J Comput Aided Mol Des (2007) 21:121–130
DOI 10.1007/s10822-007-9107-1

ORIGINAL PAPER

Design and characterization of an engineered gp41 protein from human immunodeficiency virus-1 as a tool for drug discovery

Kent D. Stewart · Kevin Steffy · Kevin Harris · John E. Harlan · Vincent S. Stoll · Jeffrey R. Huth · Karl A. Walter · Emily Gramling-Evans · Renaldo R. Mendoza · Jean M. Severin · Paul L. Richardson · Leo W. Barrett · Edmund D. Matayoshi · Kerry M. Swift · Stephen F. Betz · Steve W. Muchmore · Dale J. Kempf · Akhter Molla

Received: 3 November 2006 / Accepted: 9 January 2007 / Published online: 9 February 2007

Abstract Two new proteins of approximately 70 amino acids in length, corresponding to an unnaturally-linked N- and C-helix of the ectodomain of the gp41 protein from the human immunodeficiency virus (HIV) type 1, were designed and characterized. A designed tripeptide links the C-terminus of the C-helix with the N-terminus of the N-helix in a circular permutation so that the C-helix precedes the N-helix in sequence. In addition to the artificial peptide linkage, the C-helix is truncated at its N-terminus to expose a region of the N-helix known as the ''Trp-Trp-Ile'' binding pocket. Sedimentation, crystallographic, and nuclear magnetic resonance studies confirmed that the protein had the desired trimeric structure with an unoccupied binding site. Spectroscopic and centrifugation studies demonstrated that the engineered protein had ligand binding characteristics similar to previously reported constructs. Unlike previous constructs which expose additional, shallow, non-conserved, and undesired binding pockets, only the single deep and conserved Trp-Trp-Ile pocket is exposed in the proteins of this study. This engineered version of gp41 protein will be potentially useful in research programs aimed at discovery of new drugs for therapy of HIV-infection in humans.

Keywords Protein engineering · Protein design · Human immunodeficiency virus (HIV) · gp41 Protein

K. D. Stewart (✉)
Department of Structural Biology, Abbott Laboratories, Building AP10, 100 Abbott Park Road, Abbott Park, IL 60064, USA
e-mail: kent.d.stewart@abbott.com

K. D. Stewart · J. E. Harlan · V. S. Stoll · J. R. Huth · K. A. Walter · E. Gramling-Evans · R. R. Mendoza · J. M. Severin · P. L. Richardson · L. W. Barrett · E. D. Matayoshi · K. M. Swift · S. F. Betz · S. W. Muchmore
Department of Advanced Technologies, Abbott Laboratories, Abbott Park, IL, USA

K. Steffy · K. Harris · D. J. Kempf · A. Molla
Department of Anti-Infectives, Abbott Laboratories, Abbott Park, IL, USA

Introduction

During viral entry into the host T-cell, the transmembrane gp41 protein on the surface of the human immunodeficiency virus (HIV) virus plays an essential role in the fusion of the viral membrane with the host plasma membrane allowing the invasion of the viral components. The gp41 protein also mediates syncytium formation that occurs when infected cells fuse with neighboring uninfected cells. While the protein structural details of these processes are not well understood, the structure of a stable ''fusion active'' state of the gp41 ectodomain is believed to be a six-helix bundle of three protein molecules in which the N- and C-terminal helices, each of approximately 40 residues, are arranged into three hairpins, linked by a flexible peptide of approximately 40 residues [1, 2]. Agents that interact with gp41 and interfere with the formation of the 6-helix bundle inhibit HIV fusion in vitro and block HIV replication in vitro [3]. The prototype gp41 inhibitor, enfuvirtide (also known as T-20 or DP178), a 36-residue peptide analog of the C-helix of gp41, has recently been approved for the treatment of HIV infection. Although it is a potent antiretroviral agent, the utility of enfuvirtide is limited

by its subcutaneous mode of administration and high costs, due to its molecular size and complexity. Attention has been drawn to an evolutionarily-conserved binding pocket comprised of two N-helix units termed the ''Trp-Trp-Ile'' pocket for the three amino acids that project inwardly from the C-helix to fill this hydrophobic cavity [4–6]. This pocket represents a putative binding site for small organic molecules with potential for use in antiretroviral therapy, and efforts to discover specific ligands for this pocket have been reported [7–10]. A report describing the discovery of a small organic ligand for a related pocket of the F1 protein of respiratory syncytial virus has appeared [11].

The discovery and evaluation of small molecule ligands for the Trp-Trp-Ile pocket has been hampered, in part, by the lack of a convenient protein form of gp41 for use in screening and assay development. The available hexameric fusion-active forms of the protein do not have any well-defined solvent-exposed binding sites as the six helices pack together tightly. The three symmetry-related Trp-Trp-Ile pockets are fully occupied by residues from the C-helices. Isolated N-helix peptides form a trimeric structure with three external hydrophobic grooves that promote protein aggregation and make ligand binding studies problematic. A natural and stabilized form of gp41 in a pre-hairpin conformation has not yet been characterized. We report here the rational design of a protein-engineered form of gp41 that possesses a hexameric tertiary structure and an unoccupied Trp-Trp-Ile binding pocket. The protein's stability and solubility are sufficient for establishing ligand binding assays and permitting three-dimensional structural characterization.

