Nanomedicine Journal

Received: Aug. 5, 2014; Accepted: Sep. 22, 2014
Vol. 2, No. 1, Winter 2015, page 29-38

Online ISSN 2322-5904
http://nmj.mums.ac.ir

Original Research

Design and bioinformatics analysis of novel biomimetic peptides as nanocarriers for gene transfer

Asia Majidi[1], Maryam Nikkhah[1], Faranak Sadeghian[1], Saman Hosseinkhani[2]*

[1]Department of Nanobiotechnology, Tarbiat Modares University, Tehran, Iran
[2]Department of Biochemistry, Tarbiat Modares University, Tehran, Iran

Abstract

Objective(s): The introduction of nucleic acids into cells for therapeutic objectives is significantly hindered by the size and charge of these molecules and therefore requires efficient vectors that assist cellular uptake. For several years great efforts have been devoted to the study of development of recombinant vectors based on biological domains with potential applications in gene therapy. Such vectors have been synthesized in genetically engineered approach, resulting in biomacromolecules with new properties that are not present in nature.
Materials and Methods: In this study, we have designed new peptides using homology modeling with the purpose of overcoming the cell barriers for successful gene delivery through Bioinformatics tools. Three different carriers were designed and one of those with better score through Bioinformatics tools was cloned, expressed and its affinity for pDNA was monitored.
Results: The resultszz demonstrated that the vector can effectively condense pDNAinto nanoparticles with the average sizes about 100 nm.
Conclusion: We hope these peptides can overcome the biological barriers associated with gene transfer, and mediate efficient gene delivery.

Keywords: Biomimetic, Chimeric peptide, Gene delivery, Nanocarrier

***Corresponding author:** Saman Hosseinkhani, Department of Biochemistry, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran.*
Tel: +982182884407, Email:saman_h@modares.ac.ir

➢ *Please cite this paper as:*
Majidi A, Nikkhah M, Sadeghian F, Hoseinkhani S. Design and bioinformatics analysis of novel biomimetic peptides as nanocarriers for gene transfer, Nanomed J, 2015; 2(1): 29-38.

Introduction

Gene therapy can be defined as a clinical strategy to prevent or treat human disease by transfer of therapeutic genes into specific cells of the patients. Transferring of genetic materials towards cell and biological tissues currently is carried out through two major approaches, viral and non-viral transfer (1, 2). In viral approach, recombinant viruses which are transformed by replacing part of their genomes with a therapeutic gene are used as gene delivery vehicles (3). Although viral vectors have been used as efficient devices of gene delivery in the majority of studies (4, 5) and ongoing clinical trials (6-8), their safety concerns such as immunogenicity, oncogenicity potential, probability of reversion to wild type virions and the costs, have been limited their usefulness in human gene therapy(9).

Non-viral systems are synthetic vectors which can be divided into categories; i) cationic polymers and polypeptides, ii) cationic lipids (including cationic liposome). Despite their low transfection efficiency in comparison to viral system, they are extensively considered and investigated because of their low toxicity (2, 9).

Peptides and peptide-based vectors are synthesized in two approaches, chemical and genetic engineering.

Recombinant peptides have advantages over other chemically synthetic peptides in that they are able to be prepared as monodisperse material with well-defined monomer sequence (10). Biomimetic peptide vectors are designed to mimic the viral characteristics and perform an array of self-guided function in order to prevail over the cellular barriers. The major challenge to this approach is that, each domain of the vector can preserve its functionality and doesn't interfere with the functions of other domain (11).

In this study, we designed three biomimetic peptides containing domains which could overcome cell barriers. The peptides contained multiple domains including: i) two tandem repeating units of truncated histone H1 to condense pDNA and protection from serum nuclease degradation (12, 13), ii) amphipathic cell penetrating peptide, MPG, consisting fusion peptide sequence derived from HIV gp41 (glycoprotein) to destabilize endosomal membrane and SV40 large T-antigen NLS, which was used as a nuclear localization signal(14).

The peptide domains particularly fusion peptide and truncated histone H1 should have expected secondary structure to producing functional peptide. The first peptide of our study consists of one repeating unit of above mentioned domains and for simplicity; it will be shown as MPG-2H1.

The second one is two tandem repeating units of MPG-2H1 and it will be shown as 2TMPG-H1. Finally; the third peptide is two reverse repeating units of MPG-2H1 and it will be shown as 2RMPG-H1 (Figure 1).

