BIOINFORMATICS ORIGINAL PAPER Vol. 27 no. 24 2011, pages 3364–3370
doi:10.1093/bioinformatics/btr555

Structural bioinformatics

Advance Access publication October 11, 2011

The design of optimal therapeutic small interfering RNA molecules targeting diverse strains of influenza A virus

Mahmoud ElHefnawi[1,2,3,*], Nafisa Hassan[2], Mona Kamar[2], Rania Siam[4], Anna Lisa Remoli[5], Iman El-Azab[6], Osama AlAidy[7], Giulia Marsili[5] and Marco Sgarbanti[5]

[1]Informatics and Systems Department, Division of Engineering Research, National Research Centre, [2]Yousef Jameel Science and Technology Research Center, [3]Department of Computer Science and Engineering, Egypt-Japan University of Science and Technology (E-JUST), New Borg El-Arab City, [4]Department of Biology, American University in Cairo, Cairo, Egypt, [5]Department of Infectious, Parasitic and Immune-mediated Diseases, Istituto Superiore di Sanità, Viale Regina Elena, 299, 00161 Rome, Italy, [6]Department of Information Technology, Faculty of Computers and Information, Cairo University, Cairo, Egypt and [7]Department of Biochemistry, University of Saskatchewan, Saskatchewan, Canada

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: There is an urgent need for new medications to combat influenza pandemics.

Methods: Using the genome analysis of the influenza A virus performed previously, we designed and performed a combinatorial exhaustive systematic methodology for optimal design of universal therapeutic small interfering RNA molecules (siRNAs) targeting all diverse influenza A viral strains. The rationale was to integrate the factors for highly efficient design in a pipeline of analysis performed on possible influenza-targeting siRNAs. This analysis selects specific siRNAs that has the ability to target highly conserved, accessible and biologically significant regions. This would require minimal dosage and side effects.

Results and Discussion: First, >6000 possible siRNAs were designed. Successive filtration followed where a novel method for siRNA scoring filtration layers was implemented. This method excluded siRNAs below the 90% experimental inhibition mapped scores using the intersection of 12 different scoring algorithms. Further filtration of siRNAs is done by eliminating those with off-targets in the human genome and those with undesirable properties and selecting siRNA targeting highly probable single-stranded regions. Finally, the optimal properties of the siRNA were ensured through selection of those targeting 100% conserved, biologically functional short motifs. Validation of a predicted active (sh114) and a predicted inactive (sh113) (that was filtered out in Stage 8) silencer of the *NS1* gene showed significant inhibition of the *NS1* gene for sh114, with negligible decrease for sh113 which failed target accessibility. This demonstrated the fertility of this methodology.

Contact: mahef@aucegypt.edu

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Received on May 11, 2011; revised on September 14, 2011; accepted on October 1, 2011

*To whom correspondence should be addressed.

1 INTRODUCTION

1.1 siRNA importance

RNA interference (RNAi) is a naturally occurring endogenous biological post-transcriptional cellular mechanism that regulates against foreign genetic elements, such as viruses and inserted gene transcripts, as well as in-home gene expression regulation. Small interfering RNA (siRNA) molecules utilize this mechanism to promote homology-dependent messenger RNA (mRNA) degradation (Ichim *et al.*, 2004). siRNA triggers the formation of the RNA-induced silencing complex (RISC) in which the double-stranded siRNA is incorporated and unwound, and then binds to the target mRNA sequence, resulting in its cleavage (Meister *et al.*, 2004). The utilization of siRNA as a molecular target to silence gene expression has been extensive as a research tool in functional genomics (Taylor1 *et al.*, 2008). The unprecedented advantage of siRNA molecules, due to their ability to effectively and specifically inhibit disease-causing genes, raised great expectations in therapeutic applications and drug discovery (Chen *et al.*, 2008; Kim and Rossi, 2007; Naito *et al.*, 2007). Advantages of siRNA therapeutics include their feasibility in clinical trials and the fact that they do not rely on an intact immune system. For a review of current siRNA drugs, see Vaishnaw *et al.* (2011).

1.2 Features of effective siRNAs

The identification of structural and sequence positional reference elements that characterize highly functional siRNAs have been elucidated in previous studies (Patzel, 2007). For a large-scale feature selection study, see Klingelhoefer *et al.* (2009). First, sequence space conservation is determinative, as a single base pair change in the 5′ or central region of the siRNA may affect functionality (Birmingham *et al.*, 2007). It is worth noting that mismatches found at the 3′ end of the antisense strand siRNA duplex have minimal to no effect on functionality (Naito *et al.*, 2009). Second, the sequence of the designed siRNA duplex including antisense positional preferences and words have been determined by early classical works (e.g. Amarzguioui and Prydz,