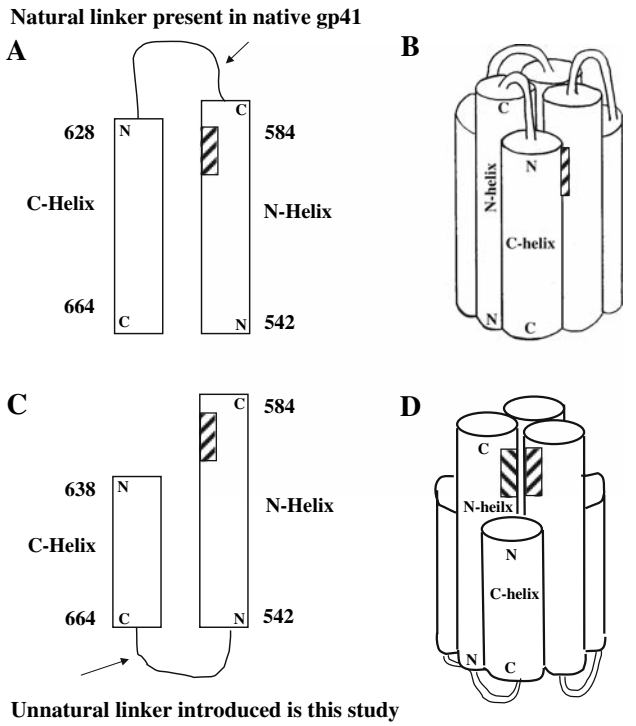

Fig. 1 Protein sequence diagram of gp41-related proteins: (**A**) The natural ectodomain of gp41 protein has a helical N-terminal domain from residues 542–584, a 44-residue linker, and a helical C-terminal domain from residues 628–664. The cross-hatched area represents the region of the N-terminal domain that forms the ''Trp-Trp-Ile'' pocket. (**B**) 6-helix quaternary structure diagram of the naturally occurring gp41 protein. The very small cross-hatched section shows that the C-helix fills the ''Trp-Trp-Ile'' pocket within the N-helical domain. (**C**) The engineered protein of this work begins with a helical domain that corresponds to residues 638–664 of the C-helix, a short 3-residue linker, and a terminal helical domain that corresponds to residues 542–584 of the N-helix. (**D**) 6-helix quaternary structure diagram of the engineered protein of this study. The cross-hatched area represents the solvent exposed region of the ''Trp-Trp-Ile'' pocket. There are three of these pockets per trimer of protein

Material and methods

Structural design

The general design strategy is illustrated in Fig. 1. The natural form of gp41 possesses a linker between the C-terminus of the N-helix and the N-terminus of the C-helix. This linker was deleted, and an unnatural linker inserted between the C-terminus of the C-helix and the N-terminus of the N-helix. This permitted a shortening of the C-helix by 10 residues. Since there are no longer any C-helix residues that occupy the ''Trp-Trp-Ile'' pocket that is formed by the N-helix, the ''Trp-Trp-Ile'' pocket is intramolecularly unoccupied.

For design of the unnatural linker, the software package Insight II (Accelrys, Inc., San Diego, CA) was used to visually inspect the coordinates of gp41 6-helix bundle (PDB entry 1AIK). Computational loop searching (Homology module, Insight II software) suggested that 4 ± 1 amino acids would suffice to link Asp 664 at the C-terminus of the C-helix with Arg542 at the N-terminus of the N-helix, as shown in Fig. 2. The sequence of designed Protein 1 is shown in Fig. 3 and was designed to have a tripeptide linker, Gly-Asp-Gly. Three additional changes were made: (1) Arg 542

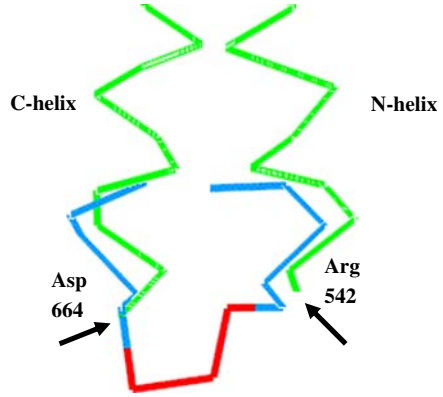

Fig. 2 Loop search: The alpha-carbon trace of the template protein (green) is shown along with the C-terminal Asp 664 of the C-helix and the N-terminal Arg 542 of the N-helix (PDB entry 1AIK). The 13-residue peptide resulting from the loop search is shown in blue and red. The residues required for overlay with the template (five residues at the start and end of the 13-mer) are shown in blue. The three residues of the linking peptide are shown in red. Similar overlays were observed for linker lengths of 4 or 5 residues

was mutated to helix-initiation residue Pro. (2) To expose the ''Trp-Trp-Ile'' pocket, the C-terminal helix was truncated so that it began with Tyr 638 (10 residues different from the Trp 628 used in the protein leading to PDB entry 1AIK). (3) This Tyr 638 residue was mutated to Met for cloning purposes. After processing of the initiator Met residue, Protein 1 would begin with a Thr residue corresponding to Thr 639 of the C-helix. Protein 2 was designed analogously to Protein 1 with the following three alterations: (1) a tetrapeptide GGGG-linker was used. (2) No mutations corresponding to either Arg 542 or Tyr 628 were made. (3) An enterokinase cleavage sequence was engineered into the N-terminus permitting residue 638 to be the natural Tyr amino acid.

Construction of expression plasmids

The nucleotide sequences encoding Protein 1 and Protein 2 are shown in Fig. 3. The expression construct for Protein 2 was generated using four overlapping oligonucleotides (5′-TACACAAGCTTG ATCCACTCTCTGATCGAAGAAAGCCAGAACC AGCAGGAAAAAAACGAACAGG-3′; 5′CCAGA CAGAAGCTGACGACCACCACCACCGTCCAGT TCTAGAAGTTCCTGTTCGTTTTTTTCCTGCTG G-3′; 5′-GGTCGTCAGCTTCTGTCTGGTATCGTT CAGCAGCAGAACAATCTGCTGCGTGCTATCG AAGCTCAGCAGCATC-3′; 5′-TTCAACAGCCAG GATACGAGCCTGAAGCTGTTTGATACCCCAA ACGGTCAGTTGCAGCAGATGCTGCTGAGCTT CGATAGCACG-3′). The oligos were phosphorylated with T4 polynucleotide kinase, annealed together, and treated with T4 DNA polymerase to fill in the single-stranded gaps followed by T4 DNA ligase [all enzymes from Invitrogen (Carlsbad CA); used according to manufacturers recommendations]. The resulting DNA molecule was used as a template for PCR-based cloning into the expression vector pET-19b (Novagen, Madison, WI). An upstream *Nco*I and downstream *Bam*HI restriction site were introduced during a PCR step utilizing the primers (5′-CATGCCATGGGCC ATCATCATCATCATCACAGCAGCGGCCATAT CGACGACGACGACAAGTACACAAGCTTGAT CCACTCTCTGATCGAAGAAAGC-3′) and (5′-CG CGGATCCTTATTCAACAGCCAGGATACGAGC C-3′). The PCR product was subcloned into the *Nco*I/*Bam*HI sites of pET-19b to generate the Protein 2 expression construct. The Protein 1 expression construct was generated in analogous fashion. In both constructs, expression of HIV-gp41 is driven by the T7lac promoter.

