Domain assembly, surface accessibility and sequence conservation in full length HIV-1 Nef

Matthias Geyer*, B. Matija Peterlin

Howard Hughes Medical Institute, University of California, San Francisco, CA 94143-0703, USA

Received 19 March 2001; revised 5 April 2001; accepted 6 April 2001

First published online 24 April 2001

Edited by Takashi Gojobori

Abstract The accessory Nef protein from human and simian immunodeficiency viruses is critical for efficient viral replication and pathogenesis. Here we present an assembly of the full length structure of HIV-1 Nef, allele NL4-3, based on the previously solved anchor and core domain structures. The center part of the 33 residue encompassing flexible loop at the C-terminus of Nef, involved in Nef internalization and CD4 endocytosis, has been modelled. The degree of sequence conservation in HIV-1 Nef proteins was determined using a total of 186 different strains from five different subtypes. The sequence conservation has been correlated with the accessible surface area and with secondary structure features for individual residues. The high amount of flexible regions in Nef accounts for the large surface and the multiple interaction sites the protein exhibits.

Key words: HIV; Nef; Structure assembly; Sequence conservation; Myristoylation

1. Introduction

Nef is a 27–35 kDa myristoylated accessory protein that is unique to the primate lentiviruses HIV-1, HIV-2 and SIV. Early experiments in rhesus macaques with SIV containing large deletions in the *nef* gene revealed that Nef is essential for maintaining optimal viral replication and the progression to AIDS and that a strong selective pressure maintains functionally intact Nef variants in the infected host [1–3]. This role of Nef as a key factor for lentiviral pathogenesis was subsequently confirmed for HIV-1 in humans [4,5].

Although the importance of Nef for the viral life cycle became clear, its function at the molecular level remains controversial (reviewed in [6]). Three functions have been attributed to Nef. First, by interacting with tyrosine and serine/threonine kinases, Nef alters cellular signalling pathways [7–9]. Second, Nef increases viral infectivity at a step after the entry of the virus into the cell [10,11]. Third, by interacting with components of the endocytic machinery, Nef decreases the expression of CD4 and major histocompatibility complex class I antigens on the surface of infected cells [12–14]. Given the high polymorphism between different *nef* alleles [15], the conservation of these genetically separable functions suggests an important role of each individual function in the viral life cycle. How these functions contribute to the importance of Nef in the infected host however still remains largely unknown and warrants further investigation.

Proteolytic experiments revealed that Nef adopts a two domain structure consisting of an N-terminal membrane anchor region and a well folded C-terminal core domain [16]. Based on this observation the structure of Nef has been determined for the core domain by nuclear magnetic resonance (NMR) spectroscopy [17] and by X-ray crystallography both in complex with the SH3 domain of the Fyn tyrosine kinase [18,19] and alone [19]. The structures are in good agreement with each other but miss a flexible loop at the C-terminus, reaching at least from residue 159 to 173. Recently, also the N-terminal membrane anchor domain structure has been solved in its myristoylated and non-myristoylated forms showing a flexible polypeptide chain with two helical structure elements [20]. These two protein fragments now render the possibility to get insights into the full length Nef protein.

Here we present the assembly of the structural domains of HIV-1 Nef and calculate the surface accessibility of individual residues. The structural features were correlated with the degree of sequence conservation based on the evaluation of 186 different strains from HIV-1 Nef and discussed for their implications on the molecular functions of Nef.

2. Materials and methods

2.1. Structure assembly and model calculation

The full length Nef (2–210) structure, strain NL4-3, has been assembled based on the structures of the myristoylated Nef anchor domain (myr2–57) ([20], PDB accession no.: 1QA5, model 1) and the truncated core domain (56–206, Δ159–172) ([21], 2NEF, model 1). To create a template file for molecular modelling of the flexible loop two additional residues at both sides of the Nef deletion site were taken out. Into this template the central region (amino acids 157–174) of the C-terminal flexible loop has been modelled using the Swiss-Model approach for automated comparative protein modelling [22]. Additional energy minimization methods have been applied using the program X-PLOR [23] in order to harmonize covalent geometry values. NMR experiments were performed on a Bruker DMX-600 spectrometer at 308 K. Highly purified myristic acid (Boehringer Mannheim) was added in increasing amounts up to 0.4 mM concentration to 0.3 mM ^{15}N-labelled Nef (40–206), solved in 10 mM Tris buffer, pH 7.85. Figures were generated with MOLSCRIPT [24]. Coordinates of the full length NL4-3 Nef model can be requested from M.G., E-mail: geyer@mpimf-heidelberg.mpg.de.