2004; Hsieh *et al.*, 2004) and refined by use of statistical and artificial intelligence models contribute to the activity of siRNAs by enhancing strand selection, target annealing and cleavage (Shabalina *et al.*, 2006). While unwanted features include repeats, low-complexity regions, immune-stimulating motifs and others (Gredell *et al.*, 2008; Klingelhoefer *et al.*, 2009; Lu and Mathews, 2008). Third, the thermodynamic properties including the structures of both the siRNA and targeted mRNA and its accessibility are highly determinant for siRNA efficacy. For example, palindromes may cause unwanted secondary structures, lowering the probability of siRNA loading onto the RISC. Factors that increase the duplex stability resulting in the inhibition of the unwinding, such as high GC content, are unfavorable. Notably, differential end stability represents the most significant single feature in determining siRNA efficacy, since the more flexible (less stable) 5′ end of the duplex is loaded onto RISC the greater the efficiency (Birmingham *et al.*, 2007; Ichihara *et al.*, 2007). Fourth, to ensure specificity to target regions, siRNA specificity can be enhanced by including a sequence alignment filter and eliminating all candidate siRNAs that have perfect or near-perfect complementarities with unintended targets (Birmingham *et al.*, 2007). siRNA can also induce moderate levels of gene knockdown by seed-mediated translation attenuation (Birmingham *et al.*, 2006). This process of off-targeting can be minimized by including design parameters that select siRNA with low 3′ untranslated region (UTR) seed complement frequencies (Jackson *et al.*, 2006b). In addition, siRNA length between 21 and 23 bp manages to overcome the non-specific interferon-mediated immune response (Jackson and Linsley, 2010).

1.3 Influenza and RNAi

The genome of influenza A viruses consists of eight single-stranded negative sense RNA segments that encode 10 or 11 viral proteins. The NS1 protein is an RNA binding protein that contributes to viral pathogenicity and plays a central role in the prevention of interferon-mediated antiviral response. NS1 protein is essential for the ability of influenza viruses to productively infect humans, and the generation of an H1N1 mutant virus bearing *NS1* deletions in its genomes results in the formation of an attenuated strain upon infection of macaca Rhesus monkeys (Baskin *et al.*, 2007). The feasibility of targeting influenza A virus was previously demonstrated in studies by Ge *et al.* (2003) and Zhou *et al.* (2007) who showed that siRNAs specific to conserved regions of the *NP* and *M* segments were successful potent and long-term inhibitors of influenza viral replication in cell lines, embryonated chicken eggs and mice. Expression plasmids that provide a more sustained siRNA production inside cells were also successfully utilized (Sui *et al.*, 2009; Zhou *et al.*, 2007, 2008). Interestingly, local successful delivery of nucleic acids to the respiratory tract was achieved using aerosols (Dreyfus *et al.*, 2007). However, targeting escape mutants and all diverse strains in addition to other drug design heralds are still to be met (McSwiggen and Seth, 2008). Although, databases for mammalian gene silencing exist (Ren *et al.*, 2006), iRNAs for pathogens silencing and therapeutic applications require careful design considerations.

In our previous work, a systematic genomic analysis of the influenza A virus was performed, which revealed genome-conserved regions, accessible regions, functional motifs, and a cross-species conservation and variability (ElHefnawi, 2011). Here, design and selection of optimal siRNAs for all diverse strains of the influenza A virus are performed by integration of the factors that affect the functionality, specificity and potency of the siRNAs. For more details on the bioinformatics applications to the design and selection of efficient siRNAs, see the chapter of ElHefnawi (2011). Efficient siRNA design is very important to increase potency and hence decrease dosage and possible side effects. Thus, a novel approach is deduced with high fertility in eliminating false positives which has been tested experimentally as a pilot phase on the *NS1* gene to test for both the siRNA that was predicted to be efficient and a siRNA that was predicted to give low efficacy due to target inaccessibility. Thus, providing a candidate best set of siRNAs for five segments of the virus recommended for animal models and clinical trials. The steps presented here include designing of all possible siRNAs using the consensus sequences of the conserved regions. This included both first-generation algorithms, which focused on sequence features, followed by second-generation algorithms, which included many other features including scoring, thermodynamics of the siRNAs and target accessibility. In Stage 7, novel multiscore-filtration was performed. Criteria were based on defining accepted threshold scores after mapping the second-generation scores to the inhibitory percent of the Heuskin/Novartis siRNA training set (Huesken *et al.*, 2005). This was followed by omitting siRNAs with off-target effects (<2 mismatches to the off-target mRNAs), siRNA undesirable sequence motifs and low-target accessibility profile. The outcome was a potential set of siRNAs for each of the studied segments, which can be potentially used as a new medication for the virus. We present the siRNA design methodology (Fig. 1), and elaborate on filtering, selection, scoring and optimizing steps. Experimental validation of two siRNAs targeting the *NS1* mRNA was performed; one that passed all stages of filtration and selection, and one that passed all stages except Stage 8 of target accessibility check, where it was predicted to be target inaccessible. This demonstrated the fertility of this approach in eliminating false positives and the marked significance of target accessibility on the functionality of siRNAs. The *NS1* gene was chosen for the validation phase because it was not tested before for silencing, as it is more variable than some of the other segments and thus being a harder target to handle (ElHefnawi, 2011). This study considered challenges and factors including: (i) viral resistance by continuous mutation (even a single mismatch) in the viral genome, which helps its escape from the targeted siRNA; (ii) viral resistance through changes in the target accessibility including changes in the RNA secondary structure—similar changes were reported with HIV (Suzuki *et al.*, 2008); (iii) wide off-target effects of siRNAs due to high sequence homology with human mRNA transcripts; and (iv) genetic re-assortment of influenza genome in the future that may lead to potential novel strains causing great pathogenicity. To overcome the first two challenges, our strategy included targeting the highly structurally conserved regions, targeting the virus with more than one siRNA to increase the chances of preventing possible mutations, and downregulating several essential biological functions. Our strategy to overcome the third challenge was achieved by completely omitting off-targeting siRNAs, since previous studies have revealed that partial complementation to cellular mRNAs and mismatches to target mRNA has a wide off-targeting effect (Jackson *et al.*, 2006a). This was overcome by adding a second phase of off-target filtering for siRNAs with seed matches to 3′ UTR of human mRNAs. To overcome the fourth challenge, we designed siRNAs that could target the highly conserved and functionally essential

