ORGANIC LETTERS

2003
Vol. 5, No. 14
2437−2440

Simple Mimetics of a Nuclear Localization Signal (NLS)

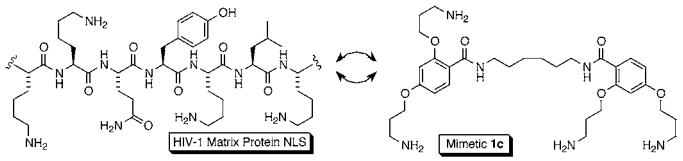
Seung Bum Park, Thai H. Ho, Brian M. Reedy, Michael D. Connolly, and Robert F. Standaert*

University of Illinois at Chicago, Department of Chemistry (M/C 111), 845 West Taylor Street, Room 4500, Chicago, Illinois 60607

rfs@uic.edu

Received April 14, 2003

ABSTRACT



Molecular modeling was used to design mimetics of the HIV-1 matrix protein nuclear localization signal (NLS) in which a scaffold of two resorcinol units joined by a diamide linker presents 3-aminopropyl ethers in place of lysine side chains. Prospective mimetics with linkers of 6, 8, 10, or 12 atoms were synthesized and compared in a competition assay for binding to the nuclear import receptor subunit karyopherin α, showing the 10-atom linker to be best and shorter ones ineffective.

The nuclear localization signal (NLS) is a short peptide motif that directs proteins into the nucleus through binding to the nuclear import receptor subunit karyopherin α (Kap α).[1] Rather than being a single sequence, the NLS comprises a heterogeneous group of peptides typically containing 3−5 clustered lysine or arginine residues,[2] with a loose minimal consensus sequence of K(K/R)X(K/R).[3] The prototype NLS, from the SV40 large T antigen (TAg), has the sequence PKKKRKV.[4] Others relevant to this work include the NLS from the c-Myc proto-oncoprotein (PAAKRVKLD)[5] and the NLS from the HIV-1 matrix protein (GKKQYKLKH).[6]

We are interested in developing nonpeptide mimetics of the NLS for two reasons: first, we seek a greater understanding of the requirements for molecular recognition in the nuclear import process; second, we seek to develop ligands that distinguish among Kap α homologues in different species or among Kap α isoforms from the same species. Little has been published in the area of NLS mimetics, and we are aware of only two classes of NLS-related molecules containing anything other than the standard coded amino acids. Gerace and co-workers prepared variants of the c-Myc NLS in which D-amino acids were used (inverso-peptide), the sequence was reversed (retro-peptide), or both (retro-inverso peptide), and showed that all were functional import signals.[7a] Nadler and co-workers have shown that a D-peptide containing dual TAg NLS sequences, earlier shown to inhibit the nuclear localization of the transcription factor NF-κB,[7b] binds Kap α and blocks its interaction with NF-κB.[7c] Conformationally constrained

(1) (a) Görlich, D.; Prehn, S.; Laskey, R. A.; Hartmann, E. *Cell* **1994**, *79*, 767−778. (b) Moroianu, J.; Hijikata, M.; Blobel, G.; Radu, A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6532−6536. (c) Imamoto, N.; Shimamoto, T.; Takao, T.; Tachibana, T.; Kose, S.; Matsubae, M.; Sekimoto, T.; Shimonishi, Y.; Yoneda, Y. *EMBO J.* **1995**, *14*, 3617−3626.

(2) Boulikas, T. *Crit. Rev. Eukaryotic Gene Expression* **1993**, *3*, 193−227.

(3) Hodel, M. R.; Corbett, A. H.; Hodel, A. E. *J. Biol. Chem.* **2001**, *276*, 1317−1325.

(4) Kalderon, D.; Richardson, W. D.; Markham, A. F.; Smith, A. E. *Nature* **1984**, *311*, 33−38.

(5) Dang, C. V.; Lee, W. M. *Mol. Cell. Biol.* **1988**, *8*, 4048−4054.