Bioinformatic analysis was then carried out for all three peptides, from sequence to structure, to evaluate the structural stability in expression and production processes. Secondary structure prediction for three peptides was done by I-TASSER server because the correct and expected secondary structures of domains leads to producing functional peptides.

One peptide with the highest suitable score in homology modeling assay and Bioinformatics tools was chosen to investigate its DNA-binding ability by analyzing the particle size and zeta potential of prepared nanoparticle (DNA/Peptide complex).

Materials and Methods

Peptides design

To protect DNA from solution to the cell nucleus, peptide-based vectors must navigate extra and intracellular barriers. Accordingly the vector should be able to: a) condense DNA for minifying DNA from a large micro-meter scale to a smaller nano-meter scale which is suitable for endocytic

uptake and protection from serum nuclease degradation and, b) enhance the release of the gene from the endosomal compartment into the cytosol, and c) facilitate the displacement of DNA from the cytosol to the nucleus. These peptides with chimeric architecture can be developed to mimic viral characteristics to facilitate efficient gene delivery (2, 15).

The genes encoding MPG-2H1, 2TMPG-2H1 and 2RMPG-2H1 were designed based on the above-mentioned motif arrangements (Figure 1) and cloned in pET-28a using *Nde*I and *Xho*I restriction sites, and under the control of T7 promotor.

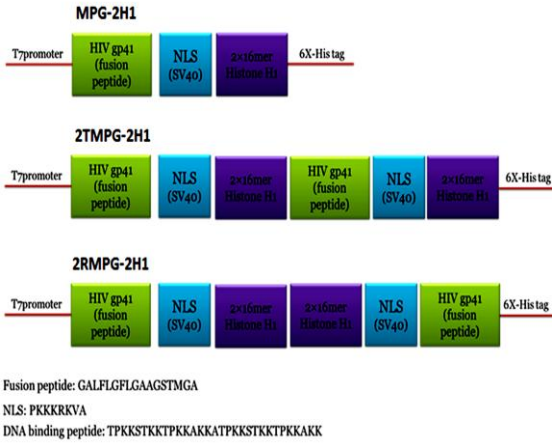
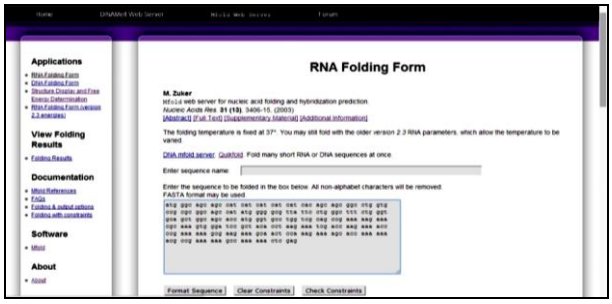


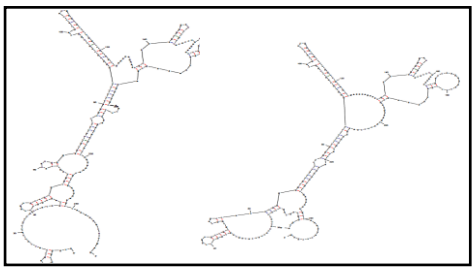
Figure 1. The schematic image of the peptides with related sequences.

The stability of the corresponding mRNA secondary structure was evaluated by *Mfold web server* which is available at http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form. *mfold* package for RNA and DNA secondary structure prediction is based on the nearest neighbor thermodynamic rules, and predicts a minimum free energy, ΔG, for folding of any particular base pair (16). This server takes the desired gene sequence as the input (Figure 2a) and displays images of all computed folded structures and their initial ΔG.

a

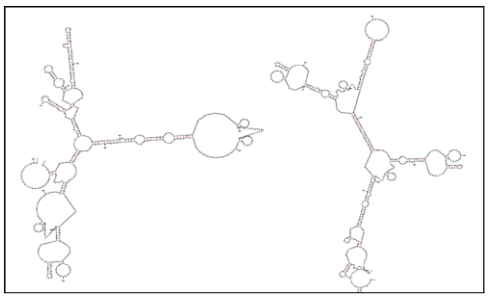


b



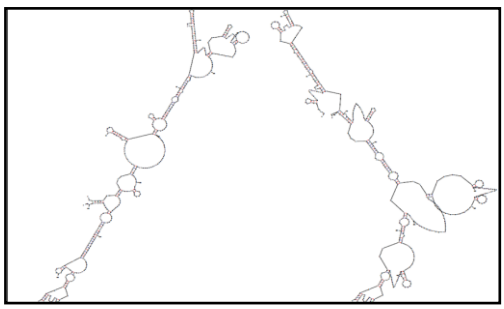
dG = -54.47 [Initially -68.00] dG = -47.69 [Initially -64.60]

c



dG = -102.05 [Initially -123.60] dG = -92.57 [Initially -117.8]

d



dG = -85.29 [Initially -113.30] dG = -82.50.69 [Initially -107.80]