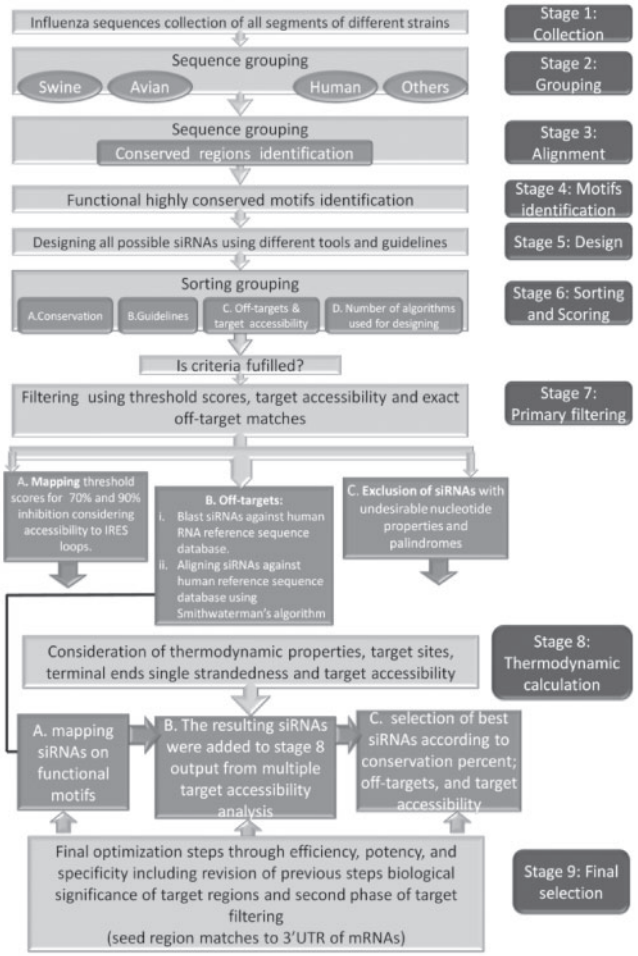

Fig. 1. Complete flow chart of our novel methodology.

regions in the genome (ElHefnawi, 2011). Also, inclusion of other species in the selection, analysis and design ensures that siRNAs remains valuable for future strains that would emerge from other hosts. The impact of the study would be to have a model for a systematic, comprehensive and novel *in silico* approach that was also experimentally tested, that could be replicated for other viruses and whose siRNA genome predictions close the gap between the accelerating fields of bioinformatics and experimental sciences. This approach is similar to the virtual screening of chemical databases for finding/optimizing leads, and we hope it will be further modified for other classes of non-coding RNA molecules like miRNA/RNA aptamers. To our knowledge, this systematic exhaustive approach of siRNA designing, filtration and selection, using an integrated set of tools and algorithms, as well as extensive bioinformatics analysis has not been previously performed.

2 METHODS

Figure 1, depicts a flow chart outlining the nine-stage approach for design ranking and selection of optimal siRNAs against the diverse strains of the influenza A virus.

Stages 1–4 were addressed in detail in ElHefnawi (2011). Influenza A virus genome segments from different hosts, different subtypes and different geographic regions were analyzed *in silico*. Genomic conserved regions across all diverse strains and hosts were extracted from the multiple sequence alignments and the conservation percentage was calculated. Additionally, analysis of interinfluenza and intrainfluenza host-origin variability in different segments and the GC percentage of the segments in the different hosts were performed. Conserved genomic regions were correlated with biologically functional motifs for appropriate functional annotation. The current work addresses Stages 5–9 as outlined in Figure 1. The workflow outlined in Figure 1 was semiautomated using BioPerl scripts.

2.1 Stage 5: siRNA design

Two siRNA datasets were designed by different tools using each of the consensus sequences of conserved regions separately as an input in order for the siRNA to target the viral mRNAs in all species. The first siRNA dataset was designed using siVirus (which uses a pre-compiled human influenza A segment sequences database) (Naito *et al.*, 2006). This was done by finding all possible siRNAs that are selected based on the conserved nucleotide region positions and then rank them based on their percent of targeting in all the studied strains. Also, it has a tool siDirect (Naito *et al.*, 2004), which checks siRNAs for off-targets.

