OJB™

Online Journal of Bioinformatics©

8 (1) : 115-138, 2007

In-silico TAT-PTD prediction of cell penetrating peptides.

Tandon C, Aggarwal A, Goel P, Sengupta D, Naik P

Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Distt.-Solan, Himachal. Pradesh, India (Pin-173215).

ABSTRACT

Tandon C, Aggarwal A, Goel P, Sengupta D, Naik P, In-silico TAT-PTD prediction of cell penetrating peptides, Online Journal of Bioinformatics, 8 (1) : 115-138, 2007. Cell Penetrating Peptides (CPP) enhance transport of membrane and secretory proteins across hydrophobic membranes. Natural or synthetic CPP water soluble peptides translocate across cell membranes and are used for intracellular delivery of large hydrophilic molecules. HIV-TAT PTD with an amino acid sequence of **YGRKKRRQRRR** is the core of TAT protein in HIV. Eight of the 11 TAT-PTD amino acids are hydrophilic yet this molecule diffuses readily through hydrophobic bi-lipid membranes. The effect of charged amino-acids in TAT-PTD on membrane transport was evaluated *In silico*. A dataset of 168 sequences was created mutating the original TAT-PTD 11-amino acid sequence. Qikprop was used to predict ADME properties and a correlation of R^2 = 73.7% between partition coefficient and skin permeability was found.

Keywords: Partition Coefficient, Skin Permeability, TAT- PTD, Cell Penetrating Peptides**,** Homology modeling.

INTRODUCTION

Protein transduction domains (PTDs) are short basic peptide sequences present in many cellular and viral proteins that mediate translocation across cellular membranes. HIV TAT-PTD with sequence YGRKKRRQRRR is also known to have such transduction activity (Green and Lowenstein, 1988). Small fragments of TAT, comprising the basic region of TAT, have been found to traverse biological membranes *in vivo* within 5-10 minutes. TAT (Trans-acting Activator of Transcription) is found in Human Immunodeficiency Virus (HIV-1) and translocates readily across mammalian cell membranes. The TAT PTD is an 11-amino acid sequence residing in the basic domain of HIV-TAT but has 6 arginine and 2 lysine residues, making the peptide highly cationic (Snyder EL and Dowdy SF, 2004). The sequence of TAT-PTD (molecular weight, 1560.81D) is:

H_3N^+-Tyr-Gly-Arg$^+$-Lys$^+$- Lys$^+$-Arg$^+$ -Arg$^+$-Gln-Arg$^+$-Arg$^+$-COO$^-$

Covalently binding the TAT peptide to proteins, oligonucleotides, phages, liposomes or even nanoparticles, allows these molecules to traverse the cell membrane (Binder H et al., 2003 and Prochiantz A, 2000). The peptide sequence is able to traverse almost all tissue cells including brain, and has a wide range of potential delivery applications. TAT-mediated delivery appears to be independent of cargo size. It has the ability to deliver therapeutic cargo, while avoiding bringing non-linked molecules into or out of the cell **(**Schwarze S et al., 1999**).** The mechanism of cellular entry is currently unknown. It is known that the basic amino acids within the TAT sequence play a central role. The chirality of the peptide backbone has no affect on the cellular uptake. The inverse and retro forms of TAT peptide, synthesized with D-amino acids or in a reversed linear orientation, are as efficient as the native peptide. (Wender PA et al., 2000)

Membrane proteins and secreted proteins must translocate across hydrophobic environments such as the lipid bilayer of biological membranes (Thoren PE et al., 2003). Because numerous proteins reside in exactly that non polar milieu or are forced across it in the process of secretion, thus there may be certain mechanism to overcome the unfavorable free energy of transferring the charges on proteins from aqueous phases into and through the bilayer (Trehin R and Merkle HP, 2004). For this transfer, many CPPs i.e. Cell Penetrating Peptides have been discovered (Richard JP, 2003). These are natural or synthetic water soluble peptides that translocate cell membranes with high efficiency and low lytic activity. Because they have successfully been used for the efficient intracellular delivery of large hydrophilic molecules like oligonucleotides and proteins, CPPs have emerged as promising tools in drug and protein delivery (Lindgren M et al., 2000). CPPs bearing the nuclear localization sequence (NLS) of SV40 large T antigen (Goldfarb DS et al., 1996) at their C-terminal and either the signal peptide of the Ig (V) light chain of *Caiman crocodylus* (SP-NLS) (Chaloin L et al., 1997) or a derivative of the fusion sequence of HIV1 gp41 protein **(**Chaloin L et al., 1998), as the hydrophobic sequences **(**FP-NLS**)** were shown to deliver oligonucleotides into mammalian cells independently of the endosome pathway (Vidal P et al,1997) . Recently, CPP derived from the FP-NLS peptide with the sequence GALFLAFLAAALSLMGLWSQPKKKRKV**,** called thereafter P $_{(a)}$, was synthesized. (Thomas P et al., 2005)

In the present study, we have made an attempt to predict the sequence requirements for 11-amino acids of HIV TAT-PTD to mediate translocation across cellular membranes.

MATERIALS AND METHODS

The basic residues in the original TAT sequence were mutated position-wise with neutral residue (alanine) and acidic residues (aspartic acid and glutamic acid) as shown:

YGRKKRRQRRR	(OriginalTAT-PTD)
YG***A**KKRRQRRR	(Mutated by Neutral amino acid at place of arginine)
YG***E**KKRRQRRR	(Mutated by Acidic amino acid at place of arginine)
YG***D**KKRRQRRR	(Mutated by Acidic amino acid at place of arginine)

168 Mutated sequences were obtained as described and are the dataset. Following its creation, the structures of these sequences were predicted using The “Prime” program (Schrödinger) based on the concept of Homology Modeling. The structural database of PDB was searched to get the templates. Viral proteins were taken as templates and percent residue identity between the query sequence and templates were checked. Model was prepared using Amber-99(Weiner et al., 1984 and Weiner at al., 1986) as the forcefield and energy was minimized using conjugate gradient scheme with an rms gradient of 0.01Å achieved. The model prepared was evaluated and analyzed on the basis of the protein report which measures dihedral angles, bond angles, bond lengths and non-bonded heavy atom contacts in polypeptides. Outliers were considered to evaluate the model and finally, the model was refined using energy minimization.

After the structures were predicted the physico-chemical descriptors were calculated. The QikProp program (Schrödinger) (Duffy EM and Jorgensen WL, 2000) has been used to predict physically significant descriptors and pharmaceutically relevant properties. All the predicted structures were neutralized before being used by Qikprop. The neutralizing step is essential, as in normal mode it is unable to neutralize a structure and no properties will be generated. The program was processed in normal mode, predicting 44 properties for the 169 molecules which consisted of principal descriptors and the physiochemical properties with a detailed analysis of the log P (Octanol/Water), log Kp, log PMDK and log S.

RESULTS AND DISCUSSION

Pharmacokinetic studies deals with properties of a drug exhibited after it enters a body (Ryosei K et al., 1998). Earlier studies, to understand the mode of delivery of drugs indicates Po/w i.e. partition coefficient is an important parameter which can be used to analyze the skin permeability factor (Kp) (Degim T, 2005) alongwith molecular weight & volume for the subject under analysis (Potts RO, Guy RH, 1992). In this study a set of probable 168 CPPs have been generated taking TAT sequence as a template (See **Appendix Tables 1A, 1B, 1C** at end of article).

As there are no crystal coordinates for the following datasets along with TAT available, so it was necessary to predict there tertiary structure for the analysis of physiochemical properties (**Figure 1a, 1b, 1c, 1d**).

