to exemplify these two modes of display. Small changes in the construct, for example, in the number of amino acids in the insert, can have large effects on the orientation and accessibility of the epitope. If the epitope is embedded in an α-helix, then a change of one amino acid position in the sequence can cause the epitope to be reoriented by 100°. In addition, changes made to the core protein around the MIR cause a change in orientation of the subunits in the dimer forming the spikes on the exterior of the core shell and this may affect the orientation and accessibility of the epitope presented.

In the S2-16 construct studied here, the inserted foreign epitope appears to be positioned in a partly helical conformation that closely mimics that of the native immunodominant core epitope. The binding of antibodies C1-5 and MA18/7, respectively, to the two kinds of shell is very similar, with the Fab fragments projecting off the "corners" of the spikes at an angle of ~45° to the radius. This overall arrangement is therefore likely to be a favorable one for strong interactions between core shell and membrane-bound B-cell receptors, exploiting an equivalent geometry. Like the native core protein, S2-16 might therefore be expected to produce a strong immune response, as indeed it does.[29] In the S1-8 construct, by contrast, the foreign epitope appears to be displayed as a less ordered feature right at the tip of the spike and the bound Fab fragment lies in a radial direction. This Fab arrangement may indicate a much less favorable geometry for interaction of the foreign epitope with the equivalent arm of a membrane-bound B-cell receptor, leading to less efficient cross-linking and a correspondingly weaker immune response than that observed against S2-16.[29] In addition, in S1-8, the epitope is presented exposed near the 2-fold axis right at the tip of the dimer spike. This might allow interactions with the partner epitope from the other member of the dimer to change the molecular surface presented. In the S2-16 construct, this could not occur. There is also the question of whether a small inserted sequence in the engineered core adopts a conformation like that of the epitope in the context of its native protein. In our case, monoclonal antibody MA18/7 was raised against native surface antigen,[30] and since both S1-8 and S2-16 react with MA18/7 and, in mice, induce production of antibodies that react with the native preS1, they must both be displaying the DPAFR epitope in something close to its native conformation. However, in S2-16, the epitope is embedded in the native-like α-helix 4a, which may have a stabilizing effect on its conformation. This more rigid and defined location and environment may also favor stronger interactions with the membrane-bound B-cell receptors. The organization of the linkers in S1-8 does not provide the same level of constraint. Different modes of insertion may have different constraints, and it may be that the SplitCore approach,[43] which provides a less constrained setting, will prove advantageous especially for larger domain inserts.

Therefore, in the design of virus-like particles for use as vaccine carriers, it is important to consider the three-dimensional (3D) structure of the entire particle. Since core particles carrying preS1 epitopes are considered as a prospective therapeutic vaccine against chronic hepatitis B,[2] our data may help in this particular case. The overall conformation changes in the spikes observed for the two different constructs were similar; hence, this information could be used when designing constructs with other epitopes. However, especially since the core particles are not strictly rigid scaffolds to which the epitopes become attached and it is also hard to predict the conformation that will be adopted by any particular inserted sequence, it will help to visualize the structures of further constructs and relate them to the results of functional assays.

## Materials and Methods

### Production of core shells

Cloning, expression, and purification of the HBV core particles S1-8 and S2-16 with DPAFR epitope inserted at the MIR were performed as previously described.[29] Briefly, *E. coli* K802 (*hsdR*, *gal*, *met*, *supE*, *mcrA*, *mcrB*) cells were grown overnight on a rotary shaker at 37 °C in 750-ml flasks containing 300 ml of M9 minimal medium supplemented with 1% casamino acids (Difco Laboratories, USA) and 0.2% glucose. An $OD_{540}$ (optical density at 540 nm) of 2–5 was usually reached. Cells were pelleted

and lysed by 30-min incubation on ice in lysis buffer containing 50 mM Tris–HCl, pH 8.0, 5 mM ethylenediaminetetraacetic acid, 50 μg/ml PMSF, and 2 mg/ml lysozyme, and then ultrasonicated 3 times for 15 s at 22 kHz. Lysates were adjusted to 10 mM $MgCl_2$ and 20 μg/ml DNase and after 5-min incubation to 0.1 M urea. After low-speed centrifugation, proteins were precipitated from the supernatant with ammonium sulfate at 33% saturation for 1–2 h at 4 °C. Ammonium sulfate precipitates were washed with phosphate-buffered saline (PBS) buffer and then dissolved in PBS containing 1.5 M urea and 0.6% Triton X-100 just before loading onto the Sepharose CL4B column (2.5 cm × 85 cm). The S1-8 and S2-16 capsids were eluted from the column with PBS buffer containing 0.25 M urea and 0.01 % Triton X-100. The presence of S1-8 and S2-16 polypeptides in fractions was tested by PAGE. Positive fractions were pooled and concentrated by ammonium sulfate precipitation at 33% saturation for 20 h at 4 °C. Pellets were resuspended in PBS or in Tris–saline buffer—10 mM Tris–HCl, pH 7.5, and 150 mM NaCl—to a final concentration of about 2 mg/ml, dialyzed overnight against 2000 volumes of the same buffer, and stored at −70 °C in 50% glycerol.

The native $T=4$ particles for labeling with the C1-5 antibody were prepared exactly as described previously, for the full-length CW isolate.[12] For isolation of $T=3$ particles, the construct HBcΔ-CW, in which the core protein is truncated after amino acid 149,[44] was subject to a further truncation after amino acid 140. This was produced by PCR using oligonucleotide primers 5′ GA CTT CAA CAT ATG ACA TTG ATC CTT ATA AAG 3′ and 5′ CCG GAA TTC TCA TAA GAT AGG GGC ATT TGG 3′. The resulting PCR product was cleaved with NdeI and EcoRI and inserted into the corresponding restriction sites in the expression plasmid PT7-SC. Preparation of $T=3$ shells was then done in the same way as for $T=4$ shells,[44] but with the modification that in the final sucrose gradient centrifugation step, the later fractions of the peak were selected.

### Antibody labeling

Fabs from the C1-5[36] or MA18/7[30] antibodies were prepared by digestion with Papain,[45] using Pierce Kit 44885 (Thermo Fisher Scientific, Cramlington, UK). Antibody fragments were mixed with core shells, at a ratio of ~1 Fab per HBV core dimer, before freezing on EM grids.

### Electron cryomicroscopy

Samples of core shells, or antibody complexes, were prepared for EM on holey carbon films, by standard methods.[12,17] Frozen hydrated samples were imaged at 300 kV or 200 kV, with a nominal magnification of 50,000×, on an FEI Tecnai F30, or F20, or 60,000× on a Hitachi HF-2000 electron microscope. Images were recorded on Kodak SO-163 film and developed in full-strength Kodak D19 for 12 min. Films were digitized using a Zeiss SCAI Scanner (Z/I imaging, Oberkochen), sampling at 7 μm per pixel. Two-by-two pixel binning was then applied, giving a resultant 14 μm per pixel. Table S1 shows a summary of imaging conditions and parameters for each structure analyzed.

### Image analysis and 3D reconstruction

Image processing, 3D map computation, and refinement of the magnification were performed as before,[12] using the MRC programs for icosahedral reconstruction.[46,47] Particles were selected manually using Ximdisp[48] or automatically using FindEM.[49] The sampling for the images of the non-native-labeled particles was reduced by a further factor of 2 by binning adjacent pixels. SPIDER[50] procedures were used for iterative alignment and centring of particles. Imagic[51] was used for multivariate statistical analysis and classification of the aligned particle sets. A significant portion of the non-native core particles suffered from poor assembly. Classes of particles forming complete, nondeformed shells were selected visually. These classes appeared unbroken and isometric. The common-lines procedure[47] was used to find the orientation of the particles directly from the class averages. The four or five best scoring class averages were used to reconstruct an initial 3D model. After this, individual particles from the selected classes were refined iteratively against projections of the current best 3D model, as before.[11,12,17] Initially, the defocus used to fit a contrast transfer function for each image was determined from the summed power spectra of the particles.[17] The contrast transfer function was refined iteratively in the final stages of the reconstruction procedure.[12] Finally, the maps of unlabeled cores were sharpened by weighting the Fourier amplitudes so as to match the rotationally averaged one-dimensional amplitude profile with the atomic model of the $T=3$ shell (as fitted), as previously described.[12] Maps of unlabeled cores were reliably determined to 8 Å resolution (Fig. S5), as defined by the Fourier shell correlation criterion 0.33.[52] The maps of Fab labeled cores were determined to beyond 15 Å, or 8 Å in the case of the native core:C1-5 Fab complex (Figs. S6 and S7. The final maps of unlabeled cores and the native core:C1-5 Fab complex were limited to 8 Å resolution by Fourier truncation for analysis, comparison, and display. The Fab complexes of S2-16 and S1-8 were similarly limited to 15 Å for analysis and display. The hand of the native $T=3$ and $T=4$ maps was assigned directly based on the distinctive handed motifs seen at the 5-fold axis (since the atomic model for $T=4$ was available). The hand of the other, lower-resolution, maps could be assigned based on comparing coarser handed features with the known $T=3$ and $T=4$ maps.

### Docking and modeling

DockEM[53] was used to fit models and fragments into 3D maps. A magnification search with atomic models and the EM maps was used to determine the magnification of the EM maps accurately. A modification of DockEM, to restrict searches to rotations about a specified pivot point, was implemented for fitting of helix fragments. *$T=3$ quasi-atomic model:* The chains from the crystal structure of the $T=4$ shell[18] were fitted as dimers, and as individual chains, into all three unique positions in the $T=3$ map, using DockEM. For docking, a density was created from the atomic models and filtered to match the resolution of

the EM map, 8 Å. Once a preliminary fit of the dimer had been obtained, the relative magnification was optimized by including a relative scale in the local search to match the dimer in the density.

A full atomic model of the $T=3$ shell was created by applying icosahedral symmetry to the fitted chains at the A,B and C,C spikes, using the CCP4[54] program PDBSET. A density map created by filtering this model to 8 Å was used to optimize the magnification scale of the other $T=3$ core shell maps, by correlation in DockEM. A scale search repeated using the native $T=3$ EM map, or a density model from the coordinates created with the tips of the spikes deleted, produced the same results. *Modeling of core particles with DPAFR epitope inserted:* Superposition of the maps indicated changes in shell region and tips of spikes. Superposition of the native $T=3$ model and maps of the modified core particles also demonstrated changes. Initially, changes at the tip were not modeled due to the weaker less defined density in this region. A series of fragments of the subunit, excluding tip region, were created and docked into the density. It was found that changes in the shell domain could be modeled as rigid-body reorientations of the subunit structure. The tip changes were modeled by docking of small helical segment representing helix 4a to the existing fitted model. The extension of helix 3 in the S1-8 map was examined by superpositioning a model helix to extend from the tip of helix 3. *Fragment of antibody variable domain (Fv) docking and occupancy analysis:* An atomic model of the core particle antibody complex was obtained by docking a model of the Fv of antibody C1-5 into the map density, using information to 8 Å resolution. The docking of the Fv at the density associated with chain C proved the most reliable, due to a combination higher occupancy and less superposition effect than the A site. Symmetry-related positions on the other three chains were obtained by applying the quasi-symmetric relationships of the underlying chains. This enabled a full atomic model of the core shell Fv complex to be obtained. The atomic model of the C1-5 Fv was obtained by submitting the sequence[37] to the Web Antibody Modeling server.[55]

A model of the core shell with complete Fabs bound was obtained using an antibody model chosen to have a compatible elbow angle[40] with the density observed, selected from the Protein Data Bank (Protein Data Bank code: 1DBA). The Fv region of this model was aligned to the positioned C1-5 Fv models. The occupancy levels at all the different sites were determined by creating a simulated density from the atomic model of this complex, for all permutations of occupancy, in steps of 10%. The simulated densities were correlated with the observed EM map of the complex using DockEM,[53] using information to 15 Å resolution, with the best match indicating the correct occupancy. On the scale used, the maximum possible occupancy per binding site on each chain is 1. However, because only 1 of the 2 sites per dimer can be occupied, the maximum possible occupancy for C+D or A+B is 1; that is, if there was 100% C site occupancy, then no D sites could label. Therefore, if the A,B and C,D spikes were 100% labeled (i.e. one Fab per spike), then the total occupancy of all the sites would add up to 2. Atomic coordinate alignments and transformations were computed using Amira Software (Visage Imaging GmbH, Berlin, Germany), as were map rendering and molecular representations for Figs. 1–5. Figure 6 was made using UCSF Chimera.[56]

## Acknowledgements

We wish to thank Prof. Wolfram H. Gerlich (Giessen) for the gift of a preparative amount of purified MA18/7 antibody. P.P. was supported by Latvian grants 2010/0261/2DP/2.1.1.1.0/10/APIA/VIAA/052 and 2010/0224/2DP/2.1.1.1.0/10/APIA/VIAA/164. This work was supported by the Medical Research Council (grant number U105184319).

## Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2012.06.032

## References

1. Pumpens, P. & Grens, E. (2001). HBV core particles as a carrier for B cell/T cell epitopes. *Intervirology*, **44**, 98–114.
2. Pumpens, P., Ulrich, R., Sasnauskas, K., Kazaks, A., Ose, V. & Grens, E. (2008). Construction of novel vaccines on the basis of the virus-like particles: hepatitis B virus proteins as vaccine carriers. In *Medical Protein Engineering* (Khudyakov, Y., ed.), pp. 205–248, CRC Press, Taylor & Francis Group, Boca Raton London, New York.
3. Whitacre, D. C., Lee, B. O. & Milich, D. R. (2009). Use of hepadnavirus core proteins as vaccine platforms. *Expert Rev. Vaccines*, **8**, 1565–1573.
4. Clarke, B. E., Newton, S. E., Carroll, A. R., Francis, M. J., Appleyard, G., Syred, A. D. *et al.* (1987). Improved immunogenicity of a peptide epitope after fusion to hepatitis B core protein. *Nature*, **330**, 381–384.
5. Milich, D. R., Hughes, J., Jones, J., Sällberg, M. & Phillips, T. R. (2002). Conversion of poorly immunogenic malaria repeat sequences into a highly immunogenic vaccine candidate. *Vaccine*, **20**, 771–788.
6. Oliveira, G. A., Wetzel, K., Calvo-Calle, J. M., Nussenzweig, R., Schmidt, A., Birkett, A. *et al.* (2005). Safety and enhanced immunogenicity of a hepatitis B core particle *Plasmodium falciparum* malaria vaccine formulated in adjuvant montanide ISA 720 in a phase I trial. *Infect. Immun.* **73**, 3587–3597.
7. Gregson, A. L., Oliveira, G., Othoro, C., Calvo-Calle, J. M., Thorton, G. B., Nardin, E. & Edelman, R. (2008). Phase I trial of an alhydrogel adjuvanted hepatitis B core virus-like particle containing epitopes of *Plasmodium falciparum* circumsporozoite protein. *PLoS One*, **3**, e1556, http://dx.doi.org/10.1371/journal.pone.0001556.
8. De Filette, M., Jou, W. M., Birkett, A., Lyons, K., Schultz, B., Tonkyro, A. *et al.* (2005). Universal

influenza A vaccine: optimization of M2-based constructs. *Virology*, **337**, 149–161.
9. Nassal, M., Skamel, C., Kratz, P. A., Wallich, R., Stehle, T. & Simon, M. M. (2005). A fusion product of the complete *Borrelia burgdorferi* outer surface protein A (OspA) and the hepatitis B virus capsid protein is highly immunogenic and induces protective immunity similar to that seen with an effective lipidated OspA vaccine formula. *Eur. J. Immunol.* **35**, 655–665.
10. Vanlandschoot, P., Cao, T. & Leroux-Roels, G. (2003). The nucleocapsid of the hepatitis B virus: a remarkable immunogenic structure. *Antiviral Res.* **60**, 67–74.
11. Crowther, R. A., Kiselev, N. A., Böttcher, B., Berriman, J. A., Borisova, G. P., Ose, V. & Pumpens, P. (1994). Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell*, **77**, 943–950.
12. Roseman, A. M., Berriman, J. A., Wynne, S. A., Butler, P. J. B. & Crowther, R. A. (2005). A structural model for maturation of the hepatitis B virus core. *Proc. Natl. Acad. Sci. USA*, **102**, 15821–15826.
13. Borisova, G. P., Kalis, J. V., Pushko, P. M., Tsibinogin, V. V., Loseva, V. J., Ose, V. P. *et al.* (1988). Genetically engineered mutants of the core antigen of the human hepatitis B virus preserving the ability for native self-assembly [Article in Russian]. *Dokl. Akad. Nauk SSSR.* **298**, 1474–1478.
14. Gallina, A., Bonelli, F., Zentilin, L., Rindi, G., Muttini, M. & Milanesi, G. (1989). A recombinant hepatitis-B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. *J. Virol.* **63**, 4645–4652.
15. Birnbaum, F. & Nassal, M. (1990). Hepatitis B virus nucleocapsid assembly: primary structure requirements in the core protein. *J. Virol.* **64**, 3319–3330.
16. Zlotnick, A., Cheng, N., Conway, J. F., Booy, F. P., Steven, A. C., Stahl, S. J. & Wingfield, P. T. (1996). Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry*, **35**, 7412–7421.
17. Böttcher, B., Wynne, S. A. & Crowther, R. A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature*, **386**, 88–91.
18. Wynne, S. A., Crowther, R. A. & Leslie, A. G. W. (1999). The crystal structure of the human hepatitis B virus capsid. *Mol. Cell*, **3**, 771–780.
19. Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J. & Steven, A. C. (1997). Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature*, **386**, 91–94.
20. Geldmacher, A., Skrastina, D., Petrovskis, I., Borisova, G., Berriman, J. A., Roseman, A. M. *et al.* (2004). An amino-terminal segment of hantavirus nucleocapsid protein presented on hepatitis B virus core particles induces and a strong and highly cross-reactive anyibody response in mice. *Virology*, **323**, 108–119.
21. Kratz, P. A., Böttcher, B. & Nassal, M. (1999). Native display of complete foreign protein domains on the surface of hepatitis B virus capsids. *Proc. Natl Acad. Sci. USA*, **96**, 1915–1920.
22. Francis, M. J., Hastings, G. Z., Brown, A. L., Grace, K. G., Rowlands, D. J., Brown, F. & Clarke, B. E. (1990). Immunological properties of hepatitis B core antigen fusion proteins. *Proc. Natl Acad. Sci. USA*, **87**, 2545–2549.
23. Brown, A. L., Francis, M. J., Hastings, G. Z., Parry, N. R., Barnett, P. V., Rowlands, D. J. & Clarke, B. E. (1991). Foreign epitopes in immunodominant regions of hepatitis B core particles are highly immunogenic and conformationally restricted. *Vaccine*, **9**, 595–601.
24. Conway, J. F., Cheng, N., Zlotnick, A., Stahl, S. J., Wingfield, P. T., Belnap, D. M. *et al.* (1998). Hepatitis B virus capsid: Localization of the putative immunodominant loop (residues 78 to 83) on the capsid surface, and implications for the distinction between c and e-antigens. *J. Mol. Biol.* **279**, 1111–1121.
25. Conway, J. F., Watts, N. R., Belnap, D. M., Cheng, N., Stahl, S. J., Wingfield, P. T. & Steven, A. C. (2003). Characterization of a conformational epitope on hepatitis B virus core antigen and quasiequivalent variations in antibody binding. *J. Virol.* **77**, 6466–6473.
26. Belnap, D. M., Watts, N. R., Conway, J. F., Cheng, N., Stahl, S. J., Wingfield, P. T. & Steven, A. C. (2003). Diversity of core antigen epitopes of hepatitis B virus. *Proc. Natl Acad. Sci. USA*, **100**, 10884–10889.
27. Harris, A., Belnap, D. M., Watts, N. R., Conway, J. F., Cheng, N., Stahl, S. J. *et al.* (2005). Epitope diversity of hepatitis B virus capsids: quasi-equivalent variations in spike epitopes and binding of different antibodies to the same epitope. *J. Mol. Biol.* **355**, 562–576.
28. Sominskaya, I., Pushko, P., Dreilina, D., Kozlovskaya, T. & Pumpen, P. (1992). Determination of the minimal length of preS1 epitope recognized by a monoclonal antibody which inhibits attachment of hepatitis B virus to hepatocytes. *Med. Microbiol. Immunol.* **181**, 215–226.
29. Borisova, G., Borschukova, O., Skrastina, D., Dislers, A., Ose, V., Pumpens, P. & Grens, E. (1999). Behaviour of a short preS1 epitope on the surface of hepatitis B core particles. *Biol. Chem.* **380**, 315–324.
30. Heermann, K. H., Goldmann, U., Schwartz, W., Seyffarth, T., Baumgarten, H. & Gerlich, W. H. (1984). Large surface proteins of hepatitis B virus containing the pre-S sequence. *J. Virol.* **52**, 396–402.
31. Germaschewski, V. & Murray, K. (1995). Screening a monoclonal antibody with a fusion-phage display library shows a discontinuity in a linear epitope within PreS1 of hepatitis B virus. *J. Med. Virol.* **45**, 300–305.
32. Pontisso, P., Ruvoletto, M. G., Gerlich, W. H., Heermann, K. H., Bardini, R. & Alberti, A. (1989). Identification of an attachment site for human liver plasma membranes on hepatitis B virus particles. *Virology*, **173**, 522–530.
33. Glebe, D., Aliakbari, M., Krass, P., Knoop, E. V., Valerius, K. P. & Gerlich, W. H. (2003). Pre-s1 antigen-dependent infection of *Tupaia* hepatocyte cultures with human hepatitis B virus. *J. Virol.* **77**, 9511–9521.
34. Böttcher, B., Vogel, M., Ploss, M. & Nassal, M. (2006). High plasticity of the hepatitis B virus capsid revealed by conformational stress. *J. Mol. Biol.* **356**, 812–822.
35. Chi, S. W., Kim, D. H., Lee, S. H., Chang, I. & Han, K. H. (2007). Pre-structured motifs in the natively unstructured preS1 surface antigen of hepatitis B virus. *Protein Sci.* **16**, 2108–2117.

36. Bichko, V., Schödel, F., Nassal, M., Gren, E., Berzinsh, I., Borisova, G. *et al.* (1993). Epitopes recognized by antibodies to denatured core protein of hepatitis B virus. *Mol. Immunol.* **30**, 221–231.
37. Skrivelis, V., Steinberg, Y., Bichko, V., Gren, E. & Tsimanis, A. (1993). The structure of the variable regions of mouse monoclonal antibodies to hepatitis B virus core antigen. *Scand. J. Immunol.* **37**, 637–643.
38. Pushko, P., Sallberg, M., Borisova, G., Ruden, U., Bichko, V., Wahren, B. *et al.* (1994). Identification of hepatitis B virus core protein regions exposed or internalized at the surface of HBcAg particles by scanning with monoclonal antibodies. *Virology*, **202**, 912–920.
39. Bowman, V. D., Chase, E. S., Franz, A. W. E., Chipman, P. R., Zhang, X., Perry, K. L. *et al.* (2002). An antibody to the putative aphid recognition site on cucumber mosaic virus recognizes pentons but not hexons. *J. Virol.* **76**, 12250–12258.
40. Lesk, A. M. & Chothia, C. (1988). Elbow motion in the immunoglobulins involves a molecular ball-and-socket joint. *Nature*, **335**, 188–190.
41. Milich, D. R. & McLachlan, A. (1986). The nucleocapsid of hepatitis B virus is both a T-cell-independent and a T-cell-dependent antigen. *Science*, **234**, 1398–1401.
42. Fehr, T., Skrastina, D., Pumpens, P. & Zinkernagel, R. M. (1998). T cell-independent type I antibody response against B cell epitopes expressed repetitively on recombinant virus particles. *Proc. Natl Acad. Sci USA*, **95**, 9477–9481.
43. Walker, A., Skamel, C. & Nassal, M. (2011). SplitCore: an exceptionally versatile viral nanoparticle for native whole protein display regardless of 3D structure. *Sci. Rep.* **1**, 5, http://dx.doi.org/10.1038/srep00005.
44. Wynne, S. A., Leslie, A. G. W., Butler, P. J. G. & Crowther, R. A. (1999). Crystallization of hepatitis B virus core protein shells: determination of cryoprotectant conditions and preliminary X-ray characterization. *Acta Crystallogr., Sect. D*, **55**, 557–560.
45. Harlow, E. & Lane, D. (1988). *Antibodies: A Laboratory Manual.*Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; pp. 628–629.
46. Crowther, R. A., Henderson, R. & Smith, J. M. (1996). MRC image processing programs. *J. Struct. Biol.* **116**, 9–16.
47. Crowther, R. A. (1971). Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. *Philos. Trans. R. Soc., B*, **363**, 221–230.
48. Smith, J. M. (1999). Ximdisp—a visualization tool to aid structure determination from electron microscope images. *J. Struct. Biol.* **125**, 223–228.
49. Roseman, A. M. (2004). FindEM—a fast, efficient program for automatic selection of particles from electron micrographs. *J. Struct. Biol.* **145**, 91–99.
50. Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. & Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199.
51. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. (1996). A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24.
52. Rosenthal, P. B. & Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745.
53. Roseman, A. M. (2000). Docking structures of domains into maps from cryo-electron microscopy using local correlation. *Acta Crystallogr., Sect. D*, **56**, 1332–1340.
54. Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D*, **50**, 760–763.
55. Whitelegg, N. R. J. & Rees, A. R. (2000). WAM: an improved algorithm for modelling antibodies on the WEB. *Protein Eng.* **12**, 819–824.
56. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.

BMC Microbiology

RESEARCH ARTICLE Open Access

# Diversity of pili-specific bacteriophages: genome sequence of IncM plasmid-dependent RNA phage M

Janis Rumnieks* and Kaspars Tars

## Abstract

**Background:** Bacteriophages of the *Leviviridae* family are small RNA viruses with linear, positive-sense, single-stranded RNA genomes that encode only four proteins. All phages of this family require bacterial pili to attach to and infect cells. *Leviviridae* phages utilizing F-pili for this purpose have been extensively studied. RNA phages specific for conjugative plasmid-encoded pili other than that of plasmid F have been isolated, but are much less understood and their relation to the F-pili-specific phages in many cases is not known.

**Results:** Phage M has the smallest known *Leviviridae* genome to date and has the typical genome organization with maturation, coat and replicase genes in the 5′ to 3′ direction. The lysis gene is located in a different position than in other known *Leviviridae* phages and completely overlaps with the replicase gene in a different reading frame. It encodes a 37 residue long polypeptide that contains a transmembrane helix like the other known lysis proteins of leviviruses. Sequence identities of M proteins to those of other phages do not exceed 25% for maturation protein, 51% for coat protein and 41% for replicase. Similarities in protein sequences and RNA secondary structures at the 3′ untranslated region place phage M together with phages specific for IncP, IncC and IncH, but not IncF plasmid-encoded pili. Phylogenetic analysis using the complete genome sequences and replicase proteins suggests that phage M represents a lineage that branched off early in the course of RNA phage specialization on different conjugative plasmids.

**Conclusions:** The genome sequence of phage M shows that it is clearly related to other conjugative pili-specific leviviruses but has an atypical location of the lysis gene. It provides a better view on the remarkable diversification of the plasmid-specific RNA phages.

**Keywords:** *Leviviridae*, RNA phage, Pili-specific phage, IncM, Conjugative plasmid, Lysis

## Background

Bacteriophages of the *Leviviridae* family are small viruses that infect several genera of Gram-negative bacteria. They have linear, positive-sense, single-stranded RNA genomes about 3500 – 4200 nucleotides in length that encode only four proteins. All *Leviviridae* phages have three genes in common – maturation, coat and replicase [1]. The replicase cistron encodes the catalytic subunit of the RNA-dependent RNA polymerase complex, which is assembled together with several bacterial proteins [2,3] and replicates phage RNA. The coat protein forms dimers, 90 of which assemble in a *T*=3 icosahedral capsid about 27 nm in diameter and encapsidate the genome [4]. A single copy of the maturation protein binds to phage RNA [5] and gets incorporated into capsids along with it. It is required for infectivity of the virions – the maturation protein binds to bacterial pili, then leaves the capsid and enters the cell as an RNA-protein complex [6].

Many of the *Leviviridae* phages are divided in two genera – leviviruses and allolevivirus. The major distinction of allolevivirus is the presence of a minor coat protein A1 in their capsid which is produced by ribosomal read-through of a leaky termination codon of the coat gene [7]. The other difference is that the maturation protein of allolevivirus also triggers cell lysis [8,9], whereas leviviruses encode a dedicated small lysis polypeptide for this purpose [10-12].

* Correspondence: j.rumnieks@biomed.lu.lv
Biomedical Research and Study Centre, Ratsupites 1, Riga LV-1067, Latvia

The ssRNA phages that infect *Escherichia coli* cells by adsorbing to F plasmid-coded pili were the first isolates of the *Leviviridae* family [13,14], and to date these "male-specific" phages, with type species MS2 and Qβ, have been the most intensively studied and best characterized of this family. However, the F plasmid is just one of the many conjugative plasmids that are present in nature. These plasmids are often highly divergent from F and are most often grouped according to their mutual compatibility. In *Enterobacteriaceae*, the conjugative plasmids form more than 20 different incompatibility (Inc) groups which are denoted by capital Latin letters [15]. All these plasmids encode conjugative pili, but the pilin subunits often share no similarity.

Several ssRNA phages specific for conjugative pili other than that of plasmid F have been discovered. Phage PRR1 [16] which adsorbs specifically to IncP plasmid-encoded pili was the first such example, and later other phages specific for Inc group C [17], D [18], H [19,20], I [21], M [22] and T [23] plasmids followed. Phages PRR1, C-1 (IncC-specific) and Hgal1 (IncH-specific) have been sequenced [24,25] and phage PRR1 capsids have also been crystallized [26], but no research has been done on the other plasmid-specific phages since their isolation.

The IncM plasmid-specific RNA phage M [22] was isolated from sewage in Pretoria, South Africa in the beginning of the 1980s. IncM plasmids have a broad host range, code for rigid pili and transfer efficiently only when bacteria are growing on solid media [27]. Likewise, the phage is able to propagate in different strains of *Escherichia*, *Salmonella*, *Klebsiella*, *Proteus* and *Serratia*, provided they contain an IncM plasmid. To obtain more insight in plasmid-specific RNA phages, we determined the genome sequence of phage M and present here its analysis and comparison to the genomes of other RNA phages of the *Leviviridae* family.

## Results and discussion

### Overall structure of the genome

The genome of phage M is 3405 nucleotides long and follows the canonical *Leviviridae* genome organization with maturation, coat and replicase cistrons following each other in the 5′-3′ direction (Figure 1). An unusual feature of the genome is that the lysis gene appears to be located in a different position than in other leviviruses, as discussed below. It is also the smallest known *Leviviridae* genome to date, about 60 nucleotides shorter than that of the group II F-specific phage GA [28]. The protein coding regions of phage M are of similar length to those of phage GA, with maturation and coat genes being a bit longer and replicase somewhat shorter; the greatest savings in M's genome come from terminal untranslated regions (UTRs), the 5′ UTR being about 45 nucleotides and the 3′ UTR about 20 nucleotides shorter.

### Identification of the lysis gene

All members of the levivirus genus encode a short polypeptide that mediates cell lysis. Amino acid sequences of lysis proteins show great variation and their only unifying feature is the existence of a hydrophobic transmembrane helix within the protein [29]. Lysis proteins have been shown to accumulate in the bacterial membrane where they presumably form pores that lead to cell lysis [30]. In all of the known *Enterobacteria*-infecting leviviruses, the lysis gene overlaps with coat and replicase genes in a different reading frame and is translationally coupled with the coat gene [1]. However, in the genome of phage M, no candidate ORFs at this location could be identified: in the +2 frame relative to the coat gene there are no termination codons until the start of replicase and in the +1 frame only a 17 amino acid long ORF that would encode a non-hydrophobic peptide is found.

Up to now, there have been two reported cases in the *Leviviridae* family where the lysis gene in is in a different location: *Acinetobacter* phage AP205 has a short lysis gene preceding the maturation gene [31], while *Caulobacter* phage φCb5 codes for a longer, two-helix protein that completely overlaps with the replicase gene [32]. To test the possibility that phage M also has a non-canonical localization of the lysis gene, we utilized the fact that the pJET1.2 plasmid, where the cDNA copies of the genome were cloned for sequencing, contains a T7 promoter that can be used to transcribe the insert. Several clones with inserts in the correct orientation with respect to the T7

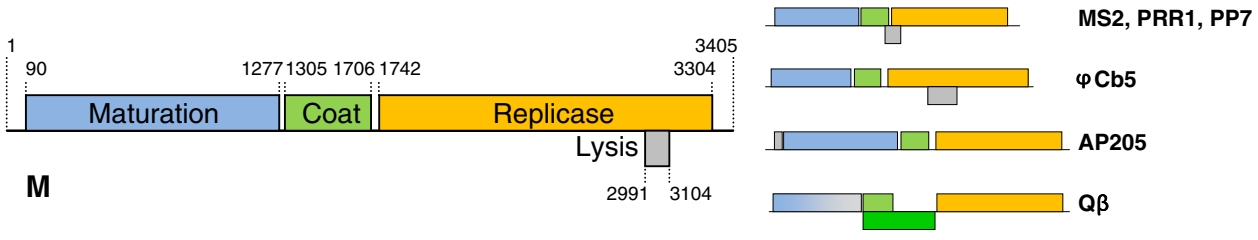



**Figure 1 Genome organization of phage M.** Start and end positions of phage genes are indicated. For comparison, the other known genome organizations of *Leviviridae* phages are represented on the right with genes color-coded as in the M genome. In phage Qβ, protein A1 (bright green) is an extended read-through variant of the coat protein and the lysis function is performed by the maturation protein.

promoter were selected and transformed to a T7 polymerase-producing *E.coli* strain. When the expression of T7 polymerase was induced, a clone containing an approximately 1000 nucleotide long fragment spanning nucleotides 2098-3129 of the phage genome resulted in a clear cell lysis. Examination of this sequence located a likely candidate for the lysis gene between nucleotides 2991-3104 (Figure 2A). This was based on several criteria: (1) it was the only ORF in the fragment with a significant length (37 amino acids; the shortest known *Leviviridae* lysis protein is that of phage AP205 with 34 amino acids); (2) according to the TMHMM server [33], the ORF-encoded protein was predicted to contain a transmembrane helix with over 95% probability; (3) although the ORF had an unusual initiation codon UUG, there was a rather strong Shine-Dalgarno (SD) sequence GAGG nine nucleotides upstream; (4) RNA secondary structure prediction using the RNAfold server [34] revealed that the initiation codon of the ORF is located on top of an AU-rich stem-loop that would presumably have sufficiently low thermodynamic stability to promote the initiation of translation [35] (Figure 2B). To verify the lytic function of the gene, the ORF together with the original SD sequence and UUG initiation codon was cloned in an inducible protein expression vector. Induction resulted in almost complete cell lysis some 45 minutes after (Figure 2C), thus demonstrating that the approximately 150 nucleotide long stretch is sufficient to encode a functional lysis protein. The above-mentioned evidence therefore lets us suggest with some confidence that this is the actual lysis gene of phage M.

### Protein similarities to other phages

The maturation proteins are very variable in *Leviviridae* phages, which is unsurprising given the vast diversity of pili they have evolved to bind. The maturation protein of phage M is most similar to those of the other plasmid-specific RNA phages, but the sequence identity is only 24.5% to phage PRR1, around 22% to C-1, Hgal1, GA and MS2 and drops to 17% when compared to allolevi-viruses SP and Qβ. The coat proteins are more conserved and here M groups clearly with phages PRR1, C-1 and Hgal1 with amino acid identities of 48-51%. The identity with F-specific phages is significantly lower and ranges from 27.1% for group II levivirus KU1 to 19% for group IV allolevivirus NL95. Notably, M coat protein shares 24.6% amino acids with that of *Pseudomonas* phage PP7, which is the only plasmid-independent phage for which the sequences could be reasonably aligned. For replicase, the trend is similar as for the maturation protein: the replicase of phage M most resembles that of PRR1 with 41% amino acid identity, followed by other plasmid-dependent phages C-1, Hgal1, MS2 and GA (33-37% identity) and alloleviviruses (27-29% identity). Again, M replicase turns out to be more closely related to that of phage PP7 (25.5% identity) than to the other plasmid-independent phages AP205 and $\varphi$Cb5 (17.7 % identity).

### Conserved RNA secondary structures

With the growing number of *Leviviridae* genomes that have been sequenced it has become clear that besides encoding proteins, the secondary and tertiary structure of the RNA itself is also very important. The complex structure of RNA provides binding sites for phage proteins [36-38], regulates their translation [1] and promotes genome packaging in capsids [39]. In many cases where nucleotide stretches from different phage genomes show no sequence similarity, the secondary

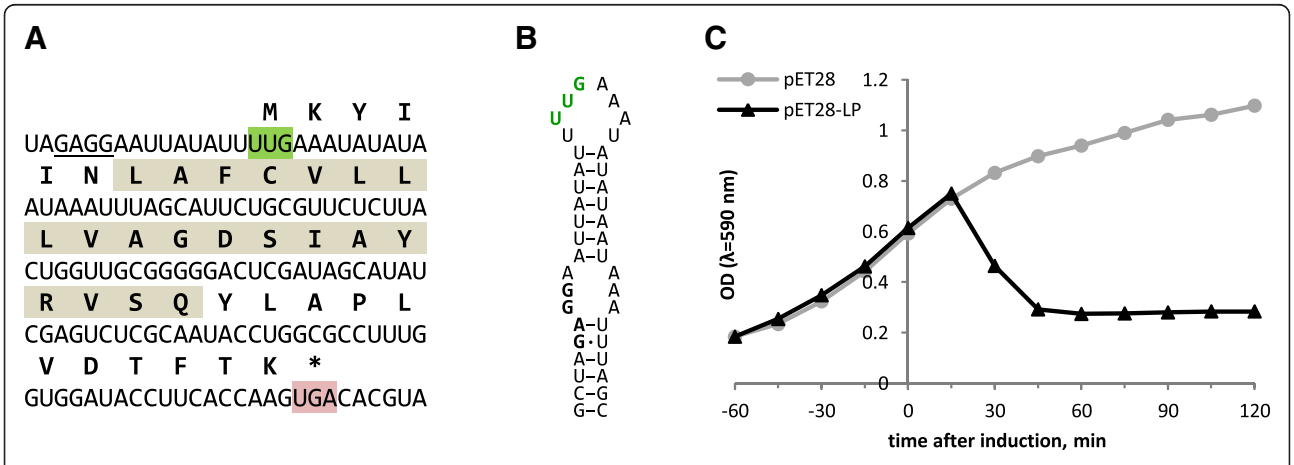



**Figure 2 Lysis protein of phage M.** (**A**) The lysis gene. The Shine-Dalgarno sequence is underlined and initiation and termination codons are indicated by green and pink shading, respectively. The translated amino acid sequence is given above the RNA sequence and the putative transmembrane helix is shaded gray. (**B**) An RNA hairpin around the initiation codon of the lysis gene. The initiation codon and the Shine-Dalgarno sequence are indicated. (**C**) Verification of the lysis gene. Growth of *E.coli* cells harboring either empty vector (pET28) or a plasmid with the cloned lysis gene (pET28-LP) before and after the induction of protein synthesis is shown.

structures they fold into are nevertheless well preserved. One such example lies at the very 5′ end of all of the sequenced ssRNA phage genomes, where there is a stable GC-rich hairpin that has been suggested to play an important role in phage RNA replication [40]. Phage M is no exception (Figure 3A). Another important RNA structure lies around the initiation codon of replicase. This approximately 20-nucleotide-long stretch folds into a hairpin structure that specifically binds the phage coat protein. This interaction acts as a translational operator to repress synthesis of replicase when enough coat protein accumulates [37] and has been suggested to play also a role in initiating specific encapsidation of the genomic RNA [41]. When the operator hairpin of phage M is compared to those of other ssRNA phages, it is evident that it groups with the conjugative pili-dependent phages PRR1, C-1, Hgal1 and MS2 (Figure 3B). An adenine residue in the loop four nucleotides upstream of the replicase initiation codon and an unpaired purine residue in the stem which are critical for RNA-protein binding in phages MS2 [42], GA [43] and PRR1 [44] are preserved also in phage M, therefore the mechanism of interaction is probably similar.