(6) The sequence given is from strain NL4-3 (Adachi, A.; Gendelman, H. E.; Koenig, S.; Folks, T.; Willey, R.; Rabson, A.; Martin, M. A. *J. Virol.* **1986**, *59*, 284−291). The exact sequence varies among strains, and a consensus of GKKKYKLKH has been proposed (Bukrinsky, M. I.; Haggerty, S.; Dempsey, M. P.; Sharova, N.; Adzhubel, A.; Spitz, L.; Lewis, P.; Goldfarb, D.; Emerman, M.; Stevenson, M. *Nature* **1993**, *365*, 666−669).

(7) (a) Saphire, A. C. S.; Bark, S. J.; Gerace, L. *J. Biol. Chem.* **1998**, *273*, 29764−29769. (b) Fujihara, S. M.; Cleaveland, J. S.; Grosmaire, L. S.; Berry, K. K.; Blake, J. J.; Loy, J.; Rankin, B. M.; Ledbetter, J. A.; Nadler, S. G. *J. Immunol.* **2000**, *165*, 1004−1012. (c) Cunningham, M. D.; Cleaveland, J.; Nadler, S. G. *Biochem. Biophys. Res. Commun.* **2003**, *300*, 403−407.

10.1021/ol034640t CCC: $25.00
Published on Web 06/17/2003

backbone-cyclic NLS peptides, in which artificial bridges were installed between backbone nitrogen atoms of the HIV-1 matrix protein NLS (Mat NLS), have been shown to inhibit nuclear import in cells.[8] Here, we report a novel class of mimetics **1**, also modeled after the NLS of the HIV-1 matrix protein, which retain only minimal peptidic character while maintaining demonstrable affinity for Kap α in vitro.

Our efforts to design NLS mimetics began before the X-ray crystal structure of Kap α in complex with an NLS peptide had been determined.[9] At that time, the HIV-1 matrix protein provided the only well-ordered NLS among known protein structures.[10] Therefore, the NMR structure of this protein, as determined by Massiah et al.,[10a] served as the basis for our design.

The Mat NLS, which spans residues 25−33, has the sequence ^{25}GKKQYKLKH^{33}; it is located in a surface-exposed loop with the four lysine side chains projecting out to solvent (Figure 1A); the tyrosine side chain (not shown) is completely buried and appears to serve as a conformational anchor. The conformation of the backbone is well ordered, while the side chains beyond C_β are disordered and presumably mobile. We sought to test the hypothesis that recognition of the NLS occurs primarily through the charged side chains by replacing the polypeptide backbone with a simple scaffold upon which aminoalkyl groups would be positioned to mimic the lysine side chains. This approach was inspired by the somatostatin and substance P mimetics developed by Hirschmann, Nicolaou, and Smith, in which suitable amino acid side chains were positioned with a β-D-glucopyranose scaffold.[11]

Examination of the average structure revealed that the four lysines of the NLS can be grouped into two close-contact pairs (Lys^{26}−Lys^{27} and Lys^{30}−Lys^{32}) in which the C_β atoms are separated by 4.2 and 5.1 Å, respectively. This spacing is closely approximated by meta substituents on an arene, and thus molecular modeling was used to superimpose two *m*-xylene units on the pairs of C_β atoms and join them with a flexible, eight-atom chain linker tracking the approximate path of the polypeptide backbone (Figure 1B). After CHarMM energy minimization,[12] the arene substituents of the resulting scaffold superimposed with an RMS deviation of 0.6 Å onto the four lysine C_β atoms. Nonetheless, the linker designed in this manner was taken only as an approximation and a starting point for experimentation.

Although this analysis is simple, we were encouraged to execute the design by the promiscuity of Kap α. Its natural ligands have no strict consensus and are flexible due to the preponderance of lysine and arginine residues. As illustrated by the work of Gerace and Nadler,[7] stereochemistry is not necessarily a critical factor, and considerable variation in the

