Predicting the Nuclear Localization Signals of 107 Types of HPV L1 Proteins by Bioinformatic Analysis

Jun Yang[1,2], Yi-Li Wang[1], and Lü-Sheng Si[1]*

[1] *The Key Laboratory of the Ministry of Education for Biomedical Information Engineering, Institute for Cancer Research, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710061, China;* [2] *Department of Pathology, Second Hospital of Medical College, Xi'an Jiaotong University, Xi'an 710004, China.*

In this study, 107 types of human papillomavirus (HPV) L1 protein sequences were obtained from available databases, and the nuclear localization signals (NLSs) of these HPV L1 proteins were analyzed and predicted by bioinformatic analysis. Out of the 107 types, the NLSs of 39 types were predicted by PredictNLS software (35 types of bipartite NLSs and 4 types of monopartite NLSs). The NLSs of the remaining HPV types were predicted according to the characteristics and the homology of the already predicted NLSs as well as the general rule of NLSs. According to the result, the NLSs of 107 types of HPV L1 proteins were classified into 15 categories. The different types of HPV L1 proteins in the same NLS category could share the similar or the same nucleocytoplasmic transport pathway. They might be used as the same target to prevent and treat different types of HPV infection. The results also showed that bioinformatic technology could be used to analyze and predict NLSs of proteins.

Key words: human papillomavirus (HPV), nuclear localization signal (NLS), bioinformatics

Introduction

Human papillomaviruses (HPVs) are small, non-enveloped DNA viruses (*1*). HPV infection is associated with more than 90% of all the cases of cervical cancer, which is the second leading cause of cancer death among women worldwide (*1, 2*). HPVs have been classified into more than 100 types based on the nucleotide sequence homology of a single molecule of 8-Kb double-stranded circular DNA. Each HPV type has different specificity for infection of skin or mucosa (*3*).

HPV virion (55–60 nm in diameter) is contained within an icosahedral capsid, which comprises L1 major and L2 minor capsid proteins (*4*). L1 proteins form pentamers (capsomeres), and 72 capsomeres assemble into a T-7d icosahedral lattice (*5, 6*). An HPV capsid comprises 360 molecules of L1 proteins. L2 proteins interact with L1 pentamers (*5*). The molar ratio of L1 and L2 proteins is estimated to be 30:1 (*7, 8*). L1 proteins can self-assemble into virus-like particles, which have the similar size, shape, and conformational epitope to native virion capsid proteins, although L2 proteins increase the efficiency of DNA encapsidation by at least 50 folds (*8–10*).

HPVs infect basal cells of epithelium through microlesions and replicate only in the differentiating cells. These cells are difficult to culture *in vitro*; hence, no tissue culture system for the large-scale propagation of HPV virions *in vitro* is available at present. The study of these viral structural proteins is behind that of the oncoproteins of their counterfeits. Consequently, little is known about the cellular and viral factors that control the switch and process of papillomavirus genome replication and viral protein expression. Many events in the papillomavirus life cycle have not been elucidated, and particularly the nuclear transport process of the viral genome and structural proteins is poorly understood.

However, at present the knowledge of L1 proteins of HPVs is understood at the molecular level in a certain extent. During the virus life cycle, L1 proteins seem to enter the nuclei of host cells twice. In the initial stage of HPV infection, immediately after the virions infect the undifferentiated proliferating epithelial cells, L1 proteins together with the vi-

*** Corresponding author.**
E-mail: slusheng@yahoo.com

ral genome are transported into the nuclei of proliferating epithelial cells. During the late stage of HPV infection, the newly synthesized L1 proteins in cytoplasm are transported into the nuclei of terminally differentiated keratinocytes to package the replicated HPV genomic DNAs and assemble into infectious virions, together with L2 proteins (*11*). This would suggest that the nuclear import of L1 proteins plays a very important role in HPV infection and production. The ability of the virus importing into the nucleus is determined by the nuclear localization signal (NLS) in the C-terminal of HPV L1 proteins, so it is important to investigate the NLSs of HPVs. To date, more than 120 HPV types have been isolated and partially characterized, and about 100 distinct HPV types have been identified and fully sequenced. But only few NLSs of HPV L1 proteins have been experimentally determined. The fact means that it is very difficult and unpractical to identify the NLSs of all HPV types by experiments. In this paper, we attempt to analyze and predict the NLSs of 107 types of HPV L1 proteins by bioinformatic analysis.