The second siRNA dataset was designed using diverse tools: siRNA software design siRNA Design Software (SDS) (Ichim *et al.*, 2004; Yiu *et al.*, 2005) version 3, the THERMO-COMPOSITION 19 & 21 (Shabalina *et al.*, 2006), DSIR (Vert *et al.*, 2006), SCALES (Matveeva *et al.*, 2007) and ISCORE (Ichihara *et al.*, 2007) programs.

The siVirus designs siRNAs based on any of the three popularly utilized basic algorithms for design (Naito *et al.*, 2006). The SDS server uses first- and second-generation algorithms in the design process, as well as algorithms of different pharmaceutical companies (Yiu *et al.*, 2005). While the ISCORE software designs and scores the efficacy of siRNAs according to many different first- and second-generation algorithms, and has its own score (Ichihara *et al.*, 2007).

2.2 Stage 6: Ranking the siRNA by scoring and sorting

The total number of siRNAs designed using siVirus software (Dataset 1) were sorted and ranked according to many criteria:

(1) Conservation percentage of the target region among the sequences.

(2) The number of guidelines satisfied by the siRNAs. The basic guidelines for siRNA designs, established by Tuschl *et al.* (2001), Reynolds *et al.*, (2004), Amarzguioui and Prydz (2004), Ui-Tei *et al.* (2004) and Hsieh *et al.* (2004), were checked in the siRNAs. Ranking based on the fulfillment of these guidelines was performed.

(3) Off targets in the human genome.

Also, the siRNAs from both datasets were scored using all available second-generation scoring algorithms, including the i-Score Designer, DSIR, siRNA SCALES, THERMO-COMPOSITION 19 & 21 (Shabalina *et al.*, 2006) and S-Biopredsi programs (Ichihara *et al.*, 2007; Ichim *et al.*, 2004), which were based on different machine learning approaches for selecting the best siRNAs. The i-Score Designer includes most of these scoring algorithms implemented. THERMO-COMPOSITION 19 & 21 is based on thermodynamics as well as compositional features like dinucleotide frequency index. The SCALES tool also implemented linear regression to derive positional preferences for the siRNA sequence (Matveeva *et al.*, 2007).

2.3 Stage 7: primary filtering using threshold scores and exact off-target matches

Stage 7 was performed using three selection criteria; mapping threshold scores for 70 and 90% experimental inhibition and filtering the siRNAs below the average threshold scores of all 11 utilized algorithms, off-targets and exclusion of siRNAs with undesirable nucleotide properties.

Well-characterized siRNAs, with a measured inhibition percent to their target genes, were utilized in order to systematically interpret the scoring

results of the different algorithms and filter the best and most potent siRNAs. The Huesken experimental Biopredsi siRNAs learning dataset (Huesken *et al.*, 2005) was utilized, scored and experimental inhibition efficiency mapped to the different algorithm scores in order to calculate the average, start and end scores for the 70 and 90% cutoff scores. The siRNAs in both datasets were sorted according to these scores and underwent successive filtration, so the remaining siRNAs exceed the threshold of all algorithms. Next, the siRNAs were locally aligned using BLAST against the human mRNA reference sequence database (Meister *et al.*, 2004) using the parameters provided in the classical Birmingham protocol (Birmingham *et al.*, 2007). A BioPerl script was used for the automation of this step. The matches were manually screened and scrutinized, and the siRNAs with less than two mismatches to a human mRNA (off-targets) were first excluded, and later other highly matching siRNAs were filtered. Finally, siRNAs with unwanted parameters were filtered. These parameters include high GC stretches, trinucleotide palindromes, more than five consecutive repeated nucleotides, internal repeats and 4G stretches as well as other undesirable properties plus additional measures outlined in the Birmingham protocol (Birmingham *et al.*, 2007).

2.4 Stage 8: thermodynamic properties and target accessibility

A number of tools such as I-Score, SFOLD and RNAXs were used in this stage for target accessibility and thermodynamics calculation and evaluation. The i-Score Designer results (Ichihara *et al.*, 2007) provided the whole and partial dG of each siRNA. The SFOLD and RNAXs tools (Ding *et al.*, 2004) were used to calculate different contributions to the target accessibility using the consensus of each segment. This is accomplished by mapping highly probable single-stranded regions in the influenza A virus segments. The measure of the binding affinity and target accessibility combined was assessed using the RNAUP tool that is part of the RNAXs tool (Muckstein *et al.*, 2006; Tafer *et al.*, 2008). The number of single-stranded base pairs at the 5′ and 3′ ends of the target mRNA has recently been shown to significantly contribute to the effectiveness of siRNAs by Patzel *et al.* (2005) in their recent patent ‘structure of active guide RNA molecules and method of selection’. The same conclusion was reached by Gredell *et al.* (2008) and Tafer *et al.* (2008), therefore this was also used as criteria for filtering with a threshold of >0.5 average probability of the five nucleotides at the 5′ and four nucleotides at the 3′ UTRs.