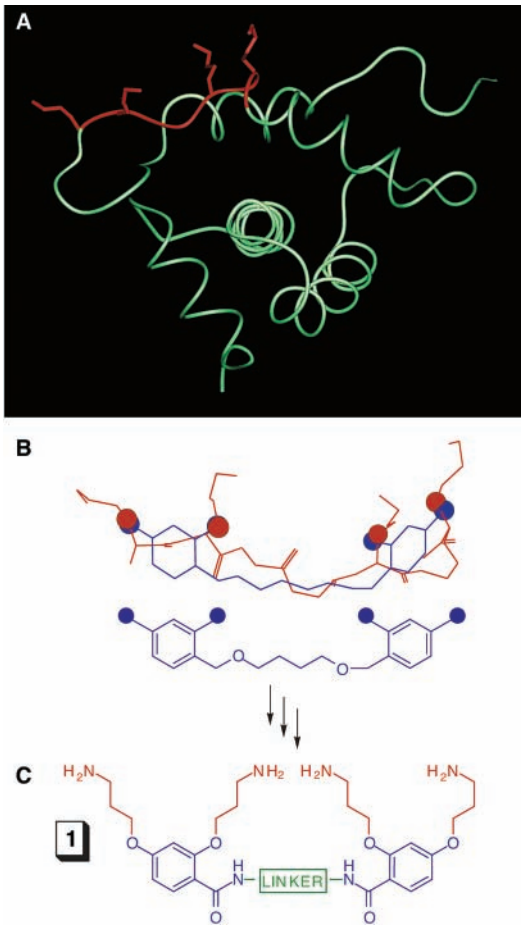


Figure 1. Design rationale for mimetics **1**. (A) Average NMR structure of the HIV-1 matrix protein[10a] showing the NLS in red. Side chains are shown only for the four lysines of the NLS. (B) Initial design of mimetics, showing superposition of the NLS (red) and a pair of linked *m*-xylene units (blue). Spheres denote C_β atoms of the lysines (protein) or corresponding arene substituents (mimetic). (C) Final design of mimetics **1**. To symplify synthesis, the diether linker was replaced with a diamide of variable length, and oxygen substituents were employed on the arene to allow installation of the side chains by alkylation.

(8) (a) Friedler, A.; Zakai, N.; Karni, O.; Broder, Y. C.; Baraz, L.; Kotler, M.; Loyter, A.; Gilon, C. *Biochemistry* **1998**, *37*, 5616−5622. (b) Hariton-Gazal, E.; Friedler, A.; Zakai, N.; Gilon, C.; Loyter, A. *Biochim. Biophys. Acta: Protein Struct. Mol. Enzymol.* **2002**, *1594*, 234−242.

(9) (a) Conti, E.; Uy, M.; Leighton, L.; Blobel, G.; Kuriyan, J. *Cell* **1998**, *94*, 193−204. (b) Kobe, B. *Nat. Struct. Biol.* **1999**, *6*, 388−397. (c) Conti, E.; Kuriyan, J. *Structure* **2000**, *8*, 329−338. (d) Fontes, M. R. M.; Teh, T.; Kobe, B. *J. Mol. Biol.* **2000**, *297*, 1183−1194.

(10) (a) Massiah, M. A.; Starich, M. R.; Paschall, C.; Summers, M. F.; Christensen, A. M.; Sundquist, W. I. *J. Mol. Biol.* **1994**, *244*, 198−223. (b) Matthews, S.; Barlow, P.; Boyd, J.; Barton, G.; Russell, R.; Mills, H.; Cunningham, M.; Meyers, N.; Burns, N.; Clark, N.; Kingsman, S.; Kingsman, A.; Campbell, I. *Nature* **1994**, *370*, 666−668.

(11) (a) Nicolaou, K. C.; Salvino, J. M.; Raynor, K.; Pietranico, S.; Reisine, T.; Freidinger, R. M.; Hirschmann, R. *Pept.: Chem., Struct. Biol., Proc. Am. Pept. Symp., 11th* **1990**, 881−884. (b) Hirschmann, R.; Nicolaou, K. C.; Pietranico, S.; Salvino, J.; Leahy, E. M.; Sprengeler, P. A.; Furst, G.; Strader, C. D.; Smith, A. B., III; et al. *J. Am. Chem. Soc.* **1992**, *114*, 9217−9218. (c) Hirschmann, R.; Nicolaou, K. C.; Pietranico, S.; Leahy, E. M.; Salvino, J.; Arison, B.; Cichy, M. A.; Spoors, P. G.; Shakespeare, W. C.; Sprengeler, P. A.; Hamley, P.; Smith, A. B. I.; Reisine, T.; Raynor, K.; Maechler, L.; Donaldson, C.; Vale, W.; Freidinger, R. M.; Cascieri, M. R.; Strader, C. D. *J. Am. Chem. Soc.* **1993**, *115*, 12550−12568. (d) Liu, J.; Underwood, D. J.; Cascieri, M. A.; Rohrer, S. P.; Cantin, L.-D.; Chicchi, G.; Smith, A. B., III; Hirschmann, R. *J. Med. Chem.* **2000**, *43*, 3827−3831. (e) Hirschmann, R.; Yao, W.; Cascieri, M. A.; Strader, C. D.; Maechler, L.; Cichy-Knight, M. A.; Hynes, J., Jr.; van Rijn, R. D.; Sprengeler, P. A.; Smith, A. B., III. *J. Med. Chem.* **1996**, *39*, 2441−2448.