Results

The full sequences of 107 types of HPV L1 proteins were obtained from available databases (see Materials and Methods). Out of the 107 types, the NLSs of 39 types were predicted by PredictNLS software (http://cubic.bioc.columbia.edu/predictNLS/). Among them, 35 types contain bipartite NLSs, where the two tight clusters of basic residues (one is KRKR, KRKRK, KRKKRK, the other is KR, RKR, KRK) are preceded, with a spacer of 10–14 amino acids. The other four types (HPV22, HPV34, HPV48, and HPV73) were predicted to contain monopartite NLSs, where these arginines and/or lysines form a tight cluster of basic residues as typified by the simian virus 40 large T antigen (SV40 T).

The NLSs of the remaining HPV types were predicted according to the characteristics and the homology of the already predicted NLSs as well as the general rule of NLSs. According to the result, the NLSs of 107 types of HPV L1 proteins were classified into 15 categories (Table 1), among which the categories XIV and XV contain monopartite NLSs. In addition, the NLSs of HPV L1 proteins 1, 6, 11, 16, 31, 33, 35, and 45 can also be obtained from the literature (*12–15*).

Discussion

In eukaryotic cells, the nucleus has a highly specialized structure that participates in the regulation of cell processes, including the regulation of cell cycle and the induction of antiviral responses (*16*). The nuclear pore complex (NPC) has a large supramolecular structure with a mass of 125 kDa in vertebrates, which is embedded in nuclear envelope as the only gateway between nucleus and cytoplasm (*17–20*).

Over the past years, a consensus model of the three-dimensional (3D) architecture of NPC shows that it is composed of an eight-fold symmetric central framework (*21*). In the course of biological evolution, NPC keeps a very high homology in eukaryotic cells, sharing a similar nuclear transport mechanism (*19, 20*).

The nuclear import of proteins typically requires the presence of NLSs, which are characteristically rich in basic amino acids (*22–24*). NLS motifs play a key role in the nuclear transport mechanism. In order to enter into nucleus, the transport of proteins with a molecular weight (MW) at 45–60 kDa must be made through NPC via an NLS or be associated with another protein via a piggyback mechanism, whereas the nuclear import of small proteins (MW<40 kDa) cross NPC via passive diffusion (*25–27*).

NLSs are subsequently found in numerous viral and nuclear proteins of eukaryotic cells. At present, they can be classified into two major categories. The first category includes the monopartite (single type) NLSs that contain 3–5 basic amino acids with the weak consensus Lys-Arg/Lys-*X*-Arg/Lys residues preceded by a helix-breaking residue, which are similar to the SV40 T NLS (pKKKRKv) (*28*). They are now referred to as classical NLSs. The second category includes the bipartite NLSs that contain two clusters of basic regions of 3–4 residues with a basic dipeptide upstream from a simple basic sequence, each separated by approximately 10 amino acids, which are similar to the nucleoplasmin NLS (KRpaatkkagqaKKKKldk) (*29, 30*). The sequences (pKKKRKv and KRpaatkkagqaKKKKldk) found in SV40 T and nucleoplasmin are prototypes for monopartite and bipartite NLSs, now known to be present in many, probably thousands of different proteins. NLSs are capable of directing a non-karyophilic protein into nucleus when conjugated genetically or chemically. However, not all experimentally known NLSs comply with the above rules (*31–33*). Several other NLS sequences have been identified, which are