Figure 2. prediction of mRNA secondary structure by mfold server: a)input, showing the nucleotide sequence of the query; MPG-2H1 gene, (b-d)output data of mfold server with the most and less related ΔG, for b) MPG-2H1; c) 2TMPG-2H1; d) 2RMPG-2H1.

Peptides sequences

The corresponding amino acid sequences for each motif of peptides are as follow; FP: GALFLGFLGAAGSTMGA; H1: TPKKSTKKTPKKAKKATPKKSTKKTP KKAKK; NLS: PKKKRKVA. So the final amino acid sequences of these constructs after expression in pET28a will be:

MPG-2H1;

MGSSHHHHHHSSGLVPRGSHMGALF LGFLGAAGSTMGAWSQPKKKRKVGS ATPKKSTKKTPKKAKKATPKKSTKKT PKKAKKLEHHHHHH

2TMPG2H1;

MGSSHHHHHHSSGLVPRGSHMGALF LGFLGAAGSTMGAWSQPKKKRKVGS ATPKKSTKKTPKKAKKATPKKSTKKT PKKAKKEFGALFLGFLGAAGSTMGA WSQPKKKRKVGSATPKKSTKKTPKK AKKATPKKSTKKTPKKAKKLEHHHH HH

2RMPG2H1;

MGSSHHHHHHSSGLVPRGSHMGALF LGFLGAAGSTMGAWSQPKKKRKVGS ATPKKSTKKTPKKAKKATPKKSTKKT PKKAKKEFKKAKKPTKKTSKKPTAKK AKKPTKKTSKKPTASGVKRKKKPQS WAGMTSGAAGLFGLFLAGLEHHHHH H

The analysis of protein or peptide sequences provides the information about the preference of amino acid residues and their distribution along the sequences which is required for understanding the secondary and tertiary structures of proteins and their functions. Several tools are available on the Internet to compute different parameters of amino acid sequences.

For example, the computation of amino acid composition of a protein is available at *http://www.expasy.ch/tools/protparam.html*

This server takes the amino acid sequence as the input (Figure 3a) and displays the amino acid composition, atomic composition, molecular weight, theoretical pI, extinction coefficients, estimated half-life (in yeast, *E. coli*, and mammalian reticulocytes) and Instability index as outputs (Figure 3b).

In the case of the above mentioned biomimetically designed peptides theoretical pI is an important parameter when purifying them by metal affinity chromatography and the instability index provides an estimate of the stability of peptides in a test tube.

The amino acid sequences of all three peptides were imported into this server (17).

Secondary structure prediction

As the biological functions of the proteins and peptides are defined by their 3D structure (which commands how the protein interacts with ligands or other protein molecules), predicting the protein structure is one of the most common motivations for applying the structural information to gain insight into the protein's biological function.

The biological utility of the predicted protein models relies on the accuracy of the structure prediction (18). The iterative threading assembly refinement (I-TASSER) server is an incorporated platform for automated protein structure and function prediction.

This server is based on the sequence-to-structure-to-function paradigm.

I-TASSER first starts from an amino acid sequence and yield three-dimensional (3D) atomic models from multiple threading alignments and iterative structural assembly simulations. In fact, this server provides a complex procedure in protein structure prediction, which combines various techniques such as threading, *ab initio* modeling and atomic-level structure refinement approaches and has been ranked as the best method for the automated protein structure prediction.