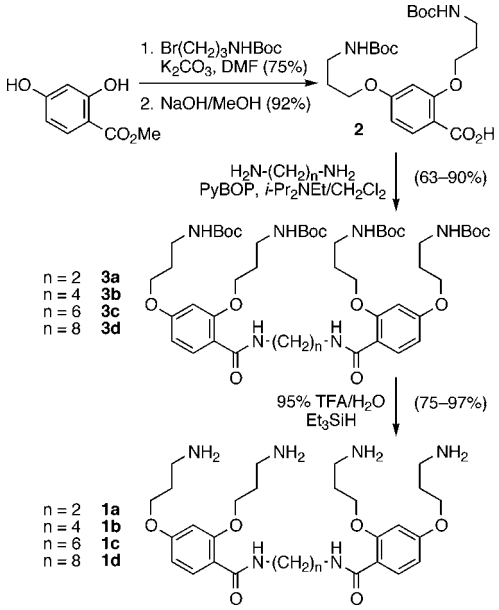
(12) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

backbone can be tolerated as well. Further, simple NLS peptides, freed from the constraints of their source protein and too short to possess well-defined secondary structures, are often fully capable of directing import.[13] Thus, a high level of conformational constraint was not expected to be necessary.

Two modifications of this original design were introduced to simplify synthesis. First, oxygen substituents on the arene would allow the aminoalkyl side chains to be introduced by simple alkylation. Second, amide bonds in the linker would allow a variety of homologues to be made from readily available resorcylic acid and α,ω-diamines. Although more conformationally restricted than the bisethers, the bisamides are still flexible, and the amides provide potential hydrogen bonding sites analogous to those in the peptide backbone, which X-ray crystal structures of Kap α in complex with NLS peptides subsequently revealed to be important.[9a,c,d] Therefore, a series of compounds of type **1** (Figure 1C) were targeted. Initially, we chose to couple the bis-aminopropylated resorcylic acid piece with C_2, C_4, C_6, and C_8 diamines, creating a range of linker lengths (6, 8, 10, and 12 atoms) that bracketed the eight-atom length of the original design.

Scheme 1 illustrates the synthetic approach to the four compounds. The key building block **2** was prepared by bis-

Scheme 1



alkylation of methyl resorcylate with $BocNH(CH_2)_3Br$ in the presence of K_2CO_3 in DMF (24 h, 75%), followed by saponification of the methyl ester in 92% yield. Coupling to the diamines using PyBOP[14] in dichloromethane proceeded smoothly to provide diamides **3a**−**d** in 63−90% yield after chromatography. Deprotection using $CF_3CO_2H/H_2O/Et_3SiH$ afforded the desired compounds **1a**−**d** in 75−97% yield; the products were >95% pure as determined by ^1H NMR and reversed-phase HPLC.

The ligands' affinities for Kap α (mouse α_2 isoform[1c]) were evaluated qualitatively using a solid-phase, competition binding assay.[15] In this method, Kap α is incubated with a ligand immobilized on TentaGel resin, and binding is detected by an enzyme-linked colorimetric assay that leaves a blue precipitate on the beads where the receptor is bound. If a soluble ligand is present, it competes for binding, thereby reducing the amount of receptor that binds the bead and diminishing color development.