Expression

The Protein 1 and Protein 2 expression constructs were transformed into *E. coli* BL21(DE3). For expression of unlabelled protein, cells were grown in LB at 37 °C until reaching an OD of 0.6, at which point expression from the T7lac promoter was induced by addition of 0.1 mM IPTG. Cells were harvested 2 h post-induction and frozen at –85 °C until purification. For NMR studies, labeled Protein 1 was expressed as follows. Cells were grown at 30 °C in minimal media containing glucose as a sole carbon source. The culture was exposed to 50 mg/L [3,-^{13}C]-α-ketobutyrate and 100 mg/L [3,3'-^{13}C]-α-ketoisovalerate 15 min prior to induction to incorporate ^{13}C at the methyl groups of leucine and valine, and the δCH_3 groups of isoleucine [12]. When the culture reached an OD of 1.9, expression was induced by the addition of 1 mM IPTG. Cells were harvested 4 h post-induction.

Purification of non-labeled protein

Five grams of cells were suspended in lysis buffer (20 mM Tris, pH 8.0 at 4 °C, 100 mM NaCl, 5% glycerol, 1 mM EDTA, 2 mM 4-(2-aminoethyl)benzenesulfonylfluoride, 1 mM sodium azide) at a ratio of 5 mL/gm of cell paste and disrupted by using a French Press. The lysate was clarified by centrifugation (Beckman JA20 for 30 min at 17,000 rpm). The resulting supernatant was treated by addition of

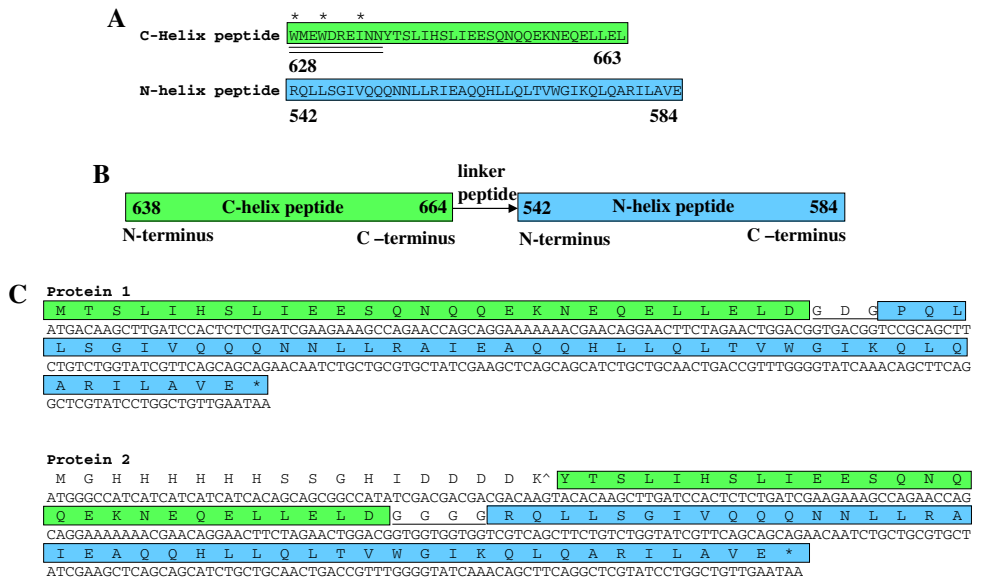

Fig. 3 (**A**) Sequences of C-helix and N-helix peptides from previous studies [1] are given for comparison purposes. The ''Trp-Trp-Ile'' residues, Trp 628, Trp 631 and Ile 635, are indicated with asterisks. In the C-helix peptide, the region that is truncated in proteins 1 and 2 is double underlined. Residue numbering refers to the natural position within full-length gp160 envelope glycoprotein. The green and blue background coloring to indicate the C-helix and N-helix regions, respectively, is retained throughout Fig. 3. (**B**) Diagram showing the designed ''reverse'' order of peptide sections used in Proteins 1 and 2. (**C**) Nucleotide and corresponding amino acid sequences of Proteins 1 and 2. The tri- and tetrapeptide linkers within proteins 1 and 2 are underlined. Protein 2 differs from Protein 1 in having one additional amino acid in the linker region, a His-tag, and enterokinase cleavage position (indicated with ^ symbol)

ammonium sulfate to 40% saturation and stirred for 30 min at 4 °C. The resulting precipitate was dissolved in 10 mL lysis buffer and dialyzed overnight versus the lysis buffer. The resulting sample was then loaded on to an UNO Q12 (Bio-Rad, Hercules, CA) 1.5 × 6.8 cm column, and eluted with a linear gradient to 1 M NaCl over 10 column volumes. The Q-pool was then dialyzed overnight at 4 °C vs. 20 mM MES, 100 mM NaCl, 1 mM EDTA, 5% glycerol, pH 6.0 in order to acid precipitate the protein. The resultant pellet was dissolved in 20 mM Hepes, 1 mM sodium azide at pH 7.6. The protein was then concentrated to 7 mg/mL. This protein had a persistent yellow color that could be removed by a final centrifugation at 450,000*g* (TLA-100 rotor, Beckman, Palo Alto, CA) for 7 min. Concentration of protein in all ensuing experiments was determined from A_{280} value and expressed as monomer concentration. The extinction coefficient was calculated from the amino acid sequence.