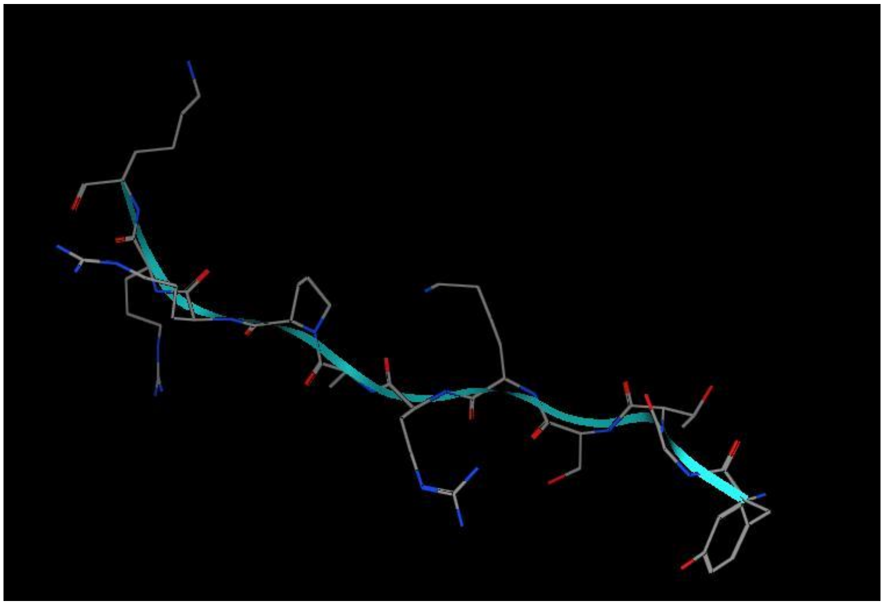

Figure 1a. Showing 3d structure of TAT-PTD, predicted using homology modeling

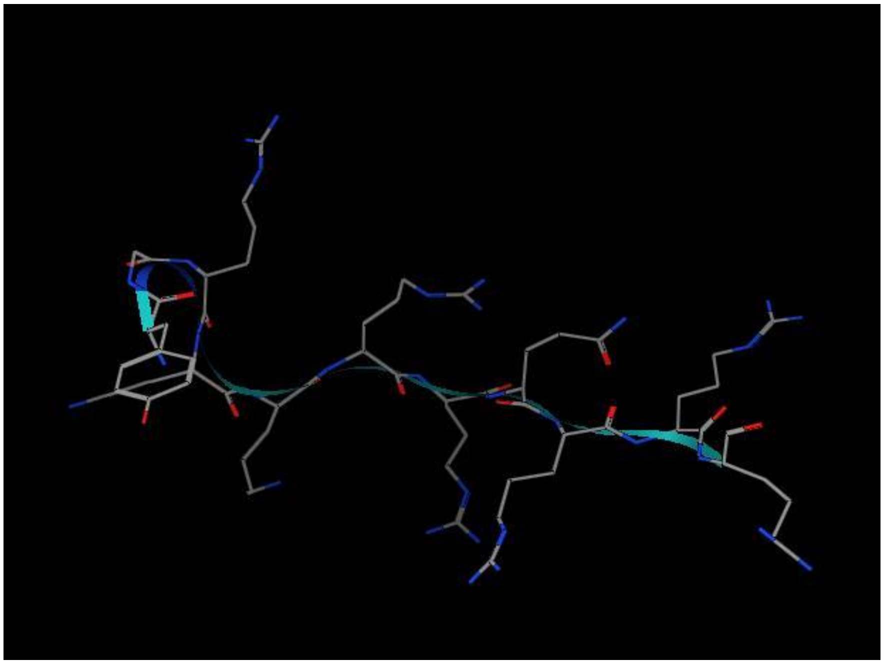

Figure 1b. Showing 3d structure of alanine mutated TAT-PTD, predicted using homology modeling

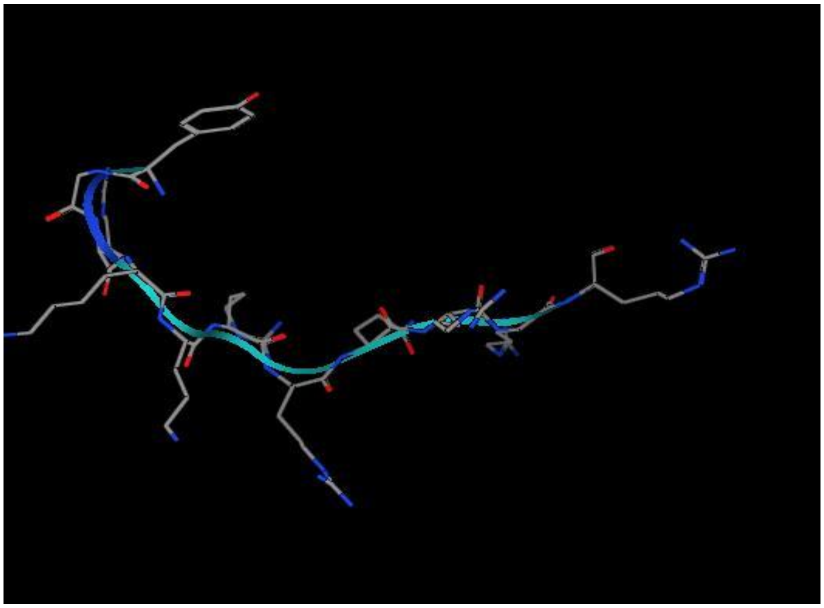

Figure 1c. Showing 3d structure of aspartic acid mutated TAT-PTD, predicted using homology modeling

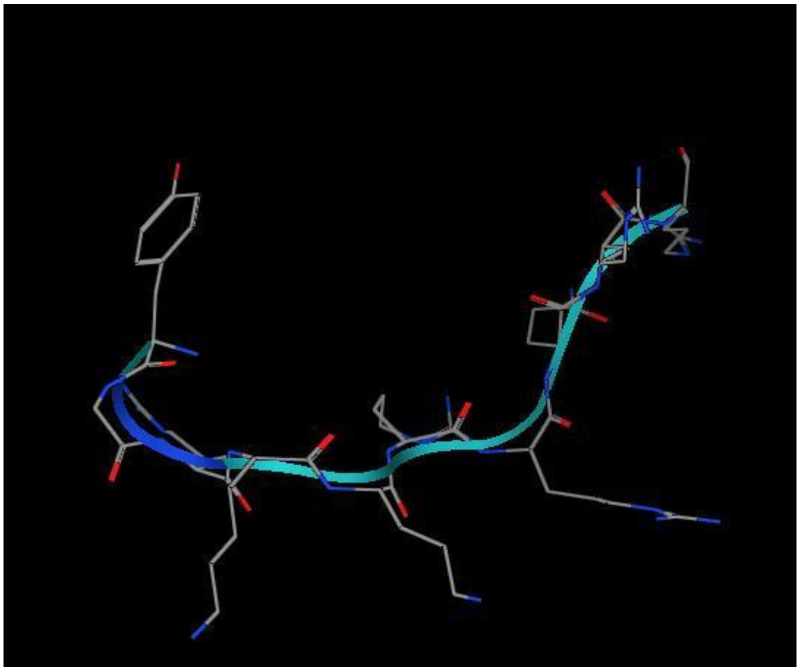

Figure 1d. Showing 3d structure of glutamic acid mutated TAT-PTD, predicted using homology modeling