2.5 Stage 9: final optimization steps

In substage 9a, the 90% potent siRNAs (from substage 7b) were mapped to functional motif regions on the influenza genome (100% conserved) as a first line of selection where the 90% siRNA set was analyzed to select the siRNAs that target functional motifs. Then, in substage 9b, the resulting siRNAs were added to Stage 8 output from multiple target accessibility analysis. Additionally, this was followed by filtering to include siRNAs which target highly accessible regions (>0.5 average probability at the 5′ and 3′ ends). This allowed us to combine highly potent siRNAs that are universally applicable for the diverse influenza strains and target biologically significant sites.

The second phase of off-targets search (implemented in 9c) filtered siRNAs with seed region complementarity to human mRNA 3′ UTR (Jackson *et al.*, 2006b). This was based on recent findings that exact complementarity at the 3′ UTR region with nucleotides 2–7 at 5′ end of the antisense strand of the siRNA molecule (seed region) enhances the chances of partial translational attenuation (Jackson and Linsley, 2010). This was done by searching the Ensemble genome browser (Hubbard *et al.*, 2009) for matches with the designed siRNAs seed region at 3′ UTR of the various human gene transcripts.

2.6 Best siRNAs

The efficiency and suitability of siRNA for use as universal silencers of influenza A virus were investigated by manual selection according to the following criteria: the final selection lists of ‘best siRNAs’ included only those siRNAs with target mRNA region conservation >90%, no first- or second-phase off-targets, low RNAUP total energy score for accessibility, and single-stranded probability score for terminal ends >0.5 using the single-stranded probability score generated by SFOLD and RNAXs.

3 EXPERIMENTAL MATERIALS & METHODS

Cell line: human embryonic kidney (HEK) cell line was grown in Dulbecco’s modified Eagle’s medium (Bio-Whittaker, Cambrex Bio Science, Verviers, Belgium), supplemented with 10% fetal bovine serum (Biological Industries, Kibbutz Beit Haemek, Israel) and antibiotics.

Plasmids: the coding sequence of the NS1 protein was obtained from the Italian National Influenza Center (ISS, Rome, Italy) belonging to an Italian isolate of the H1N1 2009 swine pandemic influenza A virus. The DNA vector expressing the above mentioned NS1 protein (pFLAG CMV2 NS1 H1N1v) was obtained. First, the NS1 coding sequence generated by *de novo* gene synthesis (GenScript USA Inc., Piscataway, NJ, USA) is provided into the pUC57 cloning vector. Then, it was subcloned, in-frame with the FLAG epitope, into the pFLAG CMV2 expression vector (Sigma-Aldrich, St Louis, MO, USA). The ‘in-frame’ insertion of the NS1 H1N1 coding sequence with the sequence of the FLAG epitope was verified by DNA sequencing. NS1-specific pU6 shRNA-expressing vectors (sh113 and sh114) as well as the luciferase-specific expressing vector (shLuc) were also obtained in the same way, which is *de novo* gene synthesis (GenScript USA Inc.), and then provided into the pUC57 cloning vector.

Transient transfection: HEK 293 cells were seeded at 1:6 dilutions from a 90% confluent plate 24 h prior to transfection. Transient transfection experiments were performed using the calcium phosphate transfection system (Life Technologies, invitrogen Corporation, Carlsbad, CA, USA). Upon 17 h, medium was replaced and cells were collected 7 h later.

Immunoblotting analysis: transfected HEK 293 cells were lysed in 10% sodium dodecyl sulfate loading dye supplemented with 5% β-mercaptoethanol. Then, whole cell extracts (WCEs) were loaded onto a 10% sodium dodecyl sulfate–polyacrylamide gel electrophoresis upon boiling at 100°C for 5 min and 30 s centrifugation to eliminate cell debris. Upon electrophoresis, WCEs were transferred onto a nitrocellulose membrane (Thermo Scientific, Rockford, IL, USA). Immunoblot analysis was performed by probing blots with a monoclonal antibody (Ab) against the FLAG epitope (Sigma-Aldrich, St Louis, MO, USA) and a polyclonal Ab against β-actin (Santa Cruz Biotechnology Inc., Santa Cruz, CA, USA), and then with anti-mouse or anti-goat horseradish peroxidase-coupled secondary antibody (1:2000 dilution) (Santa Cruz Biotechnology Inc.), respectively. The detection reaction was performed using the SuperSignal West Pico Chemioluminescent substrate (Thermo Scientific, Rockford, IL, USA).

4 RESULTS

As it is a big challenge to design universally applicable siRNAs for influenza that are potent, specific and non-toxic, we sought