In comparison with a number of other useful online structure prediction tools like *HMM-HMM*, *GenTHREADER*, *Phyre* server, *Robetta* server, *PredictProtein* server, etc., the uniqueness of the I-TASSER server is in the significant

accuracy and reliability of full- length structure prediction (19).

a

b

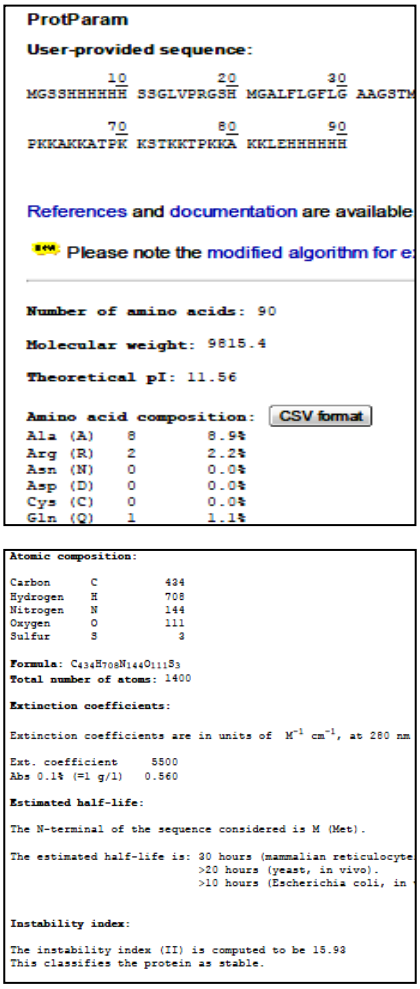


Figure 3. a) Calculation of amino acid composition by Expasy/ProtParam servers: input, showing the amino acid sequence of the query; MPG-2H1, b) output data of Expasy/ProtParam servers for MPG-2H1; showing the amino acid composition and other important data of this peptide.

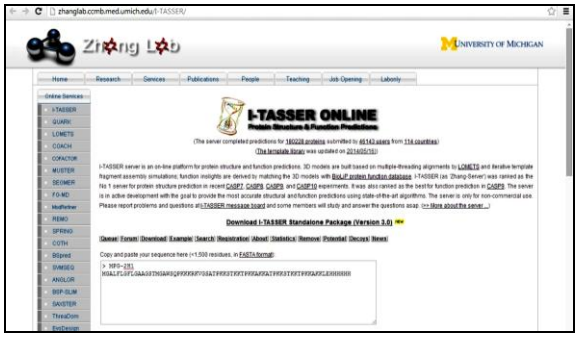
The peptides sequences in FASTA format were imported as input into the I-TASSER server which is available at http://zhanglab.ccmb.med.umich.edu/I-TASSER/ (Figure 4a). A typical secondary structure prediction contains three states: alpha helix (H), beta strand (S) and coil (C), with confidence scores for each residue which are shown in result pages of this server. Also the images of up to five predicted models, with highlighted regular secondary structures are illustrated.

This will help quickly in verifying the topology of the query protein from the modeled structure(s). As described by the I-TASSER server, C-score is a confidence score for estimating the quality of the predicted models by I-TASSER. C-score is typically in the range of -5 and 2 where a C-score of higher value signifies a model with a high confidence and vice versa.

The I-TASSER server has a standard for measuring structural similarity between two structures, which is known as TM-score. When the native structure is known, TM-score usually measures the accuracy of modelled structure. A TM-score>0.5 suggests a model of correct predicted topology and a TM-score<0.17 indicates a random similarity (19).

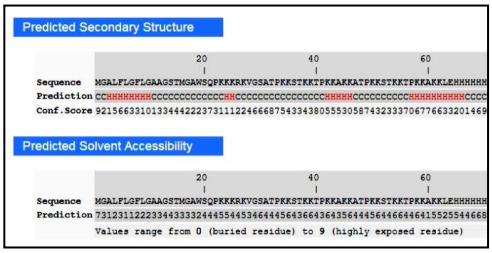
Nanoparticle preparation and characterization

To confirm DNA-binding ability of peptide, 0.5 µg plasmid with luciferase as reporter gene was mixed with increasing amounts of the peptide in several charge ratios (nitrogen to phosphate (N/P) ratio) and incubated at room temperature for 30 min. Zeta potential and size of the nanocarriers play key roles in transfection efficiency. So it is important to know about particle characteristics. The Zeta potential, average particle size and the particle size distribution of the complexes were determined by Zetasizer Nano ZS instrument (Malvern Instruments, UK). In order to do these measurements, the complexes of pDNA (500 ng) and MPG-2H1 peptide at N:P ratios (nitrogen to phosphate ratio) of 2 and 8 were prepared and measurements were done in the final volume of 1ml PBS buffer.