2.2. Surface accessibility and sequence conservation

The atomic accessible surface has been calculated using the program NACCESS (Hubbard, S.J. and Thornton, J.M. (1993) 'NACCESS', computer program, Department of Biochemistry and Molec-

*Corresponding author. UCSF Cancer Research Institute, Rm N231, HHMI Peterlin Lab, P.O. Box 0703, 2340 Sutter St., San Francisco, CA 94115, USA. Fax: (1)-415-502 1901.
E-mail: mgeyer@cc.ucsf.edu

0014-5793/01/$20.00
PII: S0014-5793(01)02394-8

ular Biology, University College London). The calculation has been performed with default van der Waals radii, including all hydrogen atoms, and a default probe size of 1.4 Å. Separation in main chain and side chain values and their relative accessibility for individual residues are based on model parameters given by Hubbard et al. [25].

Sequences of HIV-1 Nef alleles were taken from the HIV Sequence Database at the Los Alamos National Laboratory (http://hiv-web.lanl.gov/) and two studies about sequence variations published by Shugars et al. and Kirchhoff et al. [26,27]. A total of 186 sequences have been evaluated which neither exhibit a premature stop codon nor large deletion sites. The sequences were aligned with the Multalin software package (http://www.toulouse.inra.fr/) and clustered into groups of similarity for each amino acid position.

3. Results and discussion

3.1. Full length structure assembly

The most informative result towards the structure determination of Nef were proteolytic digestion experiments that revealed the modular organization of HIV-1 Nef [16]. The observation that almost one third of the protein is structurally less defined explained the oligomeric character of the protein for unspecific aggregation and the difficulties in early crystallization attempts. Subsequently, the structures of the two separable parts of Nef, a flexible myristoylated membrane anchoring region and a stable core domain, were determined using NMR spectroscopy and X-ray crystallography [17–20]. Since the experimental determination of a full length Nef structure is unlikely unless one could overcome the unspecific oligomerization properties, e.g. by complexation with a cofactor that stabilizes the flexible parts, we decided to assemble the structural fragments and model the C-terminal flexible loop which is missing in all structures. In the following the amino acid notation is based on the sequencing numbers of the NL4-3 Nef strain.

In order to assemble the two regions we first studied the effects of myristate on the structure of the core domain. NMR titration experiments were performed with free myristic acid added in increasing amounts to ^{15}N-labelled core Nef (40–206). No changes in the characteristic pattern of chemical shifts of the Nef protein were detected, indicating no interaction between the two molecules. This suggests that the myristic moiety in the myristoylated Nef protein is not hidden in the structure, but rather exposed to the solvent e.g. to direct the protein to the membrane. This result allowed us to assemble the structures of the two protein fragments with a free accessible N-terminus.

We covalently linked the myristoylated N-terminal part of Nef (myr2–57) to the core domain of Nef (56–206, Δ159–172) at the peptide bond between residue 56 and 57 preserving the dihedral angle ω. As template for the core domain we took the NMR-derived structure [21], which is most complete at its N-terminus while the two crystal structures of Nef in complex with Fyn [18,19] or Nef alone [19] start at residues 71 or 74, respectively. Next, the center part (residues 157–174) of the 33 amino acids encompassing flexible loop (148–180) was modelled to the structure and energy-minimized as described in Section 2. The full length Nef structure consists of six α-helices (α1–α6) and a β-pleated sheet of five antiparallel β-strands (β1–β5) as shown in Fig. 1A. The Nef structure is partitioned in four units: a flexible myristoylated membrane anchoring region of variable length (1–56), followed by the PxxP loop (57–80), the core domain (81–206, Δ148–180) and a C-terminal flexible loop (148–180; Fig. 1B). While the anchor and the C-terminal flexible loop are not restrained within their conformation relative to the core domain, the PxxP loop is loosely linked to the core by hydrophobic interactions.