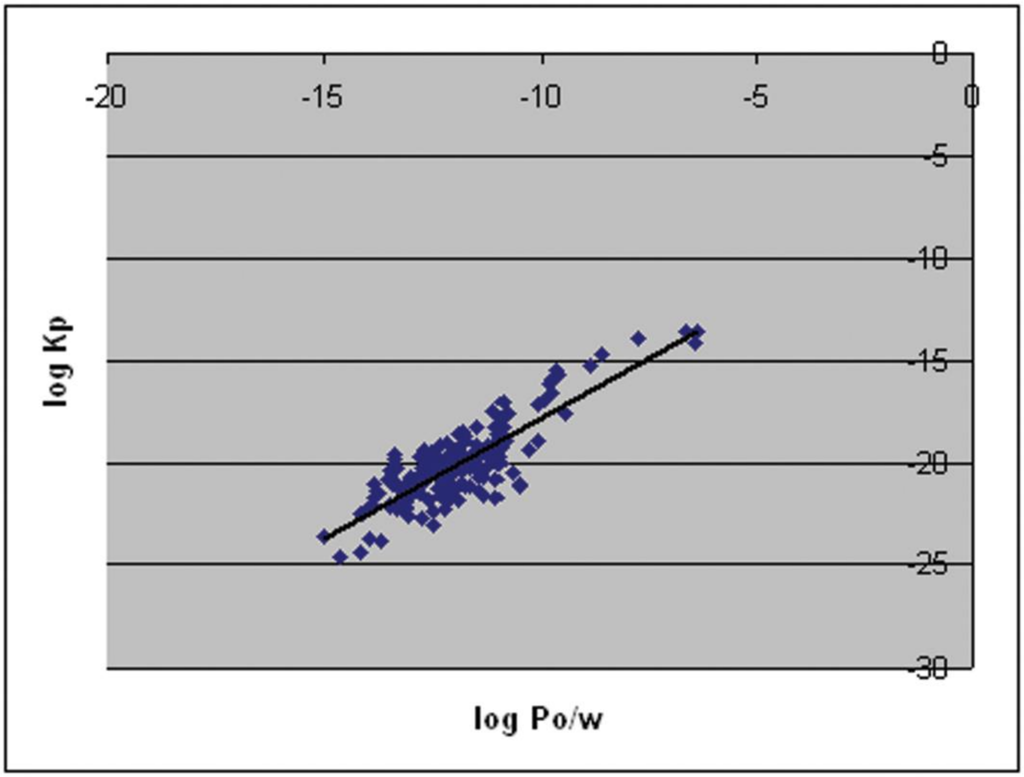

Figure 2. Linear regression plot between logKp and logPo/w of 168 mutated sequences used in the study.

In silico methodology has been used for the structure prediction of these sequences. The tertiary structure of the molecules have been generated using structure prediction module Prime (Schrödinger) and validated using procheck analysis. Further on, the important physiochemical descriptors were predicted using Qikprop program (Schrödinger) (Appendix Tables 2, 2A, 2B, 2C, 3, 3A, 3B, and 3C). Based on the properties observed for the Po/w & Kp for the 169 structures a QSAR model was built using regression analysis (Figure 2). Regression analysis between the two gave a correlation with R^2=73.7 % and helped us to the divide the dataset into various groups according to their degree of permeability (Table 4).

Following equation was observed:

$$\text{logKp} = -5.63 + 1.21\ \text{logPo/w}$$

where, Po/w is the partition coefficient and Kp is skin permeability factor.

On basis of the Kp value obtained and the plotting its value in the reference graph, the cell penetrating ability can be predicted.

Thus, the study helped us to get a set of potential CPPs from dataset of 169 structures. Using this equation, we have tried to analyze the permeability of the CPPs based on there relation with partition coefficient. Regression analysis between the two gave a correlation and helped us to the divide the dataset into various groups according to their degree of permeability. The whole analysis of the screened data obtained gave one strong observation that most of the glutamic acid and aspartic acid mutated sequences fall under category HIGH PERMEABILITY (Table 4A) or

MODERATE PERMEABILITY (Table 4B). Similarly, the original TAT-PTD having eight basic amino acid residues also fell under HIGH PERMEABILITY category. However, when basic amino acid residues were replaced by neutral amino acid i.e., alanine, resulting sequences came under LOW PERMEABILITY category (Table 4C). Based on the study it can be suggested that presence of charge is essential for transportation phenomenon using CPPs. As in the case of alanine mutated sequences the activity exhibited indicates the permeability decreases as the number of basic amino acids is being reduced (out of eight when more than two are mutated to neutral amino acid residues). Similarly, in the case of aspartate and glutamate mutations it was observed that mutation did not have much effect on the permeability ability as the total number of charged amino acid residues remained same in the sequences.

The inference can be supported based on the studies of Andre Ziegler and Joachim Seelig (2004). They suggest that for translocation, binding of the protein transduction domain (PTD) to extracellular glycosaminoglycans (GAGs) as a possible mechanism. The ubiquitous occurrence of glycosaminoglycans on the cell surface together with their tight and rapid interaction with the TAT protein transduction domain makes complex formation a strong candidate as the primary step of protein translocation. Now this complex formation can only occur when there is presence of some charge on the polypeptide sequence as GAGs themselves are charged. The enhancement of transduction by dextran sulfate suggests that electrostatic interactions play an important first step in the process by which PTDs and their cargoes traverse the plasma membrane (Mai et al., 2002).

CONCLUSIONS

This study is mainly focused on analyzing the penetration ability of TAT-PTD to transport the protein across the plasma membrane using in-silico tools. TAT-PTD is known to transport various proteins like Apactin (Tandon C and Lisle R, 2004), phosphopeptide analogs of HSP20 (Dreiza et al., 2004), cytoprotective proteins/peptides into pancreas (Klein D et al., 2005), delivery of purine nucleoside phosphorylase (Toro A and Grunebaum E, 2006) and also it is now being used in drug delivery system. The mystery arises when this TAT-PTD despite having eight basic amino acid residues is able to pass through the lipid bilayer membrane which is hydrophobic in nature. To unravel this various factors like thermodynamic properties, charge, temperature effect, and requirement of ATP etc can be taken into consideration. Based on earlier studies, we have tried to analyze the permeability of the CPPs based on there relation with partition coefficient. Regression analysis between the two gave a significant correlation and helped us to the divide the dataset into various groups according to their degree of permeability.

Our study suggests that the cell penetrating ability of a peptide can be predicted on basis of the Kp value obtained and the plotting its value in the reference graph. The oligonucleotide sequence for the peptide can then be obtained using any nucleotide to amino acid converter tool. Finally, expression vectors can be constructed to express the corresponding cell penetrating peptides fused with the protein of interest to be delivered into cells. The analysis carried out in this study can be used by biologists in future for predicting whether the given peptide sequence will act as a potent CPP or not. However, the mechanisms involved in the CPP-cargo translocation across the cells plasma membrane remain a matter of debate. Thus, whereas receptor-mediated process can be excluded, it is not clear to what extent some of the CPPs require energy to enter the cell (Trehin R and Merkle HP, 2004).

In summary, our study suggests that for translocation of such polypeptide sequences through the membrane, the sequences should have charged amino acid residues. The results demonstrate that presence of optimum number of charged amino acid residues in the sequence facilitates the permeability across the cell membrane.

References

Binder, H., and Lindblom, G. (2003). Charge-dependent translocation of the Trojan peptide penetratin across lipid membranes, Journal of Biophysics, 85, 982–995.

Chaloin, L., Vidal, P., Heitz, A., N. Van Mau, Mery J., Divita G., and Heitz, F. (1997). Conformations of primary amphipathic carrier peptides in membrane mimicking environments, Biochemistry, 36, 11179–11187.

Chaloin, L., Vidal, P.., Lory P., Mery J., Lautredou N., Divita G., and Heitz, F. (1998). Design of carrier peptide-oligonucleotide conjugates with rapid membrane translocation and nuclear localization properties, Biochemical Biophysical Research Community, 243, 601–608.

Dreiza, C.M., Brophy,C.M., Komalavilas,P., Furnish,E.J., Joshi,L., Pallero,M.A., Ullrich,J.E.M., Rechenberg, M.V., Seng, J. H., Richardson, B., Peltier,J.M., Panitch,A. (2004), Transducible heat shock protein 20 (HSP20) phosphopeptide alters cytoskeletal dynamics, The FASEB Journal express article 10.1096/fj.04-2911fje.

Duffy, E.M., Jorgensen, W.L. (2000). Prediction of Properties from Simulations : Free Energies of solvation in Hexa-decane, Octanol and Water, Journal of the American Chemical Society, 122, 2878-88.

Futaki, S., Suzuki, T., Ohashi, W., Yagami, T., Tanaka, S., Ueda, K. , Sugiura, Y. (2001). J Biol Chem (2001) **276**, 5836

Goldfarb, D. S., Gariepy, J., Schoolnik, G., and Kornberg, R. D. (1986). Synthetic peptides as nuclear localization signal's, Nature, 322, 641–644.