It is also interesting to take a look at the 3′ untranslated region of the phage genome. The configurations of 3′ UTRs vary between different phages, but nevertheless some similarities exist. In all known *Leviviridae* phages a long-distance interaction designated ld IX bridges the very 3′ terminus with a complementary nucleotide stretch upstream, forming the 3′ terminal domain [45]. The domain usually consists of at least three hairpins, denoted U1, U2 and V. In phage M, the 100-nucleotide-long 3′ UTR is made up from four hairpins U4, U3, U2 and U1 (Figure 3C). In all ssRNA phages the 3′-terminal helix U1 has a remarkably conserved nucleotide sequence in the loop: UGCUU in phages as diverse as MS2, SP and AP205, UGCUG in φCb5 and CGCUC in PP7. In the case of Qβ, this loop forms a long-distance pseudoknot with a complementary sequence approximately 1200 nucleotides upstream that is essential for phage replication [47]. In phage M, the sequence of the U1 loop is AUUGCUAUG. It has not been experimentally verified that phages other than Qβ have the pseudoknot, but in M genome a sequence AGCAA is found in the replicase gene some 1215 nucleotides upstream that could potentially basepair with UUGCU in the loop. The other notable feature of the 3′ domains, although less pronounced, is hairpin V (designated V2 in some phages) which in phages MS2, Qβ, SP and AP205 contains a large, adenine-rich loop. There is some evidence that in MS2 this might be one of the sites where the maturation protein binds to the RNA [36]. In phage φCb5, however, the candidate hairpin V lacks analogous features and in phages PRR1, C-1 and Hgal1 it does not seem to exist at all; instead, there is a stretch of unpaired nucleotides (UAUAAACA in PRR1, UAUA in Hgal1 and

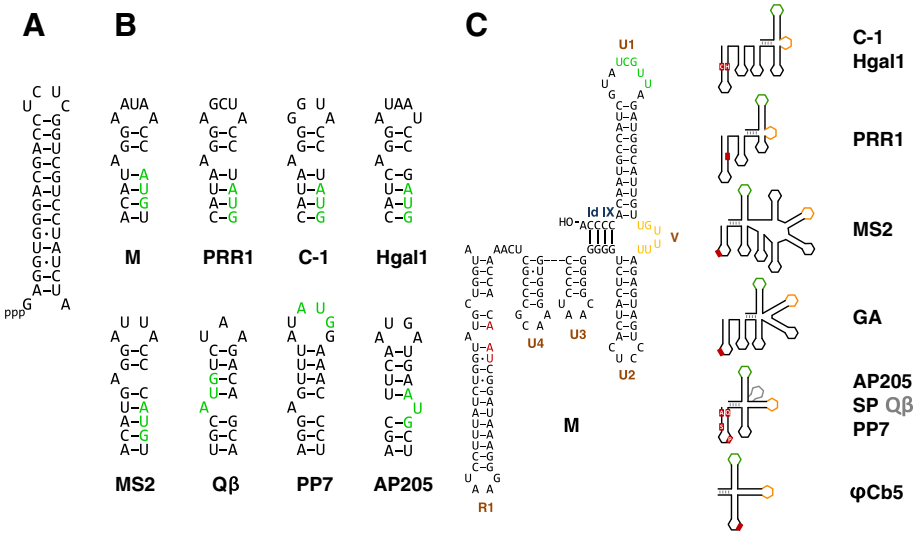



**Figure 3 RNA secondary structures in M genome.** (**A**) A stable hairpin at the very 5′ end of the genome important for phage RNA replication. (**B**) The operator hairpin around the initiation codon of replicase. The analogous hairpins from other *Leviviridae* phages are shown for comparison. Start codons of the replicase gene are colored green. (**C**) Structure of the 3′ untranslated region. The termination codon of replicase is colored dark red, the unpaired stretch corresponding to loop V or V2 in other phages in orange and the conserved nucleotide sequence in the loop of hairpin U1 that potentially forms a long-distance pseudoknot in green. On the right, schematic representations of 3′ UTRs from other phages based either on published data [31,32,45,46] or RNA secondary structure predictions are given for comparison. The 3′ UTR of phage Qβ is closely similar to that of phage SP except for a short extra helix which is depicted in gray. The locations of replicase gene termination codons are represented as red boxes. RNA secondary structures were predicted by the RNAfold server [34].

UUAAU in C-1) that connects hairpins U2 and U1 and might serve the same function as hairpin V in other phages. In phage M the situation is similar, but the loop sequence is UUUUGU and contains no adenine residues. When the overall structures of 3′ UTRs from different phages are compared (Figure 3C, right), it is evident that in the distantly related phages φCb5, AP205, PP7 and SP the 3′ domain is remarkably simple with just three hairpins, while it is considerably expanded in the plasmid-specific leviviruses, culminating in seven hairpins in phage MS2. In this respect, phages M, C-1, Hgal1 and PRR1 form their own group where the 3′ UTR adopts a characteristic fold of only two hairpins between the ld IX, a stretch of unpaired nucleotides instead of hairpin V and one or two hairpins between the terminal replicase hairpin R1 and ld IX.

### Evolutionary considerations

In many aspects, phage M is a typical representative of the *Leviviridae* family that is clearly related to other conjugative pili-dependent RNA phages. The feature that makes it unique though is the unusual location of its lysis gene. Although there are precedents of this in the distantly related phages AP205 and φCb5, it is a bit surprising to find such phenomenon also within a group of otherwise rather closely related phages. Apparently, it is relatively easy for a short ORF encoding a transmembrane helix that causes cell lysis to appear by random mutations, as several phages have arrived at the same mechanism independently. It would also suggest that the location of the lysis gene at this position is probably limited to the IncM plasmid-specific leviviruses or even to a smaller subgroup of these phages. Since M is the only IncM plasmid-specific RNA phage that has been isolated, it is not possible to address this question presently.

The high mutation rates and resulting sequence variability in RNA viruses makes reconstruction of their evolutionary history not a trivial task. Based on similarities between maturation and replicase proteins, phage M seems more related to phage PRR1, while coat protein sequences and structures of the 3′ UTRs suggest that it might be closer to phages C-1 and Hgal1. To further address this question we conducted a phylogenetic analysis of 15 representative *Leviviridae* phages using both the complete genome sequences and also the replicase protein sequences since the RNA-dependent RNA polymerases are the most conserved proteins of all positive-sense RNA viruses [48]. Both trees (Figure 4) confirm that phage M is more closely related to the IncC, IncH and IncP than to the IncF plasmid-dependent phages but they show differences in the clustering of the non-F plasmid specific phages. Although phylogenetic analysis of the coat proteins (not shown) gives the same (M(C-1 (Hgal1,PRR1))) clustering as the replicase, low bootstrap values for the IncC, IncH and IncP branches indicate that confidence in that particular branching order is not high and suggest that phages C-1, Hgal1 and PRR1 have radially diverged from a similar ancestral sequence. In both trees phage M represents a lineage that branched off early in the course of specialization on different plasmids after the separation of the IncF lineage had occurred but before the diversification on IncC, IncH and IncP plasmids took place. Both trees also support the idea that the allolevivirus lineage separated from the leviviruses before the specialization on different conjugative pili had occurred and that these phages arrived at the ability to bind to F-pili via an independent evolutionary path.

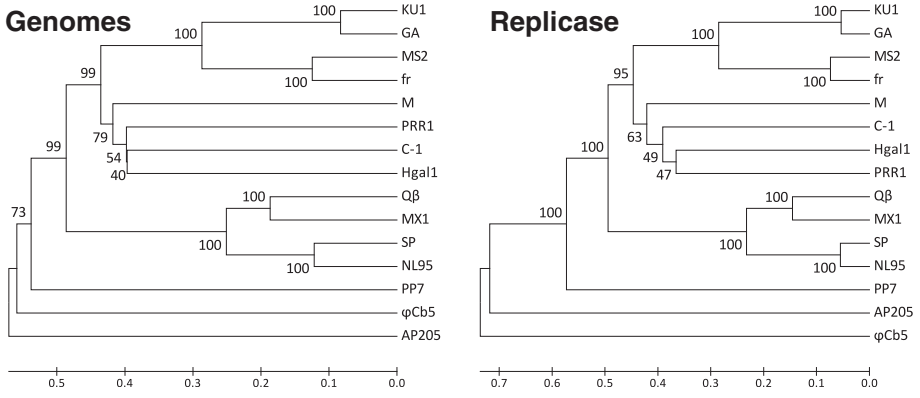



**Figure 4 Phylogeny of RNA phages.** The phylogenetic analysis was based on the complete genomic RNA sequences (left) and amino acid sequences of the replicase (right) which is the most conserved of all ssRNA phage proteins. Trees were constructed by unweighted pair group method with arithmetic mean (UPGMA) and tested using the bootstrap method with 500 replicates. The bootstrap values are expressed as percentages next to the nodes. RNA and protein sequences were aligned using MUSCLE [49] and the phylogenetic trees were constructed in program MEGA5 [50].

Although all *Leviviridae* phages use pili for attachment, there is a marked difference between the types of pili they utilize. The type IV pili used by phages AP205, φCb5 and PP7 are produced via a genome-encoded type II secretion pathway [51], whereas the plasmid-borne conjugative pili that the other phages utilize belong to a type IV secretion system [52]. Both systems share some functional similarities, like a retractable pilus and a membrane pore, but are thought to have evolved independently [53]. Therefore a jump from one to the other type of pili had to occur at some point in the *Leviviridae* history. Our phylogenetic analysis suggests that the ancestral phage infected cells via type IV pili, like phages AP205, φCb5 and PP7 are doing today and a PP7-like virus then might have evolved the ability to bind to some kind of conjugative pili and still sustain infectivity. Consequently, all of the specialized plasmid-dependent RNA phages we know today would be descendants of this ancestral virus.

## Conclusions

We have determined and characterized the genome sequence of IncM plasmid-dependent phage M and shown that it resembles the plasmid-specific leviviruses in many ways but has an atypical location of the lysis gene. It is a valuable addition to the growing number of sequenced *Leviviridae* genomes and provides a better view on the diversity and evolution within this phage family.

## Methods

### Phage propagation and purification

Bacteriophage M and its host *E.coli* J53(RIP69) were obtained from Félix d'Hérelle Reference Center for bacterial viruses, Laval University, Quebec, Canada (catalog numbers HER218 and HER1218, respectively). J53 (RIP69) cells were grown in LB medium containing 6 μg/ml tetracycline overnight at 37 °C without agitation. To propagate the phage, 0.5 ml of the host cell suspension and 10 μl of phage lysate (approximately $10^{10}$ pfu/ml) were spotted on 1.5% LB agar plates, overlaid with 15-20 ml of molten 0.7% LB agar cooled to 45 °C, mixed by swirling and incubated overnight at 30 °C. The next morning, top agar layers from several plates were scraped off, transferred to centrifuge tubes and centrifuged for 20 minutes at 18500 g. Supernatant was collected and phage particles were precipitated by addition of sodium chloride and PEG 6000 to concentrations of 0.5M and 10%, respectively, and incubation for 30 minutes at 4 °C. After centrifugation for 10 minutes at 18500 g, the supernatant was discarded and the pellet was resuspended in a small volume of distilled water. The phage preparation was then layered on top of a preformed five-step cesium chloride gradient (equal volumes of CsCl solutions in 20 mM Tris-HCl pH 7.5 with densities of 1.7, 1.6, 1.5, 1.4 and 1.3 g/ml) and centrifuged for 17 hours in a SW 40Ti rotor at 24000 rpm. 0.5 ml fractions were collected from the top of the gradient and the peak fractions containing phage were pooled and dialyzed against one liter of 20 mM Tris-HCl pH 7.5 overnight at 4 °C. The preparation was concentrated to 500 μl using Amicon Ultra 10K MW cutoff spin unit (Millipore) and used for RNA extraction.

### Isolation of genomic RNA and sequencing

500 μl of purified phage preparation was mixed with 500 μl of phenol and SDS was added to a final concentration of 0.5%. The mixture was vigorously vortexed for 60 s and centrifuged at 12000 g for 3 minutes. The aqueous phase was extracted two more times with a 1:1 phenol/chloroform mixture and once with chloroform. The RNA in the final aqueous phase was precipitated with ethanol, centrifuged and the pellet redissolved in a small volume of DEPC-treated water.

4 μg of the purified RNA was reverse-transcribed with RevertAid Premium reverse transcriptase (Fermentas) using primer 5′-GCAAATTCTGTTTTATCAGACNNNNNN-3′. Reaction products were purified using GeneJet PCR purification kit (Fermentas) and eluted in 20 μl of water. The 3′ termini of the purified first strand cDNAs were dATP-tailed using terminal deoxynucleotidyl transferase (Fermentas). The reaction products were again purified using the PCR purification kit and used as a template for second-strand PCR with primers 5′-GCAAATTCTGTTTTATCAGAC-3′ and 5′-GCGCG$(T)_{18}$-3′ and Pfu DNA polymerase (Fermentas). Reaction products were separated in a 1% agarose gel and a slice corresponding to 1000 – 3000 base pair DNA fragments was cut out. The fragments were extracted using GeneJet gel extraction kit (Fermentas) and ligated in pJET1.2/blunt vector (Fermentas).

Insert-containing clones were sequenced on an ABI Prism 3100 Genetic Analyzer using BigDye Terminator v3.1 kit (Applied Biosystems). Based on the obtained sequence data, additional reverse transcription-PCRs were performed using specific primers to fill gaps and increase coverage. Since the initial cloning procedure already involved 3′-tailing of cDNAs, it was possible to determine the 5′ end of the genome from these clones. To determine the sequence of the 3′ end, phage RNA was tailed with *E.coli* Poly(A) polymerase (Ambion), followed by reverse transcription with primer 5′-GCGCG$(T)_{18}$-3′ and PCR using primers 5′-GCGCG$(T)_{18}$-3′ and 5′-CTGGCGCCTTTGGTGGATAC-3′ corresponding to nucleotides 3072-3091 of the phage genome. Genome assembly and ORF prediction was done with the program ContigExpress from the VectorNTI Suite (Invitrogen).

The genome sequence was deposited in GenBank with accession code JX625144.

## Cloning and expression of the lysis gene

The putative lysis gene was PCR-amplified from a suitable cDNA clone using primers 5′-ATATT-CTAGACGAAGGAACAACCATTGCCG-3′ and 5′-TATGAAGCTTACTTGGTGAAGGTATCCACC-3′, the fragment was digested with XbaI and HindIII and ligated into XbaI-HindIII-digested pET28a vector (Novagen), yielding plasmid pET28-LP. To test for the lytic function of the protein, pET28-LP-containing *E.coli* BL21 AI cells (Invitrogen) were grown in LB medium supplemented with 30 μg/ml kanamycin and protein production was induced by adding arabinose to a final concentration of 0.2% and IPTG to a final concentration of 1 mM.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

JR propagated and purified the phage, sequenced the genome, cloned the lysis gene, analyzed the genome and wrote the paper. KT supervised the work, analyzed the genome sequence and wrote the paper. Both authors read and approved the final manuscript.

### Acknowledgements

This work was supported by grant 09.1294 from the Latvian Council of Science and grant 2DP/2.1.1.1.0/10/APIA/VIAA/052 from the European Regional development fund (ERDF). The publishing costs were covered by ERDF grant 2DP/2.1.1.2.0/10/APIA/VIAA/004.

Received: 13 September 2012 Accepted: 20 November 2012
Published: 24 November 2012

### References

1. Van Duin J, Tsareva N: **Single-stranded RNA phages**. In *The Bacteriophages*. 2nd edition. Edited by Calendar RL.: Oxford University Press; 2006:175–196.
2. Blumenthal T, Landers TA, Weber K: **Bacteriophage Qβ replicase contains the protein biosynthesis elongation factors EF Tu and EF Ts.** *Proc Natl Acad Sci USA* 1972, **69**:1313–1317.
3. Wahba AJ, Miller MJ, Niveleau A, Landers TA, Carmichael GG, Weber K, Hawley DA, Slobin LI: **Subunit I of Qβ replicase and 30 S ribosomal protein S1 of *Escherichia coli* Evidence for the identity of the two proteins.** *J Biol Chem* 1974, **249**:3314–3316.
4. Valegård K, Liljas L, Fridborg K, Unge T: **The three-dimensional structure of the bacterial virus MS2.** *Nature* 1990, **345**:36–41.
5. Kozak M, Nathans D: **Fate of maturation protein during infection by coliphage MS2.** *Nat New Biol* 1971, **234**:209–211.
6. Shiba T, Miyake T: **New type of infectious complex of *E.coli* RNA phage.** *Nature* 1975, **254**:157–158.
7. Weiner AM, Weber K: **Natural read-through at the UGA termination signal of Qβ coat protein cistron.** *Nat New Biol* 1971, **234**:206–209.
8. Winter RB, Gold L: **Overproduction of bacteriophage Qβ maturation (A2) protein leads to cell lysis.** *Cell* 1983, **33**:877–885.
9. Karnik S, Billeter M: **The lysis function of RNA bacteriophage Qβ is mediated by the maturation (A2) protein.** *EMBO J* 1983, **2**:1521–1526.
10. Model P, Webster RE, Zinder ND: **Characterization of Op3, a lysis-defective mutant of bacteriophage f2.** *Cell* 1979, **18**:235–246.
11. Atkins JF, Steitz JA, Anderson CW, Model P: **Binding of mammalian ribosomes to MS2 phage RNA reveals an overlapping gene encoding a lysis function.** *Cell* 1979, **18**:247–256.
12. Beremand MN, Blumenthal T: **Overlapping genes in RNA phage: a new protein implicated in lysis.** *Cell* 1979, **18**:257–266.
13. Loeb T, Zinder ND: **A bacteriophage containing RNA.** *Proc Natl Acad Sci USA* 1961, **47**:282–289.
14. Davis JE, Strauss JH, Sinsheimer RL: **Bacteriophage MS2: another RNA phage.** *Science* 1961, **134**:1427.
15. Taylor DE, Gibreel A, Lawley TD, Tracz DM: **Antibiotic resistance plasmids**. In *Plasmid biology*. Edited by Funnell BE, Philips GJ. Washington, D.C: ASM Press; 2004:473–491.
16. Olsen RH, Thomas DD: **Characteristics and purification of PRR1, an RNA phage specific for the broad host range *Pseudomonas* R1822 drug resistance plasmid.** *J Virol* 1973, **12**:1560–1567.
17. Sirgel FA, Coetzee JN, Hedges RW, Lecatsas G: **Phage C-1: an IncC group; plasmid-specific phage.** *J Gen Microbiol* 1981, **122**:155–160.
18. Coetzee JN, Bradley DE, Lecatsas G, du Toit L, Hedges RW: **Bacteriophage D: an IncD group plasmid-specific phage.** *J Gen Microbiol* 1985, **131**:3375–3383.
19. Coetzee JN, Bradley DE, Fleming J, du Toit L, Hughes VM, Hedges RW: **Phage pilHα: a phage which adsorbs to IncHI and IncHII plasmid-coded pili.** *J Gen Microbiol* 1985, **131**:1115–1121.
20. Nuttall D, Maker D, Colleran E: **A method for the direct isolation of IncH plasmid-dependent bacteriophages.** *Lett Appl Microbiol* 1987, **5**:37–40.
21. Coetzee JN, Bradley DE, Hedges RW: **Phages Iα and I2-2: IncI plasmid-dependent bacteriophages.** *J Gen Microbiol* 1982, **128**:2797–2804.
22. Coetzee JN, Bradley DE, Hedges RW, Fleming J, Lecatsas G: **Bacteriophage M: an incompatibility group M plasmid-specific phage.** *J Gen Microbiol* 1983, **129**:2271–2276.
23. Bradley DE, Coetzee JN, Bothma T, Hedges RW: **Phage t: a group T plasmid-dependent bacteriophage.** *J Gen Microbiol* 1981, **126**:397–403.
24. Ruokoranta TM, Grahn AM, Ravantti JJ, Poranen MM, Bamford DH: **Complete genome sequence of the broad host range single-stranded RNA phage PRR1 places it in the Levivirus genus with characteristics shared with Allolevivirus.** *J Virol* 2006, **80**:9326–9330.
25. Kannoly S, Shao Y, Wang IN: **Rethinking the evolution of single-stranded RNA (ssRNA) bacteriophages based on genomic sequences and characterizations of two R-plasmid-dependent ssRNA phages, C-1 and Hgal1.** *J Bacteriol* 2012, **194**:5073–5079.
26. Persson M, Tars K, Liljas L: **The capsid of the small RNA phage PRR1 is stabilized by metal ions.** *J Mol Biol* 2008, **383**:914–922.
27. Bradley DE, Taylor DE, Cohen DR: **Specification of surface mating systems among conjugative drug resistance plasmids in *Escherichia coli* K-12.** *J Bacteriol* 1980, **143**:1466–1470.
28. Inokuchi Y, Takahashi R, Hirose T, Inayama S, Jacobson AB, Hirashima A: **The complete nucleotide sequence of the group II RNA coliphage GA.** *J Biochem (Tokyo)* 1986, **4**:1169–1980.
29. Young R: **Bacteriophage lysis: mechanism and regulation.** *Microbiol Rev* 1992, **56**:430–481.
30. Goessens WH, Driessen AJ, Wilschut J, van Duin J: **A synthetic peptide corresponding to the C-terminal 25 residues of phage MS2 coded lysis protein dissipates the protonmotive force in *Escherichia coli* membrane vesicles by generating hydrophilic pores.** *EMBO J* 1988, **7**:867–873.
31. Klovins J, Overbeek GP, van den Worm SH, Ackermann HW, van Duin J: **Nucleotide sequence of a ssRNA phage from *Acinetobacter*: kinship to coliphages.** *J Gen Virol* 2002, **83**:1523–1533.
32. Kazaks A, Voronkova T, Rumnieks J, Dishlers A, Tars K: **Genome structure of *Caulobacter* phage phiCb5.** *J Virol* 2011, **85**:4628–4631.
33. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J Mol Biol* 2001, **305**:567–580.
34. Hofacker IL: **Vienna RNA secondary structure server.** *Nucl Acids Res* 2003, **31**:3429–3431.
35. de Smit MH, van Duin J: **Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis.** *Proc Natl Acad Sci USA* 1990, **87**:7668–7672.
36. Shiba T, Suzuki Y: **Localization of A protein in the RNA-A protein complex of RNA phage MS2.** *Biochim Biophys Acta* 1981, **654**:249–255.
37. Bernardi A, Spahr PF: **Nucleotide sequence at the binding site for coat protein on RNA of bacteriophage R17.** *Proc Natl Acad Sci USA* 1972, **69**:3033–3037.
38. Meyer F, Weber H, Weissmann C: **Interactions of Qβ replicase with Qβ RNA.** *J Mol Biol* 1981, **153**:631–660.
39. Basnak G, Morton VL, Rolfsson O, Stonehouse NJ, Ashcroft AE, Stockley PG: **Viral genomic single-stranded RNA directs the pathway toward a T=3 capsid.** *J Mol Biol* 2010, **395**:924–936.

40. Beekwilder J, Nieuwenhuizen R, Poot R, van Duin J: **Secondary structure model for the first three domains of Qβ RNA. Control of A-protein synthesis.** *J Mol Biol* 1996, **256**:8–19.
41. Beckett D, Wu HN, Uhlenbeck OC: **Roles of operator and nonoperator RNA sequences in bacteriophage R17 capsid assembly.** *J Mol Biol* 1988, **204**:939–947.
42. Carey J, Lowary P, Uhlenbeck OC: **Interaction of R17 coat protein with synthetic variants of its ribonucleic acid binding site.** *Biochemistry* 1983, **22**:4723–4730.
43. Gott JM, Wilhelm LJ, Uhlenbeck OC: **RNA binding properties of the coat protein from bacteriophage GA.** *Nucl Acids Res.* 1991, **19**:6499–6503.
44. Persson M, Tars K, Liljas L: **PRR1 coat protein binding to its RNA translational operator.** *Acta Crystallogr D Biol Crystallogr*, . in press.
45. Beekwilder MJ, Nieuwenhuizen R, van Duin J: **Secondary structure model for the last two domains of single-stranded RNA phage Qβ.** *J Mol Biol* 1995, **247**:903–917.
46. Olsthoorn RC, Garde G, Dayhuff T, Atkins JF, Van Duin J: **Nucleotide sequence of a single-stranded RNA phage from *Pseudomonas aeruginosa*: kinship to coliphages and conservation of regulatory RNA structures.** *Virology* 1995, **206**:611–625.
47. Klovins J, van Duin J: **A long-range pseudoknot in Qβ RNA is essential for replication.** *J Mol Biol* 1999, **294**:875–884.
48. Koonin EV, Dolja VV: **Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences.** *Crit Rev Biochem Mol Biol* 1993, **28**:375–430.
49. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucl Acids Res* 2004, **32**:1792–1797.
50. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28**:2731–2739.
51. Peabody CR, Chung YJ, Yen MR, Vidal-Ingigliardi D, Pugsley AP, Saier MH Jr: **Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella.** *Microbiology* 2003, **149**:3051–3072.
52. Lawley TD, Klimke WA, Gubbins MJ, Frost LS: **F factor conjugation is a true type IV secretion system.** *FEMS Microbiol Lett* 2003, **224**:1–15.
53. Hazes B, Frost L: **Towards a systems biology approach to study type II/IV secretion systems.** *Biochim Biophys Acta* 2008, **1778**:1839–1850.

doi:10.1186/1471-2180-12-277
**Cite this article as:** Rumnieks and Tars: **Diversity of pili-specific bacteriophages: genome sequence of IncM plasmid-dependent RNA phage M.** *BMC Microbiology* 2012 **12**:277.

# 'I had to help my child!': : The role of emotions, risk, and trust in use of nasal decongestants in children

Ieva Salmane-Kulikovska and Signe Mezinska

*J Child Health Care* 2013 17: 41 originally published online 30 December 2012
DOI: 10.1177/1367493512456108

*Article*

JCHC

Journal of Child Health Care
17(1) 41–52

DOI: 10.1177/1367493512456108
chc.sagepub.com
SAGE

# 'I had to help my child!': The role of emotions, risk, and trust in use of nasal decongestants in children

**Ieva Salmane-Kulikovska**
Department of Sociology and Psychology, Riga Stradins University, Latvia

**Signe Mezinska**
University of Latvia, Latvia

## Abstract

Recent studies suggest that a number of common cold medicines, including nasal decongestants are not suitable for small children. In Latvia nasal decongestants are primarily over-the-counter (OTC) medicines, and patient information leaflets generally permit use of these medicines for small children. The previous studies in Latvia investigating the use of medicines in cases of common cold demonstrated extensive use of decongestants in children less than six years of age, marking a necessity for more profound research. As a result a qualitative study was implemented in 2011 to identify sources of information that offer advice for the use of decongestants in children. The caregivers' perception of risks, trust, and the role of emotions associated with use of these medicines was considered. The study results suggested that emotions have a considerable effect on the decisions of caregivers. Even though caregivers are aware of the risks associated with the use of decongestants, they still frequently administer these medicines for the sake of the comfort of both the child and the caregiver. It is essential to raise a discussion about the inappropriate use of common cold medications in small children in Latvia.

## Keywords

Carers, child health, decision making, medication

## Introduction

A number of studies have suggested that several common cold medicines, including nasal decongestants, are not suitable for children (Fabi et al., 2009; KinyonMunch, 2011; Rimsza and

**Corresponding author:**
Ieva Salmane-Kulikovska, Department of Sociology and Psychology, Riga Stradins University, Dzirciema 16, Riga, LV 1007, Latvia.
Email: ieva.salmanekulikovska@gmail.com

Newberry, 2008; Ryan et al., 2008; Shefrin and Goldman, 2009). These medicines, which are sometimes referred to as remedies, contain no treatment properties, only relieve symptoms (Johnson and Helman, 2004), and are no better at treating the illness than a placebo (Smith et al., 2008). At the same time, they may cause serious side-effects for children, who experience the common cold on average six to eight times a year (Heikkinen and Järvinen, 2003). The common cold is considered to belong to the non-heavy self-limiting disease group; it is caused by numerous viruses and there are no medicines available to properly treat the common cold (Heikkinen and Järvinen, 2003). Experts suggest using alternative symptom-relieving means instead, such as fluids, humidifiers and saline water (Ryan et al., 2008).

In Latvia, the patient information leaflets (PILs) included with most of the nasal decongestants do not contain instructions to abstain from the usage of these medicines in cases of children of a certain age. A previous quantitative study carried out in Latvia showed that nasal decongestants are frequently used for relieving symptoms of rhinitis in children under six years of age (52.2% of the illness cases) (Salmane-Kulikovska et al., 2010).

The use of medicines involves many social, cultural, and psychological aspects which influence the decision to use these medicines (Haaijer-Ruskamp and Hemminki, 1993). Emotions are present in the decision-making process and are attributed to one of the stages of information processing: dealing with trial, evaluation, adoption, or rejection. These processes are influenced by emotional or life events, linked to relatives, such as children or other family members (Chetley et al., 2007). The amount of information users of medicines face nowadays is immense. The evaluation of risk and the establishment of trust are made on the basis of expert opinions, lay people, and public discourse.

The present study was aimed at clarifying the role of emotions, trust, and risk evaluation on the decision that caregivers make to use nasal decongestants in children under six years of age, based upon the information available to users of medicines.

## Research methods and sample

A qualitative study was conducted using 27 semi-structured interviews with caregivers of children ages 0–6. The significance of the individual experience and a need for an in-depth investigation of the topic led to the qualitative methodology, based upon explanatory-informative cases and purposive sampling utilizing the snowball sampling method. Interviews were performed in March 2011. The main sampling parameter was 'having used nasal decongestants in a child (children) age 0–6 for relieving symptoms of rhinitis'. Interviews were semi-structured and problem-centred (Flick, 1998), with an average length of 20–30 minutes. Interviews were recorded and each respondent signed an informed consent form prior to the interview.

There were three groups of interview questions. The first group was aimed at clarifying the health status of a child – how often the child experiences the common cold, what are the usual symptoms, does he/she have any chronic illnesses. The second group of questions focused on rhinitis – responses and activities of caregivers, choice and use of decongestants, sources suggesting the use of these medicines, care-rendering, administration of different medicines, attitude towards rhinitis, perceived properties, and risks associated with nasal decongestants. The third part of the interview examined the evaluation of sources of information about common cold medicines in Latvia, including nasal decongestants, as well as trust in these sources.

All of the respondents were females, and almost all were mothers (one respondent was a grandmother). Such sample characteristics are determined by the fact that, in Latvia, it is mostly

**Table 1.** Characteristics of the sample interviewed.

| | | Age of the child | | | | | |
|---|---|---|---|---|---|---|---|
| Respondent's ID | Age of respondent | 0–<1 | 1–<2 | 2–<3 | 3–<4 | 4–<5 | 5-6 |
| O_25 | 25 | | | x | | | |
| I_28 | 28 | | x | | | | |
| L_28 | 28 | x | | | | x | |
| N_28 | 28 | | | | | | x |
| S_30 | 30 | | x | | | | x |
| I_30 | 30 | | x | | | | |
| J_31 | 31 | | | x | | x | |
| D_31 | 31 | | x | | | | |
| E_31 | 31 | | | | | x | |
| L_31 | 31 | | | x | | | |
| J_32 | 32 | | | | | x | |
| V_32 | 32 | x | | | x | | |
| A_32 | 32 | | | x | | | x |
| I_32 | 32 | | | | x | | x |
| B_32 | 32 | | x | | | | |
| R_32 | 32 | | | x | | | |
| R_33 | 33 | | | | | x | x |
| J_33 | 33 | | x | | | | |
| I_34 | 34 | | x | | | x | x |
| E_34 | 34 | | | x | | | x |
| S_35 | 35 | | | | | x | |
| I_37 | 37 | | | | | | x |
| D_37 | 37 | | | | | x | |
| DZ_37 | 37 | | | | | | x |
| L_39 | 39 | | | x | | | |
| L_42 | 42 | | | | x | | |
| V_57 | 57 | | | | x | | |

female family members that care for sick children and make decisions about which medicines to use in the case of illness. The age of respondents was 25–57 years of age (see Table 1). The total number of children being referred to within the limits of this study was 30. The sample size was determined by data saturation. Analysis, entailing the coding of transcripts, was performed by the program Nvivo 9.0.

## Results

The analysis was based upon a thematic coding procedure and a multi-step process (Flick, 1998: 188–190). The transcripts were analysed first by providing a short description of each respondent, enclosing typical statements provided by the respondents related to the interview questions, which emerged from the aim, task, and content of the study. The statements were coded, and three categories were derived out of as a result of the grouping of codes – emotions, perceptions of risks and benefits and trust in the available information sources (see Table 2). Transcripts were checked for the presence or absence of categories (Flick, 1998).

**Table 2.** The categories and codes identified.

| Categories | Codes |
|---|---|
| Emotions of the caregivers | Empathy<br>Emotional responsibility<br>Care<br>Comfort of the carer and the child |
| Perception of risk | Perceived risks<br>• Vulnerability of a child (children)<br>• side-effects of decongestants<br>• complications from rhinitis<br>Criteria of diminished risk<br>• previous positive experience<br>• extensive information<br>• short use<br>• particular brands of medications |
| Trust in information sources | Criteria of trust<br>• successful cooperation<br>• personal attitude<br>• comprehensibility<br>• conciseness<br>• unbiasedness |

## *The role of emotions and the effect of social medication*

Emotions play an important role in the decision-making process for caregivers taking care of a sick child. This includes how they go about choosing a medicine to treat a child. During the interviews, the caregivers stressed that they felt the necessity to maintain the well-being of the child. Empathy and the perceived need to provide comfort for the child, who is experiencing symptoms of rhinitis, seemed to outweigh the perceived possible risks associated with the use of decongestants.

> It is, of course, possible to make philosophical discussions [about side-effects] (. . . ) but it is more important for me that the child feels better. (E_31)
>
> And he [the child] is in fact tired from feeling sick. (R_32)

Respondents pointed out that the perceived positive effects on the behaviour of the child were proof that they made the right choice of medication.

> I observed that, after I had given her nasal drops, she became livelier after 10–15 minutes. She was more relaxed, ready to play, full of energy. (L_39)

The comfort of the child was observed to be parallel to the comfort of the mother and the well-being of the family.

> I understand that a runny-nose – it is a certain reaction to something (. . . ) At the same time, she feels more comfortable [when the decongestant is used] and so do I. (E_31)

> I do not like that someone from my family is ill. I am not a sickly person myself, and I feel uncomfortable, indeed, if somebody is ill, and I always desperately do my best to stop it. (B_32)

These statements made by respondents show that nasal decongestants are often used to demonstrate care, to reduce the amount of inconvenience for caregivers, and to answer their emotional and moral need to improve the child's experience of illness.

## *Perceptions of risks*

The results of the study show that the risk of using nasal decongestants for children is perceived to be greater when compared to similar cases concerning adults. In these cases, information about medicine is gathered and examined more carefully.

> Regarding our own health, oh, yes, we can gather information, read something on the internet, chose something, but the health of the children ( . . . ) is different. (I_28)
>
> I do not conduct experiments ( . . . ) I may permit myself to conduct experiments on myself, but on the child – never! Concerning medicine for a child – I read everything from A to Z. (I_30)

One of the reasons for caregivers to consult several sources before making a decision is perceived vulnerability of a child.

> I try to read something. I go to the pharmacy and ask then I ask the doctor. And then I make the decision myself ( . . . ) like maybe a normal parent. (B_32)

Many of the respondents were aware of the risks potentially arising from the use of nasal decongestants. The main risk mentioned by the respondents was possible damage to the mucous membrane.

> It has a very bad influence on the mucous membrane. Intensive use may cause a habit, and the membrane dries out. (R_32)
>
> They [nasal decongestants] burn the nasal membrane. (J_32)

Other mentioned side-effects included dependence and reduced efficiency in cases of prolonged use.

> If we use those [decongestants] frequently, our body loses capacity to fight the disease itself. (D_31)

The perceived benefits of nasal decongestants included the capacity of these medicines to facilitate symptoms of illness.

> The positive influence is a possibility of sleep and no need to suffer from a stuffy nose. (DZ_37)

Some respondents stated that they were confident that they had observed the treatment properties of decongestants.

> I think they [nasal decongestants] facilitate breathing and also treat rhinitis. They seem to shorten [the illness] also. (A_32)

Use of decongestants was also justified by the assurance that there is a considerably low risk of side effects if the time of usage is not prolonged.

> If it is used without overdoing it, only when it is really needed, OK, maybe regularly, but for a short period of time, no big harm can occur. (I_30)

It was stated by some respondents that use of decongestants helps to minimize the risk of complications that could be caused by rhinitis.

> In order to not start with antibiotics and to not have an effect on their ears, our doctor suggested using Vibrocil three to four times a day. (I_32)

In evaluating the risk associated with the use of particular medicines, the respondents were basing their criteria of choice and use of decongestants on erroneous arguments.

> I think this medicine [Vibrocil] is the best to reduce dropsy. Or, maybe I have read somewhere that it has a less negative impact on the membrane ( . . . ) I think it somehow disinfects the nose. (J_33)

A tendency to associate certain decongestants with lower risk was also present:

> Vibrocil is a less strong, milder medicine, that is more suitable for children, and that's why it is not able to kill viruses instantly. (S_35)

## *Trust in information sources*

The respondents had consulted a variety of information sources prior to the use of nasal decongestants. They had varied accounts about their trust in these sources and on the quality of the information that they had obtained. Instructions received from the doctor usually did not include information about the potential side-effects of decongestants. In some cases it was stated that the doctor had suggested that they avoid the use of decongestants for a prolonged period of time. The advice of experts (doctors, pharmacists), suggesting particular decongestants and using emotionally loaded language ('the best', 'the most harmless', 'very mild', 'suitable for small children') has favoured the decision to use these medicines in children.

> Knowing my attitude, I prefer natural things, the doctor suggested to use this [the decongestant Vibrocil]. In this case, mentioning that this medicine is very mild and is not very synthetic, and so on. Meanwhile it is a real medicine. (L_39)

The same use of language was adopted by the respondents when describing certain brand names of nasal decongestants.

The respondents were found to trust the information provided by the doctor, if cooperation with the doctor was positively evaluated.

> I think that the paediatrician loves children and she listens to and understands the patient. (J_33)

In several cases trust in the doctor's advice was questioned, because it was observed that doctors have a tendency to favour certain medications. This often leads to the caregiver's suspicion that physicians could be paid to promote the medicines of certain companies.

> Of course, it is not a secret that doctors have to suggest certain medicines and they are being reimbursed for that – for the promotion of these medicines. (N_28)

It was frequently indicated that the problem in patient–doctor communication was the doctor's own limited time to meet with a patient.

> Maybe this is because of their limited time for each patient ( . . . ), so to say 'faster, faster' ( . . . ) [she] hurries up somehow – when you enter her room, you have to figure out very quickly what you should ask in order to not forget anything. (L_28)

Pharmacists were mostly evaluated to be a convenient source of information. The personal relationship between the pharmacist and patient, as well as the pharmacist's attitude and personal opinion influenced the caregiver's trust in the pharmacist's suggestions.

> A splendid auntie, indeed! I rely on her suggestions. If she is not in the pharmacy, I try not to purchase any medicine ( . . . ). She is like a member of our family. (D_31)

In several cases the lack of trust in the pharmacist's advice was based on the perceived bias in the information that they provided to the caregiver.

> I hear only about this brand ( . . . ) also in pharmacies. I think there are good sales representatives, who have convinced everybody that this decongestant is the best, and so. (J_33)

Trust in the opinions of the lay public (mainly relatives, friends, and the internet community) was a general characteristic of the respondents. This information was characterized as unbiased and objective. Having children of the same age enhanced trust in the opinions of others.