Previously,[15] we had determined that the NLS from *Xenopus laevis* nucleoplasmin is well-suited for the role of immobilized ligand. Its affinity is sufficiently high to produce a positive result in the absence of competitor yet not so strong that it fails to succumb to competition from a strong NLS such as the TAg NLS (^{126}PKKKRKV^{132}). Importantly, it resists competition from weak Kap α ligands such as the K^{128}T mutant of the TAg NLS (^{126}PK*T*KRKV^{132}). These two peptides effectively mark the ends of the biologically important affinity range: the TAg NLS, which is highly active in targeting proteins to the nucleus, has a K_d of 9 nM for yeast Kap α[16] and 17 nM for mouse Kap α_2,[17] whereas the mutant, which has very weak but observable targeting activity, has a K_d of 3 μM for yeast Kap α.[16]

Using the reported procedure,[15] we compared the ability of the mimetics **1a**−**d** and the TAg NLS to compete for binding with the immobilized nucleoplasmin NLS. As shown in Figure 2 (bottom row), the TAg peptide produces a

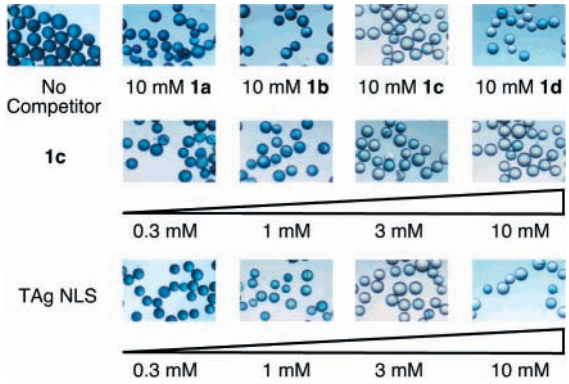


Figure 2. Competiton Binding Assay. Competition was assesed against a peptide containing the *X. laevis* nucleoplasmin NLS immobilized on TentaGel beads, which were incubated with biotinylated Kap α in the presence or absence of soluble ligands under conditions of constant ionic strength and pH as described.[15] Bound protein was detected using alkaline phosphatase-linked streptavidin and the chromogenic substrate 5-bromo-4-chloro-3-indolyl phosphate. The amount of Kap α bound to the beads is indicated by the intensity of the blue color. Soluble ligands inhibit the binding of Kap α to the beads, reducing color intensity. Top Row: Single-point determinations using no competitor (left panel) or the indicated compounds at 10 mM. Middle Row: Concentration dependence of competition by **1c**. Bottom Row: Concentration dependence of competition by a peptide (GSTPPKKKRKV) containing the TAg NLS.

(13) Goldfarb, D. S.; Gariépy, J.; Schoolnik, G.; Kornberg, R. D. *Nature* **1986**, *322*, 641−644.
(14) Coste, J.; Le-Nguyen, D.; Castro, B. *Tetrahedron Lett.* **1990**, *31*, 205−208.

concentration-dependent decrease in receptor binding over the range 0−10 mM, the highest concentration tested; the need for such high concentrations of competitor reflects the high effective concentration of the bead-bound peptide (ca. 200 mM).

Mimetics **1a**−**d** showed an interesting structure−activity relationship. The two with the shorter linkers (**1a** and **1b**, with six- and eight-atom linkers, respectively) displayed no detectable competition even at 10 mM (Figure 2, top row). Compound **1c** (10-atom linker) was a strong competitor, qualitatively comparable to the TAg NLS. It reproducibly eliminated almost all of the blue color at 10 mM and showed progressive color reduction at lower concentrations (Figure 2, middle row). Finally, mimetic **1d**, with the 12-atom linker, was intermediate, reducing the color development but not to the extent that **1c** did.

From these data, we conclude that the 10-atom linker is optimal and that compound **1c** has an affinity for Kap α comparable to those of biologically active NLS peptides. Compound **1d**, with the 12-atom linker, has an affinity that is lower but still greater than the $K^{128}T$ TAg NLS mutant (which does not compete under these conditions). Compounds **1a** and **1b** have affinities for Kap α that are comparable to or lower than the $K^{128}T$ mutant's affinity. Finally, these results show that the competition is not the result of any nonspecific effect of the compounds, as homologous compounds with equal numbers of charges have clearly distinct effects.