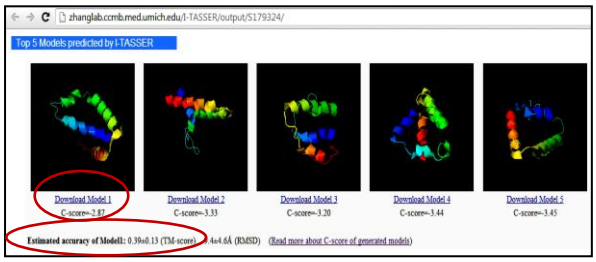


a

b



c



d

Figure 4. Prediction of peptide secondary structure by I-TASSER server: a)input, showing the peptide sequence of the query; MPG-2H1. An illustrative example of the I-TASSER result page (b–c) the page shows (b) query sequence in FASTA format; (c) Predicted secondary structure and solvent stability of the query protein The secondary structures in the model are highlighted in red (for α-helices); and (d) image of the top five predicted models and links for downloading the PDB-formatted structure files. The confidence score for estimating the model quality is reported as C-score and TM-score (red circles).

Results and Discussion

mRNA stability

The *Mfold web server* predicted eleven structures for MPG-2H1 folded mRNA with initial ΔG in the range between -68 and -64 kcal/mol and ΔG of folded structure formation between -54 and -47 kcal/mol (data not shown). Structures 1 and 11 with the most and less related ΔG are shown in figure 2b. The results showed all 11 predicted folded structure of this mRNA are stable and formed spontaneously. This server also predicted 15 fifteen structures for 2TMPG-2H1 and twenty structures for 2RMPG-2H1 folded mRNA with initial ΔG in the range between -123 and -117 kcal/mol for 2TMPG-2H1 and -113 and -107 kcal/mol for 2RMPG-2H1 (data not shown). Two examples of predicted secondary structure and the ΔG of folded structure formation for 2TMPG-2H1 and 2RMPG-2H1 are shown in figure 2c and 2d; respectively. According to calculated ΔG by this server, it seems that the mRNA secondary structures of all constructs are stable and the 2TMPG-2H1 mRNA is more stable than the others.

Analyzing of peptides sequences

According to the results predicted by *protparam* server the lysine content of the designed peptides, is 22% to 26% of the total and the peptides are basic in nature. As above mentioned theoretical pI is an important parameter when purifying these peptides by metal affinity chromatography, while the basic peptides have generally positive charge at neutral pH and their affinity decrease because of Ni^{+} Repulsive force. In these cases the pH of lysis buffer should be adjust based on pI of peptides to neutralize peptide and stable linkage between peptides and nickel-Sepharose beads of column.

The instability index (II) is computed to be in the rang between 14 and 16, so these peptides are classified as stable proteins. If the index is less than 40, then it is probably stable in the test tube. If it is greater then it is probably not stable. Finally the estimated half-life for all of them is estimated to be 30 hours in mammalian reticulocytes (in vitro), more than 20 hours in yeast (in vivo) and more than 10 hours in *Escherichia coli* (in vivo)(table 1). It means that these peptides have expression capability in all eukaryotic

and prokaryotic expression systems.The obtained result of *protparam* server for all three peprtides are shown in table 1.

Table 1. Output datas of protparam server for all three peprtides.

	Number of amino acides	Molecular weight(D)	pI	Ratio of lysine(%)	Extinction coefficient	Instability index	Estimated half life
MPG-2H1	90	9815.4	11.56	22.2	5500	15.93	30 h mammalian reticulocytes >20 h in yeast >10 h in Ecoli
2TMPG-2H1	153	16529.5	11.64	26.1	11000	15.36	30 h mammalian reticulocytes >20 h in yeast >10 h in Ecoli
2RMPG-2H1	153	16529.5	11.64	26.1	11000	14.95	30 h mammalian reticulocytes >20 h in yeast >10 h in Ecoli

Secondary structure prediction

The result page of I-TASSER includes the submitted amino acid sequence in FASTA format (Figure 4b) and the predicted secondary structures with the highest confidence score for each residue. The confidence score for each residue is shown with values ranging between 0 and 9, in which a higher score indicates a prediction with higher confidence (Figure 4c) and the images of up to five predicted models, with highlighted regular secondary structures are illustrated in Fig. 4d. The I-TASSER result page of MPG-2H1 construct, showed that the C-score for the first model is -2.87 which is greater than the other predicted models and the value of TM-score is 0.39±0.13. The C-score and TM-score are -3.62 and 0.32±0.11 for 2TMPG-2H1, -4.06 and 0.28±0.09 for 2RMPG-2H1; respecti-vely (data not shown). This server only reports the quality prediction (TM-score) for the first model of each constructs, because the correlation between C-score and TM-score is weak for other predicted models. Even though the C-score and TM-score values of all constructs are confirming the correct topology has been built up, the comparison between three construct shows that the predicted model of MPG-2H1 has higher C-score and TM-score and hence more confidenent secondary structure than the others, so this construct was selected for characterization tests.

