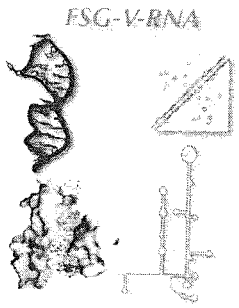


SIXTH FRAMEWORK PROGRAMME

Project no.: **503455**
Project acronym: **FSG-V-RNA**
Project title: **Functional and Structural Genomics of Viral RNA**

Instrument:
Thematic Priority:

SIXTH FRAMEWORK PROGRAMME, PRIORITY LifeSciHealth
[Life sciences, Genomics and Biotechnology for Health]
SPECIFIC TARGETED RESEARCH OR INNOVATION PROJECT

Activity Report

Final

Period covered: month 1 to 51
Date of preparation: 31-08-2008
Start date of project: 01-07-2004
Duration: 51 months

Project coordinator name: Prof. Sybren Wijmenga
Project coordinator organisation name: Radboud University Nijmegen
Revision: no

Publishable executive summary

Functional and Structural Genomics of Viral RNA

Acronym: FSG-V-RNA

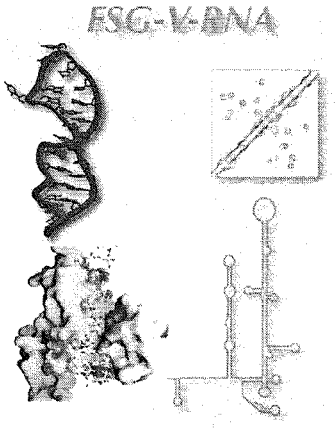
Project no: 503455

Coordinator/Contact:
Prof. dr. Sybren S. Wijmenga
Physical Chemistry - Biophysical Chemistry
Institute of Molecules and Materials (IMM)
Radboud University Nijmegen
Toernooiveld 1,
6525 ED Nijmegen,
The Netherlands

Tel: +31-24-3652678 (M.de.With, secr)
+31-24-3653384 (office)

fax: +31-24-3652112

Email: S.Wijmenga@nmr.ru.nl



Webpage of the project: www.fsgvrna.nmr.ru.nl

Project Summary and Objectives

The EU FP6 STREP FSG-V-RNA aims at developing and improving tools and approaches to facilitate the generation of new knowledge in functional and structural genomics of viral RNAs. Specifically, ***(i)*** new methods and tools for the rapid and efficient structural analysis of RNA and RNA-protein complexes will be developed, ***(ii)*** these optimised tools will be applied to essential RNA elements that are vital for the function of HBV, HCV and HIV viruses, and ***(iii)*** complementary screening techniques and structure analysis of RNA-ligand (including proteins/peptides) complexes will be implemented to promote the identification of anti-viral compounds and proteins/peptides targeting these RNA structures. The project exploits available RNA sequence data but also will also expand our knowledge of viral RNA sequence elements and their variations.

The biomedical importance of RNA as a research target is stressed by the fact that viral infections, such as HBV, HCV and HIV are major global public health problems. The outcomes of the project are expected to initiate the development of novel drugs that target viral RNA molecules and have thus strong implications for public health. The research consortium is strengthened by the involvement of an SME, which contributes facilities and expertise for the generation of small-dedicated libraries of RNA-binding small molecules.

The project involves multidisciplinary research, made possible by integrating, in an interdisciplinary fashion, the research capacities of a number of leading European labs, i.e. their equipment and complementary expertise on the structural, functional and virological analysis of RNA and RNA-ligand complexes. The innovative tools developed will be made available to other researchers throughout Europe and outside and open the way for efficient analysis of a wide range of RNA-based processes extending far beyond analysis of viral RNAs.

Participants

Participant Role*	Participant no.	Participant name	Participant short name	Country	Date enter project**	Date exit project** #
CO	1a	Stichting Katholieke Universiteit Prof. Dr. Sybren S. Wijmenga	KUN	NL	Month 1	month 51
CR	1b	Stichting Katholieke Universiteit Prof. Dr. Sylvia Speller	KUN	NL	mf onth 1	month 51
CR	1c	Stichting Katholieke Universiteit Prof.Dr. Floris Rutjes	KUN	NL	month 1	month 51
CR	2	European Molecular Biology Laboratory Dr. Michael Sattler	EMBL	GE	month 1	month 40 $
CR	3	Centre National de la Recherche Scientifique – Délégation Paris A Prof. Dr. Frédéric Dardel	CNRS- DR01	FR	month 1	month 51
CR	4	Masaryk University, Faculty of Science Prof. Dr. Vladimir Sklenar	PřF MU	CZ	month 1	month 51
CR	5	University Hospital Freiburg Prof. Dr. Michael Nassal	UKL-FR	GE	month 1	month 51
CR	6	Lund University Dr. Karin Kidd-Ljunggren	LU	SE	month 1	month 51
CR	7	Chiralix B.V*** Dr. Richard Blaauw	Chiralix	NL	month 18	month 51
CR	8	$ Forschungszentrum für Umwelt und Gesundheit Dr. Michael Sattler	GSF	GE	month 40	month 51

