Current Enzyme Inhibition, 2017, *13*, 20-26

RESEARCH ARTICLE

Targeting HIV-1 Envelope Proteins Using a Fragment Discovery All-Atom Computational Algorithm

Michael H. Peters*

Department of Chemical and Life Science Engineering, VCU Massey Cancer Center, and the Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, VA, USA

Michael H. Peters

Abstract: ***Introduction*:** HIV viral envelope proteins are targets for small inhibitor molecules aimed at disrupting the cellular entry process. Potential peptide-class inhibitor molecules (rDNA drugs) have been previously identified, with mixed results, through biomimicry and phage display experimental methods. Here we describe a new approach based on computational fragment discovery. The method has the potential to not only optimize peptide binding affinity but also to rapidly produce alternative inhibitors against mutated strains.

Methods: A comprehensive, all-atom implicit solvent method is used to bombard the C-heptad repeat unit of HIV-1 target envelope protein GP41 with single D-amino acid residues as they exist in their native state. A nascent peptide computational search process then identifies potential favorable sequences of attached ligands based on four peptide bond criteria. Finally, dynamic simulations of nascent peptides attached to host targets help refine potential peptide inhibitors for experimental HIV-1 challenge assays and testing.

Results and Discussion: Initial testing of the method was done using 64,000 total ligands of D-amino acid residues at a total computational time of 0.05 microseconds per ligand, which resulted in several thousand attached ligands. Peptide bond criteria search employing three of the four bond constraints with a tolerance of 20 percent, resulted in four potential peptide inhibitors of 5 to 6 residues in length. Only one of the four peptides demonstrated IC50 values and partial viral inhibition based on cell challenge assays using CEM-SS host cells. That peptide inhibitor also computationally demonstrated long-time attachment and stability to a helical groove in its C-heptad target. This initial testing of peptide fragment discovery against HIV-1 has helped us refine the protocols and identify key areas of improvement.

Conclusion: Our methods demonstrate the potential to design efficient peptide inhibitors to viral target proteins based on an all-atom dynamic simulation and using a ligand library as fragments of potential nascent peptides. Our methods can be greatly improved through the use of higher numbers of ligands, increased time of bombardment, and tighter constraints on the peptide bond search step. Our method may be important in the need to rapidly respond to target mutations and to advance multiple targeting methods based multiple peptide inhibitors.

ARTICLE HISTORY

Received: January 27, 2016
Revised: April 19, 2016
Accepted: May 27, 2016

DOI:
10.2174/1573408012666160725095854

Keywords: Envelope proteins, fragment discovery, HIV, implicit solvent methods, peptide inhibitors.

1. INTRODUCTION

Infectious disease from viral agents continues to represent one of the most significant health threats to society. The ability of these agents to mutate, transform, and develop across species makes them a formidable opponent to the development of therapeutics, diagnostics, and vaccines aimed at their debilitation.

Inhibitors that can bind to viral surface proteins can significantly reduce infectivity rates and, thus, function as therapeutic agents. In this study, the use of advanced molecular computational algorithms for generating small peptide inhibitor compounds that are specific to HIV-1 viral surface protein antigens has been demonstrated. Peptide therapeutics, in general, represents one of the fastest growing segments of the pharmaceutical industry, known as rDNA drugs, and often includes efficient delivery platforms and attractive pharmacokinetic profiles [1]. Advances in *ex-vivo* production methods also allow production of proteolytic-defying D-peptide drugs and provide inexpensive and large

*Address correspondence to this author at the Virginia Commonwealth University, 601 West Main St., Room 401, Richmond, VA 23284, USA; Tel:(804)828-7789; E-mail: mpeters@vcu.edu

1875-6662/17 $58.00+.00

production capability routes for protein and peptide drugs in general.