> I ask my friends. Similar problems, also small kids ( . . . ). What she has suggested, or what she says, yes, it has helped us, too. (E_34)

The internet was characterized as a source that contained plenty of information. However, it was frequently noted that this source is not trustworthy enough, particularly if the health issues of children are addressed.

> As far as the medicines for children are concerned, I do not trust this source. I myself might fall for the bait and think – I might try the product. But for the child – no! (I_30)

The lack of trust was explained by the variety and fragmented nature of information on the internet.

> It is not enough to read one source. Instead of reading one, it is necessary to read twenty more. (D_31)

Caregivers mostly paid careful attention to PILs before using nasal decongestants in children. However, there is a tendency to read PILs when using the medicine for the first time and then to not

re-read them in cases of repeated use. Generally, respondents trust the information provided by PILs, considering it to be objective, understandable, and helpful.

> Every source has its own opinion, and the information differs, I do not know what or whom I should trust. I only trust this instruction or whatever it is called. (J_31)

However, there were claims that the information provided by PILs should be more concise. It was indicated that the information available for medicines users is too contradictory, too broad, and too fragmented to be able to find the right one. This makes it more complicated to make a decision or to evaluate a risk.

> If, say, I find several similar pieces of information, I will certainly trust it. But, if every source has its own content that differs, I do not know what to trust. (J_31)

A sceptical opinion regarding the availability of unbiased and non-commercial medical information was dominant in the interviews.

> I think that the availability of independent information is limited. Let's say, available is what the producers want to be available. (D_37)

## Discussion

### *Emotions*

The perceived need for medicines in cases of minor childhood ailments is dictated by emotions and can be attributed to the empathy and moral duty of a caregiver to 'do something for them [children]' (Johnson and Helman, 2004: 101), or 'to do the right thing' (Neill, 2010: 333). Medicines possess not only a pharmacological, but also, a social efficiency. They are a means to fight illnesses, to demonstrate care, and to share ideas. A child's illness is frequently perceived as a 'moral obligation to seek medical care' (Calnan, 1987: 151). Caregivers, especially parents, feel the need to act and to prove to themselves that they are being 'a good parent'.

The previous studies suggest that gendered roles are persistent if a child within the family experiences illness symptoms. Although fathers and other family members are often involved in decision-making, women usually bear more emotional responsibility that is described in everyday language as 'mothers' worry' and 'mothers' guilt' (Neill, 2010: 337–338). Women are perceived also by themselves and others as negligent if they fail to do something in the case of a child's illness (Whyte et al., 2002). The aspect of social medication as a means of conforming the child's behaviour to certain norms, including the comfort of a child or a caregiver, has frequently been attributed to the use of OTC medicines. This pattern has been named by Allotey, Reidpath and Elisha (2004: e381–e382) as a 'pre-emptive treatment' that is used to describe situations when medicine has been given for the purpose of comforting a child, rather than for the treatment of a disease. As disclosed by the present study, the perceived main benefit, arising from the use of these medicines, was the comfort and well-being of the child and the caregiver.

## Perceived risk

Studies suggest that caregivers' knowledge about the risk of OTC medicines covers a wide range, varying from incomplete (Hughes et al., 2002) to a relatively good level of knowledge (Cuzzolin and Benoni, 2010: 97). The use of OTC medicines for minor ailments can be attributed to debates about the characteristics of a risk society – how people in the age of modernity reflect on possible risks. The study conducted by Bissell et al. (2001) demonstrated that people tend to unconsciously think that medicines should be safe if they are sold in pharmacies (Bissell et al., 2001). Users of medicines have a tendency to concentrate more on the benefits of OTC medicines than on their risks (Bissell et al., 2001; Hibbert et al., 2002). In other words, 'riskiness' is not considered to be a central characteristic of non-prescription medicines (Bissell et al., 2001: 23).

There are some differences in respect to use of medicines in children. Several studies state that children are perceived to be more vulnerable towards risk than adults (Berry, 2004; Bradley and Blenkinsopp, 1996; Walter and Britten, 2002). This was confirmed by the present study – a number of caregivers had consulted several information sources prior to the decision-making in order to gather information about possible risks related to the use of these medicines. As a result, despite awareness of certain side-effects, caregivers were generally sure that the risk was considerably mitigated if use of the medicines was not prolonged.

Several authors have remarked that, in general, perceptions of risk are influenced by individual characteristics such as previous experience and one's personal belief system (Walter and Britten, 2002). The present study also showed that repeated use of nasal decongestants in children was based upon the fact that their caregivers did not observe any side-effects when they had previously used these medicines, thereby obtaining additional proof of their non-riskiness. One more reason given for using decongestants, as disclosed by the current study, was the perceived potential risk of complications arising from rhinitis.

## Trust in information sources

The amount of information people face nowadays is extensive. Beck (1992) suggests that due to the enormous speed of dissemination of information it is now difficult to establish perceptions regarding risk and trust with any certainty. Uncertainty, the lack of sufficient evidence, and emotions place caregivers in a difficult situation in respect to decision-making (Austin, 2001: 58).

In the context of risk perception, the influence of social relationships is very important. Risk perception is shaped and influenced by a number of different, often contradictory, agendas, including different sources of expert information – as well as information provided by lay people. Knowledge shaped by these sources often challenges the authority of the expert – also known as 'official' knowledge (Giddens, 1990; Grinyer, 1995; Williams et al., 1995). Lay people are becoming experts in risk evaluation. In order to become more capable, they gather a variety of information sources and 'construct an informational basis (. . . ) via an exchange of information' (Scott and Freeman, 1995: 162). The present study has revealed that caregivers often double-check a doctor's advice against other sources of information, including lay people who have children of approximately the same age. Their trust in the advice of experts has to be re-confirmed by lay experience and mutual negotiations (Bissell et al., 2001).

The quality of information is a crucial factor in making a decision about the usage of medicine. Information about use of medicines for a particular health problem should be clear and factual (Austin, 2001). The respondents of the present study mentioned such criteria as comprehensibility,

conciseness, and lack of bias. Cooperation with the advice-giver is proven to be important (Austin, 2001: 58). The present study showed that personal attitude, devoted time and patient satisfaction are preconditions of trust in the advisor.

Unbiased information helps one to make decisions by evaluating the potential risks against the potential benefits. However, information from commercials and product advertising always put an emphasis on the benefits of the products and stress the emotional and moral duty to alleviate the condition of a sick child, and convinces people that 'a pill for every ill' is needed (Busfield, 2010; Whyte et al., 2002). Respondents frequently remarked that most information about medicine available for consumers is aimed at selling the medicine, thus maximizing the profits of the pharmaceutical companies. It was also revealed that commercial advertising strives to mitigate the risks associated with the use of nasal decongestants in children, by implying that certain nasal decongestants are 'mild' and/or 'more suitable for small children'.

## Conclusions

The role of emotions in the decision-making process regarding the use of nasal decongestants in children is evident. Caregivers are driven by the empathy and perceived need to enhance the well-being of a child. The influence of emotions is frequently exploited by advertisements, reinforcing the conviction that every symptom of illness, including those of the common cold, has to be suppressed by medication. If a doctor's advice is required in case of children's rhinitis, the wish of a caregiver 'to do something' should be taken into account. Instead of suggesting nasal decongestants, alternative symptom-relieving means can be offered, for example, saline nasal drops, humidifiers and homeopathic medicines.

According to the findings, caregivers using nasal decongestants in children put more emphasis on the benefits of these medicines – such as clear nose, comfort and reduced discharge. than on the perceived risks associated with the use of these medicines. There is also a belief that use of these medicines is unlikely to cause side-effects if usage is not prolonged. In order to stress the inappropriateness of these medicines for children and their lack of healing properties, it is necessary to provide information campaigns for healthcare professionals and users of medicines.

The studies suggest that the information on risk is more useful if it is provided taking into account different lifestyle factors, community structure (Fainzang and Haxaire, 2011) and the actual knowledge level of the patient. 'Tailored communication is more effective than general messages' (Chetley et al., 2007: 27). Knowledge about the risk associated with the use of nasal decongestants is inconsistent in Latvia, therefore experts, such as doctors, pharmacists, and nurses, should be more proactive in educating their patients about the possible side-effects of nasal decongestants and in dispelling erroneous perceptions regarding the ability of these medications to provide a healing effect.

A variety of information sources that offer fragmented information of varying quality may hinder risk evaluation and establishment of trust. It is not always necessary to ask for a doctor's advice in the case of minor ailments, like children's rhinitis. Therefore, it is essential to establish a non-commercial, interactive, and regularly updated internet source of information about medicine, including detailed information about side-effects, compatibility with different medicines, as well as other healthcare-related issues.

## Limitations of the study

The number of cases included in the study is too limited for broader generalizations, and that was not the aim of the study. The qualitative method was chosen to select cases for problem investigation, and not to acquire any representative sample. Selection of the respondents was based on the single criterion – representation of different age groups of children up to the age of six. All the caregivers were middle- or upper-class women. Opinions of a more diversified sample (including different gender, age, occupation, nationality, residence) could have provided additional perspectives to the current study.

## Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## References

Allotey P, Reidpath DD and Elisha D (2004) 'Social medication' and the control of children: a qualitative study of over-the-counter medication among Australian children. *Pediatrics* 114: e378–e383.

Austin H (2001) Parents' perceptions of information on immunisations. *Journal of Child Health Care* 5(54): 54–59.

Beck U (1992) *Risk Society: Toward a New Modernity*. London: SAGE.

Berry D (2004) Interpreting information about medication side effects: differences in risk perception and intention to comply when medicines are prescribed for adults or young children. *Psychology, Health and Medicine* 9(2): 227–233.

Bissell P, Ward PR and Noyce PR (2001) The dependent consumer: reflections on accounts of the risks of non-prescription medicines. *Health (London)* 5(1): 5–30.

Bradley C and Blenkinsopp A (1996) Over the counter drugs: the future for self medication. *British Medical Journal* 312(7034): 835–837.

Busfield J (2010) 'A pill for every ill': explaining the expansion in medicine use. *Social Science and Medicine* 70: 934–941.

Calnan M (1987) *Health and Illness: The Lay Perspective*. London and New York: Tavistock Publications.

Chetley A, Hardon A, Hodgkin C, Haaland A and Fresle D (2007) *How to Improve the Use of Medicines by Consumers*. Geneva: WHO Publications.

Cuzzolin L and Benoni G (2010) Safety of non-prescription medicines: knowledge and attitudes of Italian pharmacy customers. *Pharmacy World & Science* 32: 97–102.

Fabi M, Formigari R and Picchio FM (2009) Are nasal decongestants safer than rhinitis? A case of oxymetazoline-induced syncope. *Cardiology in the Young* 19: 633–634.

Fainzang S and Haxaire C (2011) *Of Bodies and Symptoms: Anthropological Perspectives on their Social and Medical Treatment*. Tarragona: URV Publications.

Flick U (1998) *An Introduction to Qualitative Research*. London: SAGE Publications.

Giddens A (1990) *The Consequences of Modernity*. Cambridge and Malden: Polity Press.

Grinyer A (1995) Risk, the real world and naive sociology: perceptions of risk from occupational injury in the health service. In: Gabe J (ed.) *Medicine, Health and Risk: Soiological Approaches*. Oxford: Blackwell Publishers, pp.31–52.

Haaijer-Ruskamp FM and Hemminki E (1993) The social aspects of drug use. In: Dukes MNG (ed.) *Drug Utilization Studies: Methods and Uses*. Copenhagen: WHO Regional Publications, pp.97– 124.

Heikkinen T and Järvinen A (2003) The common cold. *The Lancet* 361(9351): 51–59.

Hibbert D, Bissell P and Ward PR (2002) Consumerism and professional work in the community pharmacy. *Sociology of Health and Illness* 24(1): 46–65.

Hughes L, Whittlesea C and Luscombe D (2002) Patients' knowledge and perceptions of the side-effects of OTC medication. *Journal of Clinical Pharmacy and Therapeutics* 27(4): 243–248.

Johnson G and Helman C (2004) Remedy or cure? Lay beliefs about over-the-counter medicines for coughs and colds. *British Journal of General Practice* 54: 98–102.

KinyonMunch K (2011) What do you tell parents when their child is sick with the common cold? *Journal for Specialists in Pediatric Nursing* 16: 8–15.

Neill SJ (2010) Containing acute childhood illness within family life: a substantive grounded theory. *Journal of Child Health Care* 14(4): 327–344.

Rimsza ME and Newberry S (2008) Unexpected infant deaths associated with use of cough and cold medications. *Pediatrics* 122(2): e318–e322.

Ryan T, Brewer M and Small L (2008) Over-the-counter cough and cold medication use in young children. *Pediatric Nursing* 34(2): 174–180.

Salmane-Kulikovska I, Dobelniece S and Rungule R (2010) Use of medicines in cases of common cold in Latvia. In: *The Collection of Scientific Papers of Riga Stradins University*. Riga: Riga Stradins University, pp.96–109 (in Latvian).

Scott S and Freeman R (1995) Prevention as a problem of modernity: the example of HIV and AIDS. In: Gabe J (ed.) *Medicines, Health and Risk: Sociological Approaches*. Oxford: Blackwell Publishers, pp.151–170.

Shefrin AE and Goldman RD (2009) Use of over-the-counter cough and cold medications in children. *Canadian Family Physician* 55: 1081–1083.

Smith SM, Henman M, Schroeder K and Fahey T (2008) Over-the-counter cough and cold medicines for children: neither safe or efficacious? *British Journal of General Practice* 1(9): 757–758.

Walter FM and Britten N (2002) Patients' understanding of risk: a qualitative study of decision-making about the menopause and hormone replacement therapy in general practice. *Family Practice* 19: 576–587.

Whyte SR, van der Geest S and Hardon A (2002) *Social Lives of Medicines*. Cambridge: Cambridge University Press.

Williams G, Popay J and Bissell P (1995) Public health risks in the material world: barriers to social movements in health. In: Gabe J *Medicines, Health and Risk: Sociological Approaches*. Oxford: Blackwell Publishers, pp.113–132.

Arch Microbiol (2012) 194:461–471
DOI 10.1007/s00203-011-0785-7

ORIGINAL PAPER

# Electron transport and oxidative stress in *Zymomonas mobilis* respiratory mutants

**Inese Strazdina · Zane Kravale · Nina Galinina · Reinis Rutkis · Robert K. Poole · Uldis Kalnenieks**

Received: 27 July 2011 / Revised: 19 November 2011 / Accepted: 16 December 2011 / Published online: 7 January 2012


**Abstract** The ethanol-producing bacterium *Zymomonas mobilis* is of great interest from a bioenergetic perspective because, although it has a very high respiratory capacity, the respiratory system does not appear to be primarily required for energy conservation. To investigate the regulation of respiratory genes and function of electron transport branches in *Z. mobilis*, several mutants of the common wild-type strain Zm6 (ATCC 29191) were constructed and analyzed. Mutant strains with a chloramphenicol-resistance determinant inserted in the genes encoding the cytochrome *b* subunit of the $bc_1$ complex (Zm6-*cytB*), subunit II of the cytochrome *bd* terminal oxidase (Zm6-*cydB*), and in the catalase gene (Zm6-*kat*) were constructed. The *cytB* and *cydB* mutants had low respiration capacity when cultivated anaerobically. Zm6-*cydB* lacked the cytochrome *d* absorbance at 630 nm, while Zm6-*cytB* had very low spectral signals of all cytochromes and low catalase activity. However, under aerobic growth conditions, the respiration capacity of the mutant cells was comparable to that of the parent strain. The catalase mutation did not affect aerobic growth, but rendered cells sensitive to hydrogen peroxide. Cytochrome *c* peroxidase activity could not be detected. An upregulation of several thiol-dependent oxidative stress-protective systems was observed in an aerobically growing *ndh* mutant deficient in type II NADH dehydrogenase (Zm6-*ndh*). It is concluded that the electron transport chain in *Z. mobilis* contains at least two electron pathways to oxygen and that one of its functions might be to prevent endogenous oxidative stress.

**Keywords** *Zymomonas mobilis* · Respiratory chain · Cytochrome *bd* · Cytochrome $bc_1$

Communicated by Gregory Cook.

I. Strazdina · Z. Kravale · N. Galinina · R. Rutkis · U. Kalnenieks (✉)
Institute of Microbiology and Biotechnology, University of Latvia, Kronvalda boulv. 4, 1586 Riga, Latvia
e-mail: kalnen@lanet.lv

R. K. Poole
Department of Molecular Biology and Biotechnology, The University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, UK

## Introduction

*Zymomonas mobilis* is a facultatively anaerobic, obligately fermentative bacterium with a highly active ethanol fermentation pathway (Rogers et al. 1982, 2007; Sprenger 1996). Nevertheless, *Z. mobilis* possesses an active, constitutive aerobic respiratory chain (Strohdeicher et al. 1990; Kalnenieks et al. 1998), supporting oxygen uptake rates substantially higher than those of *Saccharomyces cerevisiae*, *Escherichia coli*, or *Pseudomonas putida* (Kalnenieks 2006; Sootsuwan et al. 2008). The physiology of aerobic electron transport in *Z. mobilis* is unusual for bacteria. Although the cytoplasmic membrane of *Z. mobilis* carries a functional $H^+$-ATP synthase complex (Reyes and Scopes 1991), and non-growing cells and membrane preparations show oxidative phosphorylation activity (Kalnenieks et al. 1993), the bacterium does not use its respiratory chain to supply energy for aerobic growth in the manner typical of aerobic and facultatively anaerobic microorganisms. The aerobic biomass yields of *Z. mobilis* do not exceed the anaerobic yields, which are typically around 8–10 g dry wt $(\text{mol glucose})^{-1}$ for batch cultures growing on rich media (Belaich and Senez 1965; Bringer et al. 1984). Indeed, respiratory metabolism is inhibitory for growth because of

the accumulation of acetaldehyde and other toxic byproducts (Viikari 1986; Viikari and Berry 1988), and a stimulation of aerobic growth is evident when respiration is inhibited by addition of cyanide at submillimolar concentrations (Kalnenieks et al. 2000), or on mutation of the respiratory NADH dehydrogenase (Kalnenieks et al. 2008). However, the reasons for the low energetic efficiency of respiration in growing *Z. mobilis* are not clear.

The organization of respiratory components and the routes for electron transfer to oxygen are not fully resolved in *Z. mobilis*, but Fig. 1 shows a current working model. There is only one functional respiratory NAD(P)H dehydrogenase in the *Z. mobilis* electron transport chain, belonging to the type II respiratory dehydrogenase (Ndh) family (Kalnenieks et al. 2008; Seo et al. 2005; Yang et al. 2009a), and only one terminal cytochrome *bd*-type quinol oxidase has been identified so far (Kalnenieks et al. 1998; Sootsuwan et al. 2008). The known *Z. mobilis* genome sequences contain genes encoding a cytochrome $bc_1$ complex and cytochrome *c* (Seo et al. 2005; Yang et al. 2009a; Kouvelis et al. 2009), yet lack sequences homologous to any known bacterial cytochrome *c* oxidase genes. Sootsuwan et al. (2008) proposed that the cytochrome $bc_1$ branch most likely is terminated by a cytochrome *c* peroxidase, as the cytochrome *c* peroxidase gene is present in the genome. In the present work, we aimed to investigate the function of cytochrome *bd* and the cytochrome $bc_1$ complex in respiration, and to find out whether the cytochrome $bc_1$ complex supplies electrons to a cytochrome *c* peroxidase.

Energy generation is the central, yet not the sole function of bacterial electron transport chains (Poole and Cook 2000). The aerobic respiration in bacteria generates reactive species of oxygen (ROS) (Gonzalez-Flecha and Demple 1995), but may also protect the cell interior from molecular oxygen (as demonstrated for nitrogen-fixing bacteria, see Kelly et al. 1990), and also from oxidative stress by diverting electrons from hydrogen peroxide-generating reactions (Korshunov and Imlay 2010). The potential respiratory-protective and ROS-protective roles of the electron transport in *Z. mobilis*, as a physiological alternative to oxidative energy generation, are of interest but so far have attracted little attention.

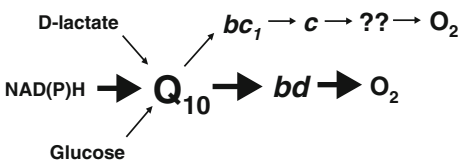



**Fig. 1** The current model of respiratory chain of *Zymomonas mobilis*. A route to oxygen via cytochrome *bd* is established and supported in this work. The nature of any other terminal oxidase is unknown

To address these problems, we have carried out a comparative study of respiration, and regulation of respiratory and ROS-protective genes in several *Z. mobilis* mutants, namely an Ndh-deficient strain, a strain deficient in the cytochrome *b* subunit of the $bc_1$ complex, a strain deficient in the subunit II of the cytochrome *bd* terminal oxidase, and a catalase-deficient strain. The results demonstrated significant effects of the knockout mutations in cytochrome *bd* and in cytochrome $bc_1$ complex on respiration, yet did not verify the presence of a cytochrome *c* peroxidase activity. At the same time, some unexpected pleiotropic effects of the respiratory gene mutations were observed.

## Materials and methods

### Bacterial strains, plasmids, and transformation

*Escherichia coli* JM109 and plasmid pGEM-3Zf(+) were purchased from Promega. Strain JM109 was used as the host for cloning of the recombinant plasmids. *Z. mobilis* ATCC 29191 (Zm6) and its mutant derivative, deficient in type II respiratory NADH dehydrogenase (the *ndh* mutant; Zm6-*ndh*) were maintained and cultivated as described previously (Kalnenieks et al. 1993, 2008). The plasmids and strains constructed and used in the present work are listed in Table 1. *E. coli* was transformed by the $CaCl_2$ procedure described by Sambrook et al. (1989). *Z. mobilis* was transformed by electroporation (Liang and Lee 1998).

### PCR, cloning techniques, and mutant construction

Genomic and plasmid DNA isolation from *Z. mobilis* were performed as before (Kalnenieks et al. 2006, 2008). The *Z. mobilis* catalase gene (*Z. mobilis* ZM4 genome sequence, locus tag ZMO 0918) was amplified by PCR using the primer pair Z.m.cat1 (AAGAGGGATCCTATGACTAGA CCCAATCTT) and Z.m.cat2 (GAAGCAGCAAGCTTT ATAACAGGCTATCGG). The engineered restriction sites for *Bam*HI and *Hind*III, respectively, are underlined. To obtain a mutant defective in the *kat* gene (Zm6-*kat*), the 1.45-kb region of the chromosomal DNA containing the *kat* ORF (ZMO 0918) was amplified, double-digested with *Bam*HI and *Hind*III, and directionally cloned between the *Bam*HI and *Hind*III restriction sites of the multiple cloning site (MCS) of plasmid pGEM-3Zf(+). In addition, the *Eco*RI site was removed from the MCS, by eliminating a 0.35-kb fragment between the *Eco*24I restriction sites of pGEM-3Zf(+) yielding plasmid pGEMcat (Table 1). A 0.8-kb fragment of plasmid pBT (Table 1), containing the 659-bp ORF of the chloramphenicol acetyltransferase gene, was amplified using a primer pair with *Eco*RI sites

**Table 1** Plasmids and strains used in the study

| Plasmid/strain | Characteristics | Source |
|---|---|---|
| pGEM-3Zf(+) | $Amp^r$ | Promega |
| pBT | $Cm^r$ | Stratagene |
| pGEMb | pGEM-Zf(+) derivative, carrying a 1.3-kb fragment of PCR-amplified genomic DNA with the ORF of the cytochrome *b* subunit gene (ZMO 0957) of the $bc_1$ complex, cloned between the *Hind*III and *Bam*HI sites of the MCS | Present work |
| pGEMb::$cm^r$ | pGEMb derivative, carrying in the *Age*I site of the cloned gene a 1.3-kb *Age*I restriction fragment of pBT, with an 0.7-kb chloramphenicol resistance ORF | Present work |
| pGEMd | pGEM-Zf(+) derivative, carrying a 1.55-kb fragment of PCR-amplified genomic DNA with part of the ORF of *cydA* (ZMO 1571) and the whole of *cydB* (ZMO 1572), cloned between the *Hind*III and *Bam*HI sites of the MCS | Present work |
| pGEMd::$cm^r$ | pGEMd derivative, carrying in the *Age*I site of the *cydB* (ZMO 1572) a 1.3-kb *Age*I restriction fragment of pBT, with an 0.7-kb chloramphenicol resistance ORF | Present work |
| pGEMcat | pGEM-Zf(+) derivative, (1) carrying a 1.4-kb fragment of PCR-amplified genomic DNA with the ORF of the catalase gene (ZMO 0918) cloned between the *Hind*III and *Bam*HI sites of the multiple cloning site, (2) lacking *Eco*RI site in the MCS in result of elimination of an 0.35-kb fragment between *Eco*24I restriction sites of pGEM-3Zf(+) | Present work |
| pGEMcat::$cm^r$ | pGEMcat derivative, carrying an 0.8-kb PCR-amplified fragment of pBT with 0.7 kb of the chloramphenicol resistance ORF, inserted in the *Eco*RI site of the cloned gene | Present work |
| Zm6 | Parent strain | ATCC 29191 |
| Zm6-*ndh* | Zm6 strain with a $cm^r$ insert in the ORF of respiratory type II NADH degydrogenase gene (*ndh*) (ZMO 1113) | Kalnenieks et al. (2008) |
| Zm6-*cytB* | Zm6 strain with a $cm^r$ insert in the ORF of the cytochrome *b* subunit gene (ZMO 0957) of the $bc_1$ complex | Present work |
| Zm6-*cydB* | Zm6 strain with a $cm^r$ insert in the ORF of the subunit II (*cydB*) gene (ZMO 1572) of the cytochrome *bd* terminal oxidase | Present work |
| Zm6-*kat* | Zm6 strain with a $cm^r$ insert in the ORF of catalase gene (ZMO 0918) | Present work |

(underlined): cm1 (TTTGCTTTCGAATTCCTGCCATTC ATCCGC) and cm2 (CACTACCGGGCGAATTCTTTGA GTTATCGAG). The PCR product was digested with *Eco*RI and inserted in the *Eco*RI site of the cloned *kat* gene, yielding plasmid pGEMcat::$cm^r$ (Table 1). This plasmid was used to transform *Z. mobilis* by electroporation, and homologous recombinants were selected on plates containing chloramphenicol (120 μg $ml^{-1}$). Several colonies were screened for the *kat::cm*$^r$ genotype by PCR on the genomic DNA template with primers Z.m.cat1 and Z.m.cat2; insertion of the chloramphenicol-resistance determinant in the catalase gene was verified by sequencing the PCR product.

The gene encoding the cytochrome *b* subunit of the respiratory chain $bc_1$ complex (*Z. mobilis* ZM4 genome sequence, locus tag ZMO 0957) was amplified using the primer pair Z.m.b1 (GAACTTATTATGAAGCTTTCAC CACCCCCT) with the engineered site for *Hin*dIII underlined, and Z.m.b2 (TTGAGCATATCAGGATCCCGT TCTTTTCTT) with the *Bam*HI site underlined. In the *Z. mobilis* ZM4 genome annotation, the expected product of ZMO 0957 is described as 'cytochrome b/b6 domain-containing protein' of the cytochrome $bc_1$ complex. ZMO 0957 is homologous to the *cytB* gene of the taxonomically related *Sphingobium japonicum*; therefore, by analogy, we named this gene '*cytB*'. To construct a *cytB* mutant defective in the cytochrome *b* subunit of the $bc_1$ complex, the 1.26-kb region of the chromosomal DNA containing the gene ZMO 0957 was amplified and cloned as described for the catalase gene construct, yielding plasmid pGEMb (Table 1). Plasmid pBT was digested with *Age*I (*Bsh*TI) to obtain three fragments of approximately 1.6, 1.3, and 0.3 kb length. The 1.3-kb fragment carried the ORF of the chloramphenicol acetyltransferase gene. The *Age*I digestion of pBT was chosen because plasmid pGEMb contained only one *Age*I restriction site that was localized in the cloned ORF of the *cytB* gene. After digestion of pGEMb with *Age*I, the 1.3-kb fragment of pBT was inserted in the

middle of the cloned gene to yield plasmid pGEMb::cm$^r$ (Table 1). Transformation of *Z. mobilis* and screening for the mutation were done exactly as described for the *kat* gene.

A 1.55-kb genomic DNA fragment, with part of the gene encoding the cytochrome *bd* subunit I and the whole of the gene encoding the cytochrome *bd* subunit II (*Z. mobilis* ZM4 genome sequence, locus tags ZMO 1571 and ZMO 1572), was amplified, using the primer pair Z.m.d1 (ATG AAGCTTCTTGGATCCTGACCCATAGTC) and Z.m.d2 (TACCCGTCACTGCTGGTAAGCTTGCCGTGG), with the engineered restriction sites for *Bam*HI and *Hind*III, respectively, underlined. In order to disrupt subunit II of the cytochrome *bd* terminal oxidase, the amplified 1.55-kb DNA fragment, containing part of ZMO 1571 (homologous to the *cydA* gene of *E. coli*) and the whole ZMO 1572 ORF (homologous to the *cydB* gene of *E. coli*, respectively), was directionally cloned in the MCS of pGEM-3Zf(+), yielding plasmid pGEMd (Table 1). The plasmid pGEMd contained a single *Age*I restriction site, which was localized in ZMO 1572. Further steps in the *cydB* mutant construction were identical to those described above for the *cytB* mutant. After transformation, several colonies were screened for the *cydB*::*cm$^r$* genotype by PCR on the genomic DNA template with primers Z.m.d1 and Z.m.d2, and the insertion of the chloramphenicol-resistance determinant in the *cydB* gene was verified by sequencing the PCR product.

Primers for PCR reactions were supplied by Operon and Invitrogen. PCRs were carried out in an Eppendorf Mastercycler, using Fermentas *Taq*DNA polymerase. Other DNA manipulations were done as described previously (Kalnenieks et al. 2008), using Qiagen kits. All DNA constructs were confirmed by DNA sequencing, done by Beckman Coulter genomics.

## Quantitative PCR analysis

Primer pairs for 14 genes (Table 2) were designed to give PCR products of 200 (±8) bp length. An RNeasy Mini kit (Qiagen) was used for mRNA purification. Reverse transcription was done with the Revert Aid Premium First Strand cDNA Synthesis kit (Fermentas), and Maxima

**Table 2** Genes and the corresponding primer pairs used for qRT-PCR analysis

| Gene product | Gene | Primers | Sequences (5′–3′) |
|---|---|---|---|
| Alkyl hydroperoxide reductase | ZMO 1732 | ahpC_f | ggattacggccaacattcaa |
| | | ahpC_r | attacgcttggcggaaatct |
| Catalase | ZMO 0918 | kat_f | agggaattgggatttagtcg |
| | | kat_r | aagaggaataccacgatcag |
| Cytochrome *bd* subunit I | ZMO 1571 | cydA_f | tatgaattcgggacgaactg |
| | | cydA_r | ccaagtcatggaaatcaagg |
| Glyceraldehyde-3-P dehydrogenase | ZMO 0177 | gapdeh_f | aagcttggcgttgatatcgt |
| | | gapdeh_r | gtgcaagatgcgttagaaac |
| NADPH-glutathione reductase | ZMO 1211 | gor_f | ttttataaagcgcgcgatcg |
| | | gor_r | gatcgggttttcggtcataa |
| Glutaredoxin 2 | ZMO 0070 | grx_f | gttataagctggtgccgatt |
| | | grx_r | aattctggcaaaggtgccat |
| D-lactate dehydrogenase | ZMO 0256 | ldh_f | tgagcaggttatctgtttgc |
| | | ldh_r | tgatacgagcataaagggtc |
| Type II NADH dehydrogenase | ZMO 1113 | ndh_f | agaaggccataaaatcagcg |
| | | ndh_r | ttgagcaatcatggttctgg |
| Putative iron-dependent peroxidase | ZMO 1573 | per_f | caaagagttaaaaggcgtgc |
| | | per_r | tctacaaaaccaatcagggc |
| Cytochrome *c* peroxidase | ZMO 1136 | perC_f | ttttcctgttcgttgatcgc |
| | | perC_r | ccatgattgggcagaagtta |
| NADH:ubiquinone oxidoreductase of RnfABCDGE type, RnfA subunit | ZMO 1814 | rnfA_f | atccaaccgaggaagcaaaa |
| | | rnfA_r | attctgattatcgcttcggc |
| Superoxide dismutase | ZMO 1060 | sod_f | ccatctttcaaaacgagcca |
| | | sod_r | ttgttcaacaatgcggcaca |
| Thioredoxin reductase | ZMO 1142 | tor_f | atgcctttgatgctggtttg |
| | | tor_r | ttattgctggggaagatagc |
| Thioredoxin | ZMO 1097 | trx_f | ttttgaaaagcagcagggtc |
| | | trx_r | taaaatcaccggtgcctgtt |

SYBR green/ROX qPCR Master Mix (Fermentas) was used for the PCR. The quantitative real-time PCRs (qRT-PCR) were carried out in duplicate in a real-time thermal cycler (Model 7300, Applied Biosystems). In order to compare gene transcription between the strains and cultivation conditions, qRT-PCR data in all cases were normalized against the respective amounts of cDNA of the glyceraldehyde-3-phosphate dehydrogenase gene (ZMO 0177). Choice of ZMO 0177 as the 'housekeeping' gene for *Z. mobilis* was justified by the data of Yang et al. (2009b), showing that transcription of this gene is constitutive, and is not affected significantly by aeration or by accumulation of metabolic end products.

### Continuous cultivation

Continuous cultivation was carried out in a Labfors fermenter (Infors) of 2.5-l working volume, containing 1 l of culture, at 30°C, pH 6.0, and a dilution rate of 0.18 $h^{-1}$. The growth medium contained glucose (20 g $l^{-1}$), yeast extract (5 g $l^{-1}$). and mineral salts, as described previously (Kalnenieks et al. 1993). Cultivations were started under anaerobic conditions and established by gassing the culture with nitrogen gas, containing only trace amounts of oxygen, at a flow rate of 1.4 l $min^{-1}$, and stirring rate 100 rpm. During the steady state under these conditions, the metabolism was strictly anaerobic: ethanol yield in the parent strain Zm6 reached 0.49–0.50 g g $glucose^{-1}$ (close to the theoretical maximum value; Rogers et al. 1982), and no acetaldehyde could be detected in the medium. When the anaerobic steady state had established, samples were taken for qRT-PCR, as well as for product analysis, and for whole-cell oxygen uptake assay, catalase, and peroxide-generation assays. After that, aeration was started with an air flow of 1.5 l $min^{-1}$ and stirring rate 300 rpm., and the chemostat was run overnight. It took typically 12–15 h to reach the aerobic steady state, and at that point, sampling was repeated. During the aerobic steady state, the parent strain Zm6 maintained $pO_2$ in the fermenter around 0–2%, ethanol yield reached 60% of the theoretical maximum, and 0.2 g acetaldehyde $l^{-1}$ was present in the medium, corresponding to a moderately aerobic metabolism of *Z. mobilis* (for review see Kalnenieks 2006).

### Preparation of cell-free extracts and membrane vesicles

For preparation of cell-free extract, cells were sedimented by centrifugation at 5,000 rpm for 15 min, washed, and resuspended in 100 mM potassium phosphate buffer, containing 2 mM magnesium sulfate, pH 6.9, to an $OD_{550}$ of about 40 (after correction for dilution; measured in a Shimadzu spectrophotometer), corresponding to a biomass concentration of 6.8–7.0 mg (dry wt) $ml^{-1}$. Cells were broken by an 8-min ultrasonic treatment with pulses of 0.5-s duration, separated by 0.5-s intervals, using a Dr. Hielscher ultrasonic processor. Typically, cell-free extracts of 4.6–4.9 mg protein $ml^{-1}$ were obtained. Subsequent removal of unbroken cells and separation of cytoplasmic membranes by ultracentrifugation were performed as before (Kalnenieks et al. 1993, 1998).

### Cytochrome spectroscopy

Room-temperature reduced minus oxidized cytochrome absorption spectra were taken using membrane samples (1 ml) at a protein concentration of 5–6 mg $ml^{-1}$ with small amounts of solid dithionite as reductant and potassium ferricyanide as oxidant. Spectra were recorded with a custom-built SDB4 dual-wavelength scanning spectrophotometer (University of Pennsylvania School of Medicine Biomedical Instrumentation Group and Current Designs, Philadelphia, PA), as described previously (Kalnenieks et al. 1998). The time-course of cytochrome *d* reduction after addition of NADH was recorded by rapid, repetitive scanning in the wavelength range between 610 and 650 nm, using the dual-wavelength scanning spectrophotometer. The degree of cytochrome *d* reduction was calculated as the average value of the absorbance differences at wavelength pairs 630/614 and 630/646 nm.

### Enzyme assays

Catalase activity in cell-free extracts was assayed spectrophotometrically, by monitoring absorbance decline at 240 nm (Gonzalez-Flecha and Demple 1994). Cytochrome *c* peroxidase activity was monitored by the decline in absorbance at the α-band of ferrocytochrome *c* at 550 nm on addition of cell-free extract and $H_2O_2$ (Ellfolk and Soininen 1970). Glutathione reductase activity was measured by decrease of NADPH absorbance at 340 nm on addition of oxidized glutathione (GSSG) in the presence of permeabilized cells (Bergmeyer et al. 1974). The cells were permeabilized as described previously (Kalnenieks et al. 2005), following a slightly modified procedure of Osman et al. (1987).

### Analytical methods

$H_2O_2$ production by cells was determined fluorimetrically by monitoring Amplex UltraRed fluorescence during its reaction with $H_2O_2$, catalyzed by horseradish peroxidase (Korshunov and Imlay 2010). Fluorescence was measured with a FluoroMax-3 spectrofluorimeter (Jobin–Yvon), using 520-nm wavelength for excitation, and 620-nm wavelength for emission. To quantitate the generated

hydrogen peroxide, fluorescence increase was calibrated by addition of 1 mM $H_2O_2$ in 5-µl increments. Ethanol and glucose in culture media were determined by HPLC (Agilent 1100 series), using a Biorad Aminex HPX—87H column. Acetaldehyde concentration was determined by the alcohol dehydrogenase assay, as described previously (Kalnenieks et al. 2000). Concentration of dissolved oxygen was monitored by Clark-type oxygen electrodes. An autoclavable Ingold electrode was used in the fermenter, and a Radiometer electrode with a thermostatted electrode cell for oxygen uptake measurements in cell or membrane vesicle suspensions. Whole-cell oxygen uptake was measured for cells taken from steady-state cultures, pelleted, washed, and resuspended in 100 mM phosphate buffer, pH 6.9, supplemented with 10 g glucose $l^{-1}$. Protein concentration in cell-free extracts and membrane samples was determined according to Markwell et al. (1978). Cell concentration was determined as $OD_{550}$, and dry cell mass of the suspensions was calculated by reference to a calibration curve. Results are means of at least three replicates. SEM values are given as error bars in the figures.

## Results

### Respiratory gene regulation in the NADH dehydrogenase-deficient strain

Previously, we have demonstrated that the *ndh* mutant (Zm6-*ndh*) grows aerobically with very low rate of respiration and high ethanol yield, reaching higher biomass concentrations than the parent strain (Kalnenieks et al. 2008). This is confirmed in the whole-cell respiration assays of Zm6-*ndh* shown in Fig. 2a. Here we studied respiratory gene expression in this mutant and showed that it differed significantly from that of the parent Zm6 and other strains. In Zm6, upon transition from anaerobic to aerobic culture conditions, the transcription of *cydA* (ZMO 1571) was down-regulated 12-fold, yet transcription of other respiratory genes was not affected significantly, changing less than twofold (not shown). In the aerobically grown Zm6-*ndh* strain, the *ldh* gene (ZMO 0256) encoding the membrane D-lactate dehydrogenase and *cydA* were up-regulated relative to Zm6 (Table 3), supporting our previous observation of the elevated activity of membrane D-lactate oxidase in an aerobically grown *ndh* mutant (Kalnenieks et al. 2008). However, the knockout of the *ndh* gene did not affect transcription of *rnfA* (ZMO 1814). In *Rhodobacter capsulatus,* this gene encodes an alternative NADH/ubiquinone oxidoreductase, which is a part of a membrane complex supplying electrons to nitrogenase (Schmehl et al. 1993). In *Z. mobilis,* its function is unknown.