Purification of ^{13}C-labeled protein

Crude Protein 1 was precipitated from cleared E. coli cell lysate with 50% ammonium sulfate at 4 °C for 30 min. Protein was resuspended in 20 mM Tris–HCl, pH = 8.0 and further purified by DEAE ion exchange chromatography using a 0–1 M NaCl gradient. Protein 1 was precipitated at pH = 4.0, washed with 20 mM Tris–HCl pH = 4.0, and resolubilized in 25 mM Tris, pH = 8.0.

Peptide synthesis

The D-peptides [5]: D10-p1-2K, Ac-KKGAC*EARHREWAWLC*AA-amide; D10-p50-2K, Ac-KKGAC*ELLGWEWAWLC*AA-amide; D10-p1, GAC*EARHREWAWLC*AA-amide, (comprised of D-configuration amino acids, C* indicates position of intramolecular disulfide linkage) were assembled on an Applied Biosystems 430A peptide synthesizer using Fastmoc® chemistry and cartridges filled with the appropriately protected Fmoc-D-amino acids. The peptides were cleaved from the resin using Reagent K and recovered by ether precipitation [13]. The crude peptides were cyclized by DMSO/air oxidation at high dilution (<1 mg/mL) in 100 mM ammonium bicarbonate monitored by RP-HPLC. When the oxidation was complete the peptides were purified by RP-HPLC on a Waters C18 column running a linear gradient of 0.1% TFA-water:acetonitrile. The fluorescently labeled peptides (TAMRA-D10-p1, mixture of Ac-K(ε-6-carboxamido-tetramethylrhodamine) KGAC*EARHREWAWLC*AA-amide and Ac-KK(ε-6-carboxamido-tetramethylrhodamine)GAC*EAR-

HREWAWLC*AA-amide; OG488-D10-p1, 5-Oregon Green- GAC*EARHREWAWLC*AA-amide; Cy5.5-D10-p1, Cy5.5- GAC*EARHREWAWLC*AA-amide) were prepared by reacting the cyclized D10-p1 peptide with the dye-succidimidyl ester in anhydrous DMSO containing 1% diisopropylamine followed by RP-HPLC purification. The identity and purity of the peptides (>95%) were verified by MALDI-TOF mass spectrometry and RP-HPLC on a YMC C18-AQ column. Concentrations of unlabeled peptides in fluorescence experiments were determined from A_{280} values using extinction coefficients of 11,200 and 16,800 M^{-1} cm^{-1}, respectively, for D10-p1-2K and D10-p5-2K. Concentrations of fluorescently labeled peptides were determined by using the dye's known extinction coefficients. Concentrations of peptides in all other experiments were determined by measuring dry weight prior to dissolution.

Sedimentation equilibrium study of Protein 2

Sedimentation equilibrium studies were done with Protein 2 at 15, 20, 25, and 30 rpm at 8 °C in 10 mM PO4, 120 mM NaCl, 2.7 mM KCl, pH 7.4 using an XL-I analytical ultracentrifuge (Beckman, Fullerton, CA). Data from the different speeds were analyzed globally by a non-linear least squares curve fitting of radial concentration profiles using the Marquardt–Levenberg algorithm as implemented in Origin 5.0 (Microcal Software, Inc., Northamptom, MA). A user-defined function describing sedimentation behavior of discrete particles was used [14]. Baselines and fixed radius signal values for each data set were allowed to vary independently; however, the molecular weight was held as a global parameter. The partial specific volume of the protein was calculated from the amino acid composition as 0.729 cm^3/gm [15]. Buffer density was measured at 8 °C at 1.01456 m/cm^3 in a Mettler-KEM Da-310 density meter (Mettler, Highstown, New Jersey).

Binding of D-peptides to Protein 1 by centrifugal enhanced affinity selection (spin screen)

Peptide D10-p5-2K (initial concentration 200 pM) and Protein 1 (initial concentration 244 pM) were mixed in the centrifuge tube, vortexed, and spun for 2 h at an average *g*-force of 309,880 (1000,000 rpm in a TLA-100 rotor, Beckman, Fullerton, CA) [16]. Five fractions, 20 μL each, were removed and examined by reverse-phase HPLC. Peptide and protein were examined alone as controls. The amount of peptide is followed by peak area of the relevant peaks on the chromatograms. The relevant peak area in each fraction was divided by the sum of the peak areas in all five fractions, and the result is plotted versus the additive volume of each fraction.

Fluorescence spectroscopy

Both direct and competition binding assays for tagged and non-tagged analogs of cyclic peptides D10-p1-2K and D10-p5-2K [5] were carried out to evaluate binding of the peptides to Protein 1. All assays gave similar results, and only the competition studies are described here. Titrations were automated using an IMx diagnostics instrument (Abbott, North Chicago) in FPIA (fluorescence polarization immunoassay) mode. A complete twofold dilution series, comprised of 20 separate 2 mL samples, was obtained by delivering appropriate individual aliquots to the first seven tubes, aliquots from an intermediate diluted stock for the next seven, and one more intermediate dilution for the final six. Binding assays were carried out at 35 °C. Dilution buffer for all samples was 120 mM sodium phosphate at pH 7.55 with 0.01% bovine gamma globulin and 0.1% sodium azide. Protein 1 was 0.26–0.77 μM (monomer concentration) and tagged peptide was 1–5 nM.

Protein nuclear magnetic resonance spectroscopy

The $^1H/^{13}C$-HSQC experiment was recorded on 0.1 mM protein samples (monomer concentration) in 25 mM Tris buffer, pH 8.0, 90:10 H_2O:D_2O. Data were recorded at 318 K on a Bruker 500 MHz spectrometer equipped with a cryoprobe with 1024 complex points in F2 and 38 points in F1 with sweep widths of 8333 and 3000 Hz, respectively. Eight scans/increment were collected. Chemical shifts of the protein were recorded in the presence of 0.05, 0.3, and 0.6 mM cyclic-D peptide, D10-p1-2K.