Green, M. and Loewenstein, P. M., (1988). Autonomous functional domains of chemically synthesized human immunodeficiency virus tat *trans*-activator protein, Cell, 55, 1179–1188.

Klein, D., Endoza, V., Pileggi, A., Molano, R.D. (2005). Delivery of TAT/PTD-Fused Proteins/Peptides to Islets via Pancreatic Duct, Cell Transplantation, 14, 241-248(8).

Lindgren, M., Hallbrink, M., Prochiantz, A., Langel, U., (2000). Cell-penetrating peptides, Trends in Pharmacological Science., 21, 99–103.

Mai, J.C., Shen,H., Watkins,S.C., Cheng,T. and Robbins, P.D. (2002). Efficiency of Protein Transduction Is Cell Type-dependent and Is Enhanced by Dextran Sulfate, Journal of Biological Chemistry, 277, 30208–30218.

Potts, R.O., Guy R.H., (1992). Predicting skin permeability, Pharmaceutical Research, 9(5), 663-9.

Prochiantz, A. (1996). Getting hydrophilic compounds into cells: lessons from homeopeptides, Current Opinion in Neurobiology, 6, 629–634.

Prochiantz, A. (2000). Messenger proteins: homeoproteins, TAT and others, Current Opinion in Cellular Biology. 12, 400–406.

Richard, J. P., Melikov, K., Vives, E., Ramos, C., Verbeure, B., Gait, M. J., Chernomordik, L.V., Lebleu, B. (2003). Cell-penetrating peptides. A re-evaluation of the mechanism of cellular uptake, Journal of Biological Chemistry, 278, 585–590.

Ryoseki, K., Diane, M., Chiaki, T., Malcom, R., (1998). Physiologically Based Pharmacokinetics of Cyclosporine A: Extension to Tissue Distribution Kinetics in Rats and Scale-up to Human, The Journal of Pharmacology and Experimental Therapeutics, 287, 457–468.

Schwarze, S., Ho, A., Vocero-Akbani, A., and Dowdy, S. F (1999). In vivo protein transduction: delivery of biologically active protein into the mouse. Science (Washington
DC), 285, 1569–1572.

Schwarze S. R., Hruska K. A., Dowdy S. F. (2000). Protein transduction: unrestricted delivery into all cells, Trends in Cell Biology, 10, 290.

Snyder E. L., Dowdy S. F. (2004). Cell penetrating peptides in drug delivery, Pharmaceutical Research, 21, 389.

Tandon C, De Lisle R. C. (2004). Apactin is involved in remodeling of the actin cytoskeleton during regulated exocytosis, European Journal of Cell Biology, 83, 79 – 89.

Thomas, P., Sylvie,B., Patrice, D., Pierre, M. and Christian, L. G. (2005). Coexistence of a Two-states Organization for a Cell Penetrating Peptide in Lipid Bilayer, Biophysical Journal, 10.1529/biophysj.105.061697

Thoren, P. E., Persson, D., Esbjorner, E. K., Goksor, M. P., Lincoln, Norden, B. (2004). Membrane binding and translocation of cell-penetrating peptides, Biochemistry. 43, 3471–3489.

Toro, A. and Grunebaum, E. (2006). TAT-mediated intracellular delivery of purine nucleoside phosphorylase corrects its deficiency in mice, Journal of Clinical Investigation, 116(10), 2717–2726.

Trehin, R., Merkle H. P. (2004). Chances and pitfalls of cell penetrating peptides for cellular drug delivery, European Journal of Pharmaceutics and Biopharmaceutics, 58,209–223.

Vidal, P., Morris, M. C., Chaloin, L., Heitz, F., Divita, G. (1997). New strategy for RNA vectorization in mammalian cells. Use of a peptide vector, Comptes Rendus de I Academie des Sciences. III, 320, 279–287.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Chio, C., Alagona, G., Profeta, S., Weiner, P. (1984). A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins, Journal of the American Chemical Society, 106, 765.

Weiner, S. J., Kollman, P. A., Case, D. A. (1986). All-Atom Force Field for Simulations of Proteins and Nucleic Acids, Journal of Computaional Chemistry, 7, 230.

Wender, P.A., Mitchell, D.J., Pattabiraman, K., Pelkey, E.T., Steinman, L., Rothbard, J.B. (2000). The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters. Proceedings of the National Academy of Sciences of the United States of America, 97,13003-8.

Ziegler A and Seelig J. (2004). Interaction of the Protein Transduction Domain of HIV-1 TAT with Heparan Sulfate: Binding Mechanism and Thermodynamic Parameters, Biophysical Journal, 86, 254–263.

MAIN

APPENDIX TABLES

TABLE 1: Dataset of sequences

A) Alanine

Seq-Id	Sequence	Seq-Id	Sequence
A1	YGAKKRRQRRR	A29	YGRKKARQARR
A2	YGRKKARQRRR	A30	YGRKKARQRAR
A3	YGRKKRAQRRR	A31	YGRKKARQRRA
A4	YGRKKRRQARR	A32	YGRKKRAQRAR
A5	YGRKKRRQRAR	A33	YGRKKRAQRRA
A6	YGRKKRRQRRA	A34	YGRKKRRQARA
A7	YGAKKARQRRR	A35	YGAKKARQARR
A8	YGRKKAAQRRR	A36	YGAKKARQRAR
A9	YGRKKRAQARR	A37	YGAKKARQRRA
A10	YGRKKRRQAAR	A38	YGRKKAAQRAR
A11	YGRKKRRQRAA	A39	YGRKKAAQRRA
A12	YGAKKAAQRRR	A40	YGRKKRAQARA
A13	YGRKKAAQARR	A41	YGAKKRAQARR
A14	YGRKKRAQAAR	A42	YGAKKRAQRAR
A15	YGRKKRRQAAA	A43	YGAKKRAQRRA
A16	YGAKKAAQARR	A44	YGAKKRRQAAR
A17	YGRKKAAQAAR	A45	YGAKKRRQRAA
A18	YGRKKRAQAAA	A46	YGRKKARQAAR
A19	YGAKKAAQAAR	A47	YGRKKARQRAA
A20	YGRKKAAQAAA	A48	YGRKKRAQRAA
A21	YGAKKAAQAAA	A49	YGAKKAAQRAR
A22	YGRAKRRQRRR	A50	YGAKKAAQRRA
A23	YGRKARRQRRR	A51	YGAKKRRQAAA
A24	YGRAARRQRRR	A52	YGAKKRAQAAR
A25	YGAKKRAQRRR	A53	YGAKKRAQRAA
A26	YGAKKRRQARR	A54	YGRKKARQAAA
A27	YGAKKRRQRAR	A55	YGAKKAAQARA
A28	YGAKKRRQRRA	A56	YGAKKAAQRAA

B) Glutamic Acid

Seq-Id	Sequence	Seq-Id	Sequence
E1	YGEKKRRQRRR	E29	YGRKKERQERR
E2	YGRKKERQRRR	E30	YGRKKERQRER
E3	YGRKKREQRRR	E31	YGRKKERQRRE
E4	YGRKKRRQERR	E32	YGRKKREQRER
E5	YGRKKRRQRER	E33	YGRKKREQRRE
E6	YGRKKRRQRRE	E34	YGRKKRRQERE
E7	YGEKKERQRRR	E35	YGEKKERQERR
E8	YGRKKEEQRRR	E36	YGEKKERQRER
E9	YGRKKREQERR	E37	YGEKKERQRRE
E10	YGRKKRRQEER	E38	YGRKKEEQRER
E11	YGRKKRRQREE	E39	YGRKKEEQRRE
E12	YGEKKEEQRRR	E40	YGRKKREQERE
E13	YGRKKEEQERR	E41	YGEKKREQERR
E14	YGRKKREQEER	E42	YGEKKREQRER
E15	YGRKKRRQEEE	E43	YGEKKREQRRE
E16	YGEKKEEQERR	E44	YGEKKRRQEER
E17	YGRKKEEQEER	E45	YGEKKRRQREE
E18	YGRKKREQEEE	E46	YGRKKERQEER
E19	YGEKKEEQEER	E47	YGRKKERQREE
E20	YGRKKEEQEEE	E48	YGRKKREQREE
E21	YGEKKEEQEEE	E49	YGEKKEEQRER
E22	YGREKRRQRRR	E50	YGEKKEEQRRE
E23	YGRKERRQRRR	E51	YGEKKRRQEEE
E24	YGREERRQRRR	E52	YGEKKREQEER
E25	YGEKKREQRRR	E53	YGEKKREQREE
E26	YGEKKRRQERR	E54	YGRKKERQEEE
E27	YGEKKRRQRER	E55	YGEKKEEQERE
E28	YGEKKRRQRRE	E56	YGEKKEEQREE