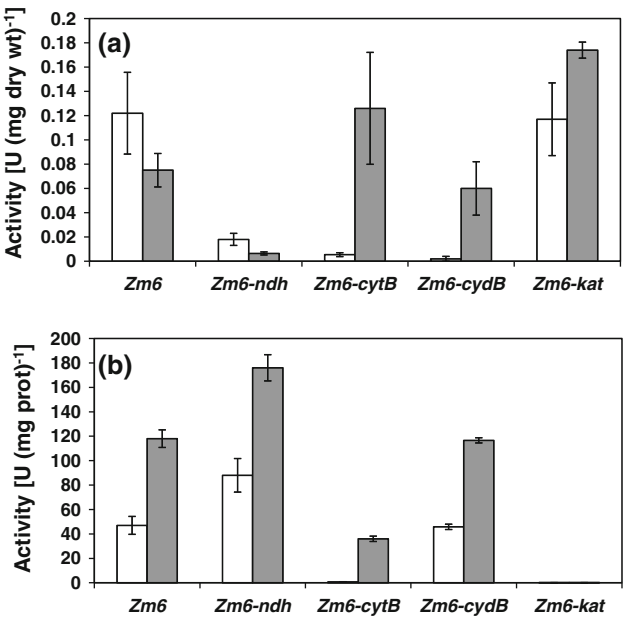



**Fig. 2** Oxygen uptake rate and catalase activity in cells, harvested from chemostat cultures under anaerobic and aerobic steady-state conditions. Cells were sedimented, washed, and resuspended in 100 mM potassium phosphate buffer (pH 6.9) supplied with glucose (10 g $l^{-1}$). *Empty bars*—anaerobically cultivated cells; *filled bars*—aerobically cultivated cells. **a** oxygen uptake rate, **b** catalase activity in cell-free extracts, obtained by ultrasonic disruption

**Table 3** The qRT-PCR ratios (mutant/Zm6) of the aerobic and anaerobic gene transcription, showing more than twofold difference between the mutant and parent strain

| | (Zm6-*ndh*/ Zm6) | | (Zm6-*cytB*/ Zm6) | | (Zm6-*cydB*/ Zm6) | | (Zm6-*kat*/ Zm6) | |
|---|---|---|---|---|---|---|---|---|
| | $+O_2$ | $-O_2$ | $+O_2$ | $-O_2$ | $+O_2$ | $-O_2$ | $+O_2$ | $-O_2$ |
| *kat* | | 0.2 | | 0.25 | 4.59 | 4.41 | n.d. | n.d. |
| *ahpC* | | | | 0.13 | 3.39 | | | 0.09 |
| *sod* | 3.2 | | 4.0 | | | 2.2 | 4.6 | |
| *per* | | | | 4.0 | | | | 0.22 |
| *perC* | | | | | | 0.18 | | |
| *rnfA* | | | 3.5 | 2.1 | 0.47 | 0.49 | | |
| *ndh* | n.d. | n.d. | | | | 0.39 | 2.5 | 0.5 |
| *ldh* | 3.0 | | | 0.38 | | | | |
| *cydA* | 6.1 | 0.27 | 2.1 | 0.23 | | | | 0.05 |
| *trx* | 4.9 | | | | | | 2.3 | 0.41 |
| *grx* | 60.0 | | | | | | 0.04 | |
| *tor* | 2.1 | | 3.0 | | | | | |
| *gor* | | | | 0.08 | 2.25 | | 0.38 | 0.5 |

*n.d.* not determined

Notably, we found some indications of oxidative stress in aerobically growing Zm6-*ndh*. The gene, encoding glutaredoxin 2 (ZMO 0070; *grx* in Table 3), was up-regulated 60-fold relative to the parent strain, the thioredoxin

gene (ZMO 1097; *trx* in Table 3) was up-regulated almost fivefold, and the thioredoxin reductase gene (ZMO 1142; *tor* in Table 3) more than twofold. As in the other mutant strains, the gene encoding superoxide dismutase (ZMO 1060; *sod*) was up-regulated more than threefold relative to Zm6. Although we did not see any change of glutathione reductase (ZMO 1211; *gor*) transcription, under aerobic growth conditions, the enzyme was more active in the Zm6-*ndh* strain: 0.030 U (mg dry wt)$^{-1}$, versus 0.009 U (mg dry wt)$^{-1}$ in Zm6 (not shown).

## The respiratory chain in the cytochrome *bd* and cytochrome $bc_1$ mutants

Anaerobically cultivated cells of the Zm6-*cydB* strain, under oxic conditions in potassium phosphate buffer supplemented with 1% glucose, had insignificant oxygen uptake activity (Fig. 2a). This finding confirmed the previously postulated key role of the *bd*-type terminal oxidase in *Z. mobilis* respiration. It is in a good agreement with the genome data (Seo et al. 2005; Yang et al. 2009a) and is supported by spectroscopic evidence (Kalnenieks et al. 1998, 2000; Sootsuwan et al. 2008) demonstrating the presence of cytochrome *bd* both in aerobically and anaerobically grown Zm6 cells. Unexpectedly, however, anaerobically grown cells of the Zm6-*cytB* strain also had a near-zero oxygen consumption rate (Fig. 2a). Furthermore, when the cultures were shifted to aerobic growth conditions, oxygen uptake both in Zm6-*cydB* and Zm6-*cytB* gradually increased (not shown). After 11–12 h of aerobic growth, the respiration rate of the mutant cells reached a level comparable to that of Zm6, as shown in Fig. 2a.

Antimycin is considered a specific inhibitor of the cytochrome $bc_1$ complex (Trumpower and Gennis 1994). Antimycin sensitivity of *Z. mobilis* respiration has been reported previously by Strohdeicher et al. (1990). Here we compared antimycin sensitivity of oxygen uptake in membrane preparations obtained from aerobically grown Zm6, Zm6-*cytB,* and Zm6-*cydB* cells. There was no significant difference in antimycin sensitivity between Zm6 and Zm6-*cydB* (not shown); however, the *cytB* mutation increased the antimycin resistance of oxygen uptake in membrane preparations with NADH as the electron donor (Fig. 3, inset). Such an effect of the *cytB* mutation upon sensitivity of respiration toward an inhibitor of the cytochrome $bc_1$ complex indicates that part of electrons in the respiratory chain of the parent strain is transported to oxygen via the cytochrome $bc_1$ branch.

Reduced minus oxidized absorption spectra of cytochromes were recorded in membranes, prepared from cells in stationary-phase batch cultures. The results presented in Fig. 3 show that cytoplasmic membranes of anaerobically cultivated Zm6-*cydB* cells, as expected, lacked the

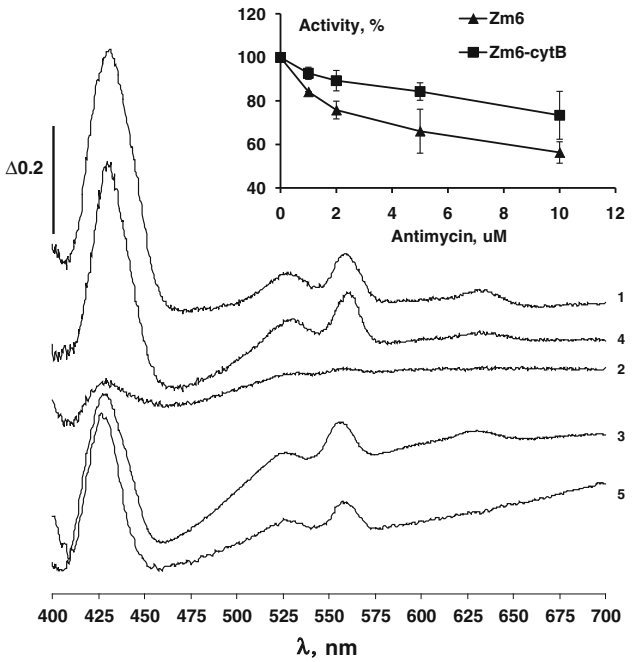



**Fig. 3** Reduced minus oxidized difference spectra of membrane preparations. (*1*) Anaerobically cultivated Zm6, (*2*) anaerobically cultivated Zm6-*cytB*, (*3*) aerobically cultivated Zm6, (*4*) aerobically cultivated Zm6-*cytB*, (*5*) anaerobically cultivated Zm6-*cydB*. *Inset* antimycin titration of oxygen uptake in aerobically cultivated cell membrane preparations of Zm6 (*filled triangle*) and Zm6-*cytB* (*filled square*)

absorbance around 630 nm, characteristic for cytochrome *d*. However, anaerobically grown Zm6-*cytB* exhibited a general deficiency of cytochromes (Fig. 3), explaining the low respiratory rates measured (Fig. 2). In comparison to anaerobically cultivated Zm6 and Zm6-*cydB*, the Soret region in Zm6-*cytB* membranes was weak; there was almost no signal in the α-region of cytochromes *c* and *b* between 550 and 560 nm, and also no cytochrome *d* absorbance at 630 nm. Surprisingly, the amount of *cydA* transcript in the anaerobically cultivated Zm6-*cytB* strain was just fourfold lower than that of the anaerobically grown Zm6 (Table 3), but still higher than in aerobically grown Zm6, in which the absorbance of cytochrome *d* around 630 nm was clearly present (Fig. 3). Likewise, anaerobically cultivated Zm6-*cytB* cells lacked catalase activity (Fig. 2b), while the transcription of *kat* was very similar to that in anaerobically cultivated Zm6-*ndh* (Table 3), in which catalase activity was high. The rise of respiratory capacity of the *cytB* mutant during aerobic cultivation was accompanied by an increase of the spectral features of cytochromes (Fig. 3) and of catalase activity (Fig. 2b).

The time-course of cytochrome *d* reduction was monitored after addition of NADH to anaerobic suspensions of membrane vesicles, prepared from aerobically cultivated Zm6, Zm6-*cytB,* and Zm6-*cydB* cells (Fig. 4). As expected,

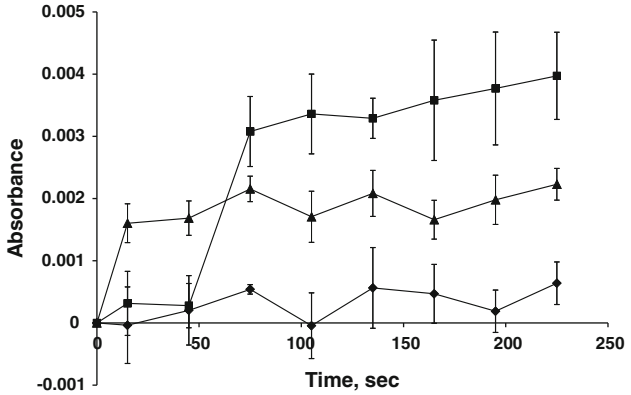



**Fig. 4** Time-course of cytochrome *d* reduction in membranes prepared form aerobically grown Zm6 (*filled triangle*), Zm6-*cytB* (*filled square*), and Zm6-*cydB* (*filled diamond*). Cytochrome *d* spectral signals were recorded by repetitive scanning in the wavelength range between 610 and 650 nm at 30-s intervals, after addition of 3.5 mM NADH to anaerobic suspension of membranes. The degree of cytochrome *d* reduction was taken as the average of the absorbance differences at wavelength pairs 630/614 and 630/646 nm

in the membranes from the Zm6-*cydB* strain, the spectral signal of reduced cytochrome *d* remained close to zero during the whole experiment. On the contrary, for Zm6-*cytB*, after a short lag time, the cytochrome *d* absorbance reached even higher level, than in Zm6. This finding indicates that (1) in Zm6-*cytB,* electrons are diverted more toward the cytochrome *bd* terminal oxidase, than in the parent strain, and that (2) in *Z. mobilis,* the cytochrome $bc_1$ complex and cytochrome *bd* are localized in different electron transport branches (as in other bacteria with branched electron transport chains; see Poole and Cook 2000).

The *cytB* mutant was the only strain in which *rnfA* (ZMO 1814) was up-regulated relative to Zm6, under both anaerobic and aerobic conditions (Table 3). However, the only functional respiratory NAD(P)H dehydrogenase of *Z. mobilis*—the type II NAD(P)H dehydrogenase *ndh*—was expressed at the same level as in the parent strain. In the *cydB* mutant, *rnfA* was down-regulated relative to Zm6, while some of the ROS-protective genes, like those of catalase, *ahpC* (ZMO 1732), and glutathione reductase (ZMO 1211), were up-regulated under aerobic growth conditions. For catalase, the fourfold increase of gene transcription did not match with the observed enzymatic activity: In Zm6-*cydB,* it was the same as in Zm6 (Fig. 2b).

## Properties of the catalase mutant and the apparent lack of cytochrome *c* peroxidase activity

Cytochrome *c* peroxidase is a periplasmic enzyme, loosely bound to the bacterial cytoplasmic membrane (Goodhew et al. 1990; Atack and Kelly 2006). During the preparation of cytoplasmic membrane vesicles, part of it may be lost, and hence, either cell-free extracts or permeabilized cells are better choices for cytochrome *c* peroxidase assays. However, *Z. mobilis* cells possess high catalase activity (Fig. 2b), which would interfere with the assay by rapidly removing the electron acceptor, $H_2O_2$. Therefore, a catalase mutant was constructed with the prime purpose of examining the activity of cytochrome *c* peroxidase in cell-free extracts. Although catalase activity was close to zero in the Zm6-*kat* strain (Fig. 2b), no cytochrome *c* peroxidase could be detected by the cytochrome *c* assay: $H_2O_2$-dependent oxidation of externally added cytochrome *c* did not occur in the cell-free extracts of the *kat* mutant (not shown).

In Fig. 5, the effect of millimolar $H_2O_2$ additions upon aerobic growth of the Zm6, Zm6-*cytB,* and Zm6-*kat* strains is presented. At the end of the exponential growth phase of shaken flask cultures, the cells were sedimented by centrifugation, resuspended in fresh culture medium (indicated by arrows) containing various concentrations of hydrogen peroxide, and aerobic cultivation was resumed. The results

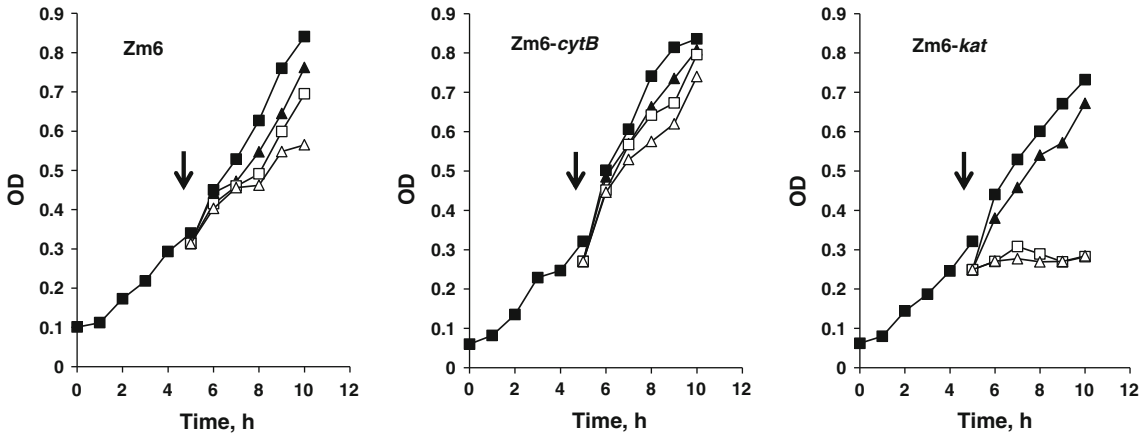



**Fig. 5** The effect of *kat* and *cytB* knockout mutations on aerobic batch growth in the presence of externally added hydrogen peroxide (*filled square*) 0 mM, (*filled triangle*) 0.5 mM, (*open square*) 1 mM, and (*open triangle*) 2 mM

show no significant difference between the sensitivities of Zm6 and Zm6-*cytB* to the added $H_2O_2$. However, Zm6-*kat* appeared to be much more vulnerable and ceased to grow in the presence of 1 and 2 mM $H_2O_2$. Monitoring of hydrogen peroxide decomposition was done for the cultures transferred to the medium with 2 mM $H_2O_2$. For that purpose, samples were taken at 5-min intervals, cells were rapidly sedimented, and $H_2O_2$ was determined in the supernatant using the Amplex UltraRed dye. In Zm6 and Zm6-*cytB*, during the 5 min after transfer of cells into the fresh medium, the concentration of $H_2O_2$ dropped from 2 mM to around 20 μM, while in the *kat* mutant, it remained 1.69 mM even after 30 min of cultivation (not shown).

Taken together, these data do not support the function of the cytochrome *c* peroxidase (ZMO 1136) gene product in the degradation of hydrogen peroxide and cast doubt on whether a functional cytochrome *c* peroxidase terminates the cytochrome $bc_1$ branch of electron transport in *Z. mobilis*. At the same time, transcription of the corresponding gene (ZMO 1136) did take place in all strains. Neither aeration nor mutations exerted significant transcriptional regulation over this gene, with one exception: In anaerobically cultivated *cydB* mutant cells, the gene encoding ZMO1136 (*perC*) was down-regulated relative to Zm6 (Table 3).

Catalase deficiency in the aerobically cultivated Zm6-*kat* cells did not give rise to any coordinated upregulation of the alternative peroxide-scavenging systems. Although *ahpC* (ZMO 1732) and the putative iron-dependent peroxidase (ZMO 1573) during the anaerobic-to-aerobic transition of Zm6-*kat* were up-regulated by factors of about 10 and 5, respectively (Table 3), neither of them exceeded the transcript levels found in aerobically cultivated parent strain. An almost fivefold higher level of superoxide dismutase and twofold higher level of thioredoxin (ZMO 1097) transcript were accompanied by a threefold down-regulation of glutathione reductase and by 25-fold down-regulation of glutaredoxin in the *kat* mutant relative to the aerobically growing Zm6.

Aerobic growth parameters of the Zm6-*kat* strain in a chemostat did not differ significantly from those of Zm6. Washed cells from an aerobic *kat* mutant culture, suspended in phosphate buffer with glucose, showed an even higher oxygen uptake rate than that of strain Zm6 (Fig. 2a), in good agreement with the >twofold aerobic upregulation of *ndh* (Table 3). Catalase deficiency did not result in any significant elevation of $H_2O_2$ production by the Zm6-*kat* strain. In all strains, the specific rate of hydrogen peroxide export from the cells was found to be in the narrow range of 4–7 nmol (mg dry wt)$^{-1}$ min$^{-1}$ and depended neither on the bulk oxygen consumption rate, nor on culture aeration (not shown).

## Discussion

Oxygen-dependent regulation of gene expression represents a rapidly evolving field of research in microbiology, focussed largely on in-depth studies of a few species of model microorganisms. The ethanol-fermenting bacterium *Z. mobilis* is not among them, and little is known about its response to aeration. However, recently a genome-scale transcriptomic and metabolomic study of aeration in wild-type *Z. mobilis* (strain ZM4) was reported by Yang et al. (2009b). The authors explored the effects of aeration on exponential and stationary-phase cultures during controlled batch cultivations. In agreement with our present findings, these authors reported aerobic downregulation of *cydA* and *cydB* and found no significant effect of aeration upon catalase, peroxidase, superoxide dismutase, and glutathione reductase gene transcription. Yet, in contrast to the report of Yang et al. (2009b), in Zm6, we did not observe aerobic upregulation of alkyl hydroperoxide reductase *ahpC* (ZMO 1732), thioredoxin *trx* (ZMO 1097), and the NADH:ubiquinone oxidoreductase subunit *rnfA* (ZMO 1814). Also, we did not see aerobic downregulation of *ndh* (ZMO 1113). In part, these discrepancies might be due to strain differences, but also different levels of aeration. For *Z. mobilis*, ethanol yield may serve as a rough estimate for culture aeration. As mentioned in 'Materials and methods', in our study, the ethanol yield of Zm6 aerobic chemostat culture was close to 60% of the theoretical maximum. Yang et al. (2009b) reported only 27% of the maximum ethanol yield, which points to more vigorous aeration in their aerobic batch culture. Finally, we compared continuously growing cultures at a moderate dilution rate (D 0.18 h$^{-1}$), while Yang et al. (2009b) mostly analyzed transcription differences that they found in stationary-phase cultures.

Our results support the general conclusion of Yang et al. (2009b) that in growing *Z. mobilis*; there is little transcriptional response to changes of aeration per se, at least concerning the respiratory chain and the ROS-protective systems. The regulation of the respiratory and oxidative stress-protective gene transcription in *Z. mobilis* is poorly understood, and probably differs from the generally established mechanisms for Gram-negative facultative anaerobes. Yang et al. (2009b) reports a strong aerobic upregulation of the alternative sigma factor *rpoH* (ZMO 0749), a MerR family regulator (ZMO 1121), and some other putative regulatory elements. At the same time, the *Z. mobilis* genome lacks homologs of the *oxyR* and *soxR* genes (Seo et al. 2005; Yang et al. 2009a).

Here we demonstrate that respiratory mutations affect the aerobic physiology of this bacterium, in some cases in an unexpected manner. For the NADH dehydrogenase mutant, showing an improved aerobic growth capacity (Kalnenieks et al. 2008), which permanently keeps its

respiration rate low under oxic growth conditions, several thiol-dependent oxidative stress-protective systems are strongly up-regulated in the presence of oxygen. Such a regulatory response to aeration has not been observed in other *Z. mobilis* strains with medium or high respiration rate under aerobic conditions. It is possible that the rapid, yet energetically inefficient respiratory chain in *Z. mobilis* helps to prevent oxidative stress in aerobically growing culture. Furthermore, a knockout mutation of the cytochrome *b* subunit of the $bc_1$ complex in *Z. mobilis* gives rise to a complex, non-respiring phenotype, when cultivated under anoxic conditions. Notably, the absence of cytochrome spectral features and catalase activity under these conditions do not result from downregulation of the corresponding genes. Further research is needed to see whether, perhaps, heme biosynthesis or assembly of heme-containing enzymes is affected in the mutant strain under oxygen-limiting conditions.

The ability of both aerobically cultivated mutants, Zm6-*cytB* and Zm6-*cydB*, to consume oxygen, as well as the difference seen in the antimycin sensitivity of their respiration, and in the kinetics of the cytochrome *d* reduction with NADH, indicates the presence of at least two branches of electron transport in the *Z. mobilis* respiratory chain. The effects of the *cydB* mutation clearly demonstrate that cytochrome *bd* is involved in electron transport. Yet, the apparent lack of genes for other terminal oxidases raises the intriguing problem of what could be the nature of the oxidase activity manifested in the Zm6-*cydB* strain. We speculate that the cytochrome *c* peroxidase gene product might have an alternative oxidase activity. Construction and study of a cytochrome *c* peroxidase/*cydB* double knockout strain could help to verify this hypothesis. We were not able to demonstrate any cytochrome *c* peroxidase activity, or any relation of the cytochrome $bc_1$ branch to the hydrogen peroxide resistance of the *Z. mobilis* cells. However, recently, Charoensuk et al. (2011) reported an electron transport chain-linked peroxidase activity in a thermotolerant *Z. mobilis* strain. We, therefore, conclude that the cytochrome *c* peroxidase activity might largely depend on the strain and culture conditions.

Obviously, catalase in *Z. mobilis* plays a minor role in the scavenging of endogenously generated $H_2O_2$ and is not critical for aerobic growth. In general, the role of bacterial catalases in endogenous $H_2O_2$ degradation varies. For example, in *Bradyrhizobium japonicum* catalase is the primary detoxifier of endogenously produced hydrogen peroxide (Panek and O'Brian 2004), and aerobic growth of the catalase-negative strain of this bacterium is severely impaired. In the pathogen *Staphylococcus aureus,* KatA and AhpC are mutually compensatory, and both enzymes are responsible for scavenging of endogenously produced $H_2O_2$ (Cosgrove et al. 2007). *Z. mobilis* seems to be more like *E. coli* in this respect. In *E. coli,* AhpC (Seaver and Imlay 2001) and possibly Fe-containing alcohol dehydrogenase (AdhE) (Echave et al. 2003) scavenge the majority of endogenous $H_2O_2$, with a small fraction degraded by catalase. As in *E. coli*, catalase in *Z. mobilis* protects the cells against exogenous hydrogen peroxide. Knowing that *Z. mobilis* is a non-pathogenic bacterium, found in sugar-rich tropical plant saps (Swings and DeLey 1977), it is likely that in the natural environment, its catalase degrades $H_2O_2$ of plant origin.

**Acknowledgments** This work was funded by grant 09.1306 of Latvian Council of Science, The Royal Society Travel Grant TG 102318 (for UK), and by Latvian ESF project 2009/0207/1DP/1.1.1.2.0/09/APIA/VIAA/128.

## References

Atack JM, Kelly DJ (2006) Structure, mechanism and physiological roles of bacterial cytochrome *c* peroxidases. Adv Microb Physiol 52:73–106

Belaich JP, Senez JC (1965) Influence of aeration and pantothenate on growth yields of *Zymomonas mobilis*. J Bacteriol 89:1195–1200

Bergmeyer HU, Gawehn K, Grassl M (1974) Glutathione reductase. In: Bergmeyer HU (ed) Methods of enzymatic analysis, vol 1. Academic Press, New York, pp 465–466

Bringer S, Finn RK, Sahm H (1984) Effect of oxygen on the metabolism of *Zymomonas mobilis*. Arch Microbiol 139:376–381

Charoensuk K, Irie A, Lertwattanasakul N, Sootsuwan K, Thanonkeo P, Yamada M (2011) Physiological importance of cytochrome *c* peroxidase in ethanologenic thermotolerant *Zymomonas mobilis*. J Mol Microbiol Biotechnol 20:70–82

Cosgrove K, Coutts G, Jonsson I-M, Tarkowski A, Kokai-Kun JF, Mond JJ, Foster SJ (2007) Catalase (KatA) and alkyl hydroperoxide reductase (AhpC) have compensatory roles in peroxide stress resistance and are required for survival, persistance, and nasal colonization in *Staphylococcus aureus*. J Bacteriol 189: 1025–1035

Echave P, Tamarit J, Cabiscol E, Ros J (2003) Novel antioxidant role of alcohol dehydrogenase E from *Escherichia coli*. J Biol Chem 278:30193–30198

Ellfolk N, Soininen R (1970) *Pseudomonas* cytochrome *c* peroxidase. Acta Chem Scand 24:2126–2136

Gonzalez-Flecha B, Demple B (1994) Intracellular generation of superoxide as a by-product of *Vibrio harveyi* luciferase expressed in *Escherichia coli*. J Bacteriol 176:2293–2299

Gonzalez-Flecha B, Demple B (1995) Metabolic sources of hydrogen peroxide in aerobically growing *Escherichia coli*. J Biol Chem 270:13681–13687

Goodhew CF, Wilson IBH, Hunter DJB, Pettigrew GW (1990) The cellular location and specificity of bacterial cytochrome *c* peroxidases. Biochem J 271:707–712

Kalnenieks U (2006) Physiology of *Zymomonas mobilis*: some unanswered questions. Adv Microb Physiol 51:73–117

Kalnenieks U, de Graaf AA, Bringer-Meyer S, Sahm H (1993) Oxidative phosphorylation in *Zymomonas mobilis*. Arch Microbiol 160:74–79

Kalnenieks U, Galinina N, Bringer-Meyer S, Poole RK (1998) Membrane D-lactate oxidase in *Zymomonas mobilis*: evidence for a branched respiratory chain. FEMS Microbiol Lett 168: 91–97

Kalnenieks U, Galinina N, Toma MM, Poole RK (2000) Cyanide inhibits respiration yet stimulates aerobic growth of *Zymomonas mobilis*. Microbiology 146:1259–1266

Kalnenieks U, Galinina N, Toma MM (2005) Physiological regulation of the properties of alcohol dehydrogenase II (ADH II) of *Zymomonas mobilis*: NADH renders ADH II resistant to cyanide and aeration. Arch Microbiol 183:450–455

Kalnenieks U, Galinina N, Toma MM, Pickford JL, Rutkis R, Poole RK (2006) Respiratory behaviour of a *Zymomonas mobilis adhB:kanr* mutant supports the hypothesis of two alcohol dehydrogenase isoenzymes catalysing opposite reactions. FEBS Lett 580: 5084–5088

Kalnenieks U, Galinina N, Strazdina I, Kravale Z, Pickford JL, Rutkis R, Poole RK (2008) NADH dehydrogenase deficiency results in low respiration rate and improved aerobic growth of *Zymomonas mobilis*. Microbiology 154:989–994

Kelly MJ, Poole RK, Yates MG, Kennedy C (1990) Cloning and mutagenesis of genes encoding the cytochrome *bd* terminal oxidase complex in *Azotobacter vinelandii*: mutants deficient in the cytochrome *d* complex are unable to fix nitrogen in air. J Bacteriol 172:6010–6019

Korshunov S, Imlay JA (2010) Two sources of endogenous hydrogen peroxide in *Escherichia coli*. Mol Microbiol 75:1389–1401

Kouvelis VN, Saunders E, Brettin TS, Bruce D, Detter C, Han C, Typas MA, Pappas KM (2009) Complete genome sequence of the ethanol producer *Zymomonas mobilis* NCIMB 11163. J Bacteriol 191:7140–7141

Liang C–C, Lee W-C (1998) Characteristics and transformation of *Zymomonas mobilis* with plasmid pKT230 by electroporation. Bioprocess Eng 19:81–85

Markwell MAK, Haas SM, Bieber LL, Talbert NE (1978) A modification of the Lowry procedure to simplify protein determination in membrane and lipoprotein samples. Anal Biochem 87:206–210

Osman YA, Conway T, Bonetti SJ, Ingram LO (1987) Glycolytic flux in *Zymomonas mobilis*: enzyme and metabolite levels during batch fermentation. J Bacteriol 169:3726–3736

Panek HR, O'Brian MR (2004) KatG is the primary detoxifier of hydrogen peroxide produced by aerobic metabolism in *Bradyrhizobium japonicum*. J Bacteriol 186:7874–7880

Poole RK, Cook GM (2000) Redundancy of aerobic respiratory chains in bacteria? Routes, reasons and regulation. Adv Microb Physiol 43:165–224

Reyes L, Scopes RK (1991) Membrane-associated ATPase from *Zymomonas mobilis*; purification and characterization. BBA 1068:174–178

Rogers PL, Lee KJ, Skotnicki ML, Tribe DE (1982) Ethanol production by *Zymomonas mobilis*. Adv Biochem Eng 23:37–84

Rogers PL, Jeon YJ, Lee KJ, Lawford HG (2007) *Zymomonas mobilis* for fuel ethanol and higher value products. Adv Biochem Eng Biotechnol 108:263–288

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor

Schmehl M, Jahn A, Meyer zu Vilsendorf A, Hennecke S, Masepohl B, Schuppler M, Marxer M, Oelze J, Klipp W (1993) Identification of a new class of nitrogen fixation genes in *Rhodobacter capsulatus*: a putative membrane complex involved in electron transport to nitrogenase. Mol Gen Genet 241:602–615

Seaver LC, Imlay JA (2001) Hydrogen peroxide fluxes and compartmentalization inside growing *Escherichia coli*. J Bacteriol 183: 7182–7189

Seo JS, Chong H, Park HS, Yoon KO, Jung C, Kim JJ, Hong JH, Kim H, Kim JH et al (2005) The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. Nat Biotechnol 23:63–68

Sootsuwan K, Lertwattanasakul N, Thanonkeo P, Matsushita K, Yamada M (2008) Analysis of the respiratory chain in ethanologenic *Zymomonas mobilis* with a cyanide-resistant *bd*-type ubiquinol oxidase as the only terminal oxidase and its possible physiological roles. J Mol Microbiol Biotechnol 14:163–175

Sprenger G (1996) Carbohydrate metabolism in *Zymomonas mobilis*: a catabolic highway with some scenic routes. FEMS Microbiol Lett 145:301–307

Strohdeicher M, Neuß B, Bringer-Meyer S, Sahm H (1990) Electron transport chain of *Zymomonas mobilis*. Interaction with the membrane-bound glucose dehydrogenase and identification of ubiquinone 10. Arch Microbiol 154:536–543

Swings J, DeLey J (1977) The biology of *Zymomonas*. Bacteriol Rev 41:1–46

Trumpower BL, Gennis RB (1994) Energy transduction by cytochrome complexes in mitochondrial and bacterial respiration: the enzymology of coupling electron transfer reactions to transmembrane proton translocation. Annu Rev Biochem 63:675–716

Viikari L (1986) By-product formation in ethanol fermentation by *Zymomonas mobilis*. Technical Research Centre of Finland. Publication 27

Viikari L, Berry DR (1988) Carbohydrate metabolism in *Zymomonas*. Crit Rev Biotechnol 7:237–261

Yang S, Pappas KM, Hauser LJ, Land ML, Chen G-L, Hurst GB, Pan C, Kouvelis VN, Typas MA, Pelletier DA, Klingeman DL, Chang YJ, Samatova NF, Brown SD (2009a) Improved genome annotation for *Zymomonas mobilis*. Nat Biotechnol 27:893–894

Yang S, Tschaplinski TJ, Engle NL, Carroll SL, Martin SL, Davison BH, Palumbo AV, Rodriguez M Jr, Brown SD (2009b) Transcriptomic and metabolomic profiling of *Zymomonas mobilis* during aerobic and anaerobic fermentations. BMC Genomics 10:34

BioInvasions Records (2013) Volume 2, Issue 1: 63–68
doi: http://dx.doi.org/10.3391/bir.2013.2.1.11


Open Access

**Short Communication**

# New records of the amphipods *Gammarus tigrinus* Sexton, 1939 and *Pontogammarus robustoides* G.O. Sars, 1894 in Latvian waters of the Baltic Sea

Evita Strode[1]*, Nadezhda A. Berezina[2], Martins Kalnins[3] and Maija Balode[1,4]

**1** *University of Latvia, Faculty of Biology, Kronvalda bulvaris 4, LV-1586 Riga, Latvia*
**2** *Zoological Institute of the Russian Academy of sciences, Universitetskaya 1, 199034 St. Petersburg, Russia*
**3** *The Entomological Society of Latvia, Kronvalda bulvaris 4, LV-1586 Riga, Latvia*
**4** *Department of Experimental Hydrobiology, Latvian Institute of Aquatic Ecology, Daugavgrivas str. 8, LV-1048, Riga, Latvia*

*E-mail: evita.strode@lhei.lv, nber@zin.ru, martins.kalnins@biology.lv , maija@hydro.edu.lv*

*****Corresponding author*

Received: 3 June 2012 / Accepted: 16 November 2012 / Published online: 21 November 2012

*Handling editor*: Vadim Panov

## Abstract

The paper presents recent records of the North-American amphipod *Gammarus tigrinus* and the Ponto-Caspian amphipod *Pontogammarus robustoides* in Latvian territorial waters of the Baltic Sea. *P. robustoides* was found for the first time in shallow waters of the Gulf of Riga near Jurmala City in June 2009. During field surveys of 2011–2012, *P. robustoides* and another invader, the North-American amphipod *G. tigrinus* (detected earlier in Estonian part of the Gulf), were recorded in different sites of the Latvian waters of the Gulf of Riga and in the coastal Lake Liepaja. Abundance ranged between 1–70 ind./m$^2$ for *P. robustoides* and 5–100 ind./m$^2$ for *G. tigrinus,* suggesting they potential role in the functioning of the benthic community in near future.

**Key words:** gammarid amphipods; invasive species; distribution; establishment; abundance; Gulf of Riga; coastal

## Introduction

Lately, the number of non-indigenous invertebrates has increased worldwide, resulting in fauna mixing and structural changes of aquatic communities (Leppäkoski et al. 2002). The Gulf of Riga is the trans-border area between Latvia and Estonia. The southern area (i.e., near Riga City and its suburbs) has environmental issues due to intensive and long-term development of major holiday resorts. The main environmental problems in the Gulf of Riga caused by humans include eutrophication, harmful algal blooms, toxic pollution and presence of invasive species. The substantial maritime traffic (both cargo ships and leisure boats), exchanging ballast water in the sea and fish stocking are important sources of transfer of non-native species to this region.

In the last few decades, at least 14 non-native macrobenthic invertebrate species have been recorded in the Gulf of Riga, including: the hydroid *Cordylophora caspia* Pallas, 1771; the polychaetes *Marenzelleria viridis* Verrill, 1873 and *M. neglecta* Sikorski and Bick, 2004; the molluscs *Mya arenaria* Linnaeus, 1758, *Dreissena polymorpha* Pallas, 1771, and *Potamopyrgus antipodarum* J.E. Gray, 1843; a barnacle *Balanus improvisus* Darwin, 1854; the crabs *Eriocheir sinensis* H. Milne Edwards, 1853 and *Rhithropanopeus harrisi* Gould, 1841; a shrimp *Palaemon elegans* Rathke, 1837, the mysids *Paramysis intermedia* Czerniavsky, 1882 and *Hemimysis anomala* G. O. Sars, 1907; and the amphipods *Pontogammarus robustoides* G.O. Sars, 1894 and *Gammarus tigrinus* Sexton, 1939 (Ojaveer 1995; Leppäkoski et al. 2002; Herkül and Kotta 2007; Blank et al. 2008; Herkül et al.2009; Kotta and Kotta, 2010; Kalinkina and Berezina, 2010; Kotta et al. 2010, 2011; Kotta and Kuprijanov 2012).

The goal of this study was to document the presence of alien amphipods in Latvian waters of the Gulf of Riga and to estimate their abundance and current distribution. We predicted the presence in the communities of this Gulf of at

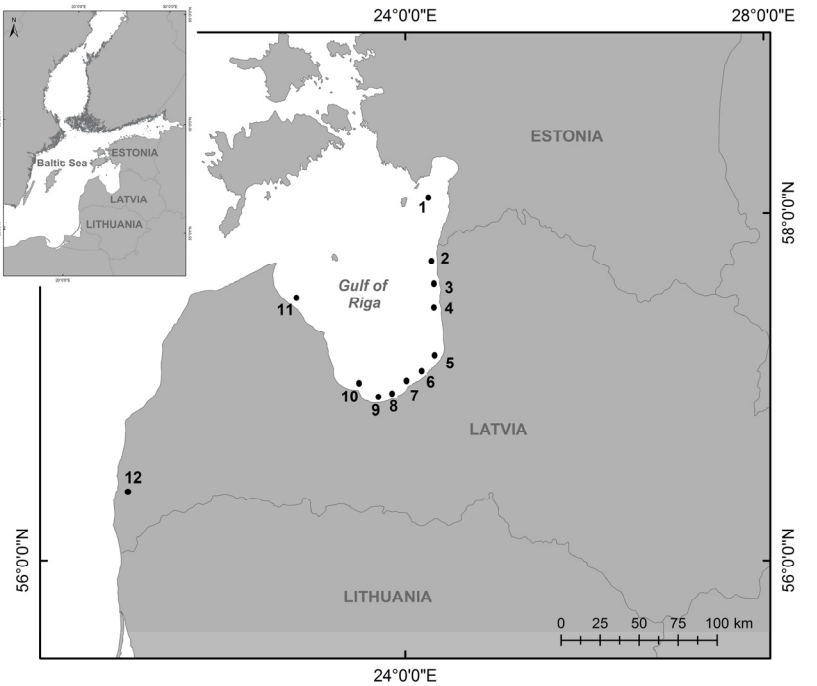



**Figure 1.** Map of the Gulf of Riga with sites of amphipod records in 2011-2012. 1 - Pärnu Bay, 2 - Ainaži, Randu meadows, 3 - Salacgrīva, 4 - Vitrupe, 5 - Lilaste, 6 - Garciems, 7 - Vecāķi, 8 - Bolderāja, 9 – Jurmala, 10 - Ragciems, 11 - Kaltene, 12 - Lake Liepaja.

least two amphipods (*P. robustoides, G. tigrinus*) that were discovered recently in the Gulf of Riga (Berezina and Kalinkina 2010; Herkül and Kotta 2007). Surveys of the bottom communities in the south-eastern part of the Gulf of Riga and the Latvian coast of the Baltic Sea were started during the Latvian-Russian collaborative program BEAST (project BONUS+Programme 2009-2011) aimed at determining the most sensitive indicators among amphipods for the control of hazardous substances in the Baltic Sea. In 2012, sampling was conducted along whole Latvian coast of the Gulf of Riga in the frame of the European Social Fund project HYDROTOX.