Protein X-ray crystallography

Protein 1 was concentrated to approximately 7 mg/ml in 20 mM Hepes buffer at pH 7.5. In this solution, the protein was stable for several months at 4 °C. Crystals form (hanging drop vapor diffusion, 17 °C) in 0.2 M magnesium chloride hexahydrate, 0.1 M Tris pH 8.5 and screening 33–40% PEG 400. An average size of $400 \times 400 \times 200$ μm is observed in approximately 1–2 weeks. The crystals were frozen at 150 K for data collection, using an Oxford cryo-system by soaking the crystals for at least 1 h in 23% PEG 400, 0.15 M

magnesium chloride hexahydrate, 0.08 M Tris pH 8.5 and 10% glycerol. Under these conditions the crystals diffract to 1.9 Å resolution at a synchrotron radiation source. Native data to 2.1 Å were collected and the crystals belong to the space group P321, with the unit cell parameters $a = b = 96.9$, $c = 72.7$, and $\alpha = \beta = 90$ and $\gamma = 120$. The structure was solved by molecular replacement using a molecular model based on the coordinates derived from Protein Data Bank entry 1ENV and using the molecular replacement program, AMORE [17]. Residues of the original 1ENV data that had no correspondence to Protein 1 were deleted for this molecular replacement analysis. The structures were refined with CNX [18] using a combination of simulated annealing maximum likelihood refinement and individual B-factor refinement. Electron density maps were inspected on a Silicon Graphics INDIGO2 workstation using the program package QUANTA 98 (Molecular Simulations Inc., San Diego, CA). The native structure was refined to 2.1 Å resolution with an $R = 28\%$ and Rfree of 34%.

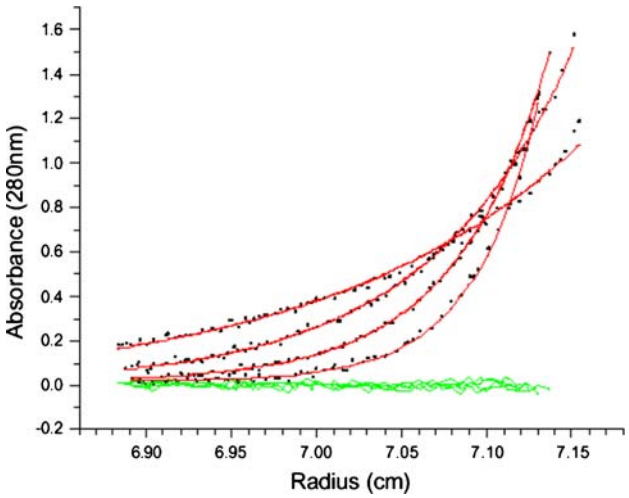

Fig. 4 Sedimentation equilibrium study of Protein 2. Data points (black squares) are the absorbance (280 nm) data at four speeds. The fit lines (red lines) refer to a single component model. Open circles (green lines) are the residuals

Results

Proteins 1 and 2 were designed to have an unnatural peptide linker between the N- and C-helical regions of the ectodomain of the gp41 protein from the human immunodeficiency virus. This design results in a circularly permuted protein, as diagrammed in Fig. 1. Modeling indicated that a linker of approximately four amino acids in length would serve to link the N- and C-helix peptides. The observed properties of the designed protein, *vide infra*, suggest that this analysis is correct. The cloning, expression, and purification of Proteins 1 and 2 were straightforward and details are available in the experimental sections.

Sedimentation studies on Protein 1 under a variety of conditions indicated that the protein was aggregated suggesting the presence of many different oligomeric states (results not shown). This kind of non-specific aggregation has been reported previously in biochemical research on engineered versions of gp41 proteins and peptides. Interestingly, we observed that the N-terminal His-tagged analog, Protein 2, yielded data that suggested the formation of a well-defined oligomeric species, and the results are shown in Fig. 4. Mathematical analysis of the sedimentation pattern indicated that a species of Mr = 33,000 is the predominant form in solution, in reasonable agreement with a calculated trimeric molecular weight of 31,000. Attempts to model the data with additional components of higher molecular weight did not significantly improve the correlation. We speculate that the presence of the additional N-terminal 18-residue peptide extension in Protein 2 containing polar His-tag and EK-cleavage sequences offset the substantial hydrophobic character of the remainder of the protein, thus providing a biophysical balance that allows trimer to be the dominant species. The importance of the sedimentation studies with Protein 2 is that they lead to the conclusion that the designed unnatural peptide linker does not seem to interfere with trimerization. Because of concern that the N-terminal extension of Protein 2 might block ligand access to the Trp-Trp-Ile pocket, all ligand binding studies were carried out with Protein 1 (which does not have any N-terminal extension).

The centrifugal enhanced affinity studies [Spin Screen, 16] gave clear indication that Protein 1 could bind ligands previously reported to bind to other engineered gp41 proteins. Two cyclic D-peptide ligands, D10-p5-2K and D10-p1-2K, were studied and gave similar results. The results for D10-p5-2K are shown in Fig. 5. The lack of sedimentation of the peptides in the absence of the protein indicated that no significant precipitation occurred, suggesting that both peptides appear to be quite soluble under the conditions examined.