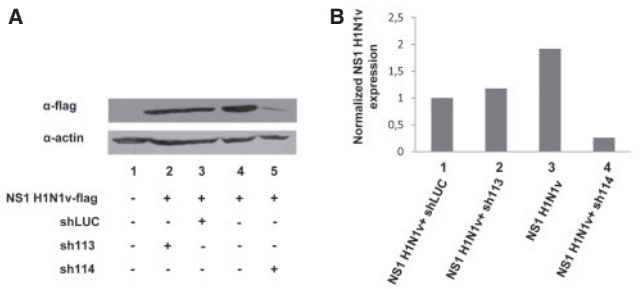

Fig. 2. NS1 expression upon co-transfection of NS1, and NS1 shRNA, expressing vectors, in HEK293 cells. (**A**) HEK 293 cells (80% confluent in 12-well plates) were transfected with 1 μg of FLAG NS1 from the H1N1 2009 pandemic virus (H1N1v) expression vector alone or together with 1 μg NS1 specific pU6 shRNA-expressing vectors (sh113 and sh114) or together with luciferase-specific pU6 shRNA-expressing vector (shLuc). Normalization for equal amount of plasmid DNA transfected was obtained by transfecting 2 and 1 μg of pFLAG empty vector, in the experimental controls were no NS1 protein and no shRNA were expressed (lane 1) or were NS1 protein was expressed but not shRNA (lane 2), respectively. Immunoblotting analysis of WCEs, collected 24 h post-transfection from HEK293 cells by using anti-FLAG and anti-actin Abs are shown. (**B**) The relative increase/decrease in the NS1 H1N1v expression upon co-expression of shRNAs was measured by densitometry analysis of the immunoblotting experiment and expressed as arbitrary units after normalization with the levels of actin.

first to exhaustively design all possible candidates, and then filter them out based on the many factors that affect siRNA efficiency. Continuing previous work that utilized siRNAs against the M and *NP* segments, the silencing of the *NS1* protein was also demonstrated experimentally as a test of our methodology in addition to checking the efficiency of the criteria by siRNAs that was filtered out in Stage 8 using the RNAXs tool as it had low single-stranded probability (Fig. 2). Moreover, we provided the methodology and results for each filtration stage in Supplementary File S1. The total siRNAs that were designed (Stage 5) from both datasets, the siRNAs remaining after the 70 and 90% inhibition cutoff scores filtration (Stages 6 and 7), the siRNAs after the target accessibility filtration (Stage 8) are all displayed sequentially in the different sheets for all the six segments. Ge *et al.* (2003) have formulated a broad inhibitor of influenza A virus H1N1 strain: NP1496 and PA 2087. Our methodology has similar extrapolated three siRNAs that target the same segments (1499, 1500 in the Stage 8 siRNAs *NP* segment and 2088 in Stage 7 siRNAs). Similarly, Zhou *et al.* (2008) have tested H5N1 targeting siRNAs.

Here, siRNAs targeting all diverse strains of influenza were sought through analyzing different hosts, finding common conserved regions and using the design power of the best recommended second-generation algorithms.

This was complemented by target accessibility analysis of the siRNAs using different approaches. Additionally, we selected siRNAs that target functional and completely conserved motifs to ensure a comprehensive exhaustive methodology for efficient potent siRNA selection. Out of the total designed siRNAs using siVirus, and second-generation tools, more than half remained after filtering the 70% inhibition scores and omitting siRNAs with less than two mismatches to human mRNAs. The RISC off-target effect was eliminated by excluding all siRNAs with two or less mismatches to human mRNAs. After mapping the siRNAs of 70% average experimental inhibition percent to various algorithms' scores, hundreds of siRNA that covered the five influenza genome remained. The highest number of siRNAs targets *PB1* (the largest segment). Cropping the list to the scores corresponding to the 90% inhibition reduced the numbers of siRNAs by more than a half, with more than a hundred siRNA for each segment. Using optimization steps, like choosing siRNAs with terminal ends having high single-stranded probability, the remaining siRNAs in each segment were in tens. Finally, using siRNAs that target functional conserved motifs, the number of siRNAs in each segment was reduced substantially to few siRNAs that we predict to be potent efficient silencers.

4.1 Experimental validation of the optimal siRNAs

As detailed earlier, two siRNAs were inserted separately into expression vectors along with controls and tested for their ability to diminish the NS1 protein production. Sh114 is the predicted potent siRNA which is seg8cr06_114 in Stage 9a dataset 2 in Supplementary File S1 after its insertion into the PUC57 vector. Sh113 is the filtered siRNA after Stage 8 target accessibility, which is seg8cr06_113 in Stage 9a dataset 2 in Supplementary File S1 after its insertion into the PUC57 vector. As illustrated in Figure 2, sh114 showed marked inhibition (0.3 relative to 2 for control) of the NS1 protein, while sh113 showed much less inhibition (1.2 relative to 2 for control).

The NS1 protein consists of two domains; the RNA-binding domain and the effector domain. It has various functions during influenza A virus infection such as protecting influenza A virus against the antiviral state, inhibiting several kinds of proinflammatory cytokines and blocking the maturation and exportation of the host cellular antiviral mRNAs. Thus, it is regarded as an appealing specific target for therapeutic intervention. This is the first demonstration of its inhibition using siRNAs.

The RNAXs tool which gave a lower accessibility for sh113 performed well in predicting target accessibility and this supports the impact of target accessibility for efficient siRNA design (Hofacker and Tafer, 2010). Measuring the level of the protein itself rather than the transcript gives a direct evidence of the silencing by siRNAs since both mRNA and protein levels are equally affected by this knock-down strategy (Tan *et al.*, 2005). Also, the rational design process that selected a few from thousands possible could further help for other tasks of therapeutic siRNA design.

The choice of the NS1 protein from the influenza A H1N1 2009 pandemic virus to test the efficacy of the siRNAs designed here has been made due to the inability of NS1 from this strain to inhibit the export of RNA polymerase II transcripts (Hale *et al.*, 2010); this is of crucial importance not to have a bias in the western blotting analysis, when the NS1 protein is expressed through an RNA polymerase II transcribed vector.