## Study area and sampling method

The Gulf of Riga (18 100 km$^2$) is a large shallow bay (maximum depth of 62 m) in the eastern part of the Baltic Sea. During spring, surface-water salinity varies between 0.5 and 2.0 PSU while bottom water salinity (i.e., in the Irbe Strait) reaches up to 7.7 PSU (Ojaveer 1995). The relatively low salinity of the Gulf of Riga is the result of high freshwater inflow from the rivers Daugava, Lielupe, Gauja, and Salaca.

Zoobenthos were sampled during field survey several times in a year, during summer, autumn and winter of 2011 and 2012 mostly from the littoral area (0-1 m) at 11 sites of the Gulf of Riga and one site at the Baltic Sea coast, from deep water stations (10 m depth) of the Pärnu Bay as well as from the Lake Liepaja (Figure 1). All sampling dates, coordinates and names of mentioned sites are illustrated in the Table 1 and 2. To record new sites of amphipod species distribution, macrobenthic organisms were collected by means of a Mini-Surber Sampler (500 µm mesh).

The samples were conserved by 4% formaldehyde and transported to the laboratory. Amphipods were identified and their abundance was expressed as number/m$^2$.

## Results and discussion

During the field survey of 2011-2012, the presence of amphipods *Pontogammarus robustoides* and *Gammarus tigrinus* was confirmed in several new locations of the Gulf of Riga (Table 1 and 2). In 2011 both species were found in the shallowest areas of the Gulf, where earlier

**Table 1.** New record of *Pontogammarus robustoides* of the Gulf of Riga.

| Site No. | Location (site name) | Record coordinates | | Record date | Number collected (ind./m$^2$) |
|---|---|---|---|---|---|
| | | Latitude, °N | Longitude, °E | | |
| 2 | Ainaži, Randu | 57°49'33" | 24°19'58" | 01.10.2011 - 04.06.2012 | 5-10 |
| 3 | Salacgrīva | 57°44'12" | 24°20'38" | 04.06.2012 | 8 |
| 4 | Vitrupe | 57°38'34" | 24°22'08" | 04.06.2012 | 11 |
| 5 | Lilaste | 57°11'38" | 24°20'01" | 04.06.2012 | 10 |
| 6 | Garciems | 57°06'55" | 24°11'25" | 04.06.2012 | 19 |
| 7 | Vecāķi | 57°04'54" | 24°06'21" | 05-09.2011-04.06.2012 | 100-6 |
| 8 | Bolderāja | 57°01'14" | 23°57'55" | 04.06.2012 | 1 |
| 9 | Jurmala | 56°57'54" | 23°33'36" | 15.06.2011 - 04.06.2012 | 10-6 |
| 11 | Kaltene | 57°27'34" | 22°53'09" | 04.06.2012 | 10 |

**Table 2**. Records of *Gammarus tigrinus* in Latvian waters of the Baltic Sea and the Gulf of Riga.

| Site No. | Location | Record coordinates | | Record date | Number collected (ind./m$^2$) |
|---|---|---|---|---|---|
| | | Latitude,°N | Longitude,°E | | |
| 1 | Pärnu Bay | 58°07'51" | 24°10'58" | 14.01.2012 | 40 |
| | | | | 07.08.2012 | 70 |
| 2 | Ainaži, Randu | 57°49'33" | 24°19'58" | 01.10.2011 | 2 |
| 9 | Jurmala | 56°58'13" | 23°45'13" | 15.06.2011 | 5 |
| 10 | Ragciems | 57°01'44" | 23°29'47" | 04.06.2012 | 1 |
| 12 | Lake Liepaja | 56°30'56" | 21°02'39" | 07.06.2011 | 50 |

(before 2000s) the Baltic *Gammarus* species were commonly recorded. From eighteen observed sites in 2012 in the littoral areas, *P. robustoides* and *G. tigrinus* were found at more than 50% places (Tables 1, 2). During January and August 2012, *G. tigrinus* was also found in deep water stations of the Pärnu Bay (in 10 m depth) where the Baltic corophiid *Corophium volutator* was detected as well.

*Pontogammarus robustoides* (Figure 2) has been recorded in 1999-2005 as an established component in the lower reaches or mouths of Latvian rivers flowing into the Baltic Sea (Grudule et al. 2007)**.** In June 2009, it was found in the Gulf of Riga near Jurmala City in sandy littoral habitats (Kalinkina and Berezina 2010). It was the first recorded report of *P. robustoides* in Latvian coastal areas of the Gulf. There is an unpublished record of this species from the same location from July 2000 (J. Ironside, personal communication). This suggests that *P. robustoides* has been present in the Jurmala district since at least 2000.

The original distribution of this Ponto-Caspian amphipod includes: the brackish and freshwater bays of the Black Sea, the Azov Sea and the Caspian Sea; coastal lakes; lagoons, lower courses and estuaries of major rivers (e.g., Volga, Don, Bug, Dnepr, Dniester, Danube, Prut, Terek, Kura, Kuban, etc.); as well as the lakes near Marmara Sea (Kalinkina and Berezina 2010). The invasion history of this species in the southern Baltic Sea basin begins in 1960 when it was introduced from the Black Sea basin to the Kaunas Reservoir (Neman River) and subsequently further north in Lithuania, Latvia and Russia (Leningrad province). By the end of 20th century, *P. robustoides* had spread in the estuaries and along the course of several large rivers (e.g., Vistula, Oder/Odra, Neman/ Nemunas), and also it penetrated several lakes and reservoirs within their basins (Rudolph 1997, Zettler 1998, Arbačiauskas 2002, 2008, Jażdżewski and Konopacka 2002; Wawrzyniak and Gruszka 2005; Ezhova et al. 2005). During the last few decades it was recorded in the eastern part of the Gulf of Finland, specifically in the estuary of the River Neva (Berezina and Panov 2003) and along Estonian coast in Narva Bay (1999) (Herkül et al. 2009).

*Gammarus tigrinus* (Figure 3) is a species native to the Atlantic seaboard of North America and introduced into Northern Ireland (known from 1931) presumably transported by ballast waters (Hynes 1955). It was firstly found in Great Britain in 1939 (Sexton and Cooper 1939).


**Figure 2.** General view of the amphipod *Pontogammarus robustoides* Sars, 1894, collected from the Vecāķi, Gulf of Riga. Photograph by E. Strode.

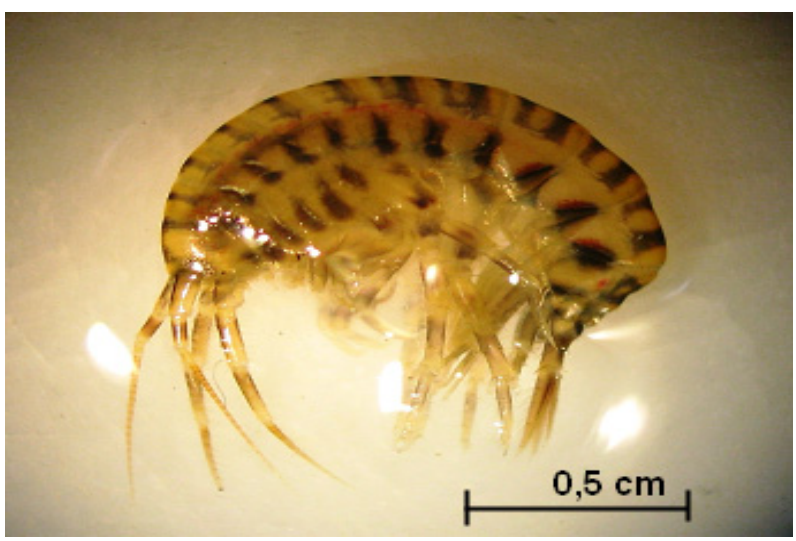



**Figure 3.** General view of the amphipod *Gammarus tigrinus* Sexton, 1939, from the Gulf of Riga near Jurmala. Photograph by E. Strode.

*G. tigrinus* became wide-spread in continental European waters after intentional introductions (as food for fish) from England to the River Werra, Germany, in 1957 (Schmitz 1960) and from Northern Ireland to the Lake Ijsselmeer, The Netherlands, in 1960 (Nijssen and Stock 1966).

Since its first discovery in 2003, *G. tigrinus* has been recorded frequently in the Estonian coastal areas of the northern Gulf of Riga (Kõiguste Bay) (Herkül and Kotta 2007). Until now, *G. tigrinus* had not been reported in the Latvian part of the Gulf of Riga when it was recorded as an abundant species in several sites of the Gulf of Riga and Lake Liepaja (Table 2).

In 1975, the species was found in the Baltic Sea, Schlei Fjord (Bulnheim 1976). In 1992 it was recorded in the Peenestrom and the Achterwasser - the River Odra estuary (Rudolph 1994). However, Wydrowska-Wawrzyniak and Gruszka (2005) showed that *G. tigrinus* was already present in the estuary - in the Szczecin Lagoon in 1988. Soon, it spread along the entire Baltic Sea shore of north-eastern Germany (Zettler 2001) and north-western Poland (Gruszka 1999; Jażdżewski and Konopacka 2000). During the next decade, the species reached Puck Bay (Szaniawska et al. 2003) and the Vistula Lagoon (Jażdżewski et al. 2002; Ezhova et al. 2005). At the Finnish coast of the Gulf of Finland, this gammarid was found in 2003 (Pienimäki et al. 2004) and later, in 2005, it was recorded in the eastern-most part of the Baltic Sea (the River Neva estuary, Berezina 2007).

Abundance of amphipods *P. robustoides* and *G. tigrinus* in our sampling sites was relatively high (Tables 1 and 2). It varied between 1-70 ind./$m^2$ in the case of *P. robustoides* and 5-100 ind./$m^2$ in the case of *G. tigrinus*, which suggests they have become an important component of the benthic community of the Gulf of Riga. Both species are key components of benthic communities in freshwater and brackish water areas where they feed upon small planktonic and benthic invertebrates (Arbačiauskas and Gumuliauskaite 2007; Berezina et al. 2011). The selective predation of invasive amphipods on native invertebrates is considered to be the main mechanism explaining the replacement of previously co-existing species and shifts in density of native invertebrates (e.g., Dick et al. 2002). As a rule, their predation pressure depends on abundance and availability of prey in particular habitats and can change seasonally. Further investigations of the distribution, reproductive success and ecological significance of these two non-native species in the Gulf of Riga are needed in order to estimate quantitatively their role in the structure and functioning of the benthic communities in Latvian waters.

## Acknowledgements

We cordially thank Dr. Joseph Ironside from Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University (UK) for communication about amphipods record in 2000, and two reviewers, Dr. Piotr Gruszka and Dr. Jonne Kotta, for valuable comments and important corrections. This research was support partially by funding from the European Community's Seventh Framework Programme (2007-2013) under grant agreement n°217246 made with the joint Baltic Sea Research and Development Programme BONUS+ (BEAST project) as well as from the European Social Fund project HYDROTOX, contract No. 2009/0226/1DP/1.1.1.2.0/09/APIA/ VIAA/080 (2010-2012). The authors thank cordially the coordinator of the HYDROTOX project Dr. Ingrida Purina for support of investigation, Mrs. Mara Newsome Satina and Dr. John Mark Hanson for English correction.

## References

Arbačiauskas K (2002) Ponto-Caspian amphipods and mysids in the inland waters of Lithuania: history of introduction, current distribution and relations with native malacostracans. In: Leppäkoski E, Gollasch S, Olenin S (eds), Invasive Aquatic Species of Europe. Distribution, Impacts and Management, Kluwer Academic Publishers, Dordrecht, Boston, London, pp 104–115

Arbačiauskas K (2008) Amphipods of the Nemunas River and Curonian Lagoon, the Baltic Sea basin: where and which native freshwater amphipods persist? *Acta Zoologica Lituanica* 18: 10–16, http://dx.doi.org/10.2478/v10043-008-0002-3

Arbačiauskas K, Gumuliauskaite S (2007) Invasion of the Baltic Sea basin by the Ponto-Caspian amphipod *Pontogammarus robustoides* and its ecological impact. In: Gherardi F (ed), Biological invaders in inland waters: profiles, distribution, and threats. Springer, pp 463–477

Berezina NA (2007) Expansion of the North American amphipod *Gammarus tigrinus* Sexton, 1939 to the Neva Estuary (easternmost Baltic Sea). *Oceanologia* 49(1): 129–135

Berezina NA, Panov VE (2003) Establishment of new gammarid species in the eastern Gulf of Finland (Baltic Sea) and their effects on littoral communities. Proceedings of the Estonian Academy of Sciences. Biology. *Ecology* 52(3): 284–304

Berezina NA, Petryashev VV, Razinkovas A, Lesutiene J (2011) Alien malacostraca in the eastern Baltic Sea: pathways and consequences. In: Galil BS, Clark PF, Carlton JT (eds), In the Wrong Place - Alien Marine Crustaceans: Distribution, Biology and Impacts. Invading Nature - Springer Series in Invasion Ecology, 6, pp 301–322

Blank M, Laine AO, Jürss K, Bastrop R (2008) Molecular identification key based on PCR/RFLP for three polychaete sibling species of the genus *Marenzelleria*, and the species' current distribution in the Baltic Sea. *Helgoland Marine Research* 62: 129–141, http://dx.doi.org/10.1007/s10152-007-0081-8

Bulnheim HP (1976) *Gammarus tigrinus*, ein neues Faunenelement der Ostseeförde Schlei. *Schriften der Naturwissenschaftlichen Vereins für Schleswig-Holstein* 46: 79–84

Dick JTA, Platvoet D, Kelly DW (2002) Predatory impact of the freshwater invader *Dikerogammarus villosus* (Crustacea: Amphipoda). *Canadian Journal of Fisheries and Aquatic Sciences* 59: 1078–1084, http://dx.doi.org/10.1139/f02-074

Ezhova E, Żmudzinski L, Maciejewska K (2005) Long-term trends in the macrozoobentos of the Vistula Lagoon, southeastern Baltic Sea: species composition and biomass distribution. *The bulletin of the Sea Fisheries Institute* 164(11): 55–73

Grudule N, Parele E, Arbačiauskas K (2007) Distribution of ponto-caspian amphipod *Pontogammarus robustoides* in Latvian waters. *Acta Zoologica Lituanica* 17(1): 28–32

Gruszka P (1999) The River Odra Estuary as a Gateway for alien Species Immigration to the Baltic Sea basin. *Acta hydrochimica et hydrobiologica* 27: 374–382, http://dx.doi.org/10.1002/(SICI)1521-401X(199911)27:5<374::AID-AHEH374>3.0.CO;2-V

Herkül K, Kotta J (2007) New records of the amphipods *Chelicorophium curvispinum*, *Gammarus tigrinus*, *G. duebeni* and *G. lacustris* in the Estonian coastal Sea. *Proceedings of the Estonian Academy of Sciences: Biology, Ecology* 56: 290–296

Herkül K, Kotta J, Püss T, Kotta I (2009) Crustacean invasions in the Estonian coastal sea. *Estonian Journal of Ecology*. 58: 313–323, http://dx.doi.org/10.3176/eco.2009.4.06

Hynes HBN (1955) Distribution of some freshwater Amphipoda in Britain, International Association of Theoretical and Applied Limnology12, pp 620–628

Jażdżewski K, Konopacka A (2000) Immigration history and present distribution of alien crustaceans in Polish waters. In: von Vaupel Klein JC, Schram FR (eds), Proceedings of the 4th International Crustacean Congress: the biodiversity crisis and Crustacea. Crustacean Issues 12(2), Brill, Leiden, pp 55–64

Jażdżewski K, Konopacka A (2002) Invasive Ponto-Caspian species in waters of the Vistula and Oder basins and the southern Baltic Sea. In: Leppäkoski E, Gollasch S, Olenin S (eds), Invasive Aquatic Species of Europe. Distribution, Impacts and Management. Kluwer Academic Publishers, Dordrecht, Boston, London, pp 384–398

Jażdżewski K, Konopacka A, Grabowski M (2002) Four Ponto-Caspian and one American gammarid species (Crustacea: Amphipoda) recently invading Polish waters. *Contributions to Zoology* 71(4): 115–122

Kalinkina NM, Berezina NA (2010) First record of *Pontogammarus robustoides* Sars, 1894 (Crustacea: Amphipoda) in the Gulf of Riga (Baltic Sea). *Aquatic Invasions* 5 (Suppl. 1): S5–S7, http://dx.doi.org/10.3391/ai.2010.5.S1.002

Kotta J, Kuprijanov I (2012) The first finding of the palaemonid shrimp *Palaemon elegans* Rathke in the Estonian coastal sea. *Estonian Journal of Ecology* 61: 148–153, http://dx.doi.org/10.3176/eco.2012.2.06

Kotta J, Orav-Kotta H, Herkül K, Kotta I (2011) Habitat choice of the invasive *Gammarus tigrinus* and the native *Gammarus salinus* indicates weak interspecific competition. *Boreal Environmental Research* 16 A: 64–72

Kotta J, Kotta I (2010) The first finding of the Ponto-Caspian mysid shrimp *Hemimysis anomala* G. O. Sars, Mysidae in the Estonian coastal sea. *Estonian Journal of Ecology* 59: 230–236

Kotta J, Orav-Kotta H, Herkül K (2010) Separate and combined effects of habitatspecific fish predation on the survival of invasive and native gammarids. *Journal of Sea Research* 64: 369–372, http://dx.doi.org/10.1016/j.seares.2010.05.006

Leppäkoski E, Gollasch S, Gruszka P, Ojaveer H, Olenin S, Panov V (2002) The Baltic – a sea of invaders. *Canadian Journal of Fisheries and Aquatic Sciences* 59: 1175–1188, http://dx.doi.org/10.1139/f02-089

Nijssen H, Stock JH (1966) The amphipod, *Gammarus tigrinus* Sexton, 1939, introduced in the Netherlands (Crustacea). *Beaufortia* 13: 197–206

Ojaveer E (1995) Large-scale processes in the ecosystem of the Gulf of Riga. In: Ojaveer E (ed), Ecosystem of the Gulf of Riga between 1920 and 1990. Estonian Academy, Tallinn, pp 268–277

Pienimäki M, Helavuori M, Leppäkoski E (2004) First findings of the North American amphipod *Gammarus tigrinus* Sexton, 1939 along the Finnish coast. *Memoranda Societatis pro Fauna et Flora Fennica* 80: 17–19

Rudolph K (1994) Erstnachweis des Amphipoden *Gammarus tigrinus* Sexton, 1939 (Crustacea: Gammaridea) im Peenestrom und Achterwasser (südliche Ostseeküste). *Naturschutzarbeit in Mecklenburg-Vorpommern* 37(2): 23–29

Rudolph K (1997) Zum Vorkommen des Flohkrebses *Pontogammarus robustoides* im Peenemündungsgebiet. *Natur und Museum* 127(9): 306–312

Schmitz W (1960) Die Einbürgerung von *Gammarus tigrinus* Sexton auf dem europäischen Kontinent. *Archiv für Hydrobiologie* 57: 223–225

Sexton EW, Cooper LHN (1939) On a new species of *Gammarus* (*G. tigrinus*) from the Droitwich District. *Journal of the Marine Biological Association of the United Kingdom* 23: 543–551, http://dx.doi.org/10.1017/S0025315400014065

Szaniawska A, Łapucki T, Normant M (2003) The invasive amphipod *Gammarus tigrinus* Sexton, 1939, in Puck Bay. *Oceanologia* 45: 507–510

Wawrzyniak-Wydrowska B, Gruszka P (2005) Population dynamics of alien gammarid species in the River Odra estuary. *Hydrobiologia* 539(1): 13–25, http://dx.doi.org/10.1007/s10750-004-3081-6

Zettler ML (1998) Zur Verbreitung der Malacostraca (Crustacea) in den Binnen- und Küsten- gewässern von Mecklenburg-Vorpommern. *Lauterbornia* 32: 49–65

Zettler ML (2001) Some malacostracan crustacean assemblages in the southern and western Baltic Sea. *Rostocker Meeresbiologische Beiträge* 9: 127–143

doi:10.1016/j.jmb.2011.10.042 J. Mol. Biol. (2012) 415, 118–127

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

# Different Binding Modes of Free and Carrier-Protein-Coupled Nicotine in a Human Monoclonal Antibody

Kaspars Tars [1]*, Svetlana Kotelovica [1], Gerd Lipowsky [2], Monika Bauer [2], Roger R. Beerli [2], Martin F. Bachmann [2] and Patrik Maurer [2]

[1]*Biomedical Research and Study Center, Ratsupites 1, Riga LV 1067, Latvia*
[2]*Cytos Biotechnology AG, CH-8952 Schlieren, Switzerland*

*Received 5 July 2011;*
*received in revised form*
*20 October 2011;*
*accepted 25 October 2011*
Available online
3 November 2011

***Edited by I. Wilson***

***Keywords:***
nicotine;
Fab fragment;
immune response;
hapten;
anti-smoking vaccine

Nicotine is the principal addictive component of tobacco. Blocking its passage from the lung to the brain with nicotine-specific antibodies is a promising approach for the treatment of smoking addiction. We have determined the crystal structure of nicotine bound to the Fab fragment of a fully human monoclonal antibody (mAb) at 1.85 Å resolution. Nicotine is almost completely (>99%) buried in the interface between the variable domains of heavy and light chains. The high affinity of the mAb is the result of a charge–charge interaction, a hydrogen bond, and several hydrophobic contacts. Additionally, similarly to nicotinic acetylcholine receptors in the brain, two cation–π interactions are present between the pyrrolidine charge and nearby aromatic side chains. The selectivity of the mAb for nicotine *versus* cotinine, which is the major metabolite of nicotine and differs in only one oxygen atom, is caused by steric constraints in the binding site. The mAb was isolated from B cells of an individual immunized with a nicotine–carrier protein conjugate vaccine. Surprisingly, the nicotine was bound to the Fab fragment in an orientation that was not compatible with binding to the nicotine–carrier protein conjugate. The structure of the Fab fragment in complex with the nicotine–linker derivative that was used for the production of the conjugate vaccine revealed a similar position of the pyridine ring of the nicotine moiety, but the pyrrolidine ring was rotated by about 180°. This allowed the linker part to reach to the Fab surface while high-affinity interactions with the nicotine moiety were maintained.



*Corresponding author. E-mail address: kaspars@biomed.lu.lv.

Present addresses: G. Lipowsky, GlycoVaxyn AG, 8952 Schlieren, Switzerland; M. Bauer, Intercell AG, CH-8952 Schlieren, Switzerland; R. R. Beerli, Intercell AG, CH-8952 Schlieren, Switzerland; P. Maurer, ESBATech LLC, CH-8952 Schlieren, Switzerland.

Abbreviations used: mAb, monoclonal antibody; VLP, virus-like particle; NLD, nicotine–linker derivative; CDR, complementarity-determining region; AChBP, acetylcholine binding protein; PEG, polyethylene glycol; ESRF, European Synchrotron Radiation Facility; PDB, Protein Data Bank.

## Introduction

Tobacco is the most widely used addictive drug in the world and leads to an estimated 5.4 million premature deaths annually caused by various forms of cancer, cardiovascular diseases, and other disorders.[1] The principal addictive compound in tobacco is nicotine. When tobacco smoke is inhaled, nicotine is distributed from the lungs into the bloodstream and from there it crosses the blood–brain barrier within a few seconds. In the brain, nicotine binds to nicotinic acetylcholine receptors

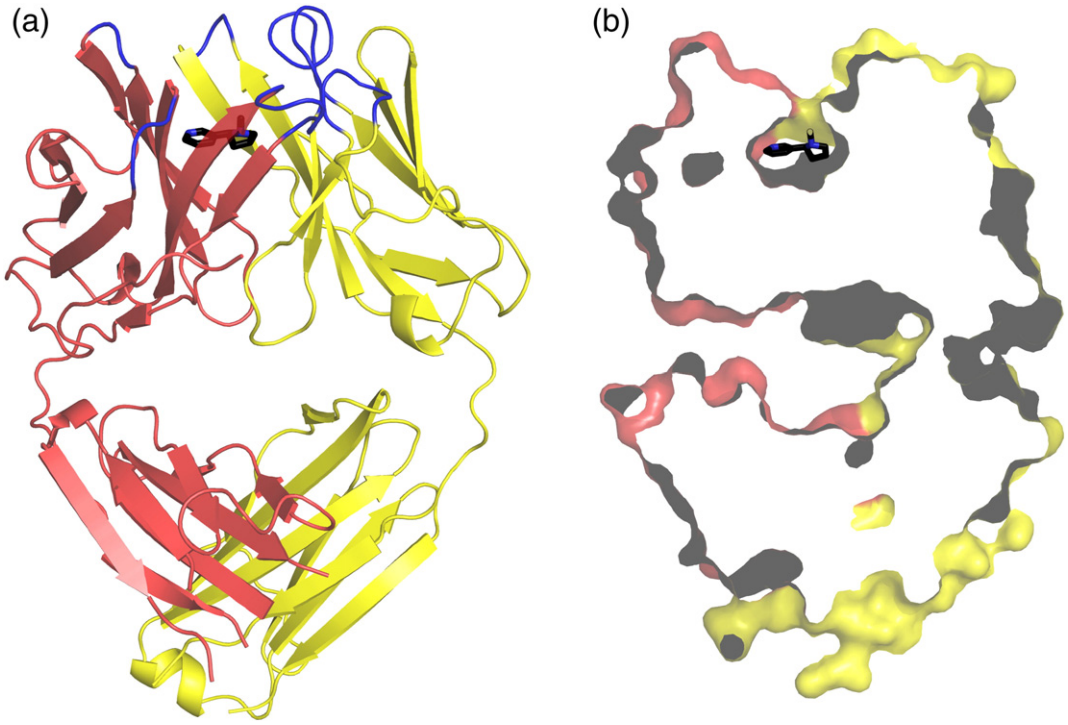



**Fig. 1.** (a) Overall structure of the nicotine molecule (black carbon atoms) in complex with the Fab fragment. The heavy chain is in red, the light chain is in yellow, and CDR loops are in blue. (b) Cross section of the accessible surface of the Fab fragment, displaying nicotine almost entirely buried in the protein. This figure and Figs. 2, 3, and 5 were prepared by PyMOL.[13]

and causes increased levels of dopamine in the mesolimbic reward system. This activation is key to the immediate rewarding signal of nicotine and the initiation and maintenance of tobacco addiction.[2] Deliberate smoking cessation attempts frequently fail due to the immense addictive power of nicotine. Even a single lapse may reinstate regular smoking behavior.[3]

One solution to the problem would be to prevent nicotine from entering the brain and binding to nicotinic receptors. During recent years, several nicotine vaccine candidates have entered clinical trials.[4] They are supposed to induce antibodies that bind nicotine in the blood, preventing it from crossing the blood–brain barrier. Consequently, there would be no rewarding effect for the smoking individual.

Since nicotine as a hapten does not raise an immune response by itself, coupling to a protein carrier is required for the induction of anti-nicotine antibodies. The nicotine vaccines tested in clinical trials differed in the way nicotine was coupled to the carrier protein as well as in the nature of the carrier protein.[5–8]

Two double-blind, placebo-controlled phase II clinical trials, one with the vaccine NIC002 based on virus-like particles (VLPs) derived from bacteriophage Qβ as carrier protein[9] and the other with the vaccine NicVAX based on exoprotein A from *Pseudomonas aeruginosa* as carrier protein,[10] demonstrated that vaccination against nicotine can significantly increase abstinence rates in smokers willing to quit smoking. Significantly, in both studies, increased abstinence rates correlated with the levels of anti-nicotine antibodies.

As an alternative to active immunization of individuals with a nicotine vaccine, nicotine-specific monoclonal antibodies (mAbs) could be used for passive immunization, as demonstrated by Pentel *et al.* in animal models.[11] Although considerably more expensive, the passive immunization approach has the following advantages: the effect is immediate and administering a precise dose of a preselected high-affinity antibody is possible. The latter is particularly important, since hapten-specific antibodies often suffer from low affinity. The three-dimensional structure of an antibody–nicotine complex might allow modifying the antibodies in order to increase their affinity to nicotine or perform other rational structure-based modifications. In the current study, we describe the structure of a NIC002-induced fully human anti-nicotine antibody Fab fragment in complex with nicotine at 1.85 Å resolution and, in addition, in complex with the nicotine–linker derivative [NLD; nicotine-11-yl-methyl-(4-ethylamino-4-oxo)-butanoate], which was used for the conjugation of nicotine to the Qβ VLP in NIC002, at 2.1 Å resolution.

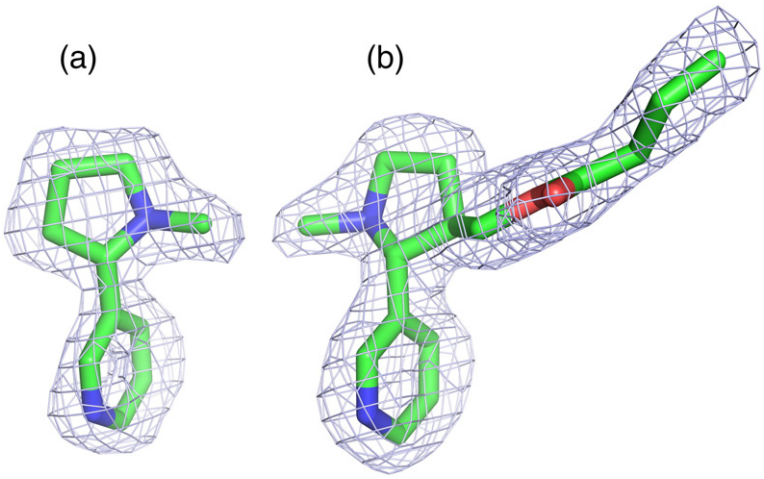



**Fig. 2.** (a) Electron density of the nicotine molecule. Features of both rings, such as the presence of the methyl group, can be clearly distinguished. (b) Electron density of NLD. $2F_o - F_c$ electron density maps were calculated in the absence of ligands and contoured at 2 σ. Orientation of (a) and (b) is the same relative to the protein.

## Results

### Structure of nicotine complex

The Fab fragment of mAb-Nic12[12] in complex with nicotine shows the typical Fab structure consisting of two constant and two variable domains from heavy and light chains (Fig. 1). It was possible to trace most of the polypeptide chains, except several loops in both constant domains and several N- and C-terminal residues. Electron density was good for the rest of the structure, including all complementarity-determining region (CDR) loops. Electron density for nicotine in the structure obtained by co-crystallization with the naturally occurring (*S*)-(−)-nicotine (Fig. 2) allowed unambiguous modeling of pyridine and pyrrolidine rings and the methyl group. Pyrrolidine and pyridine rings are in a nearly perpendicular orientation, as determined for nicotine in solution.[14]

The nicotine molecule is almost entirely enclosed by protein side chains and only the methyl group is, to some extent, exposed on the surface (accessible surface less than 1 $Å^2$ as determined by the CCP4 program AREAIMOL;[15] Fig. 1b). As quite typical for haptens, nicotine is bound in a cavity between heavy and light chains rather than being bound on the surface of the variable domains as typical for proteins (Fig. 1). The nicotine molecule has both polar and hydrophobic contacts with the Fab (Fig. 3a). The tertiary nitrogen atom of the pyrrolidine ring is protonated at physiological pH.[14,16] In the Fab structure, the tertiary ammonium group is making a charged interaction with the carboxyl group of Glu H101, which is further coordinated with a hydrogen bond to Tyr L36. Ser H50 forms a hydrogen bond with the N atom of the pyridine ring of nicotine. This

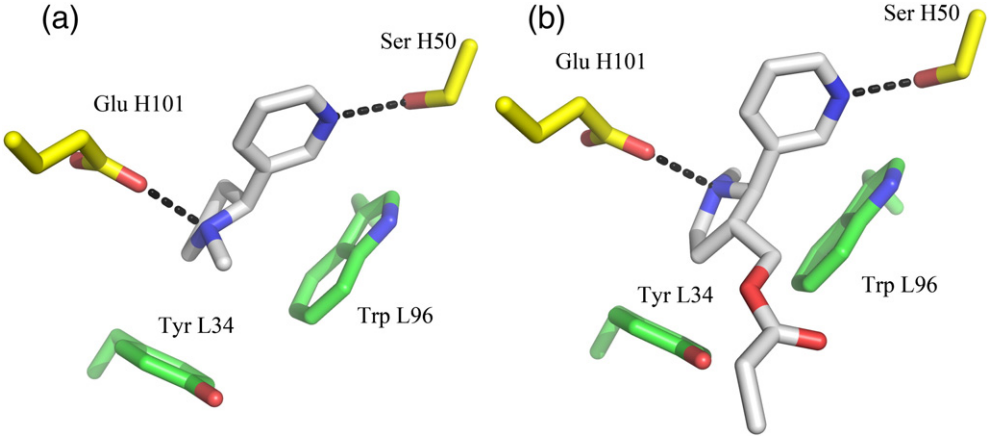



**Fig. 3.** Key interactions between the Fab fragment and (a) nicotine and (b) NLD. The specific salt bridge and the hydrogen bond are shown with black broken lines. Trp L96 makes important hydrophobic contact with both nicotine rings. Both Trp L96 and Tyr L34 are involved in cation–π interactions with the charged pyrrolidine nitrogen. Further hydrophobic interactions are not shown.

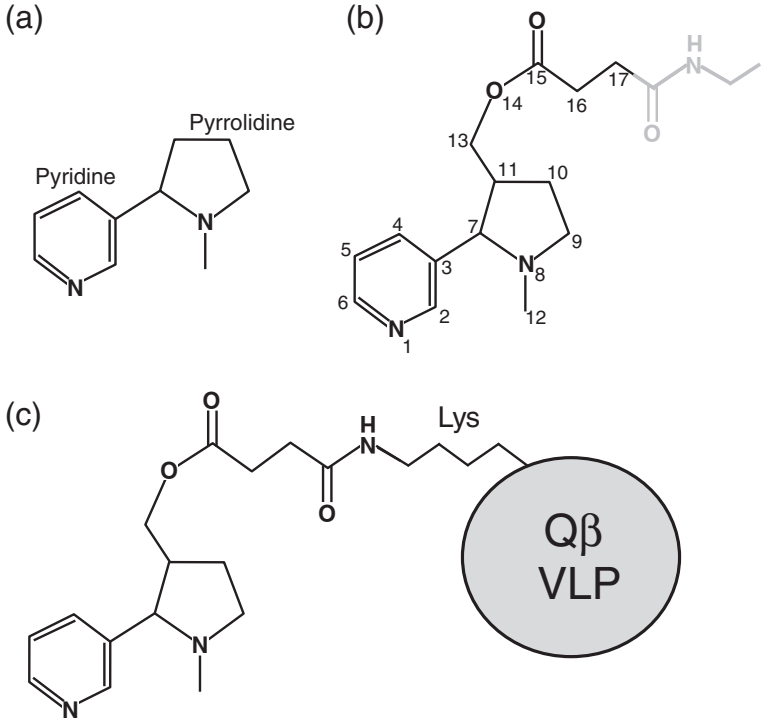



**Fig. 4.** Chemical structure of nicotine and its derivatives. (a) Nicotine. (b) NLD. The gray part of NLD was not visible in electron density. The numbering of atoms as used in the text is shown. (c) Nicotine coupled to surface lysines of Qβ VLP as used for immunization. The VLP is not drawn to scale. It has a diameter of 30 nm with about 585 nicotine molecules attached to it. Please note that at physiological pH values, the nitrogen in the pyrrolidine ring is protonated.

represents the only hydrogen bond between Fab and nicotine. Trp L98 makes hydrophobic interactions to both nicotine rings. Trp L96 and Tyr L34 are properly positioned next to the charged N atom of nicotine (4.3 and 4.4 Å distance from the center of the aromatic rings to the N atom) and are thus involved in cation–π interactions with nicotine.

The nicotine molecule is surrounded by a number of further largely hydrophobic residues, most importantly Ala L89, Phe L98, Ile H29, Ile H37, Val H93, and Trp H95.

## Structure of the NLD complex

The nicotine-specific antibody Nic12 was isolated from the blood of volunteers immunized with the nicotine vaccine NIC002 (Nicotine-Qβ).[12] In this vaccine, hydroxymethyl-nicotine was coupled to the amino groups of lysines on the surface of the Qβ VLPs using a succinic acid linker (Fig. 4). Since nicotine was entirely buried in the Fab structure and since there was no obvious "exit channel" for any attached linker, we wondered how such antibodies were induced when the vaccine initially interacted with a nicotine-specific membrane-bound antibody on a naïve B cell. In fact, C11 on the pyrrolidine ring of nicotine, to which the linker was attached to in Nicotine-Qβ, is located towards the interior of the Fab fragment, and not, as one would have been expected, towards the surface. To explain this finding, we synthesized NLD as a mimic for the nicotine–linker conjugate when attached to the VLP (Fig. 4) and solved the structure of the complex of the Fab fragment and NLD.

The electron density for the NLD was clear for both nicotine rings and atoms up to C17 of the linker

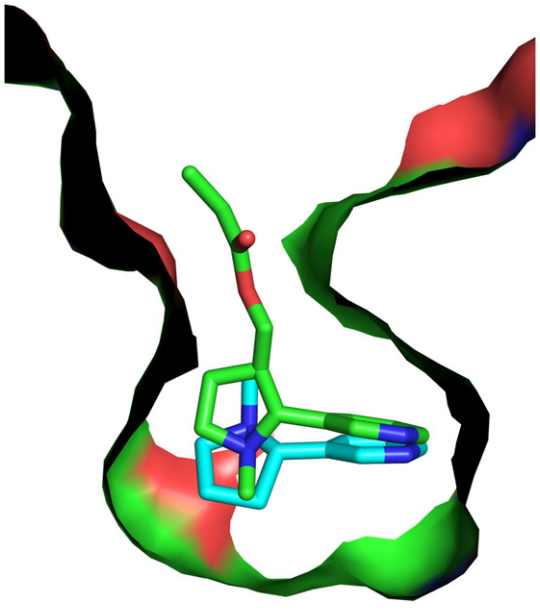


**Fig. 5.** Overlay of nicotine and NLD complex structures. Compared to nicotine, the pyrrolidine ring in the NLD complex is rotated by approximately 180° and the linker (part of which is not resolved in the structure) reaches the protein surface.

region. The rest of the linker was disordered and hence not modeled (Fig. 2). As seen from Figs. 3b and 5, NLD binds differently compared to nicotine itself. While the pyridine rings of nicotine and NLD had almost exactly the same position in both structures, the pyrrolidine ring rotated by about 180°, such that, in the NLD–Fab structure, the linker itself was indeed reaching the surface of the Fab as expected. Apparently, in the absence of a linker, it becomes energetically more favorable for the nicotine to bind in the other conformation. Despite the rotated pyrrolidine ring, NLD interacts with the same residues in a similar manner as nicotine (Fig. 3b). As in the nicotine complex, Glu H105 makes a salt bridge to the tertiary ammonium group in the pyrrolidine ring of NLD, Trp L96 makes hydrophobic interactions to both nicotine rings, and Trp L96 and Tyr L34 are involved in cation–π interactions with the charged nitrogen atom. Compared to the structure of the nicotine complex, there are no additional residues with specific interactions to the linker part of NLD. However, NLD produces a larger buried surface (225 $Å^2$), when compared to nicotine (116 $Å^2$) as determined by AREAIMOL. This might explain our observation that NLD has a higher affinity to F018 when compared to nicotine (data not shown).