Fluorescence polarization studies of binding permitted assessment of both the specificity and strength of binding of peptide ligands to Protein 1. In Fig. 6 are shown the results of competition studies indicating that non-tagged peptide ligands, D10-p1-2K and D10-p5-2K, can displace several fluorescently-tagged peptides in a specific manner. The change in fluorescence

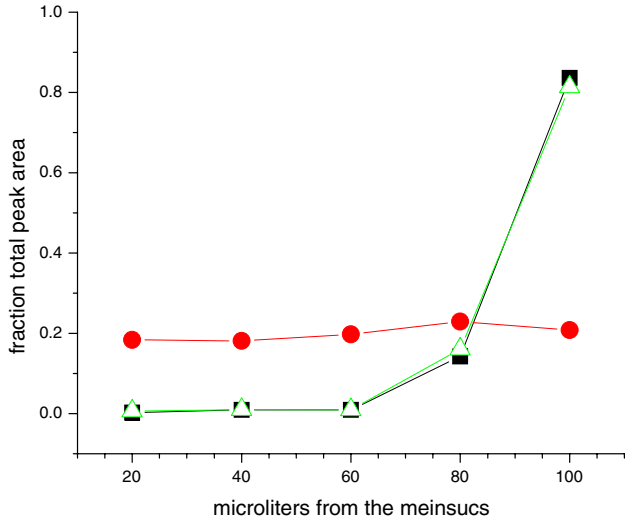

Fig. 5 Centrifugal enhanced affinity selection [Spin Screen, 16] binding of peptide D10-p5-2K to Protein 1. Closed squares (black) indicates the amount of peptide in each fraction after centrifugation in the presence of Protein 1. Close circles (red) indicate the amount of peptide in each fraction after centrifugation in the absence of Protein 1. Open triangles (green) show the distribution of Protein 1 found in each fraction after the centrifugation

polarization signal is plotted as a function of added displacing peptide during the titration. This effect is readily interpreted as a binding equilibrium between the peptide ligands and Protein 1, and K_i values that range 0.9–2.8 μM were observed. Further experiments will be required to determine if the small differences in the various titrations are significant. This magnitude of affinity is similar to the low micromolar potencies (4–46 μM) reported for cell-based assays of inhibition of syncytia formation and viral entry exhibited by these same peptides [5], but detailed comparison is prevented by the very different nature of the assays. These results suggest that the engineered Protein 1 possesses a "Trp-Trp-Ile" binding site that is not grossly perturbed relative the natural protein.

Nuclear magnetic resonance studies with Protein 1 provided additional evidence of direct binding of known gp41 ligands. Protein 1 was produced isotopically labeled with ^{13}C at the methyl groups of valine and leucine, and the δ-methyl of isoleucine. This permitted experiments to detect chemical shifts of atoms resulting from changes in the chemical environment upon ligand binding. Methyl proton chemical shift measurements were obtained from ^{13}C heteronuclear single quantum coherence (HSQC) experiments at 45 °C on a 500 MHz spectrometer equipped with a cryoprobe. To detect binding of the cyclic D-peptide to Protein 1, the NMR experiment was repeated in the presence of 0.05, 0.3, and 0.6 mM peptide. The ^{13}C HSQC spectra for the GP41 protein, in the presence of varying concentrations of peptide, are shown in Fig. 7. Three methyl resonances that give pronounced shifts upon ligand binding are indicated. These unambiguously confirm peptide binding to Protein 1. Although the exact structural nature of the binding cannot be determined from the data reported here, the fact that one large isoleucine δ-methyl chemical shift is observed is consistent with binding to the Trp-Trp-Ile pocket where one isoleucine resides (see description of X-ray results below). It is important that during the peptide titration, the peptide was found to be in fast exchange and neither doubling nor tripling of peaks was observed. In addition, nearly 100% occupancy was observed with 300 μM peptide, which is consistent with the K_D determined by fluorescent measurements (1–3 μM) and a binding site concentration of 100 μM. Overall these data suggest a stoichiometry of 3:1 (peptide monomer:Protein 1 trimer) and equivalent peptide affinities for each site.

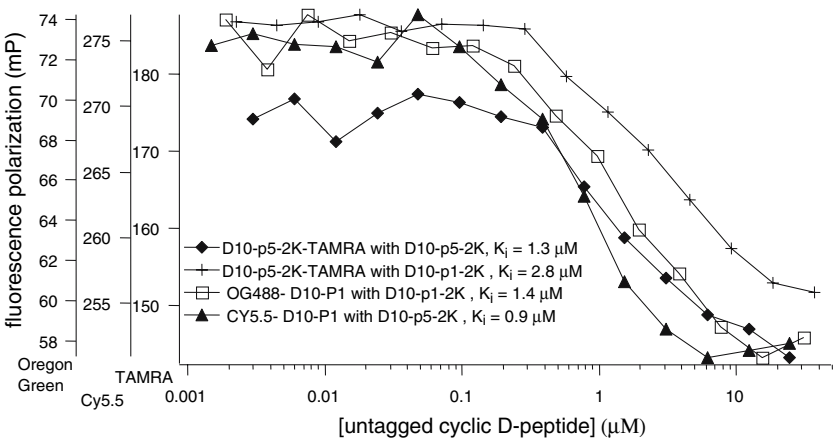

Fig. 6 Titration data showing peptide ligand displacement of variously fluorescently-labeled peptides bound to Protein 1. The different *Y*-axis scales refer to the different degrees of fluorescence polarization observed for the various peptides

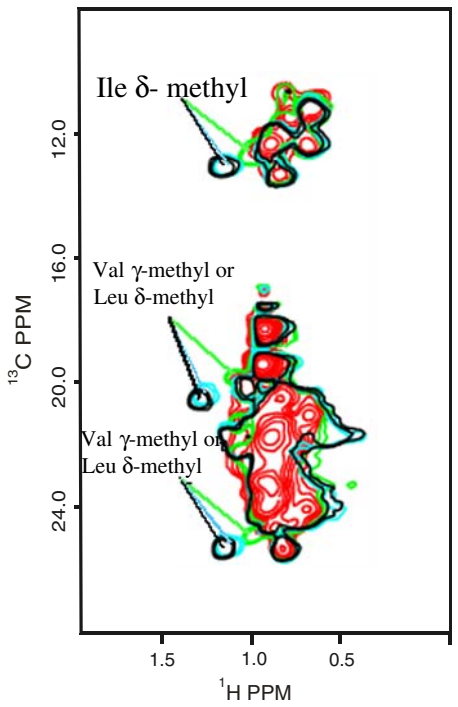

Fig. 7 $^{1}H/^{13}C$-HSQC spectra of 0.1 mM Protein 1 (monomer concentration) acquired at 500 MHz in the absence (red contours) and presence of 0.05 (green), 0.3 (cyan), and 0.6 (black) mM peptide ligand. Lines (with the same color coding as the contours) highlight the peptide-dependent chemical shifts that are most affected by Protein 1 binding