In conclusion, this study explores an exhaustive methodology to design the best siRNAs that can be tested as candidates for therapeutic treatment of the recent avian and swine influenza strains as well as emerging strains. Our study is a prototype for designing siRNA to target highly divergent viral genomes. The best siRNAs represented in Figure 3 that target five segments *PB2*, *PB1*, *NP*, *M* and *NS* are recommended as novel potent specific inhibitors of influenza A virus and the siRNA targeting the *NS1* gene was

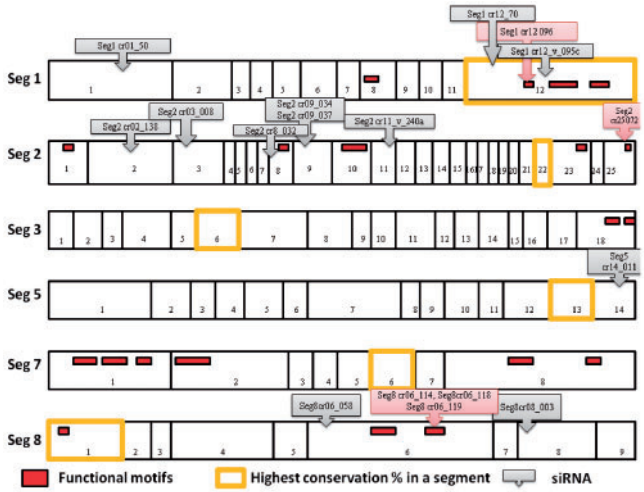

Fig. 3. Representation of the best set of siRNAs targeting functional motifs/conserved regions on accessible regions for six segments of the influenza A virus.

experimentally verified. These siRNAs were designed and analyzed using the consensus of different tools; they target highly conserved and accessible regions of the genome, and are highly specific with no seed region or near complete complementarity off-targets.

ACKNOWLEDGEMENTS

We acknowledge the kind advice, continuous discussions and scientific and networking assistance given by Dr Andrea Savarino throughout the years of the research. We acknowledge the help of Amr Aly in performing the off-target search especially the second phase; and of Mohamed Maysara for performing the conservation percent of the second dataset. We acknowledge the kind assistance of Dr Suher Zada and her support throughout. Also, we acknowledge the participation of Drs Simona Puzelli, Marzia Facchini and Isabella Donatelli from the National Influenza Center, ISS, Rome, Italy and Roberto Gilardi for ART work. We also acknowledge the assistance of Gehad AbdelRahman in revision and final editing.

Funding: Youssif-Jameel Science and Technology Research Centre (YJ-STRC); American University in Cairo (AUC); Italian Ministry of Health (grant: 'Studio e sviluppo di nuovi farmaci antivirali contro infezioni da A-H1N1').

Conflict of Interest: none declared.

REFERENCES

Amarzguioui,M. and Prydz,H. (2004) An algorithm for selection of functional siRNA sequences. *Biochem. Biophys. Res. Commun.*, **316**, 1050–1058.

Baskin,C.R. *et al.* (2007) Functional genomic and serological analysis of the protective immune response resulting from vaccination of macaques with an NS1-truncated influenza virus. *J. Virol.*, **81**, 11817–11827.

Birmingham,A. *et al.* (2006) 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods*, **3**, 199–204.

Birmingham,A. *et al.* (2007) A protocol for designing siRNAs with high functionality and specificity. *Nat. Protoc.*, **2**, 2068–2078.

Chen,Y. *et al.*. (2008) RNAi for treating hepatitis B viral infection. *Pharm. Res.*, **25**, 72–86.

Ding,Y. *et al.* (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.*, **32**, W135–W141.

Dreyfus,D.H. *et al.* (2007) Gene silencing in the therapy of influenza and other respiratory diseases: targeting to RNase P by use of External Guide Sequences (EGS). *Biologics*, **1**, 425–432.

Elbashir,S.M. *et al.* (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, **411**, 494–498.

ElHefnawi,M. *et al.* (2011) Identification of novel conserved functional motifs across most Influenza a viral strains. *Virol. J.*, **8**, 44.

Ge,Q. *et al.* (2003) RNA interference of influenza virus production by directly targeting mRNA for degradation and indirectly inhibiting all viral RNA transcription. *Proc. Natl Acad. Sci. USA*, **100**, 2718–2723.

Gredell,J.A. *et al.* (2008) Impact of target mRNA structure on siRNA silencing efficiency: a large-scale study. *Biotechnol. Bioeng.*, **100**, 744–755.

Hale,B.G. *et al.* (2010) Inefficient control of host gene expression by the 2009 pandemic H1N1 influenza A virus NS1 protein. *J. Virol.*, **84**, 6909–6922.

Hofacker,I.L. and Tafer,H. (2010) Designing optimal siRNA based on target site accessibility. *Methods Mol. Biol.*, **623**, 137–154.

Hsieh,A.C. *et al.* (2004) A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res.*, **32**, 893–901.