3.2. Surface accessibility of NL4-3 Nef

The calculation of the overall protein surface of the full length NL4-3 Nef model structure reveals an accessible surface area of 17 600 Å^2. For a protein of 206 amino acid size this can be considered a very large solvent-exposed surface area. For comparison the well folded nuclear import protein

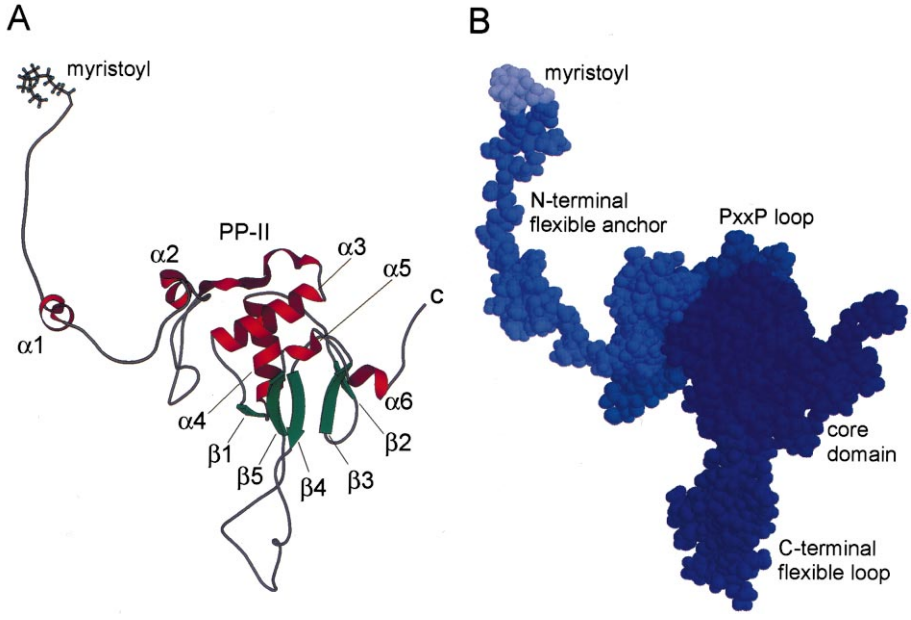

Fig. 1. Structure model of the full length, myristoylated HIV-1 Nef protein, strain NL4-3. (A) Ribbon presentation. The structure has been assembled from solution structures of the myristoylated Nef anchor domain (myr2–57; [20]) and the Nef core domain (56–206$^{\Delta 159-173}$, [21]). Coordinates of the central region (residues 157–175) of the C-terminal flexible loop have been modelled. (B) Molecular surface presentation. Three flexible regions within the Nef structure are highlighted: the myristoylated anchor domain (myr2–56, light blue), the PxxP loop (57–80, middle blue) and the C-terminal flexible loop (148–180, dark blue). The N-terminal myristoyl moiety is labelled.