Crystals of Protein 1 were obtained, and the three-dimensional structure was solved using protein X-ray crystallographic techniques, with the results shown in Fig. 8. Protein 1 crystallizes in the trigonal space group, P321. The asymmetric unit contains one full trimer of Protein 1 and one monomer located on the crystallographic threefold axis forming a crystallographic trimer. Importantly, the structure has the characteristic hexameric bundle of helices common to previously determined crystal structures of other engineered gp41 proteins. The following two comparisons are informative regarding both the overall architecture and specific Trp-Trp-Ile binding site. Comparison of 41 matched Cα-residues (out of 70 possible) between Protein 1 and an HIV gp41 monomer, (PDB code 1ENV) yields an RMSD of 0.37 Å. Comparison of 11 residues of the Trp-Trp-Ile binding site (see discussion below for the list of residues) between Protein 1 and fusion protein [5]of HIV gp41 and human GCN4, termed IQN17 (PDB code 1CZQ) yields an RMSD of 0.23 Å. While the conformations of most side chains comprising the Trp-Trp-Ile binding site were similar to previously determined crystal structures, the absence of a binding partner in the Trp-Trp-Ile binding site in the structure

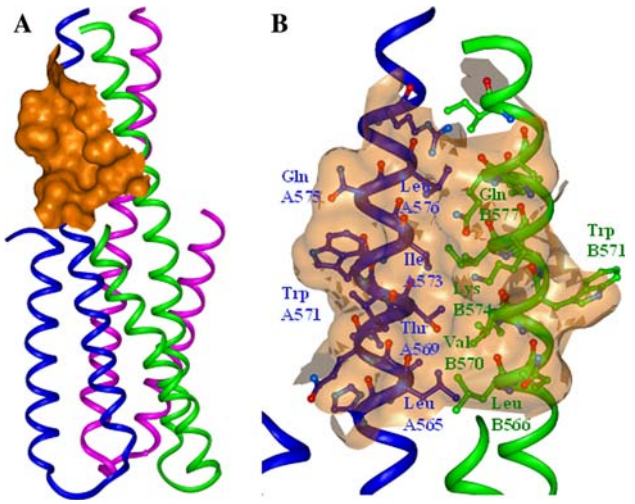

Fig. 8 X-ray Crystal Structure of Protein 1. (**A**) Ribbon-view showing the three monomers in pink, green, and blue. Three symmetry-related ''Trp-Trp-Ile'' pockets exist per trimer of Protein 1, and the surface of one is shown in orange. The designed loops of this study are visible at the bottom of the picture. (**B**) Close-up view of image A showing the ''Trp-Trp-Ile'' pocket-forming residues. The residues that form this pocket come from two monomer units and are denoted as 'Axxx' or 'Bxxx'

reported here would be expected to permit alternative conformations. Two residues, Trp 571, and Gln 575, were observed to show significant changes of approximately 120° difference in Chi1 angle, relative to that reported for IQN17 (PDB entry 1CZQ). These two residues are found at the rim of the Trp-Trp-Ile pocket. A close-up view of the exposed Trp-Trp-Ile binding pocket is shown in Fig. 8B. Residues of two monomer units comprise a single Trp-Trp-Ile binding pocket. As anticipated, there are three of these pockets per crystallographic trimer. The main chain of the designed GDG linking tripeptide which joins Asp 442 at the C-terminus of the C-helix with Arg 542 at the N-terminus of the N-helix is clearly visible in the electron density map; however the side chain of the Asp residue within this tripeptide is disordered.

Discussion

The work presented here represents a successful protein engineering effort that yielded a potentially useful drug discovery tool. Previously reported constructs of gp41 protein for anti-HIV drug discovery possessed biophysical properties that are not ideal for ligand screening efforts. These proteins have occluded desired binding sites, exposed undesired binding sites, and high

aggregation tendencies. These limitations become particularly acute during protein X-ray or protein nuclear magnetic resonance studies to evaluate ligand binding. We sought to overcome these liabilities with a re-engineered form of gp41. There are two key structural features of our design. First, we took advantage of the known close proximity of the C-terminus of the C-helix and the N-terminus of the N-helix. Artificial linker peptides of either 3 residues, in the case of Protein 1, or 4 residues, in the case of Protein 2 were inserted in this region. It is interesting to compare this aspect of our design with that reported independently in a design of a gp41 protein analog nicknamed ''5-helix'' [19]. While the ultimate goal of that design effort differed from the current work, both efforts required the engineering of an artificial linking peptide to link the C-terminus of the C-helix to the N-terminus of the N-helix. The sequences of both are compared in Fig. 9. The linkers of the current work have either 1 or 2 additional amino acid residues, Proteins 1 and 2, respectively, in their loops relative to that used in the 5-helix protein. All three linkers possess similar but non-identical sequences consisting of polar and flexible amino acids. The data currently available are not sufficient to assess if one particular loop length is optimal.