The coordinate file (in PDB format) of each constructed model has been viewed by PYMOL software which is the molecular visualizing program for the modeled protein and the predicted functional sites. The 3D models of the designed peptides represent the precise position of secondary stucture of peptides with details like the orientation of N and C-terminal of peptides, α-helix of Fusion Peptide (FP) and β-turn structures of 16 mer H1 histon motif (Figure 5).

The earlier studies about MPG showed that this peptide is chemically stable, highly soluble in physiological buffers such as PBS and exhibits a very versatile structure. CD mesearment of MPG had shown that this peptide is folded into an α-helix in the presence of trifluoroethanol (20%) or into a β-sheet structure in PBS (20). Crystal Structure of Fusion Peptide (FP) of HIV-1 gp41 indicated the occurrence of the helical form of FP in this protein (21).

α-helix structure of Fusion Peptide (FP) domain is crucial in the mechanism of endosomal pathway (22, 23).

The 3D models obtained by I-TASSER server confirmed the presence of FP α-helix structure in all the designed peptides (Figures 5a1, 5b1 and 5c1).

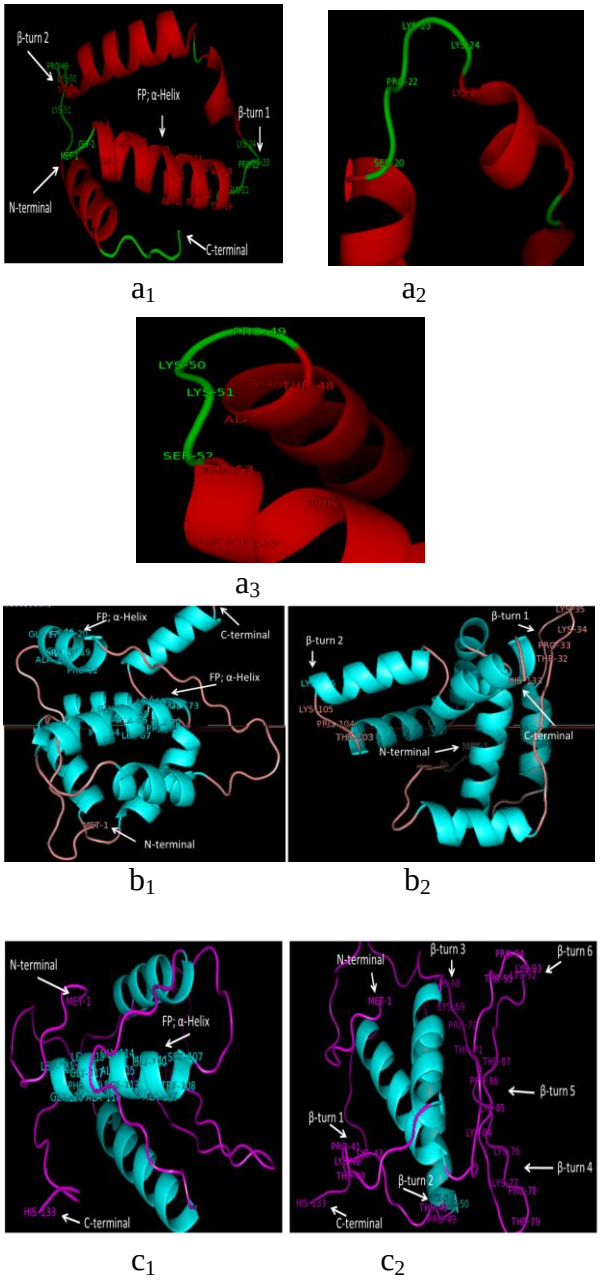


Figure 5. Illustrative 3D models of peptides obtained by I-TASSER server and opened by PYMOL software; a) MPG-2H1; a_1) the N and C-terminal of peptides, FP α-helix and β-turns structures, a_2- a_3) closer looks of β-turn1 and β-turn 2 respectively. b) 2TMPG-2H1; b_1) the N and C-terminal of peptides and two FP α-helix structures, b_2) structures of β-turns 1 and 2. c) 2RMPG-2H1; c_1) the N and C-terminal of peptides and only one FP α-helix, c_2) structures of β-turns 1 to 6.