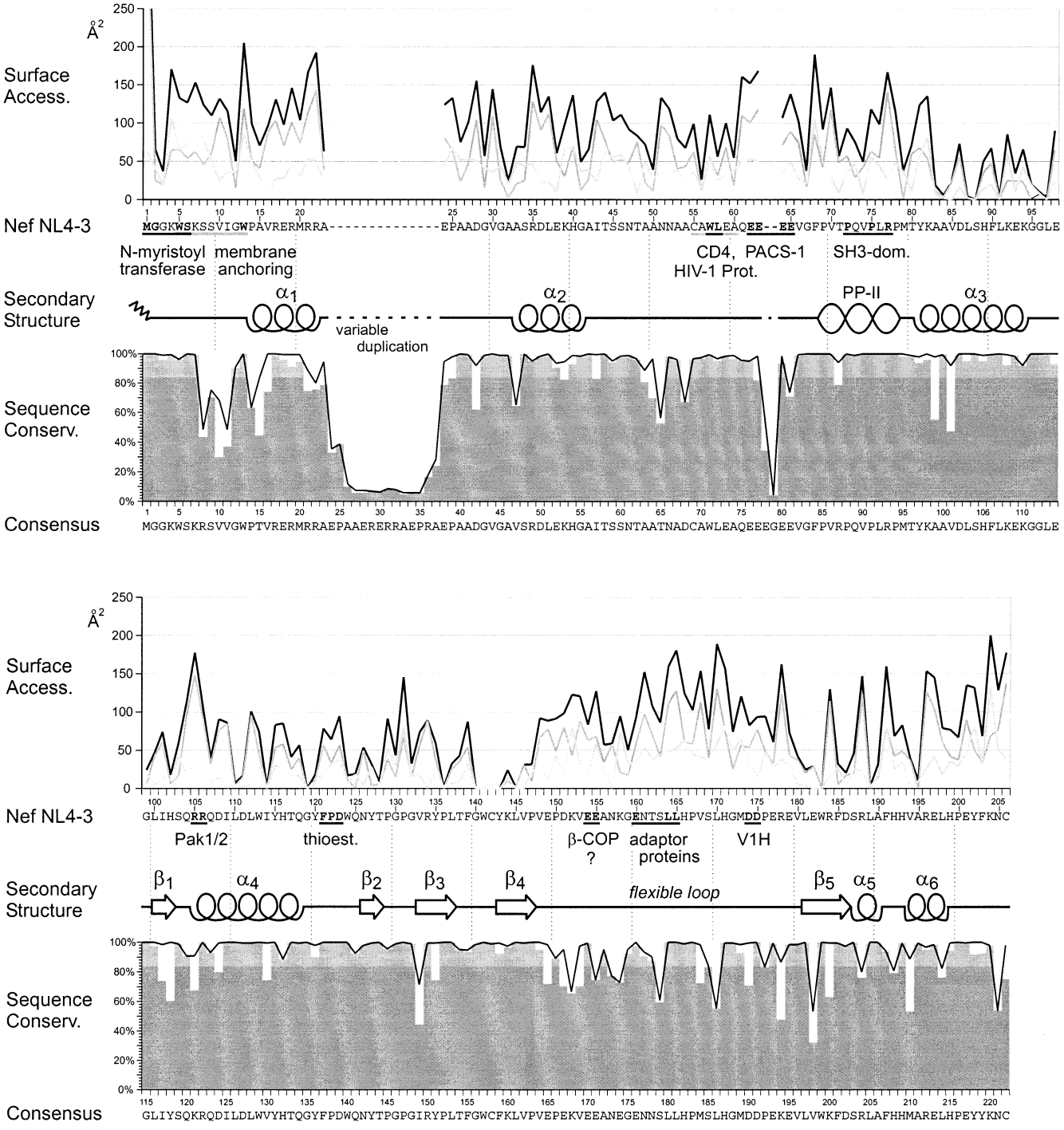

Fig. 2. Sequence conservation, secondary structure and solvent accessibility in HIV-1 Nef. (Top) Solvent accessibility area calculated for residues of the assembled NL4-3 Nef structure. The three lines indicate main chain (light gray) and side chain (gray) contributions and the total (solid line). (Middle) Secondary structure elements of Nef, displayed below the amino acid sequence of NL4-3 Nef. (Bottom) Degree of sequence conservation in HIV-1 Nef proteins determined from a total of 186 strains. Shaded blocks display the conservation of amino acid identity (average degree of conservation: 84%), the solid line indicates residue similarities. The HIV-1 Nef consensus sequence as derived from the alignment is indicated below. Sequential motifs and their respective protein interaction sites previously described are labelled at the NL4-3 sequence.

Ran, whose structure comprises the exact same number of amino acids, exhibits an accessible surface of just 9950 Å^2 as calculated here. The gain in surface area is due to the high degree of less folded, flexible regions in Nef leading to more solvent exposure. Specifically, the myristoylated anchor domain (myr2–56) encompasses 6160 Å^2 and the C-terminal flexible loop (148–180) 3470 Å^2 which together account for 55% of the entire accessible surface. The PxxP loop (57–80) as the third flexible unit contributes with another 2500 Å^2 to the overall Nef structure. Thus, in the most comprehensive view, the 93 residues of the well folded core structure (81–206, Δ148–180) form only 31% of the molecular surface.

A detailed plot of the solvent accessibility of each residue including main and side chain contribution is shown in Fig. 2. Overall, the N-terminal residues up to A83 appear very accessible, which is the first turn of helix α3 starting the core domain. It follows an alternating section until the flexible loop (V148–V180) again is homogeneously exposed. Towards the end, the C-terminal 11 residues become increasingly accessible. Only few amino acids are completely hidden in the structure as are e.g. the large aromatic side chains of W141 and W183. Both form the scaffold of the protein and assemble helices α3 and α4 to the β-pleated sheet. To mention only a few exposed residues within the core structure, the hydrophilic basic side chains of R105, R184, R188 and R196 are characteristic. Overall, the large accessible surface in combination with the high portion of flexible regions may account for the numerous protein interactions that have been reported for Nef.