It is useful to consider the activity of compounds **1c**−**d** and the design target in light of the crystal structures of the TAg NLS bound to Kap α. Both yeast[9a] and mouse[9d] Kap α proteins bind the TAg NLS in an extended conformation, with the six residues ^{127}KKKRKV^{132} (P_1−P_6) being specifically bound and their side chains residing in corresponding S_1−S_6 pockets. The clearest correlation in the Mat NLS is to residues ^{26}KKQYKL^{31}. This alignment is born out in the structures, which show that the backbone of the proposed P_1−P_5 region in the Mat NLS is similarly extended and superimposes well on the TAg NLS, with the exception of P_4 (Y^{29}), where a 60° rotation of the N−C_α torsion (ϕ) buries the side chain. The two chains diverge beginning at P_6, where the backbone of the matrix protein turns and enters an α-helix. Interestingly, the structure of the matrix protein tempts speculation that unorthodox interactions might make up for the absence of positively charged side chains at P_3 and P_4, with the side chains from K^{18} and H^{33}, respectively, being potential surrogates.

This analysis predicts that K^{32} (part of the design target) is not bound by Kap α, and in that sense, the design model was flawed. Other tenets of the design have been born out, however, notably that the Mat NLS is preordered and that the resorcinol template is a good match for any pair of consecutive C_β atoms in the P_1−P_5 region of the TAg NLS. To bring all four side chains of **1** sufficiently close to use the S_1−S_5 sites, the linker must be folded, and the need for a longer linker than predicted might be due to the difficulty of making a comfortable turn in a short linker, coupled with the reduced flexibility of the amide linkers prepared in comparison to the ether linker modeled.

Compounds **1** can fill at most four of the S_1−S_6 pockets of Kap α, and compounds that fill more pockets should have a higher affinity. Efforts to determine the structures of **1c** and related compounds bound to Kap α and to explore variables such as linker structure, number and nature of side chains, and substitution pattern are underway. These efforts should provide a clearer view of how the compounds interact with Kap α and should lead to better ligands.

In the quest for improved ligands, lack of a convenient, quantitative binding assay remains an impediment. The solid-phase binding assay used in this work is reliable, can be run either in either direct mode (immobilized ligand) or competitive mode (soluble ligand), and can be run in parallel. These features make it very useful for ligand discovery, and we have successfully used both modes for this purpose. However, it is less useful for ligand refinement because it is qualitative and insensitive to small differences in affinity, so only a rough assessment of structure−activity relationships (SARs) can be made. Recently, the first rigorously determined binding constants for several NLS peptides were reported.[16,17] Although the methods employed are not directly applicable to compounds **1**, we are actively pursuing adaptations that will be.

In conclusion, we have demonstrated that simple mimetics of the NLS, with affinities for Kap α comparable to those of functional, peptidic ligands, can be prepared using linked resorcinol units as a scaffold for four lysine-like side chains. Affinity for Kap α is strongly dependent on the length of the linker, with 10 atoms being optimal, 12 atoms suboptimal, and 6 or 8 atoms completely ineffective.

Acknowledgment. We thank Dr. Wesley Sundquist for providing coordinates of the matrix protein; Mr. Ryan Riell and Ms. Alona Umali for technical assistance; the Laboratory for Biological Mass Spectrometry at Texas A&M University, and the Research Resources Center at the University of Illinois at Chicago for determining mass spectra. We thank the Robert A. Welch Foundation (A-1332), NIGMS (GM 57543), Pfizer (summer fellowship for T.H.H.), Texas A&M University, and the University of Illinois at Chicago for financial support.

Supporting Information Available: Experimental procedures with ^{1}H NMR, ^{13}C NMR, and mass spectral data for compounds **1d**, **2**, and **3d** and the methyl ester of **2**. This material is available free of charge via the Internet at http://pubs.acs.org.

OL034640T

(15) Park, S. B.; Reedy, B. M.; Standaert, R. F. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 955−956.

(16) Fanara, P.; Hodel, M. R.; Corbett, A. H.; Hodel, A. E. *J. Biol. Chem.* **2000**, *275*, 21218−21223.

(17) Catimel, B.; Teh, T.; Fontes, M. R. M.; Jennings, I. G.; Jans, D. A.; Howlett, G. J.; Nice, E. C.; Kobe, B. J. *Biol. Chem.* **2001**, *276*, 34189−34198.