C) Aspartic Acid

Seq-Id	Sequence	Seq-Id	Sequence
D1	YGDKKRRQRRR	D29	YGRKKDRQDRR
D2	YGRKKDRQRRR	D30	YGRKKDRQRDR
D3	YGRKKRDQRRR	D31	YGRKKDRQRRD
D4	YGRKKRRQDRR	D32	YGRKKRDQRDR
D5	YGRKKRRQRDR	D33	YGRKKRDQRRD
D6	YGRKKRRQRRD	D34	YGRKKRRQDRD
D7	YGDKKDRQRRR	D35	YGDKKDRQDRR
D8	YGRKKDDQRRR	D36	YGDKKDRQRDR

D9	YGRKKRDQDRR	D37	YGDKKDRQRRD
D10	YGRKKRRQDDR	D38	YGRKKDDQRDR
D11	YGRKKRRQRDD	D39	YGRKKDDQRRD
D12	YGDKKDDQRRR	D40	YGRKKRDQDRD
D13	YGRKKDDQDRR	D41	YGDKKRDQDRR
D14	YGRKKRDQDDR	D42	YGDKKRDQRDR
D15	YGRKKRRQDDD	D43	YGDKKRDQRRD
D16	YGDKKDDQDRR	D44	YGDKKRRQDDR
D17	YGRKKDDQDDR	D45	YGDKKRRQRDD
D18	YGRKKRDQDDD	D46	YGRKKDRQDDR
D19	YGDKKDDQDDR	D47	YGRKKDRQRDD
D20	YGRKKDDQDDD	D48	YGRKKRDQRDD
D21	YGDKKDDQDDD	D49	YGDKKDDQRDR
D22	YGRDKRRQRRR	D50	YGDKKDDQRRD
D23	YGRKDRRQRRR	D51	YGDKKRRQDDD
D24	YGRDDRRQRRR	D52	YGDKKRDQDDR
D25	YGDKKRDQRRR	D53	YGDKKRDQRDD
D26	YGDKKRRQDRR	D54	YGRKKDRQDDD
D27	YGDKKRRQRDR	D55	YGDKKDDQDRD
D28	YGDKKRRQRRD	D56	YGDKKDDQRDD

TABLE 2: Physiochemical descriptors calculated for TAT-PTD with help of Qikprop simulation.

Properties of Original TAT-PTD peptide

polrz	logPC16	logPoct	logPw	logPo/w	logKp	logKhsa
131.426	61.807	123.032	107.551	-15.016	-23.588	-7.177

A) Original sequence mutated by Alanine

Seq-Id	polrz	logPC16	logPoct	logPw	logKhsa	SASA
A1	130.43	61.463	114.108	98.914	-6.616	2225.66
A2	131.912	61	114.857	100.244	-6.475	2206.58
A3	132.58	62.275	114.473	101.058	-6.529	2264.25
A4	126.502	58.089	113.297	96.498	-6.634	2085.61
A5	129.372	59.551	113.387	99.043	-6.57	2159.06
A6	124.729	56.735	112.439	97.026	-6.659	2035.92
A7	123.772	56.655	104.212	93.742	-6.105	2104.89
A8	128.945	59.315	106.851	94.419	-5.886	2197.5
A9	127.322	58.842	105.774	93.261	-5.985	2185.66
A10	123.977	56.114	104.482	89.399	-6.049	2081.69
A11	120.447	53.768	104.549	89.69	-6.086	1967.15
A12	123.468	57.115	97.437	89.784	-5.451	2174.21
A13	121.997	55.515	97.294	86.759	-5.458	2107.49
A14	120.78	54.054	96.843	86.99	-5.459	2051.54
A15	119.363	52.859	96.121	83.65	-5.49	2008.42
A16	112.562	49.241	85.681	77.466	-5.113	1940.59
A17	146.585	66.441	116.79	103.501	-6.942	2470.21
A18	110.169	46.597	85.517	74.509	-5.095	1825.09
A19	88.833	35.888	64.214	56.018	-3.601	1508.56
A20	92.704	38.497	66.307	58.832	-3.477	1604.2
A21	93.964	38.981	64.621	58.305	-3.344	1632.35

A22	92.501	38.195	63.674	57.545	-3.417	1607.62
A23	105.437	43.706	76.755	68.647	-4.626	1772.11
A24	100.762	40.695	67.628	62.017	-4.177	1720.91
A25	133.287	61.838	118.122	102.943	-6.805	2222.23
A26	136.15	64.614	120.058	106.044	-6.763	2313.35
A27	114.634	52.088	100.305	91.095	-5.825	1917.12
A28	113.835	51.434	99.791	88.585	-5.838	1894.96
A29	123.946	56.097	103.746	88.552	-6.071	2093.01
A30	123.946	56.097	103.746	88.552	-6.071	2093.01
A31	120.911	53.698	103.414	88.028	-6.091	1986.74
A32	127.148	58.633	106.089	91.947	-5.974	2172.19
A33	125.343	56.922	104.744	91.277	-6.013	2117.21
A34	125.387	57.894	105.055	92.268	-6.064	2148.62
A35	123.211	55.281	103.651	90.095	-6.062	2058.56
A36	123.465	55.449	104.821	93.467	-6.024	2047.6
A37	122.877	55.388	104.42	90.261	-6.066	2047.7
A38	119.86	53.545	96	86.27	-5.506	2039.75
A39	121.623	54.719	96.918	86.73	-5.445	2081.06
A40	118.218	51.991	95.584	87.061	-5.518	1976.33
A41	120.391	53.661	97.139	86.399	-5.45	2029.24
A42	119.626	53.352	96.937	87.088	-5.483	2015.56
A43	116.86	51.204	94.594	82.38	-5.576	1953.18
A44	121.454	54.578	96.549	85.642	-5.456	2080.8
A45	118.779	53.062	96.089	85.424	-5.539	2012.57
A46	114.239	49.397	93.958	83.055	-5.622	1875.84
A47	117.275	51.931	94.572	81.561	-5.6	1984.38
A48	116.988	51.826	94.936	81.275	-5.599	1972.64
A49	121.796	55.019	97.029	85.539	-5.449	2091.63
A50	119.85	53.709	96.504	85.468	-5.503	2037.85
A51	118.435	52.285	95.837	84.63	-5.517	1984.86
A52	110.343	47.095	84.934	76.157	-5.135	1855.29
A53	110.675	47.038	84.875	78.504	-5.113	1856.59
A54	109.319	46.128	84.629	72.802	-5.14	1815.87
A55	112.609	48.701	85.726	75.783	-5.076	1919.82
A56	111.382	47.805	86.46	76.404	-5.075	1864.4

B) Original Sequence mutated by Glutamic acid

Seq-Id	**polrz**	**logPC16**	**logPoct**	**logPw**	**logKhsa**	**SASA**
E1	125.37	57.019	116.393	101.723	-7.107	2024.94