It is important to note that for a protein as flexible as Nef, the total accessible surface area may strongly vary upon different conformations and that molecular rearrangements may occur upon complex formation. However, the full length Nef model structure represents one feasible conformation and emphasizes in particular the hidden residues.

3.3. Sequence conservation in HIV-1 Nef

The degree of sequence conservation in HIV-1 Nef proteins has been calculated using a total of 186 different strains from five different subtypes (see Section 2). The degree of sequence identity and sequence similarity for every individual position is displayed in Fig. 2, lower panel. To our surprise, Nef sequence conservation was unexpectedly high with an average sequence identity of 84% and an average similarity of 89%.

There are two main regions where sequence insertions relative to the NL4-3 allele occur. A variable duplication is found in about 36% of all sequences at position 23, following a positively charged arginine cluster (17–22). This insert is of variable length up to 14 residues and resembles the motif RxRRAEP. Most likely it will mimic the α-helical conformation of the preceding helix α1. The second region is a rare insert (in 5% of all sequences) of E and/or G residues within a cluster of glutamic acids (62–65) in the PxxP loop. Single amino acid deletions are found predominantly at three locations: From position 8 to 11 (5%), 49 to 51 (7%) and 155 to 162 (3%). These locations correlate with lower sequence conservation and structurally less defined loop regions. The latter one is of particular interest since it only occurs as deletion of six residues in combination with a mutation of a charged residue. The sequence EEANKGENTSLL$_{(154)}$ mutates to E--NN----SLL which preserves the endocytic ExxxLL motif (see e.g. strain B_NEFD).

The most homogeneously conserved part of Nef ranges from V66 to V148 which includes the polyproline helix and the β-strands. There are few examples where typical features of secondary structure elements correlate with sequence conservation. At the first turn of helix α4 the solvent-exposed residues R105 and D108 are less conserved while R106 and Q107 point towards the structural core and are highly conserved. In strand β5, E182 and R184 are on the surface and more variable while W183 is completely hidden and conserved.

3.4. Implications for molecular functions

Our analysis shows that the myristoylation site, the PxxP site for SH3 domain interaction, and the dileucine motif for internalization are the motifs with the highest degree of conservation in Nef (Fig. 2). This scheme may help to identify

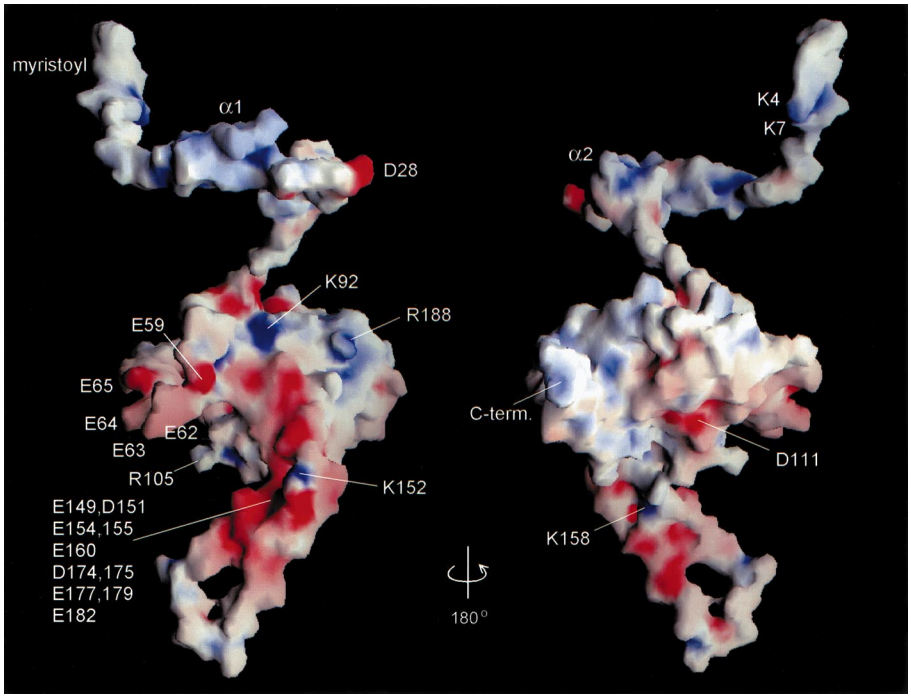

Fig. 3. Surface electrostatic potential display of NL4-3 Nef. Some characteristic charges are labelled. Note the positively charged N-terminus and the negatively charged C-terminal flexible loop creating a dipolar character of the protein. The surface was calculated using a water-probe radius of 1.4 Å, and the electrostatic potential is displayed at a scale of -18 k_BT (red) to $+18$ k_BT (blue). The figure was generated with the program GRASP [36].

structural requirements for protein formation as well as conserved exposed protein interaction sites for future binding studies.