Table 1 The Homologous Analysis of the NLSs of 107 Types of HPV L1 Proteins

HPV type	Accession No.	NLS
I (17)		[RKR]x{9,16}[KRKR]
HPV60	U31792	KRYLYQYGLLNG RKRSASSFVTKK S KTV KRKRTK *
HPV3	X74462	RKFLMQLGVGTRSSISVRKRSATTTSRTAAA KRKRTKK *
HPV94	AJ620211	RKFLLQLGVRSRSAISVRKRSATAASGSTAA KRKRTKK *
HPV10	X74465	RKFLLQLGVRSRSAVSVRKRPATSATGSTAA KRKRTKK *
HPV28	U31783	RKFLMQLGVGARSSVSVRKRPASTTRGSSAA KRKRAKK *
HPV68	X67161	RKFLLQAGVRRRPTIGPRKRTATATTTST S KH KRKRVSK *
HPV68ME180	M73258	RKFLLQAGVRRRPTIGPRKRPATATTAST S KH KRKRVSK
HPVME	P27964	RKFLLQAGVRRRPTIGPRKRPATATTAST S KH KRKRVSK
HPV39	M62849	RKFLLQARVRRRPTIGPRKRPAASTSSSSA TKH KRKRVSK *
HPV70	U21941	RKFLLQVGARRRPTIGPRKRPASAKSSSSAS KH KRKRVSK *
HPVcand85	AF131950	RKFLLQAGLRPKPTIGPRKRVASTSTATRPS KRKRTAK *
HPV2	NC_001352	RKFLLQRGAM PT VSRKRAAVSGTTPPTS KRKRVRR
HPV2a	X55964	RKFLLQRGAM PT VSRKRAAVSGTTPPTS KRKRVRR *
HPV27	X74473	RKFLLQRGTT PT VSRKRTAVGRGH
HPV27b	AB211993	RKFLLQRGTT PT VSRKRTAVAGAAAPTS KRKRVRR *
HPVJC9710	AF042837	RKFLLQRGTRVRSSPVSRKRPAPSTAPST KRKRSKRS *
HPV95	CAF05708	RRFLYQSGLING SRKRQRAITSQTATGTKRSAKRKRLSK *
II (3)		[R(K)KR]x{9,10}[KP(Q)KR]
HPV55	U31791	RKFLLQTGVQARSSVRVGRKRPASAATSSSS KPKRSRKK
HPV44	U31788	RKFLLQTGVQARSSVRVGRKRPASAATSSS KQKRSRKK
HPV74	U40822	RKFLLQTGVQARSSVRVSKKRSAPTAPSSAT KQKRSRKR
III (6)		[RKR]x{10,14}[KRKK]
HPV57	X55965	RKFLLQRGATPTVS RKRAAATAAAPTA KRKKVRR
HPV57b	AAC56600	RKFLLQRGATPTVS RKRAAATAAAPTA KRKKVRR *
HPV91(cand89)	AF436128	RKFLLQLGGRPSSVP RKRAAPVSTSKAP KRKKAKR
HPV13	X62843	RKFLLQTGVQSRSPIRVGRKRAASTSTATPTT RKKAKRK
HPV7	X74463	RKFLMQAGLRTGPKFKS RKRPAPTSSSSSGSVTPKRKKTKR *
HPV40	X74478	RKFLMQAGVRAGPRFKS RKRPAPSSSSSSKPVTPKRKKTKR *
IV (4)		[RKR]x{10,12}[KRRK(R)]
HPV84	AF293960	RKFLLQSAPRSTLVS RKRTASASTPPAS KRRKAKK *
HPV87	CAC17718	RKFLLQSAPRVSRVS RKRPASTSTASTS KRRKAKK *
HPV86	AF349909	RKFLLQSAPRVSHVS RKRPASTSTASSS KRRKTKK
HPV9	X74464	RKFLFQAGLQT RKRPIKTSVKTSKNA KRRRT *
V (7)		[RKR]x{10,15}[KR(K)K(V)K(R/V)K(R)K(R)]
HPV18	X05015	RKFLVQAGLRRKPTIGPRKRSAPSATTSSKPA KRVRVRARK
HPV45	X74479	RKFLVQAGLRRRPTIGPRKRPAASTSTASTASRPAKRVRIRSKK #
HPV59	X77858	RKFLLQLGARPKPTIGPRKRAAPAPTSTPSP KRVKRRKSSRK *
HPV71	NC_002644	RKFLLQSGTRSRPTALSRKRVAASTTSTAP KRKRVKRSR
HPV72	X94164	RKFLLQVGSRAVSVS RKRAAPPSSTSTPAPT KRKKRKK *
HPV61	U31793	RKFLLQAGPRSVSVS RKRAAPSSTPTSSPAT KRKKRKQ *
HPV83	AF151983	RKFLLQLGPRSVSVS RKRPASTAPS APS KKKVKRRK
VI (5)		[KR(K)R]x{10,14}[KRKRK(R)]
HPV50	U31790	RKFLFQTGLL KRRVRTDYTVATVSKPNKRKRTR *
HPV53	X74482	RKFLMQVGVRTKPPVSS KKRSASTTSTSAPSS KRKRK *
HPV30	X74474	RKFLMQLGVRTKPSTTT KKRSAPSSSTSTPSA KRKRR *
HPV56	X74483	RKFLMQLGTRSKPAVATS KKRSAPTSTSTP A KRKRR *
HPV66	U31794	RKFLMQLGPRPPRPKASVSASKRRAAPTS SSSSPA KRKKR *
VII (3)		[RKRR]x{7,9}[KRK(R)RR]
HPV29	U31784	RKFLLQIGARRRSVVPS RKRRTTTTAPTPAKRKRSKK *
HPV77	Y15175	RKFLLQIGARRRSVVPS RKRRAPTPSPASTKRKRSKK *
HPV63	X70828	RKFLYQSGLAQRSVPKTVNFRKRRSSNTTVA KRRRRA *