GP41 is part of an envelope glycoprotein complex of HIV-1 that binds to target cell receptors CD4 and CCR-5 or CXCR-4 [2]. GP41 is a three-stranded coiled-coil structure that is exposed during the viral entry process (prefusion state). GP41, therefore, has been a target for the development of inhibitory compounds that bind to it and disrupt the viral entry process. Each subunit of GP41 consists of an N-heptad repeat unit from its N-terminal region (NHR) and C-heptad repeat unit from the C-terminal end (CHR) arranged in an antiparallel fashion. During fusion, the subunits fold to form a six bundle helix with three NHR regions in the core stabilized by interactions with three ectodomain CHR regions. The NHR and CHR interacting regions were synthesized and structurally determined [3], as illustrated in Fig. (**1**).

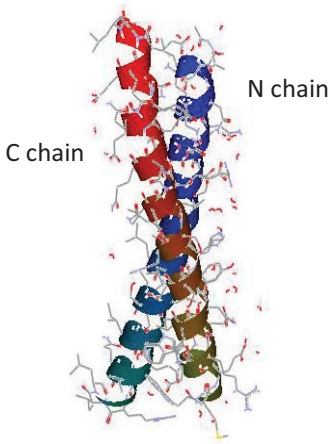

Fig. (1). N-heptad and C-heptad repeat unit from GP41; PDB ID: 1AIK [3].

Peptide sequences based on the CHR region (C peptides) potentially bind to the NHR region and vice versa [4]. C-peptides have been experimentally shown to be potent inhibitors resulting in, for example, the successful drug Fuzeon (Roche) or Enfuvirtide (T-20) [5]. Previously, we used a static all-atom energy landscape mapping algorithm [6] that yielded a 32 residue peptide sequence from the CHR region (Residue Numbers C628-C659) as the dominant energy interaction region between NHR and CHR almost exactly overlying a number of known experimental nanomolar binding peptides (C34's and SJ-2176) from the CHR region [3].

In this study we describe the development of an alternative, *ab initio* method of identifying potential peptide binding partners based on a fragment discovery algorithm. We demonstrate the ability to design peptide molecules that bind to GP41 and inhibit viral entry based on experimental HIV-1 cell challenge assays. The fragment discovery algorithm has the potential to rapidly develop inhibitory peptides and, thus, may have broader impact to the need for a robust response to viral infections and outbreaks, in general, including the rise of mutational variations.

2. MATERIALS AND METHODS

2.1. All Atom Implicit Solvent Methods

Biological macromolecular interactions, notably protein-protein interactions, are fundamentally dynamic events involving the transport and diffusion of ligands to target molecules or sites, followed by attachment or physical adsorption of the ligand to the target molecule. Physical adsorption is, in general, reversible leading to desorption and diffusion and transport of ligands away from the target site. The host solvent (water and dissolved ions) plays a number of important roles, as both a transport media (fundamental fluid mechanics) and in modifications of ligand-receptor force interactions *via* dielectric and hydrophobic effects (micro-structure considerations). Fig. (**2**) illustrates a simple ligand-receptor process that takes place in a model biological system, where the target molecule is fixed in space. This particular process is modeled in detail in our methodology, including a full accounting of molecular interactions and their manifestation to observed macroscopic behavior.

Due to its potential to significantly reduce the computational load, we have developed an all-atom implicit solvent method derived from first principles *via* the N-body Liouville equation - the fundamental equation of molecular statistical mechanics. Broadly speaking, coarse-grained methods, stochastic dynamics, Brownian dynamics, and implicit solvent methods, are closely related terms representing macromolecular dynamic methods that involve some type of averaging method for the solvent phase. Here we simply refer to our method as implicit solvent or Brownian dynamics (BD) method [7-9]. Our first-principles method is comprehensive in that it includes all possible rotational, translational, and coupled rotational-translational modes of ligands (peptide amino acid residues in this study), quantitative limits on the