**CO = Coordinator, CR = Contractor; ** Normally insert "month 1 (start of project)" and "month n (end of project)"; ***The SME Chiralix replaces the originally proposed SME Anadys as partner, entry 1 January 2006; amended form C and Technical Annex to EU in December 2005 and January 2006, and awaits approval by EU commission.*

$ Dr. Michael Nassal moved from the "European Molecular Biology Laboratory" (EMBL Heidelberg, Germany) to the "Forschungszentrum für Umwelt und Gesundheit" (GSF Munich, Germany), 1st October 2007.

A 9-month project extension has been granted by the EU commission, moving the project end datum from month 42 to month 51.

Summary of major results achieved during the project (month 1–51)

1. RNA preparation:

One of the bottlenecks of RNA structural and functional studies is the preparation of samples in adequate quantity and purity level. In the first and second reporting period emphasis has therefore been on the development of new efficient production of isotope ($^{13}C/^{15}N/^{2}H$)-labelled RNAs of defined sequence for structural and interaction studies.

In the second reporting period, protocols have now been made available (on the consortium Webpage) which are describing the production of $^{13}C/^{15}N/^{2}H$-labeled RNAs via efficient *in-vitro* enzymatic synthesis. The *in-vitro* synthesis methodology has been extended to include efficient *in-vitro* enzymatic production of $^{13}C/^{15}N/^{2}H$-labeled DNA sequences. Moreover, an efficient method for multiple segmental $^{13}C/^{15}N/^{2}H$-labelling of (large) RNAs has been developed as one of the downstream objectives of the project. Publications are in preparation to make this method available to a larger community.

Moreover a simple, cheap and general method for expressing large quantities of recombinant RNAs *in vivo*, in a stable form has been devised. The RNAs can be purified easily using an affinity tag approach conceptually similar to that currently used for proteins. This method is compatible with high-throughput techniques currently used in post-genomic programs (structural genomics, interactome...). It is based on a set of plasmids which could be licensed to companies selling molecular biology products and kits. Patent protection has been applied for (PCT/FR2007/000980, WO/2007/144508). An MTA has been signed with InvivoGen for evaluation of our technology, prior to a licensing agreement.

Finally, streamlining of RNA production was - as planned –considered in the second reporting period as well as further improvement of the developed labelling methods. This production platform constitutes a firm basis for the successful development of the new tools for fast structure determination of (large) RNAs and RNA-protein complexes. The consortium presented these results during its workshop held in March 2007 at the EMBL and during the third reporting period on the project end-meeting which was extended into symposium with external speakers.

During the third reporting period partner 1a (Wijmenga) prepared publications on the selective and multiple segmental in-vitro isotope labeling (Nelissen et al., NAR 2008) for which some experiments were carried out to complete material for publication ready data. In addition, experiments were carried out to prepare isotope labeled samples for the final determination of the structure and dynamics of the complete 62 nucleotide epsilon elements of Human and Duck HBV. Finally, these samples provided additional data to establish / test the protocols for fast structure determination of RNA based on 1H chemical shifts complemented with Residual Dipolar Couplings. Also, the method developed in the Dardel lab for preparation of RNAs was implemented in Wijmenga lab. And using this approach as a basis, a new method was developed to prepare large RNAs for structural studies. It employs a general plasmid framework that expresses the RNA of interest bracketed with hammerhead ribozymes at the 5'-end and at the 3'-end so that the final RNA of interest can be produced with high yield and with convenient ends that can be modified for ligation (a publication on this is being prepared). Partner 3 (Dardel): all deliverables already delivered at month 42 reporting period and provided their system to about 40 academic groups and 1 industrial partner in 10 countries (7 European, 2 Asian and 1 in the USA). The consortium presented the final results on the project end-meeting which was extended into symposium with external speakers in September 2008.

The protocols for labelling methods are available on the webpage of the consortium and/or can be obtained from the partners.