Table 1 *Continued*

HPV type	Accession No.	NLS
VIII (5)		[KRK]x{8,12}[KRK(R)K(R)K(R)K(H)]
HPV42	M73236	RKFLLQAGLRARPKLSVG KRKASTAKSVSS AKRKKTHK *
HPV16	U37217	RKFLLQAGFKAKPKFTLG KRKATPTTSSTSTTAKRKKRKL *#
HPV16R	N/A	RKFLLQAGFKAKPKFTLG KRKATPTTSSTSTTAKRKKRKL
HPV1aR	V01116	RKFLYQSGMTQRIATSSTTKRKTVRV STS AKRRRKA *#
HPV1a	A09292	RKFLYQSGMTQRIATSSTTKRKTVRV STS AKRRRKA #
IX (2)		[KRKR]x{14,15}[KRKRSL(I)K]
HPV4	X70827	RRFLYQSGLINGSLKRKRIISSSHAQTNTKRSAKRKRSLK *
HPV65	X70829	RRFLYQSGLINGTLKRKRTINSQAPTSI KRSAKRKRSIKQ *
X (12)		[KR]x{8,13}[KRKR]
HPV6	AY015006	RKFLLQSGYRGRSSIRTGVKRPAV SKASAAPKRKRAKTKR #
HPV6A	P69898	RKFLLQSGYRGRSSIRTGVKRPAV SKASAAPKRKRAKTKR
HPV6b	AF322411	RKFLLQSGYRGRSSIRTGVKRPAV SKASAAPKRKRAKTKR
HPV6bR	X00203	RKFLLQSGYRGRSSIRTGVKRPAV SKASAAPKRKRAKTKR
HPV11	AY541029	RKFLLQSGYRGRTSARTGIKRPAV SKPSTAPKRKRTKTKK #
HPV11R	M14119	RKFLLQSGYRGRTSARTGIKRPAV SKPSTAPKRKRTKTKK
HPV82	AAK28456	RKFLLQIGAQ RK ARPGLKRPAP SSSSSSSAKRKRVKK
HPV82 IS39/AE2	AF293961	RKFLLQIGAQ RK ARPGLKRPAP SSSSSSSAKRKRVKK
HPV51	M62877	RKFLLQVGVQ RK PRPGLKRPASSASSSSSSAKRKRVKK
HPV93	AY382778	RKFLYQANLVQ SPA KRSSSISRGT KRKRSK
HPV38	U31787	RKFLFQAGLQTARTRAV KRPLVRKSSKSV KRKRTQ
HPV24	U31782	RKFLFQAGLVQ KT SKRTS NV SK GTKRKRT
XI (11)		[KR]x{10,12}[KR(K)K(R)K]
HPV23	U31781	RKFLFQIGVQRVRSGT KRPATRKVTKTV KRKKVQL
HPV31	J04353	RKFLLQAGYRARPKFKAG KRSAPSASTTTP AKRKKTKK #
HPV43	AJ620205	RKFVMQAGLRPKPKLKTV KRSAPSSSTSAPASKRKKTKR
HPVJC9813-A	AF070938	RKFVMQAGLRPRPKLKSG KRAAPSSS SAPASKRKKTKR
HPV32	X74475	RKFLLQAGLRARPKLTAV KRTASSSQKSSSPAKRRKTRK
HPV35	P27232	RKFLLQAGLKARPNFRLG KRAAPASTSKKSSTKRRKVKS #
HPV35h	X74477	RKFLLQAGLKARPNFRLG KRAAPASTSKKSSTKRRKVKS
HPV52	X74481	RKFLLQAGLQARPKL KRPASSAPRT STKKKKVKR
HPV58	D90400	RKFLLQSGLKAKPRL KRSAPTTRAP STKRKKVKK
HPV33	M12732	RKFLLQAGLKAKPKL KRAAPTSTRT SSAKRKKVKK #
HPV67	D21208	RKFLLQAGFTAKPKL KRSSPSSSSS SSAKRKKVKR
XII (6)		[KR]x{9,17}[K(R)R(K)K(R)R(A)K(R/V)]
HPV41	X56147	RKFLFQTGITQSSSN KRVSTQSTALTTYRRP TKRRRKA
HPV48	U31789	RKFLYQTGLLNG KRARIDYTAAGSSTRS TKRRRVR *
HPV96	AY382779	RKFLFQANLQN KRVNRGVTVTGRATTSRGTKRKRR
HPV92	AF531420	RKFLFQAGLTNTSVN GLKRTRSSSQRGT KRKRKSN
HPV54	U37488	RKFLLQAGLRARPRLRPV KRAAPSSSKGTA RKRAKTKR
HPV69	AB027020	RKFMLQAGIQRRPKLGT KRPASSLSASSS ST TRKKRKLTK
XIII (4)		[KR]x{7,10}[KR(K/T)K(V)K(R)R(A)K(R)K(R)]
HPV26	X74472	RKFMLQAGIQRRPKLGT KRPLSS TSSST KRKKRKLTK
HPV22	U31780	RKFLFQSGLQRARASARVSVKRSATRKTS KTVKRRKLTS
HPV73	X94165	RKFLLQLGMRARPKLQAS KRSASATTSA TPKKKRAKRI *
HPV34	X74476	RKFLLQLGMRARPRLQAS KRSAPSSSST APKKKRAKRI *
XIV (7)		[K(R)]x{2}[K(R)]x{3,4}[K(R)]x{2,3}[K(R)]x{2}[KRKR]
HPV15	X74468	RKFIFQAGLQRRP RTIKS SVKV SKGTKRKRT
HPV80	Y15176	RKFIFQAGLQRRP KTIRS SVKV SKGTKRKRT
HPV37	U31786	RKFIFQSGLQSRP RIVRS SVKV SKGTKRKRS
HPV17	X74469	RKFIFQSGLQARP RTIRT SVKV PKGIKRKRS
HPV75	Y15173	RKFLFQAGL RRASKVTRTASVRSSSRGTKRKRT
HPVVS75L1	X79945	RKFLFQAGL RRASKVTRTASVRSSSRGTKRKRT
HPV76	Y15174	RKFLFQAGL RRASRVSRTATVRTASRGTKRKRS