The 16mer histone H1 motif in our investigated peptides has four repeating units of TPKK. The previous studies showed that the S/TPKK containing peptides adapt β-turn structures. Such a structure has been implicated in the binding of the peptide motif to the DNA molecules (24).

Another studies also suggest that the two β-turn structures present in the 16 mer (TPKK) peptide could be important in facilitating binding to different regions of duplex DNA thereby bringing about close packing and condensation (25).

In 3D model of MPG-2H1 and 2TMPG-2H1 two β-turn structures and in 2RMPG-2H1 3D model six β-turn structures are clearly seen (Figures $5a_2$, $5a_3$, $5b_2$ and $5c_2$).

Evaluation of particles characteristics

Determination of the size and charge of the nanoparticles consisting MPG-2H1 peptide and DNA at different N:P ratios indicates that nanoparticles with different average sizes can be obtained. At N:P ratio of 8, the pDNA was condensed into nanoparticles with 107.7 nm size and 18.5 mV surface charge, while at N:P ratio of 4, the size and surface charge of nanoparticles is 379.5 nm and 5.15 mV; respectively (Figure 6). According to many recent studies (11-13), the nanoparticles with above-mentioned characteristic have ability to overcome the biological barriers and mediate efficient gene transfer.

Conclusion

In coclusion, according to the results presented in this paper, we have successfully predicted that the biomimetic peptides with chimeric architecture can gain expected functional secondary structures. This can extend our knowledge for design of efficient delivery systems that can efficiently condence DNA into nanoparticles and can be fine-tuned for gene transfer needs. Such systems can be programmed to transfer genes to various cell types.

Acknowledgments

Financial support of this work was provided by Research Council of Tarbiat Modares University.

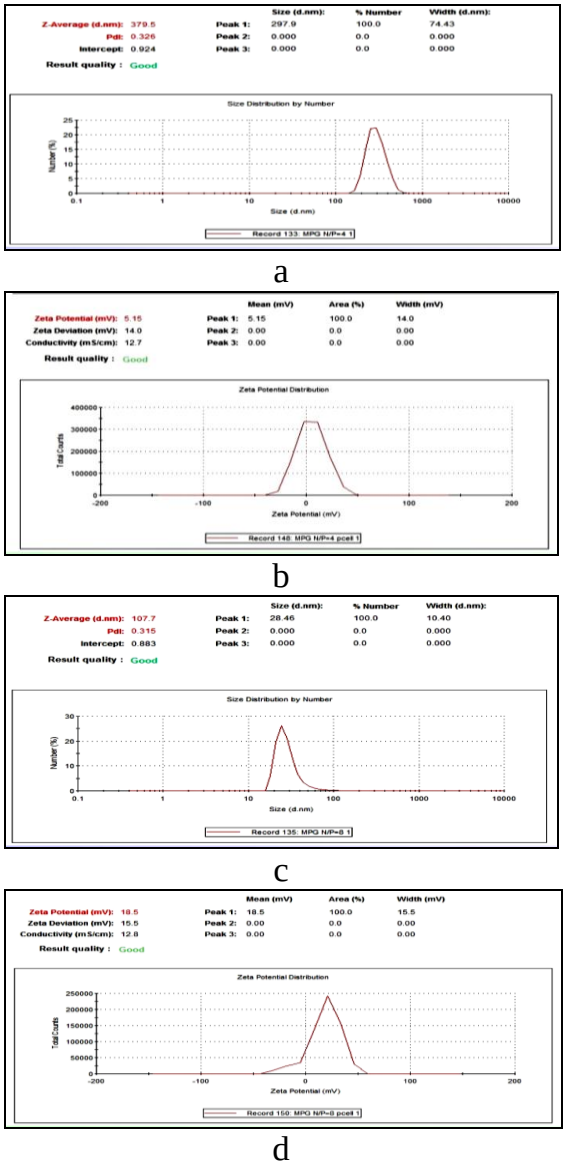


Figure 6. Size (nm) and zeta potential (mv) of nanoparticles consist of MPG-H1 peptide and DNA at N:P ratios of 4 and 8; a-b) size and zeta potential of nanoparticles N:P ratios of 4, c-d) size and zeta potential of nanoparticles N:P ratios of 8. Their related result quality and plots were shown.