A second key structural feature of the protein of the current work is the presence of a permanently exposed potential ligand-binding site comprised of residues from the N-helix. The residues of the C-helix that occupy this pocket in the natural ectodomain of gp41 are Trp 628, Trp 631 and Ile 635, and this cavity is nicknamed the ''Trp-Trp-Ile'' binding pocket. This pocket is the deepest indentation in the extensive helix–helix interfacial region between the N- and C-helices. The potential importance of this pocket to drug discovery was highlighted in 1998 when the sequence conservation of the Trp-Trp-Ile sequence and cognate binding pocket was recognized [4]. There are 11 residues from two N-helices that form this pocket. Ten residues, Leu A568 (N-helix 1), Thr A569 (N-helix 1), Val B570 (N-helix 2), Trp A571 (N-helix 1), Glu A572 (N-helix 1), Ile B573 (N-helix 2), Lys B574 (N-helix 2), Gln A575 (N-helix 1), Leu A576 (N-helix 1), Gln B577 (N-helix 2), and Arg A579 (N-helix 1) are 100% conserved. Only residue A565 shows any natural sequence variation with both Leu and Met observed at this position. The unoccupied Trp-Trp-Ile pocket observed for Protein 1 has the potential of exhibiting protein movement within the active site, relative to previous reports which studied only occupied pockets. Crystallographic data show a high correspondence of backbone positions of these 11 residues with data of previous reports, indicating that the gross shape of the pocket is unchanged and has a well-defined and conserved character, regardless of whether it is liganded or not. Most of the residues comprising the Trp-Trp-Ile binding site exhibited analogous side chain conformations, with the exception of two residues, Trp 571 and Gln 575. Retention of the native shape is further supported by the observation of binding of known ligands: cyclic D-peptides D10-p1-2K and D10-p5-2K. The structural conservation of the Trp-Trp-Ile pocket suggests that agents directed against it will have the broad utility requisite for new anti-HIV therapy.

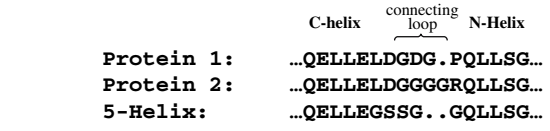

Fig. 9 Alternative sequences for linking the C-terminus of the C-Helix to the N-terminus of the N-Helix. ''5-Helix'' refers to a related engineered gp41 protein [19]

The biophysical studies reported here: sedimentation, fluorescence, NMR spectroscopy, and X-ray crystallography, are all in accord with the original design proposal diagrammed in Fig. 1. There are three significant observations that emerge from our work: (1) Studies of Protein 1 by fluorescence, centrifugation, and NMR methodologies indicate that the designed protein can bind a known ligand. Importantly, the fluorescence studies indicate that the magnitude of this affinity is in agreement with that reported previously [5]. (2) Another form of the designed protein, Protein 2, gave clear evidence of a stable trimeric form by sedimentation studies. (3) Protein X-ray crystallographic studies confirm the trimeric structure with the truncated C-helix. Importantly, the x-ray data are consistent with an exposed Trp-Trp-Ile pocket.

At this stage, the production and characterization of Proteins 1 and 2 represent a successful example of molecular design. The true utility of Proteins 1 and 2 will be realized if they play a role in the discovery of molecules that bind to the Trp-Trp-Ile pocket, prevent HIV viral fusion and exhibit anti-HIV activity. Such molecules additionally have the potential of showing inhibitory activity against T20-resistant strains of HIV. T20, a clinically-approved drug, does not interact with the Trp-Trp-Ile pocket, and mutations confering T20-resistance are located elsewhere in the gp41 protein. Studies using Proteins 1 and 2 and directed toward discovery of novel antiviral agents are ongoing in our laboratories. Our results along those lines will be the subject of a future report.

Conclusion

In this work, a re-engineered version of the gp41 protein of the HIV virus was created. A domain swapping design exercise was carried out where the N-helical and C-helical domains were switched in sequential order. The biophysical studies indicate that the protein has a 3D structure similar to previously reported constructs of gp41 with one fundamental difference: the exposure of a potential ligand binding pocket termed the Trp-Trp-Ile pocket. It is hoped that this engineered version of gp41 protein will be potentially useful in research programs aimed at discovery of new drugs for therapy of HIV-infection in humans.

Acknowledgment This manuscript is dedicated to Yvonne Connolly Martin. The precepts of molecular properties and architectures promulgated by Yvonne in 41 years at Abbott extend even to the level of protein design. All of the authors thank her for being a steadfast source of inspiration and scientific role-model of the highest degree. Crystallographic data were collected at beamline 17-ID in the facilities of the Industrial Macromolecular Crystallography Association Collaborative Access Team (IMCA-CAT) at the Advanced Photon Source. These facilities are supported by the companies of the Industrial Macromolecular Crystallography Association.

References

1. Lu M, Blacklow SC, Kim PS (1995) Nat Struct Biol 2:1075
2. Moore JP, Doms RW (2003) Proc Natl Acad Sci 100:10598
3. Kilby JM (1999) Expert Opin Investig Drugs 8:1157
4. Chan DC et al (1998) Proc Natl Acad Sci 95:15613
5. Eckert DM et al (1999) Cell 99:103
6. Mo HM et al (2004) Virology 329:319
7. Zho G et al (2000) Bioorg Med Chem 8:2219
8. Liu S, Jiang S (2004) Curr Pharm Des 10:1827
9. Jin B-S et al (2005) J Biomol Screen 10:13
10. Frey G et al (2006) Proc Natl Acad Sci 103:13938
11. Cianci C et al (2004) Proc Natl Acad Sci 101:15046
12. Hajduk P et al (2000) J Am Chem Soc 122:7898
13. King DS, Fields CG, Fields GB (1990) Int J Pept Protein Res 36:255
14. Holzman T (1994) In: Schuster TM, Lave TM (eds) Modern analytical ultracentrifugation. Birkhauser, Boston, MA, p 298
15. Cohen EJ, Edsall JT (1943) Proteins, amino acids, and peptides as ions and dipolar ions. Rheinhold, New York, ch 4, p 157
16. Harlan JE et al (2003) Assay Drug Develop Technol 1:507
17. AMORE (Collaborative Computational Project, Number 4, 1994). Acta Crystallogr D50:760
18. Brunger AT (1992) Nature 355:472
19. Root MJ, Kay MS, Kim PS (2001) Science 291:884