Table 1 *Continued*

HPV type	Accession No.	NLS
XV (15)		[K]x{3,4}[K(R)]x{3}[K(R)]x{2}[KRKRK(R)]
HPV8	M12737	RKFLFQAGLQQTTVNGT KSIS RGSVRGIKRKRKN
HPV12	X74466	RKFLFQAGLQQTTVNGTTKSSSYRSSIRGIKRKRKN
HPVR7	P88804	RKFLFQAGLQQTTVNGT KPSTYRSSVRGIKRKRKN
HPV36	U31785	RKFLFQAGLQQTTVSGI KSVSYRGFTRGIKRKRKQ
HPV21	U31779	RKFLFQAGLQQTTVNGT KTLSSRVSTRGIKRKRKN
HPV19	X74470	RKFLFQAGLQQATVNGT KTISSRVSSRGTKRKRKN
HPV49	X74480	RKFLFQAGLQRASRVS KSSAARASTRGIKRKRR
HPV25	X74471	RKFLFQAGLQQTTVNGT KTVSSRISTRGIKRKRKN
HPV14	P36734	RKFLFQAGLQQSTVNGT KTVSTRGSIKGIKRKRKN
HPV14d	X74467	RKFLFQAGLQQSTVNGT KTVSTRGSIKGIKRKRKN
HPV47	M32305	RKFLFQAGLQQTTVNGT KTTPYRGSIRGTKRKRKN
HPV20	U31778	RKFLFQAGLQQATVNGT KTVSSKLSTRGVKRKRKQQ
HPV5	P06917	RKFLFQAGLQQTTVNGT KAVSYKGSNRGTKRKRKN
HPV5b	D90252	RKFLFQAGLQHTTVNGT KAVSYKGSNRGTKRKRKN
HPV5R	M17463	RKFLFQAGLQQTTVNGT KAVSYKGSNRGTKRKRKN

*The NLSs of HPV L1 proteins predicted by the PredictNLS software; #The NLSs obtained from references. The NLS sites are in bold and underlined. Format for NLS motifs: [KR]: read K or R; [R(K)KR]: read RKR or KKR; x{3,5}: read between 3 and 5 x, where "x" stands for any amino acid.

quite different from classical NLSs, such as the NLSs discovered in hnRNP proteins, ribosomal proteins, and UsnRNPs (*34–37*).

Recent studies have identified several proteins that contain more than one NLS, including the nuclear factor 1-A (*38*), the cell division control protein mcm10 (*39*), the herpes simplex virus gene product ICP22 (*40*), the HIV preintegration complex (*41*), the Epstein-Barr virus DNase (*42*), the papillomavirus oncoprotein E6 (*43*), BRCA2 (*44*), and Rep68/78 proteins (*45*). The enzyme 5-lipoxygenase (5-LO) has three NLSs that contain dispersed basic residues, unlike the tight cluster of basic residues of the classical SV40 T NLS (*46*). It is not clear why some proteins contain multiple NLSs. One explanation is that multiple NLSs may cooperate with one another and allow more efficient nuclear import or share an alternative entry mechanism in the nuclear import, affording redundancy in proteins that require successful nuclear import, as in cell cycle proteins or viral integration proteins (*42, 47*).