2. RNA structure determination methods:

Introduction: It is urgently needed to optimise NMR methods and strategies for the analysis of large RNAs. Although, in recent years, a significant number of new RNA structures has been reported, the development of novel and efficient tools for rapid structure determination of RNAs and analysis of large RNA-protein complexes is still needed. NMR structure determination of nucleic acids is complicated by low proton density and lack of long-range structural restraints. Additional information is thus highly

desirable to improve the accuracy of RNA structures. Recently, substantial improvement of our understanding of ^{1}H chemical shift dependence on the secondary structure of DNA and RNA has been achieved[1]: ^{1}H chemical shifts can now be reliably predicted and, conversely, these shifts can serve as independent constraints to validate structures obtained from NMR data. So far, only a limited set of RNA heteronuclear chemical shifts is available in databases, which does not allow a consistent statistical analysis. The current project addresses this need: gaining insight into the chemical shift – structure relationship of ^{13}C, ^{15}N and ^{31}P nuclei by a concerted effort using both theoretical and empirical approaches. New experimental methods will be designed to measure isotropic chemical shifts of heteronuclei and their anisotropy. Theoretical studies will help to understand the physical origin of the shielding. Statistical analysis of representative data will lead to parameter sets for empirical prediction of heteronuclear chemical shifts. In a reversed approach, the structural restraints will be directly derived from the chemical shifts and used in MD refinement protocols. In addition, so-called residual dipolar couplings and residual chemical shift anisotropy in aligning media or from nucleic acids aligned directly in the magnetic field will be easily measured, providing global structure information. Moreover, AFM/STM tools for determination of topographies of larger RNAs to fill gap between NMR and larger size molecules will be developed. Finally, improved structure refinement protocols for RNA oligonucleotides will be designed.

Structure determination methods: In the first reporting period focus has on the fast recording of NMR spectra via so-called GFT methods, implemented to give access to these NMR structure parameters (partner 1a). Also, sensitivity-enhanced NMR experiments have been devised to give to quaternary heteronuclei, which provide access to the heteronuclear chemical shifts (partner 4). In addition, as a prelude on the structure determination protocols, a reliable and fast prediction method has been devised for electrostatic alignment, i.e. for predicting residual dipolar couplings and residual chemical shift anisotropies of nucleic acids in aligning media (partner 1a). The prediction of heteronuclear shifts has been started (partner 4). In this way, the important first steps have been set on the route towards (fast) structure determinations of large RNAs.

In the second and third reporting period focus has been on the development of method to support the fast structure determination of RNA structure. First, with the development of a fast recording method for NMR spectra via so-called GFT methods implemented to give access to these NMR structure parameters such as chemical shifts (partner 1a). The method has been further developed to allow for the extraction of parameters other than just the chemical shifts, but also CSA and RDC. New NMR pulse sequence has also been developed to allow for the measurement of multiple C-H based RDCs (partner 1a). Moreover, a method for fast determination of alignment of macromolecules in liquid-crystal media has been devised in prelude to structure determination (partner 1a). 1H-chemical shift refinement has also been incorporated to restrained molecular dynamic software Xplor to support the refinement of nucleic acid structures (partner 1a). Furthermore, a new (SVD) method has been developed that allows for fast structure calculation of RNAs from 1H chemical shifts and which provides very high precision and accuracy (partner 1a). Finally, the structure (and dynamics) was determined of the full-length (65 nucleotide) epsilons RNA from both human and duck HBV; they form the second largest RNA determined to date by NMR (partner 1a). These provided an ideal case for the development and testing of a structure calculation protocol of larger (domained) RNAs (partners 1a). Sensitivity-enhanced NMR experiments have been devised to give to quaternary heteronuclei, which provide access to the heteronuclear chemical shifts (partner 4). Theoretical studies have been made, concentrating on the elucidation of the effect of backbone conformation changes on the chemical shifts of heteronuclei. The derived methods allowed us to address issues of conformations that are rare or unstable in nucleic acids. Specifically, the relationship between ^{31}P chemical shift tensors and conformation of nucleic acid backbone was investigated by DFT methods (partner 4). In addition, DFT calculations were carried out on sugar pucker and glycosidic torsion angle dependence of all DNA and RNA nucleosides to determine the 13C, 15N and 1H shifts, J-couplings of atoms in base and sugar moieties (partner 1a). These data were compared experimental data and analysed. This resulted in new important important relations between these hetero-nuclear shifts and structure parameters, specifically sugar pucker and glycosidic torsion angle (partner 1a). More work needs to be done to be able to fully employ hetero-nuclear shift in the structure calculation protocols (partner 1a). An approach has been developed and experimentally tested for detecting monomer/dimer equilibria of

[1] Wijmenga SS et al. J. Biomol. NMR (1997), 10, 337-350; Dejaegere A. et al. In Modeling NMR Chemical Shifts. Gaining Insight into Structure and Environment. ACS 1999, 194-206. Cromsigt, J *et al.* J. Biomol. NMR 2001, 21, 11-29.