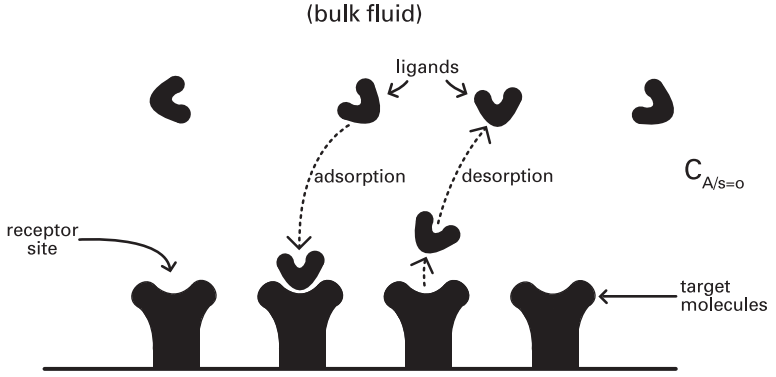

Fig. (2). Illustration of the overall dynamic process of ligand receptor interactions.

use of BD methods *via* multiple time scale perturbation theory, inclusion of any external surfaces (protein molecular targets in this study), and formal, comprehensive prescription of all implicit solvent terms *via* time force autocorrelation expressions.

In general, the BD algorithm involves a short-time averaging of the host solvent dynamics coupled to a relatively longer-time, macromolecule dynamics step. The macromolecular dynamics step requires Brownian particle diffusion terms and implicit solvent force terms that can be determined *via* the short-time behavior, or, alternatively, the diffusion and implicit solvent force terms can be approximated *via separate* analytical or computational studies for any given system, as demonstrated in proteins [8, 9]. Thus, BD has demonstrated the potential to greatly reduce the computational load required for protein dynamic simulations by reducing both the total atom-atom force computations load and increasing the integration time steps required.

Briefly, the Brownian dynamics method is in the form of dimensionless translational and rotational displacement equations describing the change in the position of the ligand center of mass, $\mathbf{R}_l$, and the orientation about the center of mass, $\boldsymbol{\phi}_l$, for the lth ligand or B-particle as [7].

$$R_{i\,l} = R^0_{i\,l} + \epsilon t\,[\,\sum_j (D^0_{ij\,Tl}\, F'^0_{j\,l} + D^0_{ij\,TRl}\, T'^0_{j\,l})] + C_{i\,l}(D^0_{ij\,l}, t), \quad (1 \le i \le 3,\ 1 \le j \le 6) \tag{1}$$

$$\phi_{i\,l} = \phi^0_{i\,l} + \epsilon t\,[\,\sum_j (D^0_{ij\,RTl}\, F'^0_{j\,l} + D^0_{ij\,Rl}\, T'^0_{j\,l})] + C_{i\,l}(D^0_{ij\,l}, t), \quad (4 \le i \le 6,\ 1 \le j \le 6) \tag{2}$$

where i and j represent the Cartesian coordinates for coupled translation and rotation of the lth particle, ϵ is a particle Stokes' number, **D** is a 6 by 6 grand diffusion tensor for the B-particle, **F** and **T** are the forces and torques acting on the B-particle as specified below, t is the time step, and the superscript (o) indicates values at the beginning of the time step. The stochastic function $C_{i\,l}(D^0_{ij\,l}, t)$ is a multivariate, Gaussian random number with zero mean and variance-covariance given by

$$< C_{i\,l}(D^0_{ij\,l}, t) C_{j\,l}(D^0_{ij\,l}, t) > = 2\, D^0_{ij\,l}\, t\, \epsilon \tag{3}$$

The force and torque terms appearing in the Brownian displacement equations above consists of two parts: (a) the local equilibrium average force and torque of the solvent acting on each ligand or B-particle (also, called the implicit solvent force) and (b) the external field force or torque due to the interactions of the ligand with the receptor molecule atoms, or in symbols, respectively

$$\mathbf{F}'_l \equiv < \mathbf{F}_{fl} >_{eq} + \sum_{kl} \sum_{ks} \mathbf{F}_{kl\text{-}ks} \tag{4}$$

$$\mathbf{T}'_l \equiv < \mathbf{T}_{fl} >_{eq} + \sum_{kl} \sum_{ks} \mathbf{T}_{kl\text{-}ks} \tag{5}$$

where k_l represents the atom of the lth B-particle and k_s represents the atom of the target receptor (*s*). These atom-atom interaction forces and associated torques are taken from the AMBER 03 force field model for all results shown here [10]. Atomic structural information (for ligand and receptor) is supplied by *.pdb or equivalent structure files.