Traditionally, in order to identify an NLS experimentally, both of the facts should be considered routinely. Firstly, the candidate should be deleted to disrupt the nuclear import of the NLS; secondly, a non-nuclear protein will be imported into the nucleus if fused to the NLS (*48*). It is very difficult and unpractical to identify the NLS motifs of more than 120 types of HPVs by experiments. Only few NLSs of HPV L1 proteins have been experimentally determined up to date.

Bioinformatic techniques perhaps could be applied to analyze and predict new NLSs, which would remedy this situation. Cokol *et al* (*49*) have found some upper boundaries. The method comprises two steps: (1) data collection: collect experimental NLS motifs from literature, and extend the motifs through close homologues; (2) generalization: refine the motifs found by shortening (for those too specific) or lengthening (for those not specific enough), and test the new motifs conceptually similar to the known motifs found in nuclear protein families. The crucial component of both steps is to accept motifs if NOT found in non-nuclear proteins. Therefore, it is feasible to discover new NLSs in HPV L1 proteins by comparing the homologues of different types.

According to Cokol's method, we analyzed the HPV L1 protein sequences for the confirmation of NLSs using PredictNLS software. Out of the 107 types of HPV L1 proteins, the NLSs of 39 types were predicted by PredictNLS. Applying PSORTII and PredictNLS could not reveal any typical NLS in the remaining 68 types.

In general, two naturally evolved proteins with more than 30% identical residues could share similar 3D structures (*50*). The sequence similarity required to infer function is much higher (*51*). It is possible to infer NLSs by comparing the homologues

of more than one protein. Due to the high homology of the NLSs among the L1 proteins of all HPV types, we subsequently found similar sequences in the C-terminal of all the remaining 68 types of HPV L1 proteins, which were similar to the NLSs already predicted by PredictNLS and collected in experimental data. According to the consensus rule of NLSs and the high homologues, the NLSs of 107 types of HPV L1 proteins were classified into 15 categories (Table 1). Among them, the categories I to XIII contain classical bipartite NLSs, while the categories XIV and XV contain classical monopartite NLSs.

However, the NLSs predicted in this paper have been proved with few experimental data. This classification cannot always consist with experimental results. The cluster of basic residues RRR in the upstream of the bipartite NLS (RRRptigpRKRpaaststastasRpaKRvRiRsKK) of HPV45 has been proved certainly to have the nuclear localization ability (*13*). The discontinuous basic amino acids K and R in the upstream of the NLSs of 107 types of HPV L1 proteins perhaps also possess the ability of nuclear localization. At the same time, the NLS of HPV33 was proposed to be bipartite, while the experimental result proved that HPV33 possibly contains a monopartite NLS (*52*). On the other hand, while the experimental results proved that many types of HPV L1 proteins (for example, HPV1, 6, 31, 33, and 35) contain NLSs, they cannot be found by PredictNLS. Whereas, it is not all the clusters of basic residues of predicted NLSs that have the nuclear localization ability, such as the NLS of the IL1 β (*53*, *54*). This instance perhaps occurs in HPV L1 proteins. Therefore, it is surely worth amending and supplementing the classification.

In conclusion, this classification would play an important role in the study of the NLSs of HPV L1 proteins. The results of this paper suggested that the different HPV types classified in the same category could share the similar or the same nucleocytoplasmic transport pathway. The NLSs in the same category would be used as a common realistic and feasible target for preventing and treating different types of HPV infection. The results also showed that bioinformatic technology could be used to analyze and predict the NLSs of proteins.

Materials and Methods

The HPV L1 protein sequences were searched from the following databases:

http://cubic.bioc.columbia.edu/db/
http://www.ncbi.nlm.nih.gov/
http://www.stdgen.lanl.gov/stdgen/virus/
http://ca.expasy.org/

Firstly, the initial sets from the literature for experimentally determined NLSs were collected. Secondly, ENTREZ, BLAST, and DNAClub software tools were used to analyze the homology of all types of HPV L1 protein sequences obtained. The useful web server (PredictNLS) for identifying potential NLSs in protein sequences is available at http://cubic.bioc.columbia.edu/predictNLS/ and was used to analyze and predict the NLSs of HPV L1 proteins. According to the characteristics and the homology of the NLSs predicted by PredictNLS, as well as the general rule of NLSs, the HPV L1 proteins were classified into 15 categories. The program also allows experimentalists to test the accuracy and coverage for new NLS motifs that they may find or suspect. This feature has already helped to experimentally unravel a novel NLS in the hairless protein (*55*).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 30300305).