References

1. Crystal RG. Transfer of genes to humans: early lessons and obstacles to success. Science. 1995; 270(5235): 404-410.
2. Pack DW, Hoffman AS, Pun S, Stayton PS. Design and development of polymers for gene delivery. Nat Rev Drug Discov. 2005; 4(7): 581-593.
3. Somia N, Verma IM. Gene therapy: trials and tribulations. Nat Rev Genet. 2000; 1(2): 91-99.
4. During MJ. Adeno-associated virus as a gene delivery system. Adv Drug Deliv Rev. 1997; 27(1): 83-94.
5. Vile RG, Tuszynski A, Castleden S. Retroviral vectors. From laboratory tools to molecular medicine. Mol Biotechnol. 1996; 5(2): 139-158.
6. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, Gross F, Yvon E, Nusbaum P, et al. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. Science. 2000; 288(5466): 669-672.
7. Sheridan C. Gene therapy finds its niche. Nature biotechnology, 2011. 29(2): 121-128.
8. Davidson BL, Breakefield XO. Viral vectors for gene delivery to the nervous system. Nat Rev Neurosci. 2003; 4(5): 353-364.
9. McCarthy HO, Wang Y, Mangipudi SS, Hatefi A. Advances with the use of bio-inspired vectors towards creation of artificial viruses. Expert Opin Drug Deliv. 2010; 7(4): 497-512.
10. Canine BF, Hatefi A. Development of recombinant cationic polymers for gene therapy research. Adv Drug Deliv Rev. 2010; 62(15): 1524-1529.
11. Wang Y, Mangipudi SS, Canine BF, Hatefi A. A designer biomimetic vector with a chimeric architecture for targeted gene transfer. J Control Release. 2009; 137(1): 46-53.
12. Soltani F, Sankian M, Hatefi A, Ramezani M. Development of a novel histone H1-based recombinant fusion peptide for targeted non-viral gene delivery. Int J Pharm. 2013; 441(1-2): 307-315.
13. Sadeghian F, Hosseinkhani S, Alizadeh A, Hatefi A. Design, engineering and preparation of a multi-domain fusion vector for gene delivery. Int J Pharm. 2012; 427(2): 393-399.
14. Morris MC, Deshayes S, Heitz F, Divita G. Cell-penetrating peptides: from molecular mechanisms to therapeutics. Biol Cell. 2008; 100(4): 201-217.
15. Morris MC, Depollier J, Mery J, Heitz F, Divita G. A peptide carrier for the delivery of biologically active proteins into mammalian cells. Nat Biotechnol. 2001; 19(12): 1173-1176.
16. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003; 31(13): 3406-3415.
17. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999; 112: 531-552.

18. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC bioinformatics. 2008; 9: 40.
19. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 2010; 5(4): 725-738.
20. Morris MC, Vidal P, Chaloin A, Heitz F, Divita G. A new peptide vector for efficient delivery of oligonucleotides into mammalian cells. Nucleic Acids Res. 1997; 25(14): 2730-2736.
21. Buzon V, Natrajan G, Schibli D, Campelo F, Kozlov MM, Weissenhorn W. Crystal structure of HIV-1 gp41 including both fusion peptide and membrane proximal external regions. PLoS Pathog. 2010; 6(5): e1000880. doi: 10.1371/journal.ppat.1000880.
22. Veldhoen S, Laufer SD, Trampe A, Restle T. Cellular delivery of small interfering RNA by a non-covalently attached cell-penetrating peptide: quantitative analysis of uptake and biological effect. Nucleic Acids Res. 2006; 34(22): 6561-6573.
23. Deshayes S, Gerbal-Chaloin S, Morris MC, Aldrian-Herrada G, Charnet P, Divita G, et al. On the mechanism of non-endosomial peptide-mediated cellular delivery of nucleic acids. Biochim Biophys Acta. 2004; 1667(2): 141-147.
24. Suzuki M, Gerstein M, Johnson T. An NMR study on the DNA-binding SPKK motif and a model for its interaction with DNA. Protein Eng. 1993; 6(6): 565-574.
25. Khadake JR, Rao MR. Condensation of DNA and chromatin by an SPKK-containing octapeptide repeat motif present in the C-terminus of histone H1. Biochemistry. 1997; 36(5): 1041-1051.