NLD is a mixture of two diastereomers with (7*S*,11*S*) and (7*R*,11*R*) configurations at its chiral centers. Both diastereomers could be fitted into the density with somewhat better preference for the (7*R*,11*R*) diastereomer, which is the one that was modeled. Apparently, the electron density to some extent represents an average of two configurations. This is compatible with our observation that mAb-Nic12 binds both (*R*)-(+)-nicotine and (*S*)-(−)-nicotine, but with a slightly higher affinity for (*R*)-(+)-nicotine ($K_d$ = 3.1 nM and 7.4 nM, respectively). The NLD structure with the modeled (7*R*,11*R*) configuration clearly indicates that (*R*)-(+)-nicotine can bind in a very similar orientation as (*S*)-(−)-nicotine. (*R*)-(+)-Nicotine is present in only small amounts in the tobacco plant, but it can be formed by the combustion process during smoking and has qualitatively similar activities on the central nervous system as (*S*)-(−)-nicotine.[17]

The protein structure is in general very similar in both complexes (RMSD of $C^\alpha$ atoms, 0.3 Å), and even side chains in direct contact with nicotine or NLD are essentially in the same conformations.

## Discussion

### Comparison to structures of Fab fragments in complex with other alkaloids

A number of Fab fragment structures in complex with alkaloids and similar synthetic compounds have previously been determined, including the recreational drugs cocaine,[18–20] morphine,[21] phencyclidine,[22] and methamphetamine.[23] Unsurprisingly, in alkaloid complexes such as morphine and, to some extent, also cocaine, the positively charged N atom was also interacting with acidic residues. However, we failed to find any other obvious similarities to the binding mode of those compounds. Nicotine also appears to be different by being almost entirely buried (>99%) in the protein and lacking any water molecules in the binding pocket. In most available alkaloid–Fab structures, the hapten is only partially buried. Even in the cocaine–Fab structure 1I7Z, where cocaine is about 92% buried, it seems that the hapten is able to enter the binding pocket without a major rearrangement of the CDR loops, which is clearly not the case for nicotine.

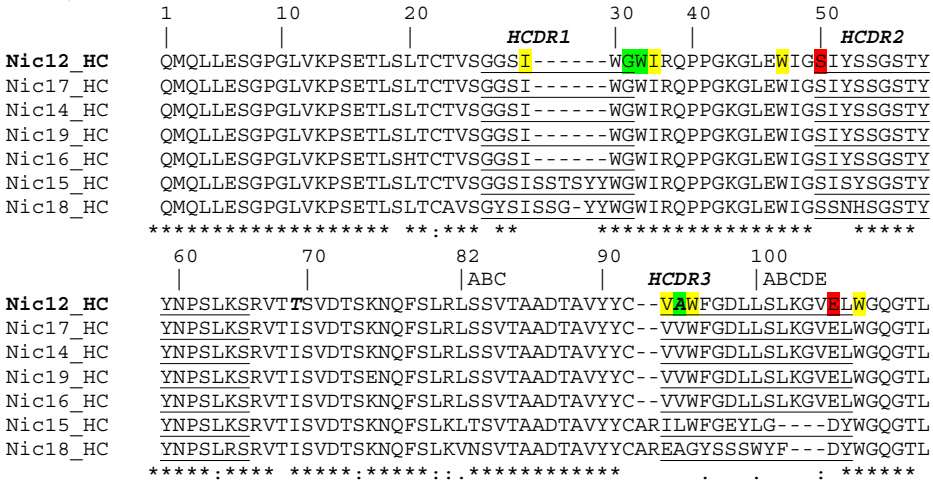
### Comparison to other structures with bound nicotine

The nicotine–Fab structure shows a novel mode of nicotine binding compared to known nicotine protein complexes, such as members of the cytochrome P450 family. These proteins catalyze the first step in the metabolism of nicotine in which nicotine is converted to cotinine by oxidation at C9 in the pyrrolidine ring. In the absence of oxygen, nicotine is bound to P450CAM in a nonproductive manner by forming a complex between the heme iron and the nitrogen of the pyridine ring (1P2Y).[24] The pyrrolidine nitrogen forms a hydrogen bond to tyrosine 96 through a bridging water molecule. This mode of binding results in only low affinities of nicotine for P450 enzymes ($K_d$ = 10–100 μM).[24,25] Upon treatment of the P450CAM–nicotine complex with CO, nicotine reorients in the active site, so that the pyridine nitrogen makes a hydrogen bond with tyrosine 96, while the pyrrolidine ring is oriented towards heme (1P7R).[24] In striking contrast to our nicotine–Fab structure, charge compensation of the pyrrolidine nitrogen was not evident in either of the P450CAM–nicotine complexes.

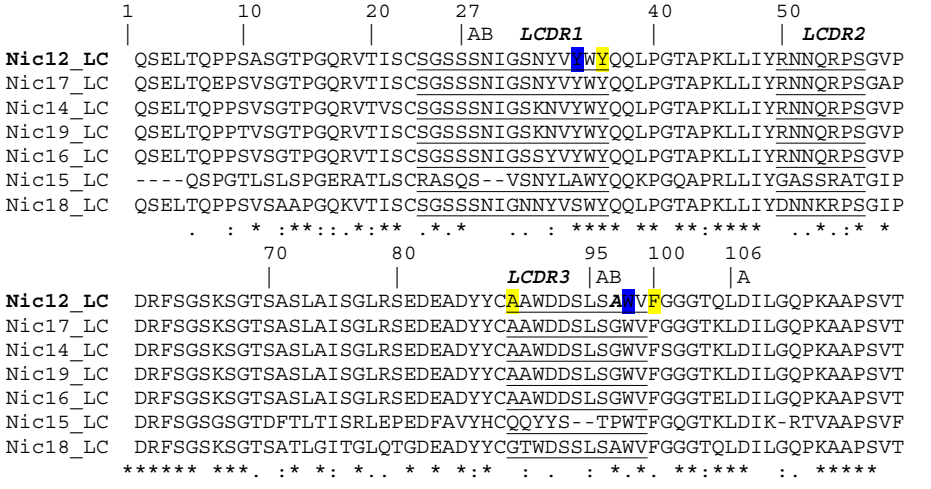
Despite extensive efforts, the structure of a nicotinic acetylcholine receptor, which belongs to the class of pentameric ligand-gated ion channels, in complex with nicotine has not been solved yet.[26] However, the pentameric molluscan acetylcholine binding protein (AChBP), which binds nicotine with moderate affinity ($K_d$ = 45 nM), is a well-established model for the nicotinic acetylcholine receptors. The structure of AChBP in complex with nicotine showed that nicotine is bound in the interface between the subunits through a cation–π interaction, two hydrogen bonds, and hydrophobic contacts (1UW6).[27,28] The tertiary ammonium group in the pyrrolidine ring of nicotine forms a cation–π interaction with a tryptophan. In addition, the

(a)

Heavy chains:




Light chains:




(b)

| Clone | Nic12 | Nic17 | Nic14 | Nic19 | Nic16 | Nic15 | Nic18 |
|---|---|---|---|---|---|---|---|
| Kd (nM) | **7.4** ± 1.2 | **30.2** ± 10.9 | **31.4** ± 13.0 | **35.6** ± 3.5 | **64.5** ± 2.2 | **89.9** ± 9.8 | **743.8** |

**Fig. 6.** (a) Sequence alignment of the variable domains of nicotine-binding antibodies. Residues of Nic12 making hydrophobic side-chain interactions with nicotine are highlighted in yellow, those with hydrophobic main-chain interactions with nicotine are highlighted in green, and the ones with polar interactions are highlighted in red. Trp L96 and Tyr L34 (blue) are involved in cation–π interactions with nicotine. CDR loop regions are underlined. Kabat numbering scheme is used. Insertions occur after residues L27, L95, L106, H82, and H100, as shown. Loop HCDR1 is shorter than 10 residues (considered minimum size in Kabat scheme); therefore, residue numbers H32–H34 are missing. (b) Affinities of antibodies to (*S*)-(−)-nicotine are as determined previously.[12]

backbone carbonyl group of the same tryptophan forms a hydrogen bond with the tertiary ammonium group of nicotine. These interactions were demonstrated to be of central importance for the affinity of nicotine.[28,29] In our Fab structure, Trp L96 and Tyr L34 are in an orientation that allows cation–π interactions (Fig. 3). However, on the opposite site of the tertiary ammonium group, the negatively charged side chain of Glu H101 is found in close proximity to the ammonium group (2.6 Å distance

between oxygen and nitrogen atoms). Accordingly, it is difficult to judge whether the cation–π interactions significantly contribute to the binding energy of the nicotine–Fab complex. In the Fab structure, the negatively charged side chain of Glu H101 and the aromatic rings of Trp L96 and Tyr L34 interact with the positive charge on the pyrrolidine nitrogen, while in the AChBP structure, the positive charge on the pyrrolidine nitrogen is only involved in a cation–π interaction. It was suggested that a partially negative charge of a backbone carbonyl group that is initiated by nearby aspartate contributes to the charge compensation.[27] A further difference is that the nicotine–Fab complex lacks any water molecules in the binding pocket, with the nitrogen atom of the pyridine ring forming a direct hydrogen bond to a serine, while in the nicotine–AChBP complex, the pyridine nitrogen atom forms a hydrogen bond through a bridging water molecule. The above factors likely contribute to the lower affinity of nicotine for AChBP compared to the Fab.

### Specificity of the mAb for nicotine

The mAb-Nic12 is specific for nicotine. No binding to acetylcholine was detectable and the major metabolic product cotinine bound with an about 1000-fold lower affinity.[8] This selectivity is a prerequisite for a therapeutic effect of an anti-nicotine mAb, as cotinine is present in the blood of smokers at a 10- to 20-fold molar excess over nicotine. Cotinine differs from nicotine by having a carbonyl oxygen attached to the C9 atom of the pyrrolidine ring. The selectivity of Nic12 for nicotine is readily obvious from the nicotine–Fab structure where the side chain of Tyr L34 would collide with the carbonyl oxygen of cotinine. Even assuming that Tyr L34 could change its position, the carbonyl oxygen of cotinine would still be in a hydrophobic environment without a compensating hydrogen bond, which is energetically unfavorable.

### Conserved sequence features within nicotine-specific mAbs

During our previous studies,[12] different anti-nicotine mAb clones were isolated and characterized. The sequence alignment of the variable regions of seven clones and their affinity to (*S*)-(−)-nicotine are shown in Fig. 6. Five of the seven mAbs (Nic12, Nic14, Nic16, Nic17, and Nic19) are derived from the same germ-line sequence, and the differences between them are presumably the result of affinity maturation steps. In general, there is remarkable sequence conservation among the residues interacting with nicotine, which suggests similar binding modes in all antibodies. Polar interactions with nicotine seem to be preserved in all cases, and hydrophobic residues forming the binding pocket are largely the same as well. mAbs Nic14, Nic16, Nic17, and Nic19 have essentially the same side chains interacting with nicotine but show a fourfold to ninefold lower affinity than Nic12. Three amino acids differ between Nic12 and the other four mAbs derived from the same germ-line sequence. An alanine in position 94 of the heavy chain is present in Nic12, while valine is found in the same position of Nic14, Nic16, Nic17, and Nic19. Nicotine forms hydrophobic interactions with the main chain of residue H94, while the side chain points in the opposite direction. In order to accommodate the larger side chain of valine (surrounding side chains are unchanged in all cases), the main chain presumably has to be slightly shifted, leaving somewhat less space for nicotine, which might result in a lower affinity. The other differences are Ala L95B in CDR3 of the Nic12 light chain, which is replaced by glycine in the other four mAbs, and Thr H69 in Nic12, which is replaced by isoleucine; however, these exchanges are not expected to have a major influence on the affinity for nicotine.

mAb Nic15 showed a 12-fold lower affinity to (*S*)-(−)-nicotine than Nic12, and mAb Nic18 even had a 100-fold lower affinity. Both are derived from different germ-line genes and appear to have somewhat different nicotine binding modes. In Nic15 and Nic18, aspartate replaces glutamate H101, but presumably this aspartate also takes part in the charged interaction with the tertiary ammonium in nicotine. Additionally, in comparison to other clones, the lengths of several CDR loops are different in Nic15 and Nic18, which certainly may lead to structural changes in the nicotine binding pocket. Trp H95 in Nic12 forms important hydrophobic interactions to nicotine. The corresponding amino acid in Nic18 is a glycine in CDR3, which also lacks any aromatic residue in the vicinity. The lack of the hydrophobic interactions with Trp H95 might thus be the major reason for the particularly low affinity of Nic18 compared to Nic12.

Nonetheless, 9 of 13 residues making side-chain interactions with nicotine are conserved between all mAbs, suggesting that major differences in binding mode are unlikely.

### Comparison of nicotine–Fab and NLD–Fab structures

The structures of the nicotine–Fab and NLD–Fab complexes reveal an unexpected plasticity in the binding of nicotine to the Fab fragment. Since nicotine as a hapten does not raise an immune response by itself, coupling to a protein carrier is required, delivering the necessary helper T cell signal. For the vaccine used in this study, coupling was achieved by the modification of nicotine at position 7 in the pyrrolidine ring, the introduction of a succinic acid linker, and the covalent attachment to

**Table 1.** Data scaling, refinement, and validation statistics

| | Fab–nicotine | Fab–NLD |
|---|---|---|
| Space group | *C*2 | *C*2 |
| *a* (Å) | 76.1 | 74.2 |
| *b* (Å) | 78.9 | 78.8 |
| *c* (Å) | 75.5 | 73.7 |
| β (°) | 116.6 | 116.1 |
| Resolution (Å) | 30–1.85 (1.9–1.85) | 30–2.1 (2.2–2.1) |
| $R_{merge}$ | 0.04 (0.23) | 0.06 (0.28) |
| Completeness (%) | 94.5 (87.7) | 100.0 (100.0) |
| $I/\sigma I$ | 8.9 (1.9) | 7.8 (2.0) |
| Multiplicity | 2.1 (2.1) | 3.7 (3.6) |
| $R_{cryst}$ | 0.22 (0.31) | 0.22 (0.23) |
| $R_{free}$ | 0.26 (0.38) | 0.27 (0.25) |
| Unique reflections collected | 30,695 | 21,216 |
| Reflections, used for refinement | 29,060 | 20,092 |
| Reflections in test set | 1635 | 1124 |
| Average *B*-factor (Å$^2$) | | |
| Overall | 28.3 | 35 |
| Nicotine/NLD | 13.5 | 23 |
| Number of protein atoms | 2772 | 2832 |
| Number of solvent atoms | 244 | 152 |
| Number of ligand atoms | 12 | 19 |
| *B*-factor from Wilson plot (Å$^2$) | 22 | 30 |
| Outliers in Ramachandran plot[38] (%) | 2.3 | 1.8 |
| RMSD bonds (Å) | 0.01 | 0.01 |
| RMSD angles (°) | 1.2 | 1.2 |

Values in parentheses are given for the highest-resolution bin.

an amino group of a lysine on the surface of the carrier protein (Fig. 4). Upon immunization, the nicotine derivative is assumed to be recognized in this orientation by membrane-bound antibodies on B cells. These B cells then become activated, differentiate, and finally secrete the soluble antibody. The NLD structure clearly shows that a linker of sufficient length is required to enable the binding to the nicotine moiety in its binding site between the variable domains of heavy and light chains. We were surprised to see that free nicotine binds in an orientation different from that of the nicotine moiety in the NLD. While the pyridine ring of free nicotine is in essentially the same position as the one of the NLD, the pyrrolidine ring is rotated by about 180°. However, the nitrogen atom of the tertiary ammonium group of nicotine is found in an only slightly shifted position compared to the corresponding nitrogen atom in the NLD structure. Accordingly, the rotated orientation of the pyrrolidine ring does not interfere with the important ion–ion interaction between Glu H101 and the tertiary ammonium group. Also, the hydrophobic interactions with Trp L96 are not disturbed and the high-affinity binding mode of the nicotine moiety in NLD is maintained for nicotine. Thus, nicotine's ability to adopt different relative orientations of pyrrolidine and pyridine rings allows free nicotine to bind in a different binding mode, but nonetheless with high affinity, to antibodies that were induced and selected *in vivo* for optimal binding to carrier-bound nicotine.

## Materials and Methods

### Expression and purification of protein

The full-length IgG2 antibody Nic12 was expressed and purified as previously described.[12] Fab fragments were obtained and purified with the Fab preparation kit from Thermo Scientific, according to the manufacturer's instructions. Briefly, 1 ml of antibody solution (7 mg/ml) in 20 mM sodium phosphate/10 mM ethylenediaminetetraacetic acid, pH 7.0, was mixed with 1.25 ml of papain agarose suspension and 1.25 ml of digestion buffer. After overnight incubation at 37 °C, Fc fragments were removed using protein A agarose columns. The resulting Fab fragments were dialyzed against 20 mM Tris–HCl, pH 8.0, and concentrated to 7 mg/ml using Amicon 10-kDa-cutoff minispin concentrators.

### Nicotine and NLD

*S*-(-)-Nicotine was purchased from Sigma and dissolved in water to 100 mM, and the pH was adjusted to 8.0 with HCl. Compound C-100096 (Cytos Biotechnology), which was used for conjugation to Qβ carrier for immunization, could not be directly used for co-crystallization due to the presence of a reactive succinimide group. To quench the succinimide moiety, we mixed C-100096 (100 mM) with ethylamine–HCl (200 mM, pH 7.0) and incubated it at room temperature overnight. The resulting NLD (Fig. 1b) was used without further purification for co-crystallization experiments.

### Co-crystallization and data collection

Five volumes of protein were mixed with one volume of nicotine or NLD stock solution. Crystallization trials were set up by sitting drop method (1 μl protein/nicotine or protein/NLD mixture + 1 μl crystallization buffer). The best crystals in the presence of nicotine appeared in 5% w/v polyethylene glycol (PEG) 3000, 0.1 M 4-morpholineethanesulfonic acid, pH 6.0, and 30% v/v PEG 200 (Emerald Cryo I screen, conditions ECryo1–13). A single crystal was picked directly from the trial plate and flash-frozen in liquid nitrogen. One hundred images at 1° oscillation were collected at European Synchrotron Radiation Facility (ESRF) beamline ID14-1.

The best crystals for the NLD complex were obtained in 0.1 M sodium cacodylate, pH 6.5, 0.2 M $MgCl_2$, and 31% PEG monomethyl ether 2000. Crystals were flash-frozen in liquid nitrogen. Ninety images at 1° oscillation were collected from a single crystal at ESRF beamline BM14U.

### Structure determination

Images were processed by MOSFLM[30] and scaled by Scala.[31] The space group was *C*2 in both cases, and the

maximum resolution was 1.85 Å for the nicotine complex and 2.0 Å for the NLD complex. The nicotine complex structure was solved by molecular replacement, using coordinates of anti-HIV-1 V3 Fab structure [Protein Data Bank (PDB) ID: 2BOS[32]], which displayed the highest sequence similarity to our structure (89% identity). Rotation and translation searches were performed in MOLREP[33] separately with variable (residues L1 to L110 and residues H1 to H116) and constant (residues L111 to L212 and residues H117 to H229) domains of anti-HIV-1 V3 Fab fragment. The first solution was obtained for the variable domain, which was fixed, and the solution for the constant domain was obtained in the second round. The model was built with O[34] and refined with REFMAC.[35] Parameters for ligands were generated with Libcheck.[36] Water molecules were added automatically by ARP/wARP[37] and inspected manually in O. The structure of the NLD complex was solved by molecular replacement, using the nicotine complex structure as model. Data scaling, refinement, and validation statistics are listed in Table 1.

### PDB accession codes

The coordinates and the structure factors have been deposited in the PDB with the nicotine–Fab complex as entry 2YK1 and the NLD–Fab complex as 2YKL.

## Acknowledgements

This study was supported by European Regional Development Foundation grant 2DP/2.1.1.1.0/10/APIA/VIAA/052. We personally thank the staff at ESRF (Grenoble, France) and MaxLab (Lund, Sweden) for their support during data collection. Dr. Pavel Plevka is acknowledged for his assistance in data collection. R. Buser is acknowledged for excellent technical assistance.

## References

1. World Health Organization (2008). *WHO Report on the Global Tobacco Epidemic, 2008: The MPOWER Package.* WHO, Geneve, Switzerland.
2. Laviolette, S. R. & van der Kooy, D. (2004). The neurobiology of nicotine addiction: bridging the gap from molecules to behaviour. *Nat. Rev., Neurosci.* **5**, 55–65.
3. Kenford, S. L., Fiore, M. C., Jorenby, D. E., Smith, S. S., Wetter, D. & Baker, T. B. (1994). Predicting smoking cessation. Who will quit with and without the nicotine patch. *JAMA*, **271**, 589–594.
4. Maurer, P. & Bachmann, M. F. (2007). Vaccination against nicotine: an emerging therapy for tobacco dependence. *Expert Opin. Invest. Drugs*, **16**, 1775–1783.
5. Maurer, P., Jennings, G. T., Willers, J., Rohner, F., Lindman, Y., Roubicek, K. *et al.* (2005). A therapeutic vaccine for nicotine dependence: preclinical efficacy, and Phase I safety and immunogenicity. *Eur. J. Immunol.* **35**, 2031–2040.
6. Hatsukami, D. K., Rennard, S., Jorenby, D., Fiore, M., Koopmeiners, J., de Vos, A. *et al.* (2005). Safety and immunogenicity of a nicotine conjugate vaccine in current smokers. *Clin. Pharmacol. Ther.* **78**, 456–467.
7. St Clair Roberts, J., Akers, C. V. & Vanhinsbergh, L. (2003). 9th Annual Meeting of the Society for Research on Nicotine and Tobacco, New Orleans, LA.
8. de Villiers, S. H., Lindblom, N., Kalayanov, G., Gordon, S., Johansson, A. M. & Svensson, T. H. (2004). Active immunization against nicotine alters the distribution of nicotine but not the metabolism to cotinine in the rat. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **370**, 299–304.
9. Cornuz, J., Zwahlen, S., Jungi, W. F., Osterwalder, J., Klingler, K., van Melle, G. *et al.* (2008). A vaccine against nicotine for smoking cessation: a randomized controlled trial. *PLoS One*, **3**, e2547.
10. Hatsukami, D. K., Jorenby, D. E., Gonzales, D., Rigotti, N. A., Glover, E. D., Oncken, C. A. *et al.* (2011). Immunogenicity and smoking-cessation outcomes for a novel nicotine immunotherapeutic. *Clin. Pharmacol. Ther.* **89**, 392–399.
11. Pentel, P. R., Dufek, M. B., Roiko, S. A., Lesage, M. G. & Keyler, D. E. (2006). Differential effects of passive immunization with nicotine-specific antibodies on the acute and chronic distribution of nicotine to brain in rats. *J. Pharmacol. Exp. Ther.* **317**, 660–666.
12. Beerli, R. R., Bauer, M., Buser, R. B., Gwerder, M., Muntwiler, S., Maurer, P. *et al.* (2008). Isolation of human monoclonal antibodies by mammalian cell display. *Proc. Natl Acad. Sci. USA*, **105**, 14336–14341.
13. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System 0.98 edit.* DeLano Scientific, San Carlos, CA.
14. Seeman, J. I. (1984). Recent studies in nicotine chemistry. Conformational analysis, chemical reactivity studies and theoretical modeling. *Heterocycles*, **22**, 165–193.
15. Briggs, P. J. (2000). Further improvements to AREAIMOL code. *CCP4 Newsletter on Protein Crystallography*, **38**.
16. Huang, X., Zheng, F., Crooks, P. A., Dwoskin, L. P. & Zhan, C. G. (2005). Modeling multiple species of nicotine and deschloroepibatidine interacting with $\alpha 4\beta 2$ nicotinic acetylcholine receptor: from microscopic binding to phenomenological binding affinity. *J. Am. Chem. Soc.* **127**, 14401–14414.
17. Hukkanen, J., Jacob, P., 3rd & Benowitz, N. L. (2005). Metabolism and disposition kinetics of nicotine. *Pharmacol. Rev.* **57**, 79–115.
18. Larsen, N. A., Zhou, B., Heine, A., Wirsching, P., Janda, K. D. & Wilson, I. A. (2001). Crystal structure of a cocaine-binding antibody. *J. Mol. Biol.* **311**, 9–15.
19. Pozharski, E., Moulin, A., Hewagama, A., Shanafelt, A. B., Petsko, G. A. & Ringe, D. (2005). Diversity in hapten recognition: structural study of an anti-cocaine antibody M82G2. *J. Mol. Biol.* **349**, 570–582.
20. Zhu, X., Dickerson, T. J., Rogers, C. J., Kaufmann, G. F., Mee, J. M., McKenzie, K. M. *et al.* (2006). Complete reaction cycle of a cocaine catalytic antibody at atomic resolution. *Structure*, **14**, 205–216.
21. Pozharski, E., Wilson, M. A., Hewagama, A., Shanafelt, A. B., Petsko, G. & Ringe, D. (2004). Anchoring a

cationic ligand: the structure of the Fab fragment of the anti-morphine antibody 9B1 and its complex with morphine. *J. Mol. Biol.* **337**, 691–697.
22. Lim, K., Owens, S. M., Arnold, L., Sacchettini, J. C. & Linthicum, D. S. (1998). Crystal structure of monoclonal 6B5 Fab complexed with phencyclidine. *J. Biol. Chem.* **273**, 28576–28582.
23. Celikel, R., Peterson, E. C., Owens, S. M. & Varughese, K. I. (2009). Crystal structures of a therapeutic single chain antibody in complex with two drugs of abuse—methamphetamine and 3,4-methylenedioxymethamphetamine. *Protein Sci.* **18**, 2336–2345.
24. Strickler, M., Goldstein, B. M., Maxfield, K., Shireman, L., Kim, G., Matteson, D. S. & Jones, J. P. (2003). Crystallographic studies on the complex behavior of nicotine binding to P450cam (CYP101). *Biochemistry*, **42**, 11943–11950.
25. DeVore, N. M., Smith, B. D., Wang, J. L., Lushington, G. H. & Scott, E. E. (2009). Key residues controlling binding of diverse ligands to human cytochrome P450 2A enzymes. *Drug Metab. Dispos.* **37**, 1319–1327.
26. Chen, L. (2010). In pursuit of the high-resolution structure of nicotinic acetylcholine receptors. *J. Physiol.* **588**, 557–564.
27. Celie, P. H., van Rossum-Fikkert, S. E., van Dijk, W. J., Brejc, K., Smit, A. B. & Sixma, T. K. (2004). Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron*, **41**, 907–914.
28. Xiu, X., Puskar, N. L., Shanata, J. A., Lester, H. A. & Dougherty, D. A. (2009). Nicotine binding to brain receptors requires a strong cation–pi interaction. *Nature*, **458**, 534–537.
29. Blum, A. P., Lester, H. A. & Dougherty, D. A. (2010). Nicotinic pharmacophore: the pyridine N of nicotine and carbonyl of acetylcholine hydrogen bond across a subunit interface to a backbone NH. *Proc. Natl Acad. Sci. USA*, **107**, 13206–13211.
30. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4+ESF-EAMCB Newsletter on Protein Crystallography*, **26**.
31. Evans, P. R. (1997). Scala. *Joint CCP4+ESF-EAMCB Newsletter on Protein Crystallography*, **33**, 22–24.
32. Stanfield, R. L., Gorny, M. K., Zolla-Pazner, S. & Wilson, I. A. (2006). Crystal structures of human immunodeficiency virus type 1 (HIV-1) neutralizing antibody 2219 in complex with three different V3 peptides reveal a new binding mode for HIV-1 cross-reactivity. *J. Virol.* **80**, 6093–6105.
33. Vagin, A. A. & Teplyakov, A (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025.
34. Jones, T. A., Bergdoll, M. & Kjeldgaard, M. (1990). O: a macromolecule modeling environment. In *Crystallographic and Modeling Methods in Molecular Design* (Bugg, C. & Ealick, S., eds), pp. 189–199, Springer-Verlag, New York, NY.
35. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255.
36. Vagin, A. A., Murshudov, G. N. & Strokopytov, B. V. (1998). BLANC: the program suite for protein crystallography. *J. Appl. Crystallogr.* **31**, 98–112.
37. Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 448–455.
38. Kleywegt, G. J. & Jones, T. A. (1996). Phi/psi-chology: Ramachandran revisited. *Structure*, **4**, 1395–1400.

Applied Organometallic Chemistry

Received: 13 July 2012 Revised: 10 November 2012 Accepted: 10 November 2012 Published online in Wiley Online Library: 19 December 2012

(wileyonlinelibrary.com) DOI 10.1002/aoc.2952

# Silyl modification of biologically active compounds. 13.† Synthesis, cytotoxicity and antibacterial action of *N*-methyl-*N*-(2-triorganylsiloxyethyl)-1,2,3,4-tetrahydro(iso) quinolinium iodides

**Alla Zablotskaya[a]*, Izolda Segal[a], Yuris Popelis[a], Solveiga Grinberga[a], Irina Shestakova[a], Vizma Nikolajeva[b] and Daina Eze[b]**

**A series of *N*-methyl-*N*-(2-triorganylsiloxyethyl)-1,2,3,4-tetrahydro(iso)quinolinium iodides has been synthesized via dehydrocondensation reaction of *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline, *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroquinoline and 4,4-dimethyl-*N*-(2-hydroxyethyl)-4-sila-1,2,3,4-tetrahydroisoquinoline with trialkyl(aryl)hydrosilanes and subsequent alkylation, and characterized by $^{1}H$, $^{13}C$ and $^{29}Si$ NMR and mass spectroscopy. The biological activity data exhibited a marked enhancement of inhibitory activity against tumour cell lines and almost all the test bacterial/fungal strains in comparison with their 2-hydroxyethyl precursors. Cytotoxicity in the microgram range against HT-1080 (human fibrosarcoma) and MG-22A (mouse hepatoma) cancer cell lines was observed for most of compounds. Copyright © 2012 John Wiley & Sons, Ltd.**

*Supporting information may be found in the online version of this article.*

**Keywords:** silicon; silylation; choline; drug research; tetrahydro(iso)quinoline; antitumour activity; antimicrobial activity

## Introduction

Functional groups containing silicon might provide lipophilicity to a drug, allowing it to pass the cell membrane by passive diffusion. This strategy has been successfully applied in the antitumour drug analogues silatecans (silicon-containing camptothecins)[1,2] and silaplatins (cisplatin analogues),[3] the HIV-1 reverse transcriptase inhibitor TSAO-T[4] and the effective anabolic 'silabolin'.[5] Some silicon-containing drugs have entered phase I or II human clinical trials.[6–10] Silylation of biologically active compounds, possessing hydrophilic functional groups, is one of the most effective methods for increasing lipophilicity, ensuring drug permeability through lipophilic barriers inside living organisms; silylation therefore can positively influence biological activity appearance or enhancement.[11,12] Trialkylsilyl derivatives of polyene macrolide antibiotic nystatin had a high level of antifungal activity against a wide set of test strains and their acute toxicity ($LD_{50}$) was two to three times lower than that of the initial antibiotic.[13] A 2′-*O*-*tert*-butyldimethylsilyl group was found to be necessary for optimal antiproliferative activity of $N^6$,5′-bis-ureidoadenosine nucleosides.[14] Its 2′,3′-bis-O-*tert*-butyldimethylsilyl derivative exhibited broad-spectrum antiproliferative activity and was accepted as a new member of the $N^6$,5′-bis-ureidoadenosine class of anticancer nucleosides.[15] Cytotoxicity study of (2*R*,3*S*)-disubstituted tetrahydropyranes bearing a *tert*-butyldimethylsilyl group at position 3 of the ring considerably induced cytotoxicity against HL60 human leukaemia cells and MCF7 breast cancer cells *in vitro* and pointed to the relevant role of the *tert*-butyldimethylsilyl group in the enhancement of cytotoxic action.[16] 5′-Triphenylsilyl modification of deoxyuridine acyclic analogues resulted in obtaining more potent inhibitors of *Plasmodium falciparum* deoxyuridine 5′-triphosphate nucleotidohydrolase, a target for the development of antimalarial drugs, in comparison with their 5′-trityloxy derivatives.[17] Within our research activity, directed at the targeted modification of biologically active compounds aimed at the improvement of their biological properties, including increased drug accumulation and prolongation of drug retention inside cells, we have found that 5′-O-*tert*-butyldimethylsilyluridine, unlike uridine, exhibits antitumour activity, suppressing the development of human lung fibrosarcoma cells,[18] and *N*-(4-phenyl-2-thiazolyl)-2-(4-trimethylsiloxypiperidino)acetamide revealed the higher cytotoxic effect on MG-22A cells in comparison with its unsilylated analogue.[19]

We have reported recently that silylation represents a plausible strategy to modulate cytotoxic and antibacterial properties in choline and colamine analogues. A first series of silylated choline analogues was synthesized.[20] It was demonstrated that some

* *Correspondence to: Alla Zablotskaya, Latvian Institute of Organic Synthesis, Aizkraukles 21, Riga, LV-1006, Latvia. E-mail: aez@osi.lv*

† *Dedicated to the memory of Professor E. Lukevics.*

a *Latvian Institute of Organic Synthesis, Riga, LV-1006, Latvia*

b *Biological Faculty, University of Latvia, Riga LV-1586, Latvia*

silyl-modified aliphatic ethanolamines are low toxic compounds and reveal biological activity, inhibiting tumour growth and possessing antimicrobial properties, in comparison with their unsilylated precursors. Rather small structural modifications (for instance, silyl group nature) can significantly influence their activity. Inspired by the results, and to make more general conclusions, we have synthesized a new series of silylated derivatives of *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline, which could be considered heterocyclic choline analogues, possessing an ambivalent nature, namely, lipophilic and hydrophilic fragment combination in one molecule, and investigated their biological properties against tumour HT-1080 and MG-22A and normal NIH 3 T3 cell lines, and also against some Gram-positive – *Bacillus cereus* MSCL 330 (BC) and *Staphylococcus aureus* MSCL 334 (SA) – and Gram-negative microbial strains – *Proteus mirabilis* MSCL 590 (PM), *Escherichia coli* MSCL 332 (EC) and *Pseudomonas aeruginosa* MSCL 331 (PA) – and fungi – *Candida albicans* MSCL 378 (CA) – in comparison with their 2-hydroxyethyl precursors. Choline is recognized as a metabolic marker of active tumour tissue due to its ability to accumulate in cancer cells.[21] There are some structural requirements of choline derivatives for 'conversion' of pneumococcal amidase, and this amino alcohol has been identified as an allosteric ligand necessary for recognition and degradation of cell wall by the enzyme.[22]

On the other hand, quinoline derivatives, including tetrahydro(iso)quinoline ones, are biologically active compounds possessing a wide spectrum of biological activity. Structural fragments of quinoline and its derivatives are included in the molecules of antitumour drugs, for instance in amsacrine, bruneomicine and vinblastine.[5,23,24] The tetrahydroisoquinoline ring system is an important structural motif[25] that is commonly encountered in naturally occurring alkaloids, with interesting biological activities. Typical examples include saframycin-B,[26] narciclasine[27] and ecteinascidin-743.[28] In this regard, tetrahydroisoquinoline has become widely identified as a 'privileged' structure, with representation in several medicinal agents of diverse therapeutic action, and are potential drug candidates.[29,30] Tetrahydroisoquinoline derivatives possess antitumour properties.[31–33]

## Experimental

### Chemicals and Instrumentation

$^1$H, $^{13}$C and $^{29}$Si NMR spectra were obtained on Varian Mercury 200 and Varian Mercury 400 spectrometers with $CDCl_3$ or DMSO-$d_6$ as solvent and $CDCl_3$ ($\delta = 7.25$ ppm for $CHCl_3$) as internal standard for compounds containing a silicon atom in the molecule, and with hexamethyldisiloxane ($\delta = 0.055$ ppm) as internal standard for the other compounds. Heteronuclear single-quantum correlation and heteronuclear multiple-bond correlation NMR 2-D correlation spectra were taken for compounds **1–4**, **8**, **9d**, **12d**, **12 g** and **12 h** for an unambiguous assignment of $^1$H and $^{13}$C signals. Mass spectra under electron impact conditions were recorded on an Agilent Technologies 5975C mass spectrometer (GC 7890A, 70 eV) and on Waters 3100 mass spectrometer (LC Alliance Waters 2695). Elemental analyses (C, H, N) were performed on a Carlo Erba 1108 elemental analyser. Elemental analysis results agreed with calculated values. Melting points were determined on a Boetius melting point apparatus and were taken uncorrected. Analytical thin-layer chromatography (TLC) was performed on 60 $F_{254}$ (Merck) silica gel plates and Machery-Nagel silica gel plastic plates, with visualization under UV (254 nm). Column chromatography was performed using Merck silica gel (0.040–0.063 nm). Solvents and reagents used in this study were purchased from Fluka, Acros and Aldrich. Syntheses involving air-sensitive compounds were carried out under dry argon. All solvents used were freshly dried using standard techniques and all glassware was oven dried.

### Synthesis

*N*-(2-Hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline **(1)**,[34] *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroquinoline **(2)**,[34], *N*-(2-hydroxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium iodide **(3)**,[34] *N*-(2-hydroxyethyl)-*N*-methyl-1,2,3,4-tetrahydroquinolinium iodide **(4)**,[34] bromomethyldimethyl(2-methylphenyl)silane **(5)**,[34] (2-bromomethylphenyl)bromomethyl-dimethylsilane **(6)**,[34] 4,4-dimethyl-*N*-(2-hydroxyethyl)-4-sila-1,2,3,4-tetrahydroisoquinoline **(7)**[34] and 4,4-dimethyl-*N*-(2-hydroxyethyl)-*N*-methyl-4-sila-1,2,3,4-tetrahydroisoquinolinium iodide **(8)**[34] were characterized by comparing their $^1$H NMR spectra with those reported in the literature.

An outline is given in Scheme 1.

*N*-(2-Diethylmethylsiloxyethyl)-1,2,3,4-tetrahydroisoquinoline (**9a**)

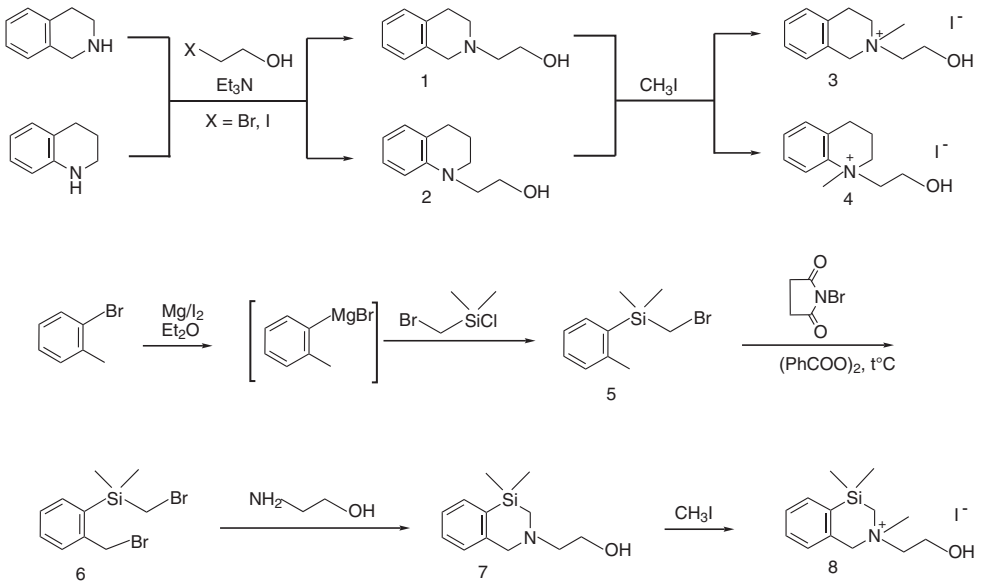
Diethylmethylsilane (0.59 g, 5.77 mmol) was added to *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline (0.88 g, 4,97 mmol) and the reaction mixture was heated at 70 °C under stirring for 4 h in the presence of a trace amount of metallic sodium. The course of the reaction was followed using data provided by gas chromatography–mass spectrometry (GC-MS). If necessary (presence of unreacted alcohol detected by TLC), more hydrosilane was added. Once the dehydrocondensation reaction had reached completion, the reaction mixture was cooled to room temperature and the solid traces were filtered off. The resulting crude residue was purified by column chromatography on silica gel eluted with chloroform–methanol (4:1) to give the desired diethylmethylsilyl ether as a clear light-yellow liquid.