References

1. zur Hausen, H. 1996. Papillomavirus infections—a major cause of human cancers. *Biochim. Biophys. Acta* 1288: F55-78.
2. zur Hausen, H. 2000. Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis. *J. Natl. Cancer Inst.* 92: 690-698.
3. zur Hausen, H. 1991. Human papillomaviruses in the pathogenesis of anogenital cancer. *Virology* 184: 9-13.
4. Doorbar, J. and Gallimore, P.H. 1987. Identification of proteins encoded by the L1 and L2 open reading frames of human papillomavirus 1a. *J. Virol.* 61: 2793-2799.
5. Chen, X.S., *et al.* 2000. Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. *Mol. Cell* 5: 557-567.
6. Hagensee, M.E., *et al.* 1994. Three-dimensional structure of vaccinia virus-produced human papillomavirus type 1 capsids. *J. Virol.* 68: 4503-4505.
7. Hagensee, M.E., *et al.* 1993. Self-assembly of human papillomavirus type 1 capsids by expression of the L1 protein alone or by coexpression of the L1 and L2 capsid proteins. *J. Virol.* 67: 315-322.

8. Kirnbauer, R., *et al.* 1993. Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles. *J. Virol.* 67: 6929-6936.
9. Le Cann, P., *et al.* 1994. Self-assembly of human papillomavirus type 16 capsids by expression of the L1 protein in insect cells. *FEMS Microbiol. Lett.* 117: 269-274.
10. Zhou, J., *et al.* 1991. Expression of vaccinia recombinant HPV 16 L1 and L2 ORF proteins in epithelial cells is sufficient for assembly of HPV virion-like particles. *Virology* 185: 251-257.
11. Nelson, L.M, *et al.* 2002. Nuclear import strategies of high risk HPV16 L1 major capsid protein. *J. Biol. Chem.* 277: 23958-23964.
12. Zhou, J., *et al.* 1991. Identification of the nuclear localization signal of human papillomavirus type 16 L1 protein. *Virology* 185: 625-632.
13. Merle, E., *et al.* 1999. Nuclear import of HPV11 L1 capsid protein is mediated by karyopherin alpha2beta1 heterodimers. *J. Cell. Biochem.* 74: 628-637.
14. Nelson, L.M., *et al.* 2003. The L1 major capsid protein of human papillomavirus type 11 interacts with Kap beta2 and Kap beta3 nuclear import receptors. *Virology* 306: 162-169.
15. Nelson, L.M., *et al.* 2000. Nuclear import and DNA binding of human papillomavirus type 45 L1 capsid protein. *J. Cell. Biochem.* 79: 225-238.
16. Rowland, R.R. and Yoo, D. 2003. Nucleolar-cytoplasmic shuttling of PRRSV nucleocapsid protein: a simple case of molecular mimicry or the complex regulation by nuclear import, nucleolar localization and nuclear export signal sequences. *Virus Res.* 95: 23-33.
17. Fried, H. and Kutay, U. 2003. Nucleocytoplasmic transport: taking an inventory. *Cell. Mol. Life Sci.* 60: 1659-1688.
18. Vasu, S.K. and Forbes, D.J. 2001. Nuclear pores and nuclear assembly. *Curr. Opin. Cell Biol.* 13: 363-375.
19. Fahrenkrog, B. and Aebi, U. 2003. The nuclear pore complex: nucleocytoplasmic transport and beyond. *Nat. Rev. Mol. Cell. Biol.* 4: 757-766.
20. Fahrenkrog, B., *et al.* 2004. The nuclear pore complex: a jack of all trades? *Trends Biochem. Sci.* 29: 175-182.
21. Stoffler, D., *et al.* 2003. Cryo-electron tomography provides novel insights into nuclear pore architecture: implications for nucleocytoplasmic transport. *J. Mol. Biol.* 328: 119-130.
22. Dingwall, C. and Laskey, R.A. 1991. Nuclear targeting sequences-a consensus? *Trends Biochem. Sci.* 16: 478-481.
23. Garcia-Bustos, J., *et al.* 1991. Nuclear protein localization. *Biochim. Biophys. Acta* 1071: 83-101.
24. Makkerh, J.P., *et al.* 1996. Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. *Curr. Biol.* 6: 1025-1027.
25. Leslie, D.M., *et al.* 2004. Characterization of karyopherin cargoes reveals unique mechanisms of Kap121p-mediated nuclear import. *Mol. Cell. Biol.* 24: 8487-8503.
26. Johnson, H.M., *et al.* 2004. Trafficking and signaling pathways of nuclear localizing protein ligands and their receptors. *Bioessays* 26: 993-1004.
27. Wendler, P., *et al.* 2004. The bipartite nuclear localization sequence of Rpn2 is required for nuclear import of proteasomal base complexes via karyopherin alpha-beta and proteasome functions. *J. Biol. Chem.* 279: 37751-37762.
28. Kalderon, D., *et al.* 1984. A short amino acid sequence able to specify nuclear location. *Cell* 39: 499-509.
29. Dingwall, C., *et al.* 1982. A polypeptide domain that specifies migration of nucleoplasmin into the nucleus. *Cell* 30: 449-458.
30. Robbins, J., *et al.* 1991. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* 64: 615-623.
31. Hsieh, J.C., *et al.* 1998. Novel nuclear localization signal between the two DNA-binding zinc fingers in the human vitamin D receptor. *J. Cell. Biochem.* 70: 94-109.
32. Truant, R. and Cullen, B.R. 1999. The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. *Mol. Cell. Biol.* 19: 1210-1217.
33. Irie, Y., *et al.* 2000. Molecular cloning and characterization of Amida, a novel protein which interacts with a neuron-specific immediate early gene product arc, contains novel nuclear localization signals, and causes cell death in cultured cells. *J. Biol. Chem.* 275: 2647-2653.
34. Izaurralde, E. and Adam, S. 1998. Transport of macromolecules between the nucleus and the cytoplasm. *RNA* 4: 351-364.
35. Mattaj, I.W. and Englmeier, L. 1998. Nucleocytoplasmic transport: the soluble phase. *Annu. Rev. Biochem.* 67: 265-306.
36. Moroianu, J. 1999. Nuclear import and export pathways. *J. Cell. Biochem.* 32-33: 76-83.
37. Rollenhagen, C., *et al.* 2003. Importin beta-depending nuclear import pathways: role of the adapter proteins in the docking and releasing steps. *Mol. Biol. Cell.* 14: 2104-2115.
38. Imagawa, M., *et al.* 2000. Two nuclear localization signals are required for nuclear translocation of nuclear factor 1-A. *FEBS Lett.* 484: 118-124.
39. Burich, R. and Lei, M. 2003. Two bipartite NLSs mediate constitutive nuclear localization of Mcm10. *Curr. Genet.* 44: 195-201.