2.2. Solvent Considerations

The local equilibrium average force of the solvent on the ligand consists of polar and apolar contributions as described below.

2.2.1. Polar Implicit Solvent Forces

Polar implicit solvent forces lead to the so-called dielectric behavior of the solvent. In the calculations given here, following the work of Ramstein and Lavery [11] as reviewed by Smith and Pettitt [12], we employ a distance dependent dielectric of sigmoidal shape with a decay constant of $0.5A^{-1}$.

2.2.2. Apolar Implicit Solvent Forces

Following previous studies [9], the apolar implicit solvent potential was taken to be proportional to the solvent accessible surface area with the surface tension parameter set to $0.5kcal/molA^2$. The solvent accessible surface area was calculated at the beginning of each Brownian dynamics step using the method of Hasel *et al*. [13].

2.2.3. Diffusion Tensor for the Ligand

For the purposes of calculating the diffusion tensor, each ligand is modeled as a sphere in a continuum with an effective diameter based on their respective van der Waals volume [9]. The hydrodynamic interaction of the ligand with the receptor is neglected in all calculations given here. We note that any of these assumptions can be relaxed through the incorporation of short-time force autocorrelation or analytical studies.

Now, although the general algorithm can also include the flexibility of both ligand and receptor [9], the computational costs are currently prohibitive for this application, which necessarily involves the dynamic simulation of tens of thousands of ligands interacting with the target molecule. Consequently, all results given here are restricted to rigid ligands and receptors.

2.3. Multiple Processing

Fig. (**3**) shows the basic simulation "box" where the "ligands" are, in general, peptide fragments from a compound library. Here we take the ligand library of compounds to be the 20 amino acid residues as they naturally appear in a protein, *i.e*., the residues contain one hydrogen and one oxygen attached to the N and C atoms, respectively (Fig. **4**). These amino acid residue structure files are taken from random protein structure files from the protein data bank. The center of mass of a single target receptor is placed at the center of the box, and a fixed lab frame Cartesian coordinate system is assigned to it (Fig. **4**). Ligands are initially placed randomly, including position and orientation, within the available volume of the box. The "available volume" excludes the box volume occupied by the receptor, and the center of mass of any ligand must reside within the box and not outside.

The BD method traces the trajectory of each ligand in real time as it interacts with the molecular target. The target receptor is "bombarded" with multiple copies of each type of ligand. Since ligands and their copies are not to interact with each other, each trajectory can be done independently and on a separate processor. Thus, the method is highly amenable to multiprocessor strategies [14], as graphically illustrated in Fig. (**3**).

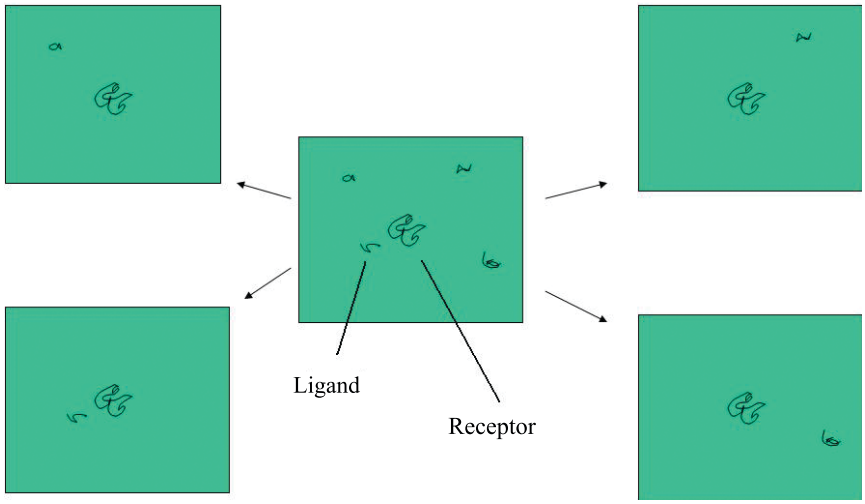

Fig. (3). Illustration of the control volume and multiprocessor computational strategy for ligand-receptor interaction dynamics.