Yield 0.77 g (56%). LC-MS (*m/z*, %): 278 ($M^+$+1, 100), 277 ($M^+$, 58), 262 ($M^+ - CH_3$, 9), 248 ($M^+ - C_2H_5$, 12), 174 (32), 161 ($M^+ - OSi(C_2H_5)_2CH_3 + 1$, 52), 144 (26). GC-MS (*m/z*, %): 277 ($M^+$, 2), 262 ($M^+ - CH_3$, 3), 248 ($M^+ - C_2H_5$, 5), 146 ($M^+ - CH_2CH_2OSi(C_2H_5)_2CH_3$, 100), 132 (26). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 012 (3H, s, $SiCH_3$), 0.66 (4H, m, $SiCH_2$), 1.00 (6H, m, C&bond;$CH_3$), 2.76 (2H, m, $\alpha$-$CH_2N$), 2.86 and 2.97 (total 4H, m and m, 3- and 4-$CH_2N$), 3.75 (2H, m, 1-$CH_2N$), 3.88 (2H, t, $J = 6.6$ Hz, $OCH_2$), 7.0–7.2 (4H, m, 5-, 6-, 7- and 8-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −5.01 ($SiCH_3$), 6.16 ($SiCH_2$), 6.70 ($SiCCH_3$), 28.99 (4-$CH_2$), 50.66 (3-$CH_2N$), 55.74 (1-$CH_2N$), 58.07 ($OCH_2$), 59.17 ($\alpha$-$CH_2N$), 125.63 (7-C), 125.99 (8-C), 126.46 (6-C), 128.60 (5-C), 134.12 (9-C), 134.40 (10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +19.90. Anal. calcd for $C_{16}H_{27}NOSi$: C, 69.26; H, 9.81; N, 5.05; found: 68.97; H, 9.73; N, 5.11.

*N*-(2-Triethylsiloxyethyl)-1,2,3,4-tetrahydroisoquinoline (**9b**)

Compound **9b** was obtained following the procedure described for **9a**, from 1.80 g (10.16 mmol) *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline (**1**) and 1.20 g (10.32 mmol) triethylsilane, as a clear light-yellow liquid.

Yield 2.25 g (77%). LC-MS (*m/z*, %): 292 ($M^+$+1, 51), 291 ($M^+$, 49); 290 ($M^+ - 1$, 100), 262 ($M^+ - C_2H_5$, 32), 176 ($M^+ - Si(C_2H_5)_3$, 5), 158 (88), 131 (50). GC-MS (*m/z*, %): 291 ($M^+$, 1), 262 ($M^+ - C_2H_5$, 3), 146 ($M^+ - CH_2OSi(C_2H_5)_3$, 100), 132 ($M^+ - CH_2CH_2OSi(C_2H_5)_3 - 1$, 16). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.63 (6H, m, $SiCH_2$), 0.97 (9H, m, $SiCCH_3$), 2.71 (2H, t, $J = 6.5$ Hz, $\alpha$-$CH_2N$), 2.81 and 2.89 (total 4H, m and m, 3- and 4-$CH_2N$), 3.69 and 3.71 (total 2H, d and d,




**Scheme 1.** Synthesis of compounds **1–8**.

$J = 5.2$ Hz, 1-$CH_2N$), 3.84 (2H, t, $J = 6.5$ Hz, $OCH_2$), 6.9–7.2 (4H, m, 5-, 6-, 7- and 8-H). $^{13}$C NMR ($CDCl_3$, δ, ppm): 4.41 ($SiCH_2$), 6.75 ($SiCCH_3$), 29.00 (4-$CH_2$), 51.60 (3-$CH_2N$), 56.65 (1-$CH_2N$), 60.23 ($OCH_2$), 61.22 (α-$CH_2N$), 125.51 (7-C), 126.03 (8-C), 126.51 (6-C), 128.61 (5-C), 134.16 (9-C), 134.82 (10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +19.62. Anal. calcd for: $C_{17}H_{29}NOSi$: C, 70.07; H, 10.03; N, 4.80; found: C, 69.78; H, 10.04; N, 4.72.

*N*-(2-Di-*n*-butylethylsiloxyethyl)-1,2,3,4-tetrahidroisoquinoline (**9c**)

Compound **9c** was obtained following the procedure described for **9a**, from 0.89 g (5.00 mmol) *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline (**1**) and 1.21 g (7.02 mmol) dibutylethylsilane, as a clear light-yellow liquid.

Yield 1.31 g (75%). LC-MS (*m/z*, %): 348 ($M^+$+1, 26), 347 ($M^+$, 100), 160 ($M^+ - OSi(C_4H_9)_2C_2H_5$, 25). GC-MS (*m/z*, %): 347 ($M^+$, 2), 318 ($M^+ - C_2H_5$, 3), 290 ($M^+ - C_4H_9$, 5), 146 ($M^+ - CH_2OSi(C_4H_9)_2C_2H_5$, 100), 132 ($M^+ - CH_2CH_2OSi(C_4H_9)_2C_2H_5 - 1$, 27). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.52–0.71 (6H, m, $SiCH_2$), 0.93–1.10 (15H, m, $SiCCH_3$, $SiCCH_2CC$ and $SiCCCCH_3$), 1.79 (4H, m, $SiCCCH_2C$), 2.71 (2H, t, $J = 5.6$ Hz, α-$CH_2N$), 2.80 and 2.91 (total 4H, m and m, 3- and 4-$CH_2$), 3.69 (2H, s, 1-$CH_2N$), 3.70 (2H, t, $J = 5.6$ Hz, $OCH_2$), 7.0–7.2 (4H, m, 5-, 6-, 7- and 8-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): 4.09 ($SiCH_2$) and 8.36 ($SiCCH_3$), 22.13, 25.33, 25.97, 26.42 ($CCH_2$ $CH_2$ $CH_3$), 29.06 (4-$CH_2$), 50.72 (3-$CH_2N$), 55.79 (1-$CH_2N$), 58.12 ($OCH_2$), 59.27 (α-$CH_2N$), 125.53 (7-C), 126.05 (8-C), 126.53 (6-C), 128.63 (5-C), 134.16 (9-C), 134.42 (10-C). $^{29}$Si NMR($CDCl_3$, $\delta$, ppm): +16.74. Anal. calcd for: $C_{21}H_{37}NOSi$: C, 72.56; H, 10.73; N, 4.03; found: C, 72.80; H, 10.65; N, 4.10.

*N*-(2-Di-*n*-heptylmethylsiloxyethyl)-1,2,3,4-tetrahydroisoquinoline (**9d**)

Compound **9d** was obtained following the procedure described for **9a**, from 0.89 g (5.00 mmol) *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline (**1**) and 1.46 g (6.02 mmol) diheptylmethylsilane, as clear yellow liquid.

Yield 1.19 g (57%). LC-MS (*m/z*, %): 417 ($M^+$, 71), 416 ($M^+ - 1$, 100), 318 ($M^+ - C_7H_{15}$, 15), 161 ($M^+ - OSi(C_7H_{15})_2CH_3 + 1$, 98), 160 ($M^+ - OSi(C_7H_{15})_2CH_3$, 57). GC-MS (*m/z*, %): 318 ($M^+ - C_7H_{15}$, 9), 158 ($M^+ - OSi(C_7H_{15})_2CH_3 - 2$, 2), 146 ($M^+ - CH_2OSi(C_7H_{15})_2CH_3$, 100), 132 ($M^+ - CH_2CH_2OSi(C_7H_{15})_2CH_3 - 1$, 27). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.08 (3H, s, $SiCH_3$), 0.55 (4H, m, $SiCH_2$), 0.88 (6H, t, $J = 6.8$ Hz, $2CH_3$), 1.26 (20H, m, $10CH_2$), 2.70 (2H, t, $J = 6.4$ Hz, α-$CH_2N$), 2.80 and 2.88 (total 4H, m and m, 3- and 4-$CH_2$), 3.71 (4H, m, 1-$CH_2N$ and $OCH_2$), 7.1–7.20 (4H, m, 5-, 6-, 7- and 8-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −4.10 ($SiCH_3$), 14.12 ($CH_3$), 15.02 ($SiCH_2$), 22.70, 23.19, 29.03, 29.05, 31.82 and 33.52 ($CH_2$, 4-$CH_2$), 50.68 (3-$CH_2N$), 55.78 (1-$CH_2N$), 58.08 ($OCH_2$), 59.16 (α-$CH_2N$), 125.52 (7-C), 126.05 (8-C), 126.54 (6-C), 128.61 (5-C), 134.18 (9-C), 134.46 (10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +18.20. Anal. calcd for: $C_{26}H_{47}NOSi$: C, 74.75; H, 11.34; N, 3.35; found: C, 75.04; H, 11.40; N, 3.29.

*N*-(2-*n*-Dimethyloctylsiloxyethyl)-1,2,3,4-tetrahydroisoquinoline (**9e**)

Compound **9e** was obtained following the procedure described for **9a**, from 0.88 g (4.95 mmol) *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline (**1**) and 1.28 g (7.43 mmol) dimethyloctylsilane, as a clear yellow oil.

Yield 1.32 g (77%). LC-MS (*m/z*, %): 348 ($M^+$+1, 100), 347 ($M^+$, 62), 332 ($M^+ - CH_3$, 10). GC-MS (*m/z*, %): 332 ($M^+ - CH_3$, 5), 234 ($M^+ - C_8H_{17}$, 12), 146 ($M^+ - CH_2OSi(CH_3)_2C_8H_{17}$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.10 (6H, s, $SiCH_3$), 0.60 (2H, m, $SiCH_2$), 0.87 (3H, t, $J = 6.8$ Hz, $CH_3$), 1.25 (12H, bs, &bond;$CH_2$&bond;), 2.70 (2H, t, $J = 6.6$ Hz, α-$CH_2N$), 2.80 and 2.89 (2H and 2H, t and t, $J = 5.9$ Hz, 3- and 4-$CH_2$), 3.69 (2H, s, 1-$CH_2N$), 3.81 (2H, t, $J = 6.6$ Hz, $OCH_2$), 6.9–7.0 (1H, m, 8-H), 7.0–7.2 (3H, m, 5-, 6- and 7-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −2.13 ($SiCH_3$), 14.07 ($CH_3$), 16.31 ($SiCH_2$), 22.72, 23.24, 23.33, 29.25, 29.32, 29.39, 32.0, 33.43 ($CH_2$), 29.02 (4-$CH_2$), 51.48 (3-$CH_2N$), 56.61 (1-$CH_2N$), 60.06 ($OCH_2$), 60.92 (α-$CH_2N$), 125.51 (7-C), 126.05 (8-C), 126.54 (6-C), 128.57 (5-C), 134.22 (9-C), 134.83 (10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +18.32. Anal. calcd for $C_{21}H_{37}NOSi$: C, 72.56; H, 10.73; N, 4.03; found: C, 72.28 ; H, 10.63; N, 4.10.

*N*-(2-*n*-Decyldimethylsiloxyethyl)-1,2,3,4-tetrahidroisoquinoline (**9f**)

Compound **9f** was obtained as clear yellow oil following the procedure described for **9a**, from 5.05 g (28.5 mmol) *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline (**1**) and 6.28 g (31.2 mmol) decyldimethylsilane, by heating the reagents at 80 °C under stirring for 15 h.

Yield 8.04 g (75%). LC-MS (*m/z*, %): 376 ($M^+$+1, 100), 375 ($M^+$, 52), 160 ($M^+ - OSi(CH_3)_2C_{10}H_{21}$, 10). GC-MS (*m/z*, %): 375 ($M^+$, 2), 360 ($M^+ - CH_3$, 4), 234 ($M^+ - CH_3$&bond;$C_{10}H_{21}$, 9), 146 ($M^+ - CH_2OSi(CH_3)_2C_{10}H_{21}$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.10 (6H, s, $SiCH_3$), 0.59 (2H, m, $SiCH_2$), 0.83 (3H, t, $J$ = 6.8 Hz, $CH_3$), 1.2–1.4 (16H, m, $CH_2$), 2.71 (2H, t, $J$ = 5.4 Hz, $\alpha$-$CH_2N$), 2.81 and 2.90 (2H and 2H, t, $J$ = 5.8 Hz, 3- and 4-$CH_2$), 3.70 (2H, s,1-$CH_2N$), 3.71 (2H, t, $J$ = 5.4 Hz, $OCH_2$), 7.0–7.05 (1H, m, 8-H), 7.05–7.2 (3H, m, 5-, 6- and 7-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −2.10 ($SiCH_3$), 14.09 ($CH_3$), 16.31 ($SiCH_2$), 22.66 23.16, 29.07, 29.34, 29.57, 29.64, 31.89 and 33.43 ($CH_2$), 28.98 (4-$CH_2$), 50.73 (3-$CH_2N$), 55.71 (1-$CH_2N$), 58.09 ($OCH_2$), 59.32 ($\alpha$-$CH_2N$), 125.66 (7-C), 126.23 (8-C), 126.49 (6-C), 128.65 (5-C), 134.15 (9-C), 134.46 (10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +18.36. Anal. calcd for $C_{23}H_{41}NOSi$: C, 73.54; H, 11.00; N, 3.73; found: C, 73.28; H, 10.54; N, 3.78.

*N*-(2-*n*-Dimethylundecylsiloxyethyl)-1,2,3,4-tetrahidroisoquinoline (**9 g**)

Compound **9 g** was obtained as a clear yellow oil following the procedure described for **9a**, from 0.76 g (3.54 mmol) *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline (**1**) and 0.91 g (4.25 mmol) dimethylundecylsilane, by heating the reagents at 90 °C under stirring for 15 h.

Yield 0.68 g (49%). LC-MS (*m/z*, %): 390 ($M^+$+1, 100). GC-MS (*m/z*, %): 390 ($M^+$+1, 1), 147 ($M^+ - CH_2OSi(CH_3)_2C_{11}H_{23} + 1$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.07 (6H, s, $SiCH_3$), 0.53 (2H, m, $SiCH_2$), 0.88 (3H, t, $J$ = 6.8 Hz, $CH_3$), 1.25 (18H, m, $CH_2$), 2.71 (2H, t, $J$ = 5.8 Hz, $\alpha$-$CH_2N$), 2.79 and 2.88 (2H and 2H, t and t, $J$ = 5.9 Hz, 3- and 4-$CH_2$), 3.70 (4H, m, 1-$CH_2N$ + $OCH_2$), 7.0–7.2 (4H, m, 5-, 6-, 7- and 8-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −2.21 ($SiCH_3$), 14.03 ($CH_3$), 17.72 ($SiCH_2$), 22.63, 23.11, 29.28, 29.62, 29.74, 31.82 and 33.43 ($CH_2$), 28.92 (4-$CH_2$), 50.71 (3-$CH_2N$), 55.70 (1-$CH_2N$), 58.12 ($OCH_2$), 59.34 ($\alpha$-$CH_2N$), 125.62 (7-C), 126.2 1(8-C), 126.39 (6-C), 128.63 (5-C), 134.02 (9-C), 134.31 (10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +18,23. Anal. calcd for $C_{24}H_{43}NOSi$: C, 73.97; H, 11.12; N, 3.59; found: C, 74.21; H, 11.19; N, 3.52.

*N*-(2-*n*-Hexadecyldimethylsiloxyethyl)-1,2,3,4-tetrahidroisoquinoline (**9h**)

Compound **9 h** was obtained as a clear yellow oil following the procedure described for **9a**, from 0.80 g (4.50 mmol) *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline (**1**) and 2.56 g (9.0 mmol) hexadecyldimethylsilane, by heating the reagents at 85 °C under stirring for 15 h.

Yield 0.78 g (38%). LC-MS (*m/z*, %): 460 ($M^+$+1, 100), 459 ($M^+$, 5), 160 ($M^+ - OSi(CH_3)_2C_{16}H_{33}$, 96). GC-MS (*m/z*, %): 459 ($M^+$, 2), 444 ($M^+ - CH_3$, 3), 234 ($M^+ - C_{16}H_{33}$, 9), 146 ($M^+ - CH_2OSi(CH_3)_2C_{16}H_{33}$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.10 (6H, s, $SiCH_3$), 0.60 (2H, m, $SiCH_2$), 0.87 (3H, t, $J$ = 6.8 Hz, $CH_3$), 1.24 (28H, m, $CH_2$), 2.71 (2H, t, $J$ = 6.0 Hz, $\alpha$-$CH_2N$), 2.80 and 2.87 (2H and 2H, m and m, 3- and 4-$CH_2$), 3.71 (2H, s, 1-$CH_2N$), 3.81 (2H, t, $J$ = 6.0 Hz, $OCH_2$), 7.00 (1H, m, 8-H), 7.10 (3H, m, 5-, 6- and 7-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −4.42 ($SiCH_3$), 14.13 ($CH_3$), 16.41 $SiCH_2$), 23.25, 24.42, 29.38, 29.61, 29.73, 31.94, 33.21 and 33.52 ($CH_2$), 29.02 (4-$CH_2$), 51.62 (3-$CH_2N$), 56.81 (1-$CH_2N$), 60.23 ($OCH_2$), 60.92 ($\alpha$-$CH_2N$), 125.51 (7-C), 126.13 (8-C), 126.52 (6-C), 128.69 (5-C), 134.2 2 (9-C), 134.83 (10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +18.33. Anal. calcd for $C_{29}H_{53}NOSi$: C, 75.75; H, 11.62; N, 3.05; found: C, 75.98; H, 11.69; N, 3.01.

*N*-(2-Methyldiphenylsiloxyethyl)-1,2,3,4-tetrahydroisoquinoline (**9i**)

Compound **9i** was obtained as a clear yellow oil following the procedure described for **9a**, from 0.63 g (3.8 mmol) *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline (**1**) and 1.0 g (5.0 mmol) methyldiphenylsilane.

Yield 1.22 g (91%). LC-MS (*m/z*, %): 374 ($M^+$+1, 100), 296 ($M^+ - C_6H_5$, 98). GC-MS (*m/z*, %): 372 ($M^+ - 1$, 2), 295 ($M^+ - C_6H_5$ 1, 9), 146 ($M^+ - CH_2OSi(C_6H_5)_2CH_3$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.55 (3H, s, $SiCH_3$), 2.64 (4H, m, 3- and 4-$CH_2$), 2.74 (2H, t, $J$ = 6.0 Hz, $\alpha$-$CH_2N$), 3.52 (2H, s, 1-$CH_2N$), 3.81 (2H, t, $J$ = 6.0 Hz, $OCH_2$), 6.84 (1H, m, 8-H), 6.98 (3H, m, 5-,6- and 7-H), 7.2–7.5 (10H, m, $C_6H_5$-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −2.98 ($SiCH_3$), 29.04 (4-$CH_2$), 51.49 (3-$CH_2N$), 56.41 (1-$CH_2N$), 60.02 ($OCH_2$), 61.68 ($\alpha$-$CH_2N$), 125.42 (7-C), 126.03 (8-C), 126.51 (6-C), 128.66 (5-C), 134.12 (9-C), 134.71 (10-C), 127.81, 129.77, 134.32, 137.51 ($C_6H_5$-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): −2.39. Anal. calcd for $C_{24}H_{27}NOSi$: C, 77.16; H, 7.28; N, 3.75; found: C, 76.85; H, 7.20; N, 3.81.

*N*-(2-*n*-Decyldimethylsiloxyethyl)-1,2,3,4-tetrahydroquinoline (**10**)

Compound **10** was obtained as clear yellow oil following the procedure described for **9a**, from 3.24 g (18.3 mmol) *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroquinoline (**2**) and 4.02 g (20.1 mmol) decyldimethylsilane, by heating the reagents at 80 °C under stirring for 26 h.

Yield 1.66 g (24%). LC-MS (*m/z*, %): 376 ($M^+$+1, 50), 375 ($M^+$, 100). GC-MS (*m/z* %): 375 ($M^+$, 11), 234 ($M^+ - C_{10}H_{21}$, 3), 146 ($M^+ - CH_2O$-$SiMe_2C_{10}H_{21}$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.07 (6H, s, $SiCH_3$), 0.57 (2H, m, $SiCH_2$), 0.87 (3H, t, $J$ = 6.8 Hz, $CH_3$), 1.25–1.30 (16H, m, $CH_2$), 1.94 (2H, m, 3-$CH_2N$), 2.77 (2H, m, 4-$CH_2$), 3.32 (2H, m, 2-$CH_2N$), 3.39 (2H, t, $J$ = 6.8 Hz, $\alpha$-$CH_2N$), 3.73 (2H, t, $J$ = 6.8 Hz, $OCH_2$),. 6.58, 6.67, 6.93 and 7.02 (total 4H, m, dd, m and m, 5-, 6-,7- and 8-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −2.14 ($SiCH_3$), 14.11 ($CH_3$), 16.30 ($SiCH_2$), 22.24 (3-$CH_2$), 22.69, 23.18, 29.34, 29.35, 29.60, 29.67, 31.93 and 33.22 ($CH_2$), 28.18 (4-$CH_2$), 50.42 (2-$CH_2N$), 53.63 ($\alpha$-$CH_2N$), 59.13 ($OCH_2$), 110.40, 115.49, 127.06, 129.16 (5-, 6-, 7-, 8-C), 122.06 (9-C) and 145.24 (10-C). $^{29}$Si ($\delta$, ppm): +18.62 Anal. calcd for $C_{23}H_{41}NOSi$: C, 73.54; H, 11.00; N, 3.73; found: C, 73.28; H, 11.04; N, 3.68.

*N*-(2-*n*-Decyldimethylsiloxyethyl)-4,4-dimethyl-4-sila-1,2,3,4-tetrahidroisoquinoline (**11**)

Compound **11** was obtained as a clear yellow oil following the procedure described for **9a**, from 0.4 g (1.4 mmol) *N*-(2-hydroxyethyl)-4,4-dimethyl-4-sila-1,2,3,4-tetrahydroisoquinoline (**7**) and 0.31 g (1.54 mmol) of decyldimethylsilane.

Yield 0.13 g (22%). LC-MS (*m/z*, %): 420 ($M^+$+1, 100). GC-MS (*m/z*, %): 389 ($M^+ - 2CH_3$, 33), 388 ($M^+ - 2CH_3 - 1$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.09 (6H, s, $SiCH_3$-acycl.), 0.27 (6H, s, $SiCH_3$-cycl.), 0.57 (2H, m, $SiCH_2$), 0.87 (3H, t, $J$ = 6.8 Hz, $CH_3$), 1.25 (16H, m, $CH_2$), 2.21 (2H, s, $SiCH_2N$), 2.72 (2H, t, $J$ = 5.2 Hz, $\alpha$-$CH_2N$), 3.65 (2H, m, $OCH_2$), 3.67 (2H, s, 1-$CH_2N$), 7.0–7.5 (4H, m, 5-, 6-, 7- and 8-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −2.12 ($SiCH_3$), −2.01 ($SiCH_3$-cycl.), 14.1 ($CH_3$), 16.32 ($SiCH_2$), 22.66, 23.18, 29.34, 29.37, 29.59, 29.66, 31.91 and 33.20 ($CH_2$), 44.15 ($SiCH_2N$), 58.1 (1-$CH_2N$), 61.2 ($OCH_2$), 63.5 ($\alpha$-$CH_2N$), 125.14, 126.61, 128.94, 133.61 (5-, 6-, 7- and 8-C), 134.35, 145.42 (9,10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): −13.50, +18.06. Anal. calcd for $C_{24}H_{45}NOSi_2$: C, 68.67; H, 10.80; N, 3.33; found: C, 68.91; H, 10.73; N, 3.25.

*N*-(2-Diethylmethylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium iodide (**12a**)

A solution of diethylmethylsilyl ether **9a** (0.84 g, 3.03 mmol) in hexane (5 ml) was heated at 60 °C for 5 h with methyl iodide (2.22 g, 15.60 mmol). After cooling to room temperature the reaction mixture was filtered, and the solid was washed with hexane

and dried for 6 h *in vacuo* to give the desired methiodide as a yellow powder.

Yield 0.41 g (33%), m.p. 138–139 °C. LC-MS (*m/z* %): 293 ($M^+ - I + 1$, 19), 292 ($M^+ - I$, 100), 197 (2), 188 (9), 183 (5). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.08 (3H, s, $SiCH_3$), 0.59 (4H, q, $J = 7.9$ Hz, $SiCH_2$), 0.91 (6H, t, $J = 7.9$ Hz, SiC-$CH_3$), 3.24 (2H, bt, $J = 6.1$ Hz, 4-$CH_2$, 3.54 (3H, s, $N^+CH_3$), 3.9–4.3 (6H, m, α-$CH_2N^+$, 3-$CH_2N^+$ and $OCH_2$), 4.74 and 5.07 (total 2H, d and d, $J = 15.1$ Hz, 1-$CH_2N^+$), 7.1–7.4 (4H, m, 5-, 6-, 7- and 8-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −5.12 ($SiCH_3$), 5.77 ($SiCH_2$), 6.65 (SiC&–$CH_3$), 23.86 (4-C), 49.06 ($N^+CH_3$), 57.23 ($OCH_2$), 59.80 (3-$CH_2N^+$), 62.87 (1-$CH_2N^+$), 63.63 (α-$CH_2N^+$), 125.96, 127.34, 127.81, 128.54, 128.90 and 129.06 (5-,6-,7-,8-,9- and 10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +24.23 ($^2J_{SiCH} = 6.0$ Hz). Anal. calcd for $C_{17}H_{30}INOSi$: C, 48.68; H, 7.21; N, 3.34; found: C, 48.42; H, 7.15; N 3.41.

*N*-Methyl-*N*-(2-triethylsiloxyethyl)-1,2,3,4-tetrahydroisoquinolinium iodide (**12b**)

A solution of triethylsilyl ether **9b** (1.44 g, 8.13 mmol) and of methyl iodide (5.75 g, 4.05 mmol) in ethyl ether (23 ml) was heated at 40 °C for 3 h, the precipitate was filtered off, washed with 50 ml hexane and dried for 6 h *in vacuo* to give the desired product as a light yellow powder.

Yield 1.86 g (67%), m.p. 93–95 °C. LC-MS (*m/z*, %): 307 ($M^+ - I + 1$, 25), 306 ($M^+ - I$, 100), 203 ($M^+ - CH_3I - 3C_2H_5 - 1$, 79), 179 (22), 147 (39). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.64 (6H, q, $J = 7.8$ Hz, $SiCH_2$), 0.94 (9H, t, $J = 7.8$ Hz, SiC&bond;$CH_3$), 3.27 (2H, bt, $J = 5.8$ Hz, 4-$CH_2$), 3.58 (3H, s, $N^+CH_3$), 3.9–4.3 (6H, m, α-$CH_2N^+$, 3-$CH_2$ and $OCH_2$), 4.79 and 5.11 (total 2H, d and d, $J = 15.8$ Hz, 1-$CH_2N^+$), 7.15 (1H, m, 8-H), 7.2–7.4 (3H, m, 5-, 6- and 7-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): 3.95 ($SiCH_2$), 6.62 (SiC-$CH_3$), 23.80 (4-C), 49.07 ($N^+CH_3$), 57.3 ($OCH_2$), 59.62 (3-C), 62.66 (1-$CH_2N^+$), 63.45 (α-$CH_2N^+$), 125.98, 127.22, 127.62, 128.56, 128.79 and 128.87 (5-,6-,7-,8-,9- and 10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +23.84. Anal. calcd for $C_{18}H_{32}INOSi$: C, 49.88; H, 7.44; N, 3.23; found: C, 49.59; H, 7.37; N 3.29.

*N*-(2-Di-*n*-butylethylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium iodide (**12c**)

Compound **12c** was obtained following the procedure described for **12b**, from 0.90 g (2.58 mmol) dibutylethylsilyl ether **9c**, as a light-yellow oily powder.

Yield 0.52 g (41%), m.p. 124–126 °C. LC-MS (*m/z*, %): 363 ($M^+ - I$, 72), 362 ($M^+ - I - 1$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.63 (6H, m, $SiCH_2$), 0.92 (13H, m, SiC&bond;$CH_3$, $SiCCH_2CC$ and $SiCCCCH_3$), 1.74 (4H, m, $SiCCCH_2C$), 3.26 (2H, bt, $J = 6.2$Hz, 4-$CH_2$), 3.55 (3H, s, $N^+CH_3$), 4.00–4.15 (6H, m, α-$CH_2N^+$, 3-$CH_2N^+$, $OCH_2$), 4.79 and 5.06 (total 2H, d and d, $J = 15.1$ Hz, 1-$CH_2N^+$), 7.1–7.3 (4H, m, 5-, 6-, 7- and 8-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): 6.04 ($SiCH_2$) and 6.89 ($SiCCH_3$), 23.65, 24.04 and 26.21 ($CCH_2$ $CH_2$ $CH_3$ and 4-C), 48.94 ($N^+CH_3$), 57.14 ($OCH_2$), 59.62 (3-C), 62.72 (1-$CH_2N^+$), 63.63 (α-$CH_2N^+$), 125.92, 127.21, 127.65, 128.52, 128.81 and 128.91 (5-,6-,7-,8-,9- and 10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +21.49. Anal. calcd for $C_{22}H_{40}INOSi$: C, 53.98; H, 8.24; N, 2.86; found: C, 53.70; H, 8.16; N 2.93.

*N*-(2-Di-*n*-heptylmethylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium iodide (**12d**)

Compound **12d** was obtained following the procedure described for **12b**, from 0.65 g (1.56 mmol) diheptylmethylsilyl ether **9d**, as a yellow oily product.

Yield 0.52 g (44%), m.p. 104–106 °C. LC-MS (*m/z*, %): 433 ($M^+ - I + 1$, 40), 432 ($M^+ - I$, 100). GC-MS (*m/z*, %): 417 ($M^+ - CH_3I$, 2), 402 ($M^+ - CH_3I - CH_3$, 2), 318 ($M^+ - CH_3I - C_7H_{15}$, 5); 146 ($M^+ - CH_3I - CH_2OSiCH_3(C_7H_{15})_2$, 100), 132 ($M^+ - CH_3I - CH_2CH_2OSiCH_3(C_7H_{15})_2$, 26). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.08 (3H, s, $SiCH_3$), 0.60 (4H, m, $SiCH_2$), 0.86 (6H, t, $J = 6.8$Hz, $2CH_3$), 1.24 (20H, m, $10CH_2$), 3.25 (2H, m, 4-$CH_2$), 3.56 (3H, s, $N^+CH_3$), 3.99–4.2 (6H, m, α-$CH_2N^+$, 3-$CH_2N^+$, $OCH_2$), 4.76 and 5.10 (total 2H, d and d, $J = 15.0$ Hz, 1-$CH_2N^+$); 7.1–7.4 (4H, m, 5-, 6-, 7- and 8-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −4.08 ($SiCH_3$), 14.12 ($CH_3$), 15.03 ($SiCH_2$), 23.12, 28.89, 31.73 and 33.42 ($CH_2$), 23.92 (4-$CH_2$), 49.21 ($N^+CH_3$), 57.24 ($OCH_2$), 59.78 (3-$CH_2N^+$), 62.73 (1-$CH_2N$), 63.41 (α-$CH_2N^+$), 126,11, 127.32, 127.70, 128.6, 129.02 and 129.91 (5-, 6-, 7-, 8-, 9- and 10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +22.35. Anal. calcd for $C_{27}H_{50}INOSi$: C, 57.94; H, 9.00; N, 2.50; .found: C, 57.68; H, 8.86; N 2.58.

*N*-(2-*n*-Dimethyloctylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium iodide (**12e**)

Compound **12e** was obtained following the procedure described for **12b**, from 0.96 g (2.76 mmol) of dimethyloctylsilyl ether **9e**, as a yellow powder.

Yield 0.36 g (27%), m.p. 144–145 °C. LC-MS (*m/z*, %): 363 ($M^+ - I + 1$, 100); 346 ($M^+ - CH_3I - 1$, 5). GC-MS (*m/z*, %): 332 ($M^+ - CH_3I - CH_3$, 5), 146 ($M^+ - CH_3I - CH_2OSi(CH_3)_2C_8H_{17}$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.11 (6H, s, $SiCH_3$), 0.58 (2H, m, $SiCH_2$), 0.87 (3H, t, $J = 7.0$ Hz, $CH_3$), 1.2–1.3 (12H, m, $CH_2$), 3.2 (2H, m, 4-$CH_2$), 3.56 (3H, s, $N^+CH_3$), 3.9–4.2 (6H, m, α-$CH_2N^+$, 3-$CH_2N^+$, $OCH_2$), 4.77 and 5.07 (total 2H, d and d, $J = 14.9$ Hz, 1-$CH_2N^+$), 7.11 (1H, d, 8-H), 7.2–7.4 (3H, m, 5-, 6- and 7-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −2.57 ($SiCH_3$), 13.93 ($CH_3$), 15.78 ($SiCH_2$), 22.52, 22.96, 29.02, 29.16, 31.78 and 33.27 ($CH_2$), 23.84 (4-$CH_2$), 49.13 ($N^+CH_3$), 56.91 ($OCH_2$), 59.59 (3-$CH_2N^+$), 62.54 (1-$CH_2N^+$), 63.42 (α-$CH_2N^+$), 126,03, 127.24, 127.58, 128.62, 128.77 and 128.85 (5-, 6-, 7-, 8-, 9- and 10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +22.37. Anal. calcd for $C_{22}H_{40}INOSi$: C, 53.98; H, 8.24; N, 2.8; found: C, 53.79; H, 8.15; N 2.93.

*N*-(2-*n*-Decyldimethylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium iodide (**12f**)

Compound **12f** was obtained following the procedure described for **12b**, from 0.23 g (0.80 mmol) decyldimethylsilyl ether **9f**, as a yellow oily precipitate.

Yield 0.28 g (68%), m.p. 142 °C. LC-MS (*m/z*, %): 391 ($M^+ - I + 1$, 15), 390 ($M^+ - I$, 55), 389 ($M^+ - I - 1$, 95), 376 ($M^+ - CH_3I + 1$, 100). GC-MS (*m/z*, %): 391 ($M^+ - I + 1$, 1), 234 ($M^+ - CH_3I - C_{10}H_{21}$, 4), 146 ($M^+ - CH_3I - CH_2OSi(CH_3)_2C_{10}H_{21}$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.11 (6H, s, $SiCH_3$), 0.58 (2H, m, $SiCH_2$), 0.85 (3H, t, $J = 6.8$ Hz, $CH_3$), 1.24 (16H, m, $CH_2$), 3.26 (2H, m, 4-$CH_2$), 3.54 (3H, s, $N^+CH_3$), 4.0–4.2 (6H, m, α-$CH_2N^+$ 3-$CH_2N^+$, $OCH_2$), 4.76 and 5.08 (total 2H, d and d, $J = 15.3$ Hz, 1-$CH_2N^+$), 7.11 (1H, d, 8-H), 7.1–7.4 (3H, m, 5-, 6- and 7-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −2.43 ($SiCH_3$), 14.01 ($CH_3$), 15.78 ($SiCH_2$), 22.61, 23.02, 29.24, 29.38, 29.49, 31.85 and 33.22 ($CH_2$), 23.82, (4-C), 49.21 ($N^+CH_3$), 57.03 ($OCH_2$), 59.67 (3-C), 62.71 (1-$CH_2N^+$), 63.52 (α-$CH_2N^+$), 126.02 (9-C), 127.31 (8-C), 127.68 (5-C), 128.57 (6-C), 128.92 (7-C), 129.03 (10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +22.50. Anal. calcd for $C_{24}H_{44}INOSi$: C, 55.69; H, 8.57; N, 2.71; found: C, 55.84; H, 8.66; N 2.55.

*N*-(2-*n*-Dimethylundecylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium iodide (**12g**)

Compound **12 g** was obtained following the procedure described for **12b**, from 0.11 g (0.28 mmol) dimethylundecylsilyl ether **9 g**, as a yellow oily precipitate.

Yield 0.06 g (40%), m.p. 92–94 °C. LC-MS (*m/z*, %): 405 ($M^+ - I + 1$, 11), 404 ($M^+ - I$, 58), 390 ($M^+ - CH_3I + 1$, 100), 346 (2), 181 (2), 160

(42). GC-MS (*m/z*, %): 530 ($M^+ - 1$, 3), 374 ($M^+ - CH_3I - CH_3$, 2), 146 ($M^+ - CH_3I - CH_2OSi(CH_3)_2C_{11}H_{23}$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.10 (6H, s, $SiCH_3$), 0.57 (2H, m, $SiCH_2$), 0.85 (3H, t, $J = 6.8$Hz, $CH_3$), 1.23 (18H, m, $CH_2$), 3.24 (2H, m, 4-$CH_2$), 3.54 (3H, s, $N^+CH_3$), 4.0–4.2 (6H, m, $\alpha$-$CH_2N^+$, 3-$CH_2N^+$, $OCH_2$), 4.76 and 5.08 (total 2H, d and d, $J = 14.3$ Hz, 1-$CH_2N^+$), 7.1–7.4 (4H, m, 5-, 6-, 7- and 8-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −2.40 ($SiCH_3$), 14.03 ($CH_3$), 15.82 ($SiCH_2$), 22.61, 23.1, 29.32, 29.48, 29.62, 31.81 and 33.34 ($CH_2$), 23.82 (4-$CH_2$), 49.12 ($N^+CH_3$), 57.03 ($OCH_2$), 59.76 (3-$CH_2N^+$), 62.79 (1-$CH_2N^+$), 63.65 ($\alpha$-$CH_2N^+$), 126.01 (9-C), 127.28 (8-C), 127.82 (5-C), 128.63 (6-C), 128.87 (7-C) and 129.02 (10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +22.67 ($^2J_{SiCH} = 6.8$ Hz). Anal. calcd for $C_{25}H_{46}INOSi$: C, 56.48; H, 8.72; N, 2.64; found: C, 56.72; H, 8.83; N 2.57.

*N*-(2-*n*-Hexadecyldimethylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium iodide (**12 h**)

Compound **12 h** was obtained following the procedure described for **12b**, from 0.67 g (1.5 mmol) hexadecyldimethylsilyl ether **9 h**, as a yellow powder.

Yield 0.80 g (89%), m.p. 59 °C. LC-MS (*m/z*, %): 475 ($M^+ - I + 1$, 100), 474 ($M^+ - I$, 67), 460 ($M^+ - CH_3I + 1$, 49). GC-MS (*m/z*, %): 460 ($M^+ - CH_3I + 1$, 3), 445 ($M^+ - CH_3I - CH_3 + 1$, 4), 234 ($M^+ - CH_3I$ $C_{16}H_{33}$, 11), 146 ($M^+ - CH_3I - CH_2OSi(CH_3)_2C_{16}H_{33}$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.11 (6H, s, $SiCH_3$), 0.58 (2H, m, $SiCH_2$), 0.86 (3H, t, $J = 7.0$ Hz, $CH_3$), 1.24 (28H, bs, $CH_2$), 3.25 (2H, m, 4-$CH_2$), 3.56 (3H, s, $N^+CH_3$), 3.9–4.2 (6H, m, $\alpha$-$CH_2N^+$, 3-$CH_2N^+$, $OCH_2$), 4.78 and 5.07 (total 2H, d and d, $J = 15.4$ Hz, 1-$CH_2N^+$), 7.11 (1H, d, 8-H), 7.1–7.4 (3H, m, 5-, 6- and 7-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −2.4 ($SiCH_3$), 14.01, 15.79, 22.57, 22.98, 23.83, 29.24, 29.48, 29.54, 29.6, 31.81, 33.26 (4-C, $CH_2$, $CH_3$, $SiCH_2$), 49.10 ($N^+CH_3$), 57.00 ($OCH_2$), 59.68 (3-C), 62.65 (1-$CH_2N^+$), 63.49 ($\alpha$-$CH_2N^+$), 126.01 (9-C), 127.29 (8-C), 127.68 (5-C), 128.58 (6-C), 128.83 (7-C), 128.94 (10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +22.49. Anal. calcd for $C_{30}H_{56}INOSi$: C, 59.88; H, 9.38; N, 2.33; found: C, 59.63; H, 9.32; N, 2.40.