In addition to that, display of the electrostatic surface provides information about charged areas that may form potential interaction sites (Fig. 3). Very remarkable is the strong negatively charged cluster formed by eight acidic residues at the basis of the C-terminal flexible loop (E149–E155 and D174–E179). These acidic amino acids at both ends of the flexible loop are opposite to each other, thereby creating a repulsive force that might be required to prevent formation of a regular secondary structure element but result in a flexible loop. Also by that, the protein adopts a dipolar character with a positively charged N-terminus formed by K4, K7 and the helices α1 and α2, and a negatively charged flexible loop, opposing each other. Other characteristic positive charges are located at the C-terminus, namely R184, R188, R196 and K206 supported by the three histidines 192, 193, 199. Arginine 184 has been proposed to be involved in binding of Nef to RNA stem-loop structures (Pierre Boulanger, personal communication) which correlates with the high accessibility of these residues (Fig. 2).

Besides the known interacting proteins, two issues in Nef modifications deserve further considerations. A recent study suggests the dimerization of Nef, driven by electrostatic interactions of two charged residues, R105 and D123 [28]. The oligomerization properties of Nef in cells were described earlier [29] and backbone dynamic measurements determined the dimerization constant to 0.6–0.7 mM [21]. Analysis of crystallographic models provides a molecular basis for dimer and trimer formation of the Nef core domain which indeed shows helix α4 and the flanking loop required for this interaction [30]. Another Nef modification to further investigate is phosphorylation. Nef has been demonstrated to be phosphorylated on serine and to a lesser degree on threonine residues [31–33]. Threonine 15 was identified as one possible phosphorylation site, this residue is however not well conserved and the biological relevance of this modification is thus questionable. Phosphorylated serines were roughly mapped between amino acids 1–35 and 36–86, respectively, however await further functional characterization. Since phosphorylation is an attractive mechanism for the regulation of the pleotropic functions of Nef, this topic remains an important area of research.

Interestingly, Nef contains a second PxxP motif that is solvent accessible located at the beginning of the C-terminal flexible loop (P147, P150) opening the speculation of another SH3 domain binding site [34]. This sequence element does not form the structural polyproline helix and also the SH3 binding motif requirements PxφPxR are only partially fulfilled with a predominance of glutamic or aspartic acid at position 149 (98%) and the arginine homologue lysine residue at positions 152 (66%). Nevertheless, the accessibility and just the sequential divergence might make this motif a suitable target for small N-substituted peptoids [35] adding additional specificity to possible inhibitors.

Acknowledgements: We thank Hans Robert Kalbitzer for discussions, Giovanna Musco for NMR experiments and Oliver Fackler and Stephan Grzesiek for discussions and critical reading of the manuscript. M.G. acknowledges support by an EMBO fellowship.