40. Stelz, G., *et al.* 2002. Identification of two nuclear import signals in the alpha-gene product ICP22 of herpes simplex virus 1. *Virology* 295: 360-370.
41. Haffar, O.K., *et al.* 2000. Two nuclear localization signals in the HIV-1 matrix protein regulate nuclear import of the HIV-1 pre-integration complex. *J. Mol. Biol.* 299: 359-368.
42. Liu, M.T., *et al.* 1998. Epstein-Barr virus DNase contains two nuclear localization signals, which are different in sensitivity to the hydrophobic regions. *Virology* 247: 62-73.
43. Tao, M., *et al.* 2003. Signals that dictate nuclear localization of human papillomavirus type 16 oncoprotein E6 in living cells. *J. Virol.* 77: 13232-13247.
44. Yano, K., *et al.* 2000. Nuclear localization signals of the BRCA2 protein. *Biochem. Biophys. Res. Commun.* 270: 171-175.
45. Cassell, G.D. and Weitzman M.D. 2004. Characterization of a nuclear localization signal in the C-terminus of the adeno-associated virus Rep68/78 proteins. *Virology* 327: 206-214.
46. Luo, M., *et al.* 2004. Multiple nuclear localization sequences allow modulation of 5-lipoxygenase nuclear import. *Traffic* 5: 847-854.
47. Caron, M.C. and Caruso, M. 2005. A nuclear localization signal in the matrix of spleen necrosis virus (SNV) does not allow efficient gene transfer into quiescent cells with SNV-derived vectors. *Virology* 338: 292-296.
48. Tinland, B., *et al.* 1992. The T-DNA-linked VirD2 protein contains two distinct functional nuclear localization signals. *Proc. Natl. Acad. Sci. USA* 89: 7442-7446.
49. Cokol, M., *et al.* 2000. Finding nuclear localization signals. *EMBO Rep.* 1: 411-415.
50. Rost, B. 1999. Twilight zone of protein sequence alignments. *Protein Eng.* 12: 85-94.
51. Devos, D. and Valencia, A. 2000. Practical limits of function prediction. *Proteins* 41: 98-107.
52. Florin, L, *et al.* 2002. Assembly and translocation of papillomavirus capsid proteins. *J. Virol.* 76: 10009-10014.
53. Jans, D.A. and Hassan, G. 1998. Nuclear targeting by growth factors, cytokines, and their receptors: a role in signaling? *Bioessays* 20: 400-411.
54. Song, L. and Shen, B. 2001. Does putative nuclear localization sequence in intra cellular domain of gp-130 possess function of nuclear import? *Chin. J. Cancer* 20: 11-14.
55. Djabali, K., *et al.* 2001. Hairless is translocated to the nucleus via a novel bipartite nuclear localization signal and is associated with the nuclear matrix. *J. Cell Sci.* 114: 367-376.