E2	126.384	57.572	117.365	101.14	-7.063	2033.77
E3	125.981	57.648	116.667	102.557	-7.119	2044.27
E4	126.53	58.045	116.39	100.085	-7.132	2067.75
E5	121.243	53.504	114.884	99.835	-7.056	1879.79
E6	132.194	62.153	118.887	107.189	-7.035	2223.47
E7	124.235	56.409	111.965	100.006	-7.046	2032.13
E8	125.527	56.749	112.44	97.285	-6.983	2047.49
E9	123.627	55.057	111.085	94.716	-6.997	1991.29
E10	121.705	54.449	110.633	97.585	-7.067	1965.95
E11	123.02	54.89	110.881	98.115	-7.024	1983.45
E12	117.073	50.743	104.079	89.188	-7.021	1873.42
E13	119.569	53.511	105.574	90.278	-7.076	1964.59
E14	117.124	51.136	104.718	88.939	-7.03	1873.57
E15	120.396	53.448	107.55	96.002	-6.948	1937.8
E16	121.48	54.14	103.482	93.572	-6.882	2010
E17	115.41	49.721	100.35	86.824	-6.969	1848.91
E18	107.803	45.36	98.346	86.086	-6.998	1648.32
E19	114.558	49.817	96.087	83.184	-7.01	1884.96
E20	115.139	50.159	97.589	84.939	-6.951	1876.9
E21	107.463	44.989	89.458	77.424	-7.087	1728.76
E22	127.851	58.826	120.869	107.565	-7.397	2061.2
E23	127.584	56.805	120.276	104.107	-7.246	1996.63
E24	127.443	57.631	120.171	104.851	-7.529	2021.82
E25	123.341	55.105	110.84	95.761	-7.025	1999.13
E26	125.475	57.347	112.711	95.521	-7.021	2071.4
E27	127.116	58.235	113.501	97.627	-6.973	2097.22
E28	120.291	52.941	110.284	97.594	-7.018	1900.37
E29	124.035	55.632	111.315	96.964	-7.015	2011.79
E30	123.761	55.24	111.4	96.099	-6.993	1994.6
E31	118.875	52.426	109.778	97.619	-7.065	1876.73
E32	123.121	55.342	111.627	99.233	-7.03	1988.81
E33	121.39	53.335	110.187	94.117	-7	1923.94
E34	124.595	56.397	112.917	99.217	-6.988	2023.22
E35	123.174	55.439	108.069	96.3	-6.936	2028.59
E36	117.145	51.529	104.553	93.479	-7.073	1889.2
E37	121.353	53.406	106.324	95.379	-6.937	1962.22
E38	117.729	51.965	105.298	92.044	-7.052	1898.11
E39	118.741	52.794	106.346	92.898	-7.031	1917.25
E40	122.025	54.117	106.971	93.78	-6.937	1981.47

E41	120.863	53.117	106.304	93.625	-6.942	1945.6
E42	124.542	56.53	108.636	94.479	-6.927	2063.8
E43	120.917	53.359	106.44	93.366	-6.954	1953.9
E44	118.186	51.885	106.305	92.649	-6.976	1887.19
E45	119.608	52.967	105.571	92.942	-7.027	1943.19
E46	123.727	55.821	107.641	92.412	-6.952	2048.14
E47	124.37	56.454	108.313	94.48	-6.942	2064.54
E48	117.647	52.531	105.866	92.055	-7.086	1910.02
E49	118.4	52.686	101.348	89.41	-7.022	1963.87
E50	117.256	51.355	100.115	88.799	-7.018	1932.02
E51	118.661	51.845	101.936	89.785	-6.914	1928.15
E52	119.924	53.142	103.345	90.358	-6.9	1959.71
E53	117.889	51.416	101.625	87.496	-6.935	1906.57
E54	119.183	51.949	102.311	87.669	-6.881	1925.63
E55	114.643	49.154	96.302	86.548	-6.942	1854.49
E56	119.343	53.46	98.001	85.632	-6.943	2028.44

C) Original Sequence mutated by Aspartic acid

Seq-Id	**polrz**	**logPC16**	**logPoct**	**logPw**	**logKhsa**	**SASA**
D1	115.462	52.638	103.009	88.562	-7.107	1919.17
D2	123.516	56.164	115.135	102.744	-7.063	1990.91
D3	123.794	55.245	115.461	97.866	-7.119	1959.27
D4	122.81	55.098	114.074	98.452	-7.132	1963.34
D5	123.241	54.965	115.171	99.976	-7.056	1942.12
D6	125.837	57.974	116.49	102.525	-7.035	2054.08
D7	121.816	55.109	109.56	97.375	-7.046	1995.88
D8	122.674	55.015	110.219	96.456	-6.983	1993.46
D9	120.894	55.14	109.205	94.473	-6.997	1999.46
D10	118.593	52.147	108.903	96.135	-7.067	1871.67
D11	121.026	54.182	110.574	95.382	-7.024	1938.2
D12	118.481	52.741	104.308	91.998	-7.021	1933.99
D13	117.365	52.529	104.088	91.038	-7.076	1924.54
D14	112.65	48.45	101.888	87.632	-7.03	1769.56
D15	115.515	50.482	103.474	91.025	-6.948	1839.23
D16	116.032	51.551	99.03	87.519	-6.882	1924.28
D17	114.643	50.966	99.171	88.447	-6.969	1890.18
D18	109.444	47.295	96.866	85.166	-6.998	1748.92
D19	108.304	47.63	91.728	79.518	-7.01	1806.86
D20	107.387	46.685	88.106	79.392	-6.951	1794.26

D21	131.577	62.487	121.973	108.463	-7.087	2202.16
D22	127.623	58.73	120.079	104.732	-7.397	2065.29
D23	127.365	59.826	119.462	107.861	-7.246	2107.29
D24	123.375	55.587	110.004	94.854	-7.529	2013.56
D25	120.747	53.741	109.033	94.793	-7.025	1944.33
D26	122.428	55.364	110.56	96.534	-7.021	1995.07
D27	121.995	55.797	109.681	95.655	-6.973	2027.2
D28	123.457	56.665	110.913	98.164	-7.018	2044.88
D29	121.627	55.007	110.602	97.579	-7.015	1975.32
D30	120.682	54.624	109.885	98.76	-6.993	1969.08
D31	117.964	52.284	108.334	94.021	-7.065	1882.58
D32	123.872	55.912	109.858	92.92	-7.03	2032.14
D33	118.598	53.577	109.419	98.395	-7	1916.68
D34	123.995	57.388	106.531	94.985	-6.988	2115.11
D35	117.914	52.677	103.789	92.249	-6.936	1939.44
D36	118.405	53.285	104.481	91.485	-7.073	1955.23
D37	114.736	51.024	103.169	89.102	-6.937	1861.15
D38	116.79	51.436	104.011	91.463	-7.052	1876.82
D39	116.352	51.527	103.452	89.422	-7.031	1889.14
D40	121.385	55.369	105.853	93.478	-6.937	2034.54
D41	120.042	54.364	105.092	94.293	-6.942	1994.2
D42	114.466	49.681	102.764	88.951	-6.927	1817.78
D43	119.922	54.754	105.932	94.458	-6.954	1995.99
D44	118.837	53.103	105.161	92.689	-6.976	1936.28
D45	112.316	49.332	101.55	90.418	-7.027	1809.18
D46	118.249	53.317	105.03	93.598	-6.952	1946.85
D47	118.957	52.629	105.517	93.019	-6.942	1913.42
D48	111.948	48.951	97.49	84.98	-7.086	1821.24
D49	115.872	51.997	98.877	90.285	-7.022	1946.14
D50	112.128	49.186	98.073	83.93	-7.018	1823.47
D51	112.338	48.978	97.96	84.615	-6.914	1826.74
D52	115.043	51.27	99.853	88.748	-6.9	1895.49
D53	113.308	50.523	98.565	87.863	-6.935	1877.67
D54	113.582	50.548	94.454	86.191	-6.881	1912.3
D55	111.14	49.117	93.293	84.984	-6.942	1857.17
D56	104.938	43.661	77.47	71.797	-4.635	1754.22

TABLE 3: Physical descriptors calculated for TAT-PTD with help of Qikprop simulation.