References

[1] Kestler, H.W., Ringler, D.J., Mori, K., Panicali, D.L., Sehgal, P.K., Daniel, M.D. and Desrosiers, R.C. (1991) Cell 65, 651–662.
[2] Sawai, E.T., Khan, I.H., Montbriand, P.M., Peterlin, B.M., Cheng-Mayer, C. and Luciw, P.A. (1996) Curr. Biol. 6, 1519–1527.
[3] Khan, I.H., Sawai, E.T., Antonio, E., Weber, C.J., Mandell, C.P., Montbriand, P. and Luciw, P.A. (1998) J. Virol. 72, 5820–5830.
[4] Deacon, N.J. et al. (1995) Science 270, 988–991.
[5] Kirchhoff, F., Greenough, T.C., Brettler, D.B., Sullivan, J.L. and Desrosiers, R.C. (1995) N. Engl. J. Med. 332, 228–232.
[6] Cullen, B.R. (1998) Cell 93, 685–692.
[7] Baur, A.S., Sawai, E.T., Dazin, P., Fantl, W.J., Cheng-Mayer, C. and Peterlin, B.M. (1994) Immunity 1, 373–384.
[8] Iafrate, A.J., Bronson, S. and Skowronski, J. (1997) EMBO J. 16, 673–684.
[9] Hanna, Z., Kay, D.G., Rebai, N., Guimond, A., Jothy, S. and Jolicoeur, P. (1998) Cell 95, 163–175.
[10] Aiken, C. and Trono, D. (1995) J. Virol. 69, 5048–5056.
[11] Schwartz, O., Marechal, V., Danos, O. and Heard, J.M. (1995) J. Virol. 69, 4053–4059.
[12] Garcia, J.V. and Miller, A.D. (1991) Nature 350, 508–511.
[13] Aiken, C., Konner, J., Landau, N.R., Lenburg, M.E. and Trono, D. (1994) Cell 76, 853–864.
[14] Schwartz, O., Marechal, V., Le Gall, S., Lemonnier, F. and Heard, J.M. (1996) Nat. Med. 2, 338–342.
[15] Delassus, S., Cheynier, R. and Wain-Hobson, S. (1991) J. Virol. 65, 225–231.
[16] Freund, J., Kellner, R., Houthaeve, T. and Kalbitzer, H.R. (1994) Eur. J. Biochem. 221, 811–819.
[17] Grzesiek, S., Bax, A., Clore, G.M., Gronenborn, A.M., Hu, J.-S., Kaufman, J., Palmer, I., Stahl, S.J. and Wingfield, P.T. (1996) Nat. Struct. Biol. 3, 340–345.
[18] Lee, C.-H., Saksela, K., Mirza, U.A., Chait, B.T. and Kuriyan, J. (1996) Cell 85, 931–942.
[19] Arold, S., Franken, P., Strub, M.-P., Hoh, F., Benichou, S., Benarous, R. and Dumas, C. (1997) Structure 5, 1361–1372.
[20] Geyer, M., Munte, C.E., Schorr, J., Kellner, R. and Kalbitzer, H.R. (1999) J. Mol. Biol. 289, 123–138.
[21] Grzesiek, S., Bax, A., Hu, J.-S., Kaufman, J., Palmer, I., Stahl, S.J., Tjandra, N. and Wingfield, P.T. (1997) Protein Sci. 6, 1248–1263.
[22] Peitsch, M.C. (1996) Biochem. Soc. Trans. 24, 274–279.
[23] Brünger, A.T. (1992) X-PLOR Version 3.1: A system for X-ray crystallography and NMR, Yale University Press, New Haven, CT.
[24] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.
[25] Hubbard, S.J., Campbell, S.F. and Thornton, J.M. (1991) J. Mol. Biol. 220, 507–530.
[26] Shugars, D.C., Smith, M.S., Glueck, D.H., Nantermet, P.V., Seilliert-Moiseiwitsch, F. and Swanstrom, R. (1993) J. Virol. 67, 4639–4650.
[27] Kirchhoff, F., Easterbrook, P.J., Douglas, N., Troop, M., Greenough, T.C., Weber, J., Carl, S., Sullivan, J.L. and Daniels, R.S. (1999) J. Virol. 73, 5497–5508.
[28] Liu, L.X. et al. (2000) J. Virol. 74, 5310–5319.
[29] Kienzle, N., Freund, J., Kalbitzer, H.R. and Mueller-Lantzsch, N. (1993) Eur. J. Biochem. 214, 451–457.
[30] Arold, S., Hoh, F., Domergue, S., Birck, C., Delsuc, M.-A., Jullien, M. and Dumas, C. (2000) Protein Sci. 9, 1137–1148.
[31] Bandres, J.C., Luria, S. and Ratner, L. (1994) Virology 201, 157–161.
[32] Bodeus, M., Marie-Cardine, A., Bougeret, C., Ramos-Morales, F. and Benarous, R. (1995) J. Gen. Virol. 76, 1337–1344.
[33] Coates, K., Cooke, S.J., Mann, D.A. and Harris, M.P.G. (1997) J. Biol. Chem. 272, 12289–12294.
[34] Saksela, K., Cheng, G. and Baltimore, D. (1995) EMBO J. 14, 484–491.
[35] Nguyen, J.T., Turck, C.W., Cohen, F.E., Zuckermann, R.N. and Lim, W.A. (1998) Science 282, 2088–2092.
[36] Nicholls, A., Sharp, K.A. and Honig, B. (1991) Proteins 11, 281–296.