RNA hairpins in solution (partner 2). A new pulse sequence has been developed for correlating pyrimidine H5 protons with purine nitrogens across base pairs (partner 2). In addition, new spin-labelled compounds have been devised (partner 1c) which can be incorporated into the RNA sequence using the segment labelling methods developed by partner 1a (partner 1c/1a).

In conclusion, several important steps have been set on the route towards (fast) structure determination of larger RNAs and usable protocols are available for fast structure determination of large RNA and for assessment of their (domain) dynamics.

RNA structure determinations: A number of large (> 40 nt) viral RNAs of relevance have been investigated and their structure determined by means of NMR by partner 1a: the apical-loop of ε-RNA of human HBV, the complete ε-RNA of human HBV (65-nt), the apical loop and primer loop of ε-RNA of Duck HBV as well as the complete ε-RNA of Duck HBV. Also the dynamics of the domains and of the full duck and human epsilon were assessed. The full-length (62 nt) epsilon from duck and human HBV each form to date the second-largest RNAs for which structure (and dynamics) have been determined (partner 1a). Partner 2 has determined the structures of RNAs involved in gene regulation (see below)

AFM/SPM: One of the ultimate objectives of the project is to develop new tools for structural studies of larger RNAs (> ca. 60 nt) by means of NMR with AFM/STM providing complementary lower resolution structural information. The AFM/STM studies have been able to derive images of different viral element such as: epsilon loop of HBV, domain III of IRES (HCV), RNA from the Phi-29 bacteriophage, HIV RNA dimer, CCMV virus. To this purpose, methods were further developed to attach the naked RNA via a flexible linker to flat surfaces, to prevent interaction forces between the naked RNA and the charged surface to unduly affect the natural (single-stranded) RNA conformation. In RNAs, larger than ca. 60 nt, one of the objectives for NMR studies in this project, the AFM/STM can thus provide complementary rough 3D information. High potential appears to lie also in single-molecule atomic force studies of such naked RNAs, e.g. in folding/unfolding studies of the naked RNAs or in the delineation of interactions within single e RNA:RNA or RNA:protein or RNA:ligand complex.

It is evident that the development of new tools, as planned in the project and for which significant steps have now been established in the second and third reporting periods, will indeed have a major impact on the field of structural biology of RNA.

3. Structural analysis of RNA-protein and RNA-ligand interactions:

It is urgently needed to (i) optimise NMR methods and strategies for the analysis of large RNA-protein complexes and (ii) to develop efficient tools for the characterization of molecular interfaces that involve RNA sequence variants.

During the second reporting period, NMR has been used to determine the three-dimensional structures of a number of RNA-protein complexes. Sattler's group (partner 2) has developed an efficient protocol for quaternary structure determination of protein-protein and protein-RNA complexes. The protocol is based on measuring orientational restraints from residual dipolar couplings (RDCs), which are used to orient individual domains, for which an atomic resolution structure is already available. For the structural analysis of variant RNA/protein complexes, they have recorded chemical shift data for different BPS RNA-Splicing Factor 1 complexes and have analyzed these in the context of the structure and NMR data available for the wild type complex. Sattler's group (partner 2) also has used an analysis by chemical shift perturbation to study the interaction of a branch point sequence-U2 snRNA duplex with the spliceosomal p14 protein alone and when bound to a binding epitope in Splicing Factor 3b (SF3b).

4. RNA virology:

Over the 1st and 2nd reporting period, naturally occurring HBV strains were identified by partner 6 and isolated from clinical samples. To date, 850 strains have been isolated, characterized and analysed. The pre-S and S gene sequences of approximately 400 HBV strains have also been sequences and analyzed. The analysis of naturally occurring HCV strains, particularly the IRES region, is ongoing.

Beyond previously functionally characterized DHBV ε-RNA variants, partner 5, after joint discussion of NMR-data obtained by partner Wijmenga, has begun to characterize a set of 12 ε variants. Currently assays are established to exactly determine the DNA initiation sites, both in vitro and in replicating viral

genomes. Major efforts went also into characterization of the chaperone-dependent formation of ε-RNA/reverse transcriptase (RT) complexes for structural characterization. Biochemical analyses clearly defined which chaperones are essential (Stahl, Retzlaff, Nassal, Beck, NAR 2007) and provided first insights into their mechanism of action (Stahl, Beck, Nassal, J Virol 2007).