Properties of Original TAT-PTD peptide

polrz	Mol wt.	Mol vol	logPo/w	logKp
131.426	1560.81	123.032	-15.016	-23.588

A) Original sequence mutated by Alanine

Seq-Id	MW	Mol Vol	logPo/w	logKp
A1	1474.73	4388.58	-13.48	-22.146
A2	1474.73	4425.17	-13.048	-21.724
A3	1474.73	4441.86	-13.126	-21.944
A4	1474.73	4290.42	-13.44	-21.06
A5	1474.73	4362.14	-13.004	-20.753
A6	1474.73	4246.11	-13.324	-20.262
A7	1389.63	4141.56	-12.153	-19.883
A8	1389.63	4271.06	-12.038	-21.4
A9	1389.63	4230.27	-12.118	-20.957
A10	1389.63	4146.66	-11.984	-19.635
A11	1389.63	4058.68	-12.385	-19.769
A12	1304.52	4053.72	-11.291	-20.755
A13	1304.52	4017.22	-11.245	-20.27
A14	1304.52	3986.56	-11.056	-19.473
A15	1304.52	3951.15	-10.936	-18.733
A16	1219.41	3700.99	-9.927	-16.924
A17	1588.83	4808.31	-13.501	-22.049
A18	1219.41	3641.41	-9.804	-16.082
A19	978.113	2929.63	-6.437	-11.847
A20	978.113	3025.48	-6.624	-13.587
A21	978.113	3055.92	-6.395	-14.13
A22	978.113	3019.68	-6.363	-13.506
A23	1134.3	3443.2	-8.855	-15.176
A24	1049.19	3245.7	-7.766	-13.847
A25	1502.75	4459.75	-13.187	-21.767
A26	1502.75	4531.29	-13.681	-23.876
A27	1289.47	3800.91	-11.454	-20.085
A28	1289.47	3780.94	-11.351	-19.586

A29	1389.63	4146.11	-11.783	-19.034
A30	1389.63	4146.11	-11.783	-19.034
A31	1389.63	4070.51	-11.896	-18.566
A32	1389.63	4226.34	-12.203	-21.177
A33	1389.63	4180.8	-11.868	-19.747
A34	1389.63	4181.9	-12.256	-20.651
A35	1389.63	4127.54	-11.73	-18.739
A36	1389.63	4133.87	-12.03	-19.704
A37	1389.63	4119.17	-12.092	-19.614
A38	1304.52	3963.58	-10.959	-18.868
A39	1304.52	4007.63	-11.008	-19.585
A40	1304.52	3922.55	-10.88	-18.237
A41	1304.52	3976.83	-11.143	-19.636
A42	1304.52	3957.73	-11.24	-19.625
A43	1304.52	3888.6	-10.792	-17.526
A44	1304.52	4003.62	-10.915	-19.269
A45	1304.52	3936.58	-11.187	-19.131
A46	1304.52	3823.14	-10.871	-16.99
A47	1304.52	3899.24	-10.891	-17.824
A48	1304.52	3892.27	-11.056	-18.198
A49	1304.52	4011.95	-11.078	-19.799
A50	1304.52	3963.56	-11.16	-19.412
A51	1304.52	3927.98	-10.976	-18.547
A52	1219.41	3645.55	-9.751	-15.871
A53	1219.41	3654.06	-9.606	-15.627
A54	1219.41	3620.19	-9.681	-15.43
A55	1219.41	3702.38	-9.75	-16.567
A56	1219.41	3671.5	-10.069	-17.125

B) Original sequence mutated by Glutamic acid

Seq-Id	**MW**	**Mol Vol**	**logPo/w**	**logKp**
E1	1532.77	4322.79	-13.776	-21.434
E2	1532.77	4344.94	-13.878	-22.031
E3	1532.77	4333.92	-13.841	-21.672
E4	1532.77	4347.6	-13.723	-21.453

E5	1532.77	4215.58	-13.395	-19.534
E6	1532.77	4491.58	-13.954	-23.7
E7	1505.7	4258.27	-12.938	-21.408
E8	1505.7	4291.55	-12.769	-21.447
E9	1505.7	4243.63	-12.426	-20.041
E10	1505.7	4199.45	-12.747	-20.293
E11	1505.7	4228.01	-12.507	-20.033
E12	1478.63	4056.12	-11.454	-18.285
E13	1478.63	4110.25	-11.973	-20.015
E14	1478.63	4053.02	-11.717	-18.922
E15	1478.63	4133.44	-12.233	-21.279
E16	1451.56	4129.07	-11.286	-21.556
E17	1451.56	3977.8	-10.871	-18.78
E18	1451.56	3784.15	-11.088	-17.441
E19	1424.48	3921.42	-10.317	-19.312
E20	1424.48	3935.81	-10.652	-20.502
E21	1397.41	3714.47	-9.461	-17.53
E22	1560.78	4380.84	-14.174	-22.491
E23	1560.78	4378.64	-13.41	-20.853
E24	1561.73	4338.42	-13.31	-22.332
E25	1505.7	4240.14	-12.499	-20.133
E26	1505.7	4295.93	-13.109	-22.301
E27	1505.7	4330.99	-13.046	-22.584
E28	1505.7	4161.6	-12.545	-19.526
E29	1505.7	4253.27	-12.554	-20.421
E30	1505.7	4247.47	-12.537	-20.382
E31	1505.7	4124.51	-12.666	-19.365
E32	1505.7	4230.22	-12.834	-20.909
E33	1505.7	4187.88	-12.296	-19.16
E34	1505.7	4271.58	-13.108	-22.154
E35	1478.63	4205.97	-12.184	-21.874
E36	1478.63	4051.1	-11.79	-18.978
E37	1478.63	4156.24	-11.564	-19.741
E38	1478.63	4065.93	-11.96	-19.629
E39	1478.63	4088.75	-12.191	-20.513
E40	1478.63	4171.09	-11.756	-20.39
E41	1478.63	4141.88	-11.6	-19.686
E42	1478.63	4234.12	-12.218	-22.246
E43	1478.63	4143.98	-11.712	-19.97

E44	1478.63	4082.85	-12.093	-20.357
E45	1478.63	4110.52	-11.761	-19.584
E46	1478.63	4214.33	-11.974	-21.34
E47	1478.63	4229.39	-12.156	-21.993
E48	1478.63	4063.88	-12.344	-20.54
E49	1451.56	4048.22	-11.185	-20.135
E50	1451.56	4026.06	-10.808	-18.941
E51	1451.56	4059.06	-10.997	-20.037
E52	1451.56	4087.01	-11.394	-21.389
E53	1451.56	4035.47	-10.981	-19.706
E54	1451.56	4068.47	-10.936	-20.047
E55	1424.48	3922.23	-10.099	-18.924
E56	1424.48	4042.83	-10.526	-21.206

C) Original sequence mutated by Aspartic acid

Seq-Id	**MW**	**volume**	**logPo/w**	**logKp**
D1	1362.56	3933.92	-11.991	-19.56
D2	1518.74	4256.23	-13.822	-21.098
D3	1518.74	4269.88	-13.539	-20.798
D4	1518.74	4239.37	-13.389	-19.833
D5	1518.74	4249.47	-13.487	-20.409
D6	1518.74	4314.22	-14.069	-22.367
D7	1477.65	4165.28	-12.851	-20.756
D8	1477.65	4192.28	-12.758	-21.003
D9	1477.65	4145.06	-13.089	-21.024
D10	1477.65	4087.1	-12.674	-19.734
D11	1477.65	4146	-13.057	-21.33
D12	1436.55	4033.77	-12.058	-20.556
D13	1436.55	4008.59	-12.311	-20.853
D14	1436.55	3892.48	-11.77	-18.427
D15	1436.55	3960.05	-11.989	-19.694