Hubbard,T.J. *et al.* (2009) Ensembl 2009. *Nucleic Acids Res.*, **37**, D690–D697.

Huesken,D. *et al.* (2005) Design of a genome-wide siRNA library using an artificial neural network. *Nat. Biotechnol.*, **23**, 995–1001.

Ichihara,M. *et al.* (2007) Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. *Nucleic Acids Res.*, **35**, e123.

Ichim,T.E. *et al.* (2004) RNA interference: a potent tool for gene-specific therapeutics. *Am. J. Transplant.*, **4**, 1227–1236.

Jackson,A.L. *et al.* (2006a) Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *RNA*, **12**, 1197–1205.

Jackson,A.L. *et al.* (2006b) Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA*, **12**, 1179–1187.

Jackson,A.L. and Linsley,P.S. (2010) Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nat. Rev. Drug Discov.*, **9**, 57–67.

Kim,D.H. and Rossi,J.J. (2007) Strategies for silencing human disease using RNA interference. *Nat. Rev. Genet.*, **8**, 173–184.

Klingelhoefer,J.W. *et al.* (2009) Approximate Bayesian feature selection on a large meta-dataset offers novel insights on factors that effect siRNA potency. *Bioinformatics*, **25**, 1594–1601.

Lu,Z.J. and Mathews,D.H. (2008) Efficient siRNA selection using hybridization thermodynamics. *Nucleic Acids Res.*, **36**, 640–647.

Matveeva,O. *et al.* (2007) Comparison of approaches for rational siRNA design leading to a new efficient and transparent method. *Nucleic Acids Res.*, **35**, e63.

McSwiggen,J.A. and Seth,S. (2008) A potential treatment for pandemic influenza using siRNAs targeting conserved regions of influenza A. *Expert Opin. Biol. Ther.*, **8**, 299–313.

Meister,G. *et al.* (2004) Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. *RNA*, **10**, 544–550.

Muckstein,U. *et al.* (2006) Thermodynamics of RNA-RNA binding. *Bioinformatics*, **22**, 1177–1182.

Naito,Y. *et al.* (2007) Optimal design and validation of antiviral siRNA for targeting HIV-1. *Retrovirology*, **4**, 80.

Naito,Y. *et al.* (2006) siVirus: web-based antiviral siRNA design software for highly divergent viral sequences. *Nucleic Acids Res.*, **34**, W448–W450.

Naito,Y. *et al.* (2004) siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference. *Nucleic Acids Res.*, **32**, W124–W129.

Naito,Y. *et al.* (2009) siDirect 2.0: updated software for designing functional siRNA with reduced seed-dependent off-target effect. *BMC Bioinformatics*, **10**, 392.

Patzel,V. (2007) In silico selection of active siRNA. *Drug Discov. Today*, **12**, 139–148.

Patzel,V. *et al.* (2005) Design of siRNAs producing unstructured guide-RNAs results in improved RNA interference efficiency. *Nat. Biotechnol.*, **23**, 1440–1444.

Reynolds,A. *et al.* (2004) Rational siRNA design for RNA interference. *Nat. Biotechnol.*, **22**, 326–330.

Ren,Y. *et al.* (2006) siRecords: an extensive database of mammalian siRNAs with efficacy ratings. *Bioinformatics*, **22**, 1027–1028.

Shabalina,S.A. *et al.* (2006) Computational models with thermodynamic and composition features improve siRNA design. *BMC Bioinformatics*, **7**, 65.

Sui,J. *et al.* (2009) Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat. Struct. Mol. Biol.*, **16**, 265–273.

Suzuki,K. *et al.* (2008) Closed chromatin architecture is induced by an RNA duplex targeting the HIV-1 promoter region. *J. Biol. Chem.*, **283**, 23353–23363.

Tafer,H. *et al.* (2008) The impact of target site accessibility on the design of effective siRNAs. *Nat. Biotechnol.*, **26**, 578–583.

Taylor1,D.L. *et al.* (2008) High content screening: a powerful approach to systems cell biology and drug discovery. In Gregory Moore, K. *et al.* (eds) *Methods in Molecular Biology Part III*. Humana Press, United States, NY, pp. 245–251.

Vaishnaw,A.K. *et al.* (2011) A status report on RNAi therapeutics. *Silence*, **1**, 14.

Tan,P.H. *et al*. (2005) Gene knockdown with intrathecal siRNA of NMDA receptor NR2B subunit reduces formalin-induced nociception in the rat. *Gene Ther.*, **12**, 59.

Ui-Tei,K. *et al*. (2004) Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res.*, **32**, 936–948.

Vert,J.P. *et al.* (2006) An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics*, **7**, 520.

Yiu,S.M. *et al.* (2005) Filtering of ineffective siRNAs and improved siRNA design tool. *Bioinformatics*, **21**, 144–151.

Zhou,H. *et al.* (2007) Effective small interfering RNAs targeting matrix and nucleocapsid protein gene inhibit influenza A virus replication in cells and mice. *Antiviral Res.*, **76**, 186–193.

Zhou,K. *et al.* (2008) RNA interference of avian influenza virus H5N1 by inhibiting viral mRNA with siRNA expression plasmids. *J. Biotechnol.*, **135**, 140–144.