*N*-Methyl-*N*-(2-methyldiphenylsiloxyethyl)-1,2,3,4-tetrahydroisoquinolinium iodide (**12i**)

Compound **12i** was obtained following the procedure described for **12b**, from 1.22 g (3.3 mmol) methyldiphenylsilyl ether **9i**, as a white powder.

Yield 0.41 g (24%), m.p. 96 °C. LC-MS (*m/z*, %): 388 ($M^+ - I$, 100). GC-MS (*m/z*, %): 516 ($M^+$+1, 1), 373 ($M^+ - CH_3I$, 4), 295 ($M^+ - CH_3I$ $C_6H_5 - 1$, 12), 146 ($M^+ - CH_3I - CH_2OSiCH_3(C_6H_5)_2$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.67 (3H, s, $SiCH_3$), 3.17 (2H, m, 4-$CH_2$), 3.50 (3H, s, $N^+CH_3$), 4.0–4.2 (6H, m, $\alpha$-$CH_2$, 3-$CH_2$, $OCH_2$), 4.73 and 5.02 (total 2H, d and d, $J = 14.9$ Hz, 1-$CH_2N^+$), 6.9–7.2 (4H, m, 5-, 6-, 7- and 8-H), 7.3–7.5 (10H, m, $C_6H_5$-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −3.60 ($SiCH_3$), 23.65 (4-$CH_2$), 49.12 ($N^+CH_3$), 57.41 $OCH_2$), 59.49 (3-$CH_2N^+$), 62.22 (1-$CH_2N^+$), 63.13 ($\alpha$-$CH_2N^+$), 125.81, 127.09, 127.41, 128.42, 128.63 and 128.68 (5-, 6-, 7-, 8-, 9- and 10-C), 128.14, 130.42, 133.61, 133.65, 134.04 ($C_6H_5$&bond;C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): −0.36. Anal. calcd for $C_{25}H_{30}INOSi$: C, 58.25; H, 5.87; N, 2.72; found: C, 58.52; H, 5.78; N 2.64.

*N*-(2-*n*-Decyldimethylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroquinolinium iodide (**13**)

Compound **13** was obtained following the procedure described for **12b**, from 1.574 g (4.20 mmol) decyldimethylsilyl ether **10**, as an oil.

Yield 0.772 g (35%). LC-MS (*m/z*, %): 391 ($M^+ - I + 1$, 17), 390 ($M^+ - I$, 54), 389 ($M^+ - I$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.11 (6H, s, $SiCH_3$), 0.58 (2H, m, $SiCH_2$), 0.87 (3H, t, $J = 6.8$ Hz, $CH_3$), 1.24 (16H, m, $CH_2$), 2.25 and 2.39 (total 2H, m and m, 3-$CH_2N$), 3.00 and 3.06 (total 2H, m and m, 4-$CH_2$), 3.84 (3H, s, $N^+CH_3$), 3.92 and 4.59 (total 2H, m and m, 2-$CH_2N^+$), 3.92 and 4.13 (total 2H, m and m, $\alpha$-$CH_2N^+$), 4.33 and 4.42 (total 2H, m and m, $OCH_2$), 7.3–7.5 (3H, m, 5-, 6- and 7-H), 7.91 (1H, d, $J = 8.3$, 8-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −2.42 ($SiCH_3$), 14.10 ($CH_3$), 15.82 ($SiCH_2$), 17.63 (3-$CH_2$), 22.67, 23.65, 28.69 , 29.28, 29.65, 31.85 and 33.30 ($CH_2$), 25.82 (4-C), 56.28 ($\alpha$-$CH_2N^+$), 57.25 ($N^+CH_3$), 62.35 (2-C), 65.87 ($OCH_2$), 121.98, 128.65, 129.31, 131.39 (5-, 6-, 7-, 8-C), 130.80 (9-C) and 141.69 (10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): +22.63. Anal. calcd for $C_{24}H_{44}INOSi$: C, 55.69; H, 8.57; N, 2.71; found: C, 55.52; H, 8.51; N 2.75.

*N*-(2-*n*-Decyldimethylsiloxyethyl)-4,4-dimethyl-*N*-methyl-4-sila-1,2,3,4-tetrahidroisoquinolinium iodide (**14**)

Compound **14** was obtained following the procedure described for **12b**, from 0.13 g (0.31 mmol) decyldimethylsilyl ether **11**, as a dark-yellow powder.

Yield 0.04 g (24%), m.p. 46 °C. LC-MS (*m/z*, %): 434 ($M^+ - I$, 100). $^1$H NMR ($CDCl_3$, $\delta$, ppm): 0.10 (6H, s, $SiCH_3$-acycl.), 0.47 (6H, m, $SiCH_3$-cycl.), 0.57 (2H, m, $SiCH_2$), 0.86 (3H, t, $J = 6.8$ Hz, $CH_3$), 1.24 (16H, m, $CH_2$), 3.42 (3H, s, $N^+CH_3$), 3.52 and 3.68 (1H and 1H, d and d, $J = 14.8$ Hz, $SiCH_2N^+$), 3.88 (2H, t, $J = 4.7$ Hz, $\alpha$-$CH_2N^+$), 4.18 (2H, m, $OCH_2$), 4.86 and 4.96 (1H and 1H, d and d, $J = 14.5$ Hz, 1-$CH_2N^+$), 7.4–7.8 (4H, bm, 5-, 6-, 7- and 8-H). $^{13}$C NMR ($CDCl_3$, $\delta$, ppm): −2.42 ($SiCH_3$), −1.25 ($SiCH_3$-cycl.), 14.00 ($CH_3$), 15.82 ($SiCH_2$), 22.55, 22.96, 23.82, 29.20, 29.45, 29.52, 31.78, and 33.23 ($CH_2$), 53.22 ($N^+CH_3$), 55.45 ($SiCH_2N^+$), 55.72 ($OCH_2$), 67.41 ($ArCH_2N^+$), 67.95 ($\alpha$-$CH_2N^+$ ), 129.33, 129.77, 131.23, 131.28, 133.96, 135.17 (5-, 6-, 7-, 8-, 9- and 10-C). $^{29}$Si NMR ($CDCl_3$, $\delta$, ppm): −13.46, +22.23. Anal. calcd for $C_{25}H_{48}INOSi_2$: C, 53.45; H, 8.61; N, 2.49; found: C, 53.20; H, 8.50; N 2.57.

## Biological Tests

*Cytotoxicity*

Monolayer tumour cell lines HT-1080 (human fibrosarcoma), MG-22A (mouse hepatoma) and normal mouse fibroblasts (NIH 3 T3) were cultivated for 72 h in Dulbecco's modified Eagle's medium (DMEM) standard medium (Sigma) without an indicator and antibiotics.[35]

Tumour cell lines were taken from the European Collection of Cell Culture (ECACC). After the ampoule was thawed not more than four passages were performed. The control cells and cells with tested substances in the range of $2–5 \times 10^4$ cells $ml^{-1}$ concentration (depending on line nature) were placed on separate 96-well plates. The volume of each plate was 200 µl. Solutions containing test compounds were diluted and added to wells to give final concentrations of 50, 25, 12.5 and 6.25 µg $ml^{-1}$. Control cells were treated in the same manner but in the absence of test compounds. The plates were incubated for 72 h at 37 °C and 5% $CO_2$. The number of surviving cells was determined using tri(4-dimethylaminophenyl) methyl chloride (crystal violet, CV) or 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2*H*-tetrazolium bromide (MTT) coloration, which was assayed by multiscan spectrophotometer. The quantity of live cells on the control plate was taken in calculations for 100%.[35,36] The $LC_{50}$ was calculated using the Graph Pad Prism® 3.0 program, $r < 0.05$. Concentration of NO was determined according to the procedure described in Fast *et al.*[36]

*Antimicrobial activity*

For the determination of antimicrobial activity, several reference microbial strains, received from the Microbial Strain Collection

of Latvia (MSCL), Riga, Latvia, were used: *Staphylococcus aureus* MSCL 334 (SA), *Bacillus cereus* MSCL 330 (BC), *Proteus mirabilis* MSCL 590 (PM), *Escherichia coli* MSCL 332 (EC), *Pseudomonas aeruginosa* MSCL 331 (PA) and *Candida albicans* MSCL 378 (CA). All bacteria were cultivated on Plate Count Agar (Sanofi Diagnostics Pasteur, France) at 37 °C for 24 h. *Candida albicans* was cultivated on Difco™ Malt Extract Agar (Becton, Dickinson and Co., UK) at 37 °C for 48 h. Antimicrobial activity was determined by agar well diffusion method.[37] The agar diffusion test was performed on Mueller–Hinton (Carl Roth GmbH + Co. KG, Germany) agar for bacteria and Malt Extract Agar for yeast. Suspensions of 18–24 h microbial cultures of turbidity $A_{540} = 0.16 \pm 0.20$ were used and uniformly spread on Petri plates. Aliquots of 70 μl of each test sample solution, corresponding solvent and reference antimicrobial drugs solutions were added to 6.0 μm diameter agar wells. Gentamicin (KRKA, Slovenia) and fluconazole (Diflucan, Pfizer Ltd, UK), 10 mg ml$^{-1}$ and 5 mg ml$^{-1}$, were used as reference antibiotics. The antimicrobial activity was evaluated based on the diameter of zone of inhibition. After incubation at 37 °C for 24 h for bacteria and 48 h for yeast under aerobic conditions, the diameter of the clear zone (no growth) around the well in the bacterial lawn was measured. The inhibition zone diameter was measured in millimetres (mm). The tests were performed in duplicate. The results were expressed as the arithmetic average. The observed zones of growth inhibition are presented in Table 4.

## Results and Discussion

### Chemistry

The synthesis of the compounds used in this study is outlined in Schemes 1 and 2.

*N*-Methyl-*N*-(2-triorganylsiloxyethyl)-1,2,3,4-tetrahydroisoquinolinium iodides (**12a–i**) were envisioned as being prepared from the common precursor alcohol **1**[34] in moderate to good overall yields (24–89%). *N*-(2-Decyldimethylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroquinolinium iodide (**13**) was prepared analogously from **2**. *N*-(2-Decyldimethylsiloxyethyl)-4,4-dimethyl-*N*-methyl-4-sila-1,2,3,4-tetrahydroisoquinolinium iodide (**14**) was synthesized by a series of sequenced reactions starting from 2-bromotoluene (Schemes 1 and 2). The reaction of dehydrocondensation,[38] a special feature of hydrosilanes in their ability to undergo alcoholysis leading to alkoxysilanes and gaseous hydrogen, was used in coupling the silyl group to the organic substrates **1**, **2** and **7**.

The interaction of *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline (**1**) with triorganylhydrosilanes $R^1R^2R^3SiH$ resulted in 38–91% yield. It was revealed using GC-MS monitoring that under the same reaction conditions the yield of the reaction products depends on both the nature of hydrosilane and the nature of the heterocycle.

The influence of nature of the heterocyclic base on the reaction result was more evident, provided the same triorganylsilane was used and the same reaction conditions: *N*-(2-decyldimethysiloxyethyl)-1,2,3,4-tetrahydroisoquinoline (**9f**) was obtained with a yield of 75% and the yield of *N*-(2-decyldimethysiloxyethyl)-1,2,3,4-tetrahydroquinoline (**10**), in its turn, was significantly lower: 24%. In the latter case the reaction proceeded more slowly and it took a longer time in comparison with the *N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinoline: 26 and 15 h, correspondingly, to obtain the product with the highest yield. During dehydrocondensation of **1** with different trialkylhidrosilanes, the tendency of the reaction yield decrease has been observed with carbon chain elongation of the alkyl substituent at the Si atom. The yield of compounds **9a–f**, possessing $C_2$–$C_{10}$ alkyl chain length, ranged within 56–77%; for compounds **9 g** ($C_{11}$) and **9 h** ($C_{16}$) it was lower – 49% and 38%, respectively – and the reaction time for **9f** and **9 h** was longer.

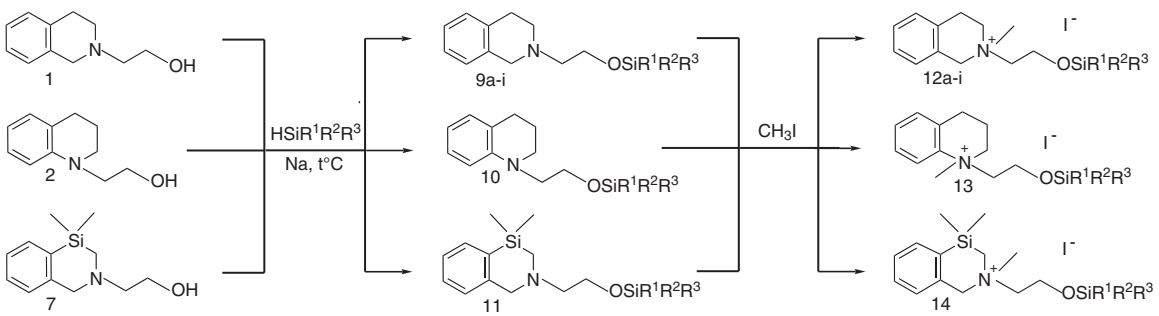
The compounds synthesized were characterized by multinuclear NMR data. The values of $^{29}$Si NMR chemical shifts are presented in Tables 1 and 2. $^{29}$Si NMR resonance depends upon the substituent at the silicon atom, and qualitatively similar changes for $^{29}$Si chemical shifts of tetrahydroisoquinoline derivatives **9a–h** and their methiodides **12a–h** during transition along the sequence should be noted:

$$EtBu_2 < Hp_2Me < OctMe_2 \leq DcMe_2 \leq (C_{11}H_{23})Me_2 \leq (C_{16}H_{33})Me_2 < Et_3 < Et_2Me$$

The low field shifting of $^{29}$Si NMR resonance from 16.74–19.90 to 21.49–24.23 ppm and the narrowing of its interval from 3.16 to 2.74 ppm have been noted for compounds containing positively charged nitrogen (**12a–h**) in comparison with proper *N*-(2-triorganylsiloxyethyl)-1,2,3,4-tetrahydroisoquinolines (**9a–h**). The highest changes in $\delta$ ($^{29}$Si) values (4.75 ppm) under nitrogen quaternization were observed for compound **12c** with ethyldibutyl substituent, revealing stronger N. . .Si interaction in the appropriate silane.

### Biological evaluation

*In vitro* antitumour and antimicrobial properties of the compounds synthesized were investigated. A comparison of **8**, **12a–i** and **13** with **3** and **4** provided evidence that the latter compounds




$R^1R^2R^3$: $Et_2Me$ (9a, 12a); $Et_3$ (9b, 12b); $EtBu_2$ (9c, 12c); $MeHp_2$ (9d, 12d); $Me_2Oct$ (9e, 12e); $Me_2Dc$ (9f, 10, 11, 12f, 13, 14); $Me_2(C_{11}H_{23})$ (9g, 12g); $Me_2(C_{16}H_{33})$ (9h, 12h); $MePh_2$ (9i, 12i)

**Scheme 2.** Synthesis of compounds **9–14**.

**Table 1.** $^{29}$Si NMR resonance of, *N*-(2-triorganosiloxyethyl)-1,2,3,4-tetrahydroisoquinolines (**9a–i**), *N*-(2-decyldimethylsiloxyethyl)-1,2,3,4-tetrahydroquinoline (**10**) and *N*-(2-decyldimethylsiloxyethyl)-4,4-dimethyl-4-sila-1,2,3,4-tetrahydroisoquinoline (**11**) $R^1R^2R^3SiOCH_2CH_2–Het_N$

| Compound | $R^1$ | $R^2$ | $R^3$ | $\delta$ $^{29}$Si (ppm) |
|---|---|---|---|---|
| **9a** | $CH_3$ | $C_2H_5$ | $C_2H_5$ | +19.90 |
| **9b** | $C_2H_5$ | $C_2H_5$ | $C_2H_5$ | +19.62 |
| **9c** | $C_2H_5$ | $C_4H_9$ | $C_4H_9$ | +16.74 |
| **9d** | $CH_3$ | $C_7H_{15}$ | $C_7H_{15}$ | +18.20 |
| **9e** | $CH_3$ | $CH_3$ | $C_8H_{17}$ | +18.32 |
| **9f** | $CH_3$ | $CH_3$ | $C_{10}H_{21}$ | +18.36 |
| **9 g** | $CH_3$ | $CH_3$ | $C_{11}H_{23}$ | +18.24 |
| **9 h** | $CH_3$ | $CH_3$ | $C_{16}H_{33}$ | +18.33 |
| **9i** | $CH_3$ | $C_6H_5$ | $C_6H_5$ | −2.39 |
| **10** | $CH_3$ | $CH_3$ | $C_{10}H_{21}$ | +18.62 |
| **11** | $CH_3$ | $CH_3$ | $C_{10}H_{21}$ | −13.5, +18.06 |

**Table 2.** $^{29}$Si NMR resonance of, *N*-methyl-*N*-(2-triorganosiloxyethyl)-1,2,3,4-tetrahydroisoquinolinium (**12a–i**), *N*-(2-decyldimethylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroquinolinium (**13**) and *N*-(2-decyldimethylsiloxyethyl)-4,4-dimethyl-*N*-methyl-4-sila-1,2,3,4-tetrahydroisoquinolinium iodides (**14**) $[R^1R^2R^3SiOCH_2CH_2–Het_NMe]^+I^-$

| Compound | $R^1$ | $R^2$ | $R^3$ | $\delta$ $^{29}$Si (ppm) |
|---|---|---|---|---|
| **12a** | $CH_3$ | $C_2H_5$ | $C_2H_5$ | +24.23 |
| **12b** | $C_2H_5$ | $C_2H_5$ | $C_2H_5$ | +23.84 |
| **12c** | $C_2H_5$ | $C_4H_9$ | $C_4H_9$ | +21.49 |
| **12d** | $CH_3$ | $C_7H_{15}$ | $C_7H_{15}$ | +22.35 |
| **12e** | $CH_3$ | $CH_3$ | $C_8H_{17}$ | +22.37 |
| **12f** | $CH_3$ | $CH_3$ | $C_{10}H_{21}$ | +22.50 |
| **12 g** | $CH_3$ | $CH_3$ | $C_{11}H_{23}$ | +22.67 |
| **12 h** | $CH_3$ | $CH_3$ | $C_{16}H_{33}$ | +22.49 |
| **12i** | $CH_3$ | $C_6H_5$ | $C_6H_5$ | –0.36 |
| **13** | $CH_3$ | $CH_3$ | $C_{10}H_{21}$ | +22.63 |
| **14** | $CH_3$ | $CH_3$ | $C_{10}H_{21}$ | –13.46, +22.23 |

possessed lower biological activity, practically did not reveal cytotoxic properties and in general were less active concerning the microbial strains examined. The experimental evaluation of cytotoxic properties is presented in Table 3.

In this study, two monolayer tumour cell lines – HT-1080 (human fibrosarcoma) and MG-22A (mouse hepatoma) – and normal mouse fibroblasts (NIH 3 T3) have been employed to evaluate the antiproliferative activity of the synthesized tetrahydro(iso)quinoline compounds in culture. *N*-(2-Hydroxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium iodide (**3**) and 4,4-dimethyl-*N*-(2-hydroxyethyl)-*N*-methyl-4-sila-1,2,3,4-tetrahydroisoquinolinium iodides (**8**) did not exhibit cytotoxic action at all. Thus this study clearly demonstrates that human fibrosarcoma and mouse hepatoma cells are highly responsive to some *N*-methyl-*N*-(2-triorganylsiloxyethyl)-1,2,3,4-tetrahydroisoquinolinium iodides. The following sequence of organosilicon substituents in the cytotoxicity display concerning HT-1080 tumour cell lines has been revealed:

$$Me_2(C_{11}H_{23}) < MePh_2 < Me_2(C_{16}H_{33}) < Et_3 < MeHp_2 < EtBu_2$$

Almost the same sequence has been determined concerning MG-22A tumour cell lines:

$$Me_2Dc < Et_2Me < MePh_2 \leq Me_2(C_{16}H_{33}) < Et_3 < MeHp_2 < Me_2(C_{11}H_{23}) < EtBu_2$$

The present investigation shows that compounds **12c** and **12d** induced growth inhibition of HT-1080 and MG-22A cells, and compound **12 g** exhibits selective cytotoxic action against MG-22A cells at a significant level. In general, *N*-methyl-*N*-(2-triorganylsiloxyethyl)-1,2,3,4-tetrahydroisoquinolinium iodides screened had high NO generation activity, being most active for **12d** in both the HT-1080 and MG-22A tests: 300% and 250% respectively (Table 3). All the compounds synthesized possess low toxicity or are non-toxic compounds ($LD_{50}$ = 450–3200 mg kg$^{-1}$).

The antibacterial and antifungal activity of compounds **12a–i**, **13** and **14** have been investigated in dimethyl sulfoxide against two Gram-positive – *Bacillus cereus* MSCL 330 (BC) and *Staphylococcus aureus* MSCL 334 (SA) – and three Gram-negative microbial strains – *Proteus mirabilis* MSCL 590 (PM), *E. coli* MSCL 332 (EC) and *Pseudomonas aeruginosa* MSCL 331 (PA) – and fungi – *Candida albicans* MSCL 378 (CA) – in comparison with 2-hydroxyethyl precursors **3**, **4** and **8**, using the agar ditch diffusion method.[37] The results of the study are presented in Table 4.

The results were compared with those of standard antibacterial (gentamicin) and antifungal (fluconazole) drugs. *N*-(2-Hydroxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium iodide (**3**) showed low activity against fungi and Gram-negative bacterial strains and was inactive against the Gram-positive bacterial strains tested. Biological activity data exhibited enhancement of inhibitory activity against the tested pathogen microorganisms for *N*-methyl-*N*-(2-triorganylsiloxyethyl)-1,2,3,4-tetrahydro(iso,silaiso)quinolinium iodides **8**, **12** and **14**. In general, Gram-negative bacterial and fungal strains were more resistant to the synthesized compounds than were Gram positive strains. The most effective action of compounds studied has been demonstrated against Gram-positive *Bacillus cereus* MSCL 330 and *Staphylococcus aureus* MSCL 334 and Gram-negative *Pseudomonas aeruginosa* MSCL 331. Compound **12b** was found to be more active against *Bacillus cereus* MSCL 330 and *Staphylococcus aureus* MSCL 334 than gentamicin. The highest antifungal potency was revealed for *N*-(2-hydroxyethyl)-*N*-methyl-1,2,3,4-tetrahydroquinolinium iodide (**4**). Selectivity of antibacterial action was revealed concerning unsilylated **3**, **4** and **8**. Contrary to *N*-(2-hydroxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium iodide (**3**), which is active only against Gram-negative strains, compound **4** is mostly potent against Gram-positive antibacterial strains. Compound **8**, in turn, revealed a wide spectrum of antibacterial activity. Among the heterocyclic choline derivatives **12f**, **13** and **14** bearing the same substituents at the silicon atom (dimethyldecylsilyl group) and their precursors **3**, **4** and **8**, silatetrahydroisoquinoline derivatives **14** and **8** possessed a wide spectrum of biological activity and were the most potent compounds. Contrary to tetrahydroisoquinoline derivatives **12f** and **3**, which exhibited higher activity against Gram-negative antibacterial strains, tetrahydroquinoline derivatives **13** and **4** were mostly potent against Gram-positive strains. The degree of antibacterial activity increased among the number of tetrahydroisoquinoline derivatives with the silyl group nature in the following order concerning Gram-positive strains:

**Table 3.** *In vitro* cell cytotoxicity and intracellular NO generation caused by N-methyl-N-(2-triorganylsiloxyethyl)-1,2,3,4-tetrahydroisoquinolinium (**12a-i**), N-(2-hydroxyethyl)-N-methyl-1,2,3,4-tetrahydroisoquinolinium (**3**) and 4,4-dimethyl-N-(2-hydroxyethyl)-N-methyl-4-sila-1,2,3,4-tetrahydroisoquinolinium (**8**) iodides

| Compound/test | HT-1080 | | | MG-22A | | | NIH 3 T3 | |
|---|---|---|---|---|---|---|---|---|
| | $LC_{50}$ (μg ml$^{-1}$) | | NO (%) | $LC_{50}$ (μg ml$^{-1}$) | | NO (%) | $LC_{50}$ (μg ml$^{-1}$) | $LD_{50}$ (mg kg$^{-1}$) |
| | CV | MTT | CV | CV | MTT | CV | NR | |
| **3** | ** | ** | 2 | 100 | ** | 4 | 1000 | 2202 |
| **12a** | ** | ** | 5 | 73 | 100 | 9 | 827 | 2391 |
| **12b** | 8 | 16 | 200 | 6 | 10 | 150 | 171 | 1214 |
| **12c** | <1 | <1 | 150 | 1 | 1 | 150 | 24 | 539 |
| **12d** | 3 | 3 | 300 | 3 | 3 | 250 | 14 | 448 |
| **12e** | ** | ** | 2 | ** | ** | 3 | 1340 | 3182 |
| **12f** | ** | ** | 4 | 100 | 100 | 6 | ** | >2000 |
| **12 g** | 30 | 14 | 200 | 2 | 2 | 150 | 19 | 532 |
| **12 h** | 17 | 16 | 200 | 13 | 16 | 150 | 85 | 1083 |
| **12i** | 19 | 18 | 100 | 13 | 17 | 100 | 41 | 722 |
| **8** | ** | ** | 2 | ** | ** | 4 | 452 | 1671 |

CV, crystal violet coloration; MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2*H*-tetrazolium bromide coloration; NR, neutral red coloration; NO, nitrogen oxide concentration degree, determined according to the procedure.[36]

* $LC_{50}$, concentration of compound providing 50% cell killing effect.

** No cytotoxic effect.

**Table 4.** *In vitro* antibacterial and antifungal activity data of *N*-methyl-*N*-(2-triorganylsiloxyethyl)- 1,2,3,4-tetrahydroisoquinolinium (**12a–i**), *N*-(2-*n*-decyldimethylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroquinolinium (**13**) and *N*-(2-*n*-decyldimethylsiloxyethyl)-4,4-dimethyl-*N*-methyl-4-sila-1,2,3,4-tetrahidroisoquinolinium (**14**) iodides given in concentrations of 5 mg ml$^{-1}$ (0.35 mg per disc) and 10 mg ml$^{-1}$ (0.7 mg per disc)

| Compound | Diameter of zones showing complete inhibition of growth (mm) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SA | | BC | | EC | | PA | | PM | | CA | |
| | 5 | 10 | 5 | 10 | 5 | 10 | 5 | 10 | 5 | 10 | 5 | 10 |
| **3** | ** | ** | 11 | ** | ** | 13 | 12 | 16 | 13 | 15 | 11 | 10 |
| **12a** | 22 | 26 | 15 | 16 | 9 | 15 | 12 | 20 | 16 | 20 | 11 | 9 |
| **12b** | 25 | 30 | 22 | 23 | 11 | 19 | 19 | 19 | 14 | 18 | 11 | 11 |
| **12c** | 19 | 20 | 19 | 19 | 11 | 14 | 18 | 20 | 15 | 16 | 16 | 12 |
| **12d** | 9 | 8 | 10 | 9 | 9 | 13 | 17 | 18 | 14 | 17 | 13 | 11 |
| **12e** | 16 | 18 | 15 | 15 | 12 | 14 | 14 | 20 | 11 | 18 | 11 | ** |
| **12f** | 11 | 11 | 11 | 10 | 9 | 13 | 11 | 18 | 14 | 20 | 11 | 10 |
| **12 g** | ** | 12 | 13 | ** | 11 | 13 | 11 | 20 | 10 | 15 | 11 | 9 |
| **12 h** | ** | ** | 10 | ** | 9 | 13 | 10 | 18 | 10 | 19 | 13 | 9 |
| **12i** | 13 | 15 | 13 | 14 | 8 | 12 | 17 | 21 | 14 | 18 | 12 | 11 |
| **4** | 8 | 11 | 8 | 10 | 8 | 9 | ** | 8 | ** | ** | 13 | 18 |
| **13** | 12 | 13 | 12 | 13 | 8 | 10 | ** | ** | ** | ** | 11 | 11 |
| **8** | 18 | 14 | 17 | 15 | 10 | 10 | 10 | 14 | 9 | ** | 13 | 13 |
| **14** | 19 | 21 | 17 | 19 | 11 | 12 | 10 | 12 | ** | ** | 13 | 14 |
| Gentamicin | 25 | 27 | 19 | 22 | 21 | 23 | 40 | 42 | 20 | 21 | n | n |
| Fluconazole (2/0.2 mg ml$^{-1}$) | n/n | | n/n | | n/n | | n/n | | n/n | | 10/** | |

n, non-determined.

** No inhibiting effect.

$$Me_2(C_{16}H_{33}) < MeHp_2 < Me_2Dc < Me_2(C_{11}H_{23}) < MePh_2 < Me_2Oct \leq HO(Si_{cycl}Me_2) < Me_2Dc(Si_{cycl}Me_2) < EtBu_2 < Et_2Me < Et_3$$

and in the following order concerning Gram-negative strains:

$$MePh_2 < HO(Si_{cycl}Me_2) < Me_2Dc(Si_{cycl}Me_2) < Me_2(C_{16}H_{33}) \leq Me_2Dc \leq MeHp_2 < Me_2(C_{11}H_{23}) < EtBu_2 < Me_2Oct < Et_2Me < Et_3$$

(*E. coli* MSCL 332)

**Table 5.** *In vitro* minimum inhibitory concentration (MIC) and minimum bactericidal concentration (MBC) of compounds **13** and **14**

| Compound | MIC ($\mu g\,ml^{-1}$) | | | | | MBC ($\mu g\,ml^{-1}$) |
|---|---|---|---|---|---|---|
| | SA | BC | EC | PA | CA | SA |
| **13** | 256 | >256 | >256 | >256 | >256 | — |
| **14** | 8 | 32 | >256 | >256 | 32 | 32 |

$Me_2Dc(Si_{cycl}Me_2) < HO(Si_{cycl}Me_2) < Me_2(C_{16}H_{33}) \leq Me_2Dc < MeHp_2 < Et_3 < Me_2(C_{11}H_{23}) < Et_2Me < Me_2Oct < EtBu_2 < MePh_2$ (*Pseudomonas aeruginosa* MSCL 331) and

$Me_2Dc(Si_{cycl}Me_2) \leq HO(Si_{cycl}Me_2)Me_2 < (C_{11}H_{23}) < EtBu_2 < MeHp_2 < Me_2Oct < MePh_2 \leq Et_3 < Me_2(C_{16}H_{33}) < Me_2Dc < Et_2Me$ (*Proteus mirabilis* MSCL 590)

Compounds **8** and **14** in general were more active concerning the Gram-positive strains and less active against the Gram-negative strains.

*In vitro* minimum inhibitory concentration (MIC) and minimum bactericidal concentration (MBC) were determined for compounds **13** and **14**. The data obtained (Table 5) show that compound **14** possessed not only bacteriostatic ($8\,\mu g\,ml^{-1}$) but also marked bactericidal action ($32\,\mu g\,ml^{-1}$) against the *Staphylococcus aureus* strain.

## Conclusions

A series of *N*-methyl-*N*-(2-triorganosiloxyethyl)-1,2,3,4-tetrahydroisoquinolinium, *N*-(2-decyldimethylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroquinolinium and *N*-(2-decyldimethylsiloxyethyl)-4,4-dimethyl-*N*-methyl-4-silatetrahydroisoquinolinium iodides has been synthesized as potential pharmacological agents. $^{29}Si$ NMR resonance of *N*-(2-triorganosiloxyethyl)tetrahydro(iso,silaiso)quinolines and *N*-methyl-*N*-(2-triorganosiloxyethyl)-1,2,3,4-tetrahydro(iso,silaiso)quinolinium iodides depended upon the substituent at the silicon atom.

At the basis of $LC_{50}$ data, the structure–antitumour activity relationship was discussed in terms of the nature of the heterocycle and silyl group, which was present in the amphiphilic molecule of heterocyclic choline analogues. It has been revealed that all the synthesized compounds possessed selective cytotoxic action in relation to the studied tumour cells. *N*-(2-Di-*n*-butylethylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium and *N*-(2-di-*n*-heptylmethylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium iodides were the most effective in this respect. A specificity of the cytotoxic effect towards MG-22A line was displayed by *N*-(2-*n*-dimethylundecylsiloxyethyl)-*N*-methyl-1,2,3,4-tetrahydroisoquinolinium iodide. All the compounds synthesized had high NO generation activity, exhibited low toxicity or were non-toxic compounds.

The biological activity data exhibited enhancement of inhibitory activity against the tested pathogen microorganisms for *N*-methyl-*N*-(2-triorganylsiloxyethyl)-1,2,3,4-tetrahydroisoquinolinium iodides in comparison with their 2-hydroxyethyl precursors. The degree of antibacterial activity of tetrahydroisoquinoline derivatives depends on the silyl group nature.

The results obtained are in agreement with our previously published data for aliphatic choline analogues and allow one to assume that blocking of the hydroxyl group of *N*-methyl-*N*-(2-hydroxyethyl)-1,2,3,4-tetrahydroisoquinolinium iodides with triorganylsilyl group is essential for improvement of their biological properties.

### Supporting information

Supporting information may be found in the online version of this article.

### Acknowledgement

This work was supported by the Latvian Council of Sciences (Pr. 09.1321).

## References

[1] A. M. van Hattum, H. M. Pinedo, H. M. M. Schluper, F. N. Hausheer, E. Boven, *Int. J. Cancer* **2000**, *88*, 260.
[2] A. K. Bence, C. A. Mattingly, T. G. Burke, V. R. Adams, *Cancer Chemother. Pharmacol.* **2004**, *54*, 354.
[3] W. K. Anderson, R. Kasliwal, D. M. Houston, Y.-S. Wang, V. L. Narayanan, R. D. Haugwitz, J. Plowman, *J. Med. Chem.* **1995**, *38*, 3789.
[4] M.-C. Bonache, C. Chamorro, S. Velazquez, E. de Clercq, J. Balzarini, F. Rodriquez-Barrios, F. Gago, M.-J. Camarasa, A. San-Felix, *J. Med. Chem.* **2005**, *48*, 6653.
[5] M. D. Mashkovsky, Drugs (12th edn), Medicina, Moscow, **1993**.
[6] W. Bains, R. Tacke, *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 526.
[7] D. Schirlin, J. N. Collard, J. M. Hornspreger, P. R. Keshary, Acetylcholinesterase inhibitors. Merrel Pharmaceuticals, US-05693668, **1997**.
[8] F. H. Hausheer, K. Haridas, P. Seetharamulu, D. G. Reddy, S. Yao, P. N. Petluru, M. Dhanabalan, Highly lipophilic camptothecin derivatives. Bionumeric Pharmaceuticals, WO-09835940, **1998**.
[9] D. P. Curran, H. Josien, D. Bom, T. G. Burke, Camptotecin analogs and methods of preparation thereof. University of Pittsburgh, WO-0035924, **2000**.
[10] S. Farkas, S. Foldeak, E. Karpati, P. Hegyes, J. Kreidl, L. Szporny, L. Czibula, S. Petofi Vassne, Organosilicon derivatives, pharmaceutical compositions containing them and process of preparing same. Richter Gedeon Vegyeszeti Gyar RT, EP-00468825 (**1992**).
[11] J. S. Millership, M. L. Shanks, *Int. J. Pharm.* **1986**, *28*, 1–9.
[12] E. Lukevics, A. Zablotskaya, *Metalloorg. Khim.* **1993**, *6*, 263.
[13] V. V. Belakhov, Y. D. Shenin, *Pharm. Chem. J.* **2008**, *42*, 322.
[14] M. A. Peterson, M. Oliveira, M. A. Christiansen, C. E. Cutler, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6775.
[15] J. R. Shelton, S. R. Burt, M. A. Peterson, *Bioorg. Med. Chem. Lett.* **2011**, *21*, 1484.
[16] O. J. Donadel, T. Martin, V. S. Martin, J. Villar, J. M. Padron, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3536.
[17] C. Nguyen, G. F. Ruda, A. Schipani, G. Kasinathan, I. Leal, A. Musso-Buendia, M. Kaiser, R. Brun, L. M. Ruiz-Perez, B.-L. Sahlberg, N. G. Johansson, *J. Med. Chem.* **2006**, *49*, 4183.
[18] E. Lukevics, I. Segal, I. Birgele, A. Zablotskaya, *Chem. Heterocycl. Compd.* **1998**, *34*, 1076.
[19] A. Zablotskaya, I. Segal, S. Germane, I. Shestakova, I. Domracheva, A. Nesterova, A. Geronikaki, E. Lukevics, *Chem. Heterocycl. Compd.* **2002**, *38*, 859.
[20] A. Zablotskaya, I. Segal, Yu. Popelis, E. Lukevics, S. Baluja, I. Shestakova, I. Domracheva, *Appl. Organomet. Chem.* **2006**, *20*, 721.
[21] M. Lemort, L. S. Chao, M. Radermecker, R. Demeure, *MAGNETOM Flesh* **2007**, 38.
[22] J. M. Sanz, R. Lopez, J. L. Garcia, *FEBS Lett.* **1988**, *232*, 308.
[23] G. W. A. Milne (Ed), Ashgate Handbook of Antineoplastic Agents, Gower Publishing, Aldershot, UK, **2000**.
[24] H. Malonne, G. Atassi, *Anticancer Drugs* **1997**, *8*, 811.
[25] J. D. Scott, R. M. Williams, *Chem. Rev.* **2002**, *102*, 1668.
[26] Y. Mikami, K. Yokoyama, H. Tabeta, K. Nakagaki, T. Arai, *J. Pharm. Dyn.* **1981**, *4*, 282.
[27] G. R. Pettit, V. Gaddamidi, D. L. Herald, S. B. Singh, G. M. Cragg, J. M. Schmidt, F. E. Boettner, M. Williams, Y. Sagawa, *J. Nat. Prod.* **1986**, *49*, 995.
[28] E. Izbicka, R. Lawrence, E. Raymond, G. Eckhardt, G. Faircloth, J. Jimeno, G. Clark, D. D. Von Hoff, *Ann. Oncol.* **1998**, *9*, 981.

[29] A. Kleeman, J. Engel, B. Kutscher, D. Reichert. Pharmaceutical Substances (4th edn), Thieme, New York, **2001**.
[30] H. R. Lin, M. K. Safo, D. J. Abraham, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2581.
[31] C. M. Tarby, R. F. Kaltenbach III, T. Huynh, A. Pudzianowski, H. Shen, M. Ortega-Nanos, S. Sheriff, J. A. Newitt, P. A. McDonnell, N. Burford, C. R. Fairchild, W. Vaccaro, Z. Chen, R. M. Borzilleri, J. Naglich, L. J. Lombardo, M. Gottardis, G. L. Trainor, D. L. Roussell, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2095.
[32] C. Jiang, Q. You, F. Liu, W. Wu, Q. Guo, J. Chern, L. Yang, M. Chen, *Chem. Pharm. Bull.* **2009**, *57*, 567.
[33] A. Zablotskaya, I. Segal, E. Lukevics, S. Belyakov, H. Spies, *Appl. Organomet. Chem.* **2007**, *21*, 288.
[34] I. Segal, A. Zablotskaya, E. Lukevics, *Chem. Heterocycl. Comp.* **2005**, *41*, 613.
[35] P. J. Freshney, Culture of Animal Cells: A Manual of Basic Technique, Wiley-Liss, New York, **1994**, pp. 296.
[36] D. J. Fast, R. C. Lynch, R. W. Leu, *J. Leukocyte Biol.* **1992**, *52*, 255.
[37] A. Wanger, in Antimicrobial Susceptibility Testing Protocols, (Eds: R. Schwalbe, L. Steele-Moore, A. C. Goodwin), CRC Press, Boca Raton, FL, **2007**, p. 53.
[38] C. Eaborn, Organosilicon Compounds, Butterworths, London, **1960**, pp. 89.