D16	1395.45	3924.42	-11.405	-20.744
D17	1395.45	3890.3	-11.7	-21.136
D18	1395.45	3761.44	-11.45	-19.133
D19	1354.35	3684.23	-11.014	-19.915
D20	1313.25	3617.07	-10.486	-21.033
D21	1546.76	4457.66	-14.624	-24.674
D22	1546.76	4359.82	-14.122	-22.451
D23	1533.67	4304.44	-14.151	-24.409
D24	1477.65	4200.28	-12.597	-20.575
D25	1477.65	4138.44	-12.571	-19.866
D26	1477.65	4182.11	-13.023	-21.516
D27	1477.65	4175.15	-13.106	-21.43
D28	1477.65	4206.73	-13.226	-22.187
D29	1477.65	4162.37	-13.174	-21.694
D30	1477.65	4141.75	-13.177	-21.389
D31	1477.65	4071.65	-12.783	-19.682
D32	1477.65	4212.98	-12.508	-20.451
D33	1477.65	4085.14	-13.321	-21.116
D34	1436.55	4174.04	-12.455	-23.053
D35	1436.55	4022.61	-12.105	-20.458
D36	1436.55	4035.29	-12.349	-21.235
D37	1436.55	3941.38	-12.371	-20.242
D38	1436.55	3992.03	-12.067	-20.211
D39	1436.55	3983.28	-12.112	-20.104
D40	1436.55	4111.21	-12.473	-22.436
D41	1436.55	4072.72	-12.326	-21.592
D42	1436.55	3938.89	-11.886	-19.205
D43	1436.55	4071.25	-12.743	-22.669
D44	1436.55	4042.7	-12.301	-21.344
D45	1436.55	3884.86	-12.136	-18.99
D46	1436.55	4031.94	-12.565	-21.803
D47	1436.55	4046.29	-12.193	-21.259
D48	1395.45	3823.41	-11.411	-19.626
D49	1395.45	3923.46	-11.594	-21.113
D50	1395.45	3829.38	-11.605	-20.234
D51	1395.45	3841.94	-11.488	-20.134
D52	1395.45	3902.27	-11.876	-21.785
D53	1395.45	3860.52	-11.846	-21.086
D54	1354.35	3816.95	-11.044	-21.709

D55	1354.35	3754	-11.022	-20.856
D56	1134.3	3430.27	-9.204	-15.956

TABLE 4: Screening of TAT derived CPP's

High permeability

Seq-Id	Sequnce	Seq-Id	Sequence
D21	YGDKKDDQDDD	D36	YGDKKDRQRDR
D23	YGRKDRRQRRR	A32	YGRKKRAQRAR
A26	YGAKKRRQARR	D33	YGRKKRDQRRD
E6	YGRKKRRQRRE	D2	YGRKKDRQRRR
D34	YGRKKRRQDRD	A4	YGRKKRRQARR
D43	YGDKKRDQRRD	D9	YGRKKRDQDRR
E27	YGEKKRRQRER	D8	YGRKKDDQRRR
E22	YGREKRRQRRR	A9	YGRKKRAQARR
D22	YGRDKRRQRRR	E32	YGRKKREQRER
D40	YGRKKRDQDRD	E23	YGRKERRQRRR
D6	YGRKKRRQRRD	D13	YGRKKDDQDRR
E24	YGREERRQRRR	D3	YGRKKRDQRRR
E26	YGEKKRRQERR	D7	YGDKKDRQRRR
E42	YGEKKREQRER	A5	YGRKKRRQRAR
D28	YGDKKRRQRRD	A34	YGRKKRRQARA
E34	YGRKKRRQERE	D24	YGRDDRRQRRR
A1	YGAKKRRQRRR	D12	YGDKKDDQRRR
A17	YGRKKAAQAAR	E48	YGRKKREQREE
E2	YGRKKERQRRR	E39	YGRKKDDQRRD
E47	YGRKKERQREE	D35	YGDKKDRQDRR
A3	YGRKKRAQRRR	D32	YGRKKRDQRDR
E35	YGEKKERQERR	E29	YGRKKERQERR
D46	YGRKKDRQDDR	D5	YGRKKRRQRDR
D52	YGDKKRDQDDR	E30	YGRKKERQRER
A25	YGAKKRAQRRR	E44	YGEKKRRQEER
A2	YGRKKARQRRR	E10	YGRKKRRQEER
D29	YGRKKDRQDRR	A6	YGRKKRRQRRA
E3	YGRKKREQRRR	D37	YGDKKDRQRRD
D41	YGDKKRDQDRR	D38	YGRKKDDQRDR
D26	YGDKKRRQDRR	E25	YGEKKREQRRR
E4	YGRKKRRQERR	D39	YGRKKDDQRRD

E8	YGRKKEEQRRR	E9	YGRKKREQERR
E1	YGEKKRRQRRR	E11	YGRKKRRQREE
D27	YGDKKRRQRDR	E13	YGRKKEEQERR
E7	YGEKKERQRRR	A7	YGAKKARQRRR
A8	YGRKKAAQRRR	D25	YGDKKRDQRRR
D30	YGRKKDRQRDR	D4	YGRKKRRQDRR
D44	YGDKKRRQDDR	A11	YGRKKRRQRAA
E46	YGRKKERQEER	A33	YGRKKRAQRRA
D11	YGRKKRRQRDD	D10	YGRKKRRQDDR
E15	YGRKKRRQEEE	D47	YGRKKDRQRDD
A36	YGAKKARQRAR	E5	YGRKKRRQRER
D15	YGRKKRRQDDD	E28	YGEKKRRQRRE
D31	YGRKKDRQRRD	E31	YGRKKERQRRE
A10	YGRKKRRQAAR	D42	YGDKKRDQRDR
E38	YGRKKEEQRER	E33	YGRKKREQRRE
A37	YGAKKARQRRA	D45	YGDKKRRQRDD
D1	YGDKKRRQRRR	A31	YGRKKARQRRA

Medium permeability

Seq-Id	**Sequence**
D54	YGRKKDRQDDD
E16	YGEKKEEQERR
E52	YGEKKREQEER
E56	YGEKKEEQREE
D17	YGRKKDDQDDR
D49	YGDKKDDQRDR
D53	YGDKKRDQRDD
D20	YGRKKDDQDDD
D55	YGDKKDDQDRD
A12	YGAKKAAQRRR
D16	YGDKKDDQDRR
E20	YGRKKEEQEEE
E40	YGRKKREQERE
A13	YGRKKAAQARR
D50	YGDKKDDQRDR
E49	YGEKKEEQRER
D51	YGDKKRRQDDD
A27	YGAKKRRQRAR
E54	YGRKKERQEEE
E51	YGEKKRRQEEE

Low permeability

Seq-ID	**Sequence**	**Seq-Id**	**Sequence**
E43	YGEKKREQRRE	A35	YGAKKARQARR
D19	YGDKKDDQDDR	A15	YGRKKRRQAAA
A49	YGRKKARQRAA	A51	YGAKKRRQAAA
E37	YGEKKERQRRE	D14	YGRKKRDQDDR
E53	YGEKKREQREE	E12	YGEKKEEQRRR
E41	YGEKKREQERR	A40	YGRKKRAQARA
A41	YGAKKRAQARR	A48	YGRKKRAQRAA
D48	YGRKKRDQRDD	A47	YGRKKARQRAA
A42	YGAKKRAQRAR	E21	YGEKKEEQEEE
A28	YGAKKRRQRRA	A43	YGAKKRAQRRA
A39	YGRKKAAQRRA	E18	YGRKKREQEEE
E45	YGEKKRRQREE	A56	YGAKKAAQRAA
A14	YGRKKRAQAAR	A46	YGRKKARQAAR
A50	YGAKKAAQRRA	A16	YGAKKAAQARR
E19	YGEKKEEQEER	A55	YGAKKAAQARA
A44	YGAKKRRQAAR	A18	YGRKKRAQAAA
D18	YGRKKRDQDDD	A52	YGAKKRAQAAR
A45	YGAKKRRQRAA	A53	YGAKKRAQRAA
A29	YGRKKARQARR	A54	YGRKKARQAAA
A30	YGRKKARQRAR	A23	YGRKARRQRRR
E36	YGEKKERQRER	D56	YGDKKDDQRDD
E50	YGEKKEEQRRE	A21	YGAKKAAQAAA
E55	YGEKKEEQERE	A24	YGRAARRQRRR
E14	YGRKKREQEER	A20	YGRKKAAQAAA
A38	YGRKKAAQRAR	A22	YGRAKRRQRRR
E17	YGRKKEEQEER	A19	YGAKKAAQAAR