5. RNA targeting compounds:

During the second reporting period P2, the company Chiralix (partner 7) stepped in the project in replacement of the company Anadys which was finally invalidated by the EU as industrial partner. Chiralix has worked on developing new molecules for testing their binding/activity against specific RNA targets. They developed synthesis routes for multigram scale preparation of 2-deoxystreptamine that is used as scaffold for the development of new ligands library.

Screening for RNA affinity of the synthesized molecules was performed by interaction screening with NMR. The objectives are first to assemble a specifically designed library of small compounds (*elementary pharmacophores*) and then to use this library to screen against various viral RNA fragments, using a robotised flow-injection NMR platform. This approach allows for the identification of primary binding elements, which will be combined and improved using both structural and functional studies.

The group of Dardel (partner 3) has been for instance working on (i) assembling custom library of "RNA friendly" fragments (current library contains 150 specifically designed short peptides, 60 commercially available molecules, 60 proprietary compounds, devised by partners 3, 1c, 7) ; (ii) producing the target RNA, either unlabelled or ^{15}N enriched, for NMR screening ; (iii) performing the RNA-ligand screening itself, to find potential new interesting ligands. The process resulted in finding four candidate ligands that are specific to HIV primer (2 short peptides and 2 small chemical molecules, kynuramin (commercial) and DACP (proprietary). Other specific ligands against HIV-1 DIS RNA have also been identified, which are currently being analyzed. Further screening against other targets and using new compounds is still underway.

Partner 5 has made available in vitro assays for DHBV used for functional testing of anti-HBV antivirals. Other systems such as transient transfection systems and stable cell lines for HBV and DHBV, and primary hepatocytes from *tupaias* (for HBV) and ducks (for DHBV), plus limited resources for in vivo testing in ducks, were also made available. Various anti-DHBV ε PMOs have been tested in a cell-free in vitro system established by partner 5 and found to sequence-specifically inhibit ε-dependent priming activity of the DHBV reverse transcriptase. PMOs conjugated to cell-permeability enhancing peptides displayed antiviral activity in a stable DHBV expressing cell line, but turned out to be toxic for primary hepatocytes. Finally, functional test systems based on bivalent oligos and small RNA ligands including HBV and DHBV expressing cell lines have been devised and made available by partner 5:

Exploitable knowledge and its use/Dissemination plan

The current project is in fact only possible by setting up a trans-European consortium, which involves excellent researchers that are experts in different areas of RNA-related research. The proposal also involves in part extremely expensive high-tech equipment which in this combination is available only on the European but not on a national level. The combination of complementary expertise from structural biologists, chemists and biologists allows to set up a multidisciplinary research project to ***(i)*** develop novel research tools and ***(ii)*** address biomedical questions in a unique way that extends beyond current research programmes, especially of individual groups. The research tools to be developed will be available to the European scientific community and thus will stimulate and enhance basic research of RNA and RNA molecular recognition. The tools will be available to understand the wide range of RNA-based processes even beyond analysis of viral RNAs and promote structural research on the various important RNA-based processes that are involved in the regulation of gene expression.

The results obtained by the consortium are and will be made available to the scientific community by publications: to date ca. **40** articles have been published in relation to the project, and **17** other publications are in preparation or have been submitted for review. Information about the consortium is also made available on its official website. This web page is used for communication and data exchange within the consortium, but as the project progresses consolidated results and tools generated will be made

available for other researchers. For instance, a protocol for the *in-vitro* synthesis of isotope ($^{13}C/^{15}N/^{2}H$) labelled RNAs has been placed on the webpage. The progress reports, that are part of the regular management meetings at the different locations of the partner labs, have been announced locally and can be attended by scientists at the respective institutions.

A patent protection has been applied for the technology developed by partner 3 on the in-vitro expression of RNA (PCT/FR2007/000980, WO/2007/144508) and a MTA has been signed with InvivoGen for evaluation of this technology, and partner 3 provided the system to about 40 academic groups and 1 industrial partner in 10 countries (7 European, 2 Asian and 1 in the USA).

In September 2008 we organised an end-meeting expanded into a symposium. Project partners and scientists from outside the project, but working on related topics, attended the meeting/symposium. Here project results were presented in the form of oral and poster contributions together with results from invited speakers via oral contributions and via poster contributions from other attendees. This makes the available to a broader audience and places the project results in a broader context. Also potential for extension of project via new applications under FP7 was discussed.