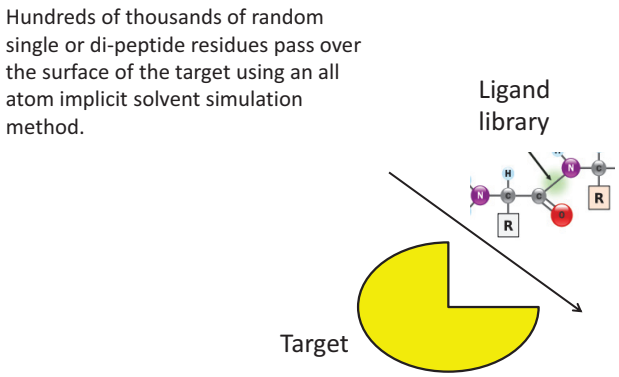

Fig. (4). Illustration of fragment passing over specified target.

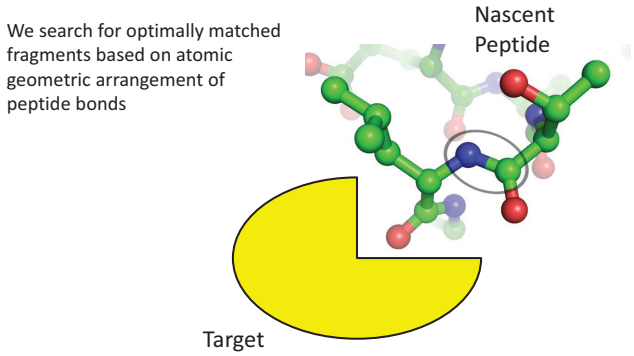

Fig. (5). Illustration of peptide bond searching for bound fragments.

2.4. Peptide Discovery

After a sufficient simulation time on the order of 0.1 microseconds here, a large number of amino acid residues will remain attached to the surface target by virtue of favorable atom-atom interaction energies. For these attached amino acid residues, we have developed an additional code to search for the possibility of patching residues to create a peptide of at least 6-12 attached residues in order to potentially obtain a strong binding, approximately linear peptide. In brief we search over attached residues to find potentially favorable peptide bonds based on the orientation and position of attached residues. These residues are then connected *via* peptide bonds to create the "birth" structure file of the peptide or peptide candidates (Fig. **5**).

2.5. Peptide Bond Criteria

We have utilized four peptide bond criteria in order to determine potential peptide binding segments as given below:

1. *Angle* $(O_i - C_i - N_{i+1}) = 123.5^o$
2. *Angle* $(C_i - N_{i+1} - C_{\alpha,i+1}) = 120^o$
3. *Distance* $(C_i - N_{i+1}) = 1.46A$
4. *Distance* $(C_{\alpha,i} - C_{\alpha,i+1}) = 1.51A$

Where the subscript i is the residue number and C_α, N, and C, and O are the carbon atom, backbone nitrogen, backbone carbon, and backbone oxygen of the amino acid residue. The nascent peptide can then be studied dynamically using the flexible implicit solvent simulation [9] in order to obtain its more natural state and examine its stability or continued binding ability with the target surface. We note that all results given here are restricted to finding linear peptides; however, the general method could be extended to secondary structures, such as helices from some known ensemble of potential helical fragment candidates.

3. RESULTS AND DISCUSSION

We sought to initially demonstrate the potential feasibility of the method using a limited number of ligands and constraints. For our amino acid segment library we used the D-form here, based on its enhanced pharmacokinetics, taken randomly from protein data bank structure files for D-amino acid proteins. However, the method is not restricted to this form and, for example; the natural L-form residue library is also easily generated. With the C-heptad repeat as the target, 64,000 total ligands (3200 per residue) bombarded it at a total time of 0.05 microseconds for each ligand using a Sun Ultra Sparc III microprocessor. This resulted in approximately 3000 attached residues. From these 3000 attached residues 4 linear peptide segments of 5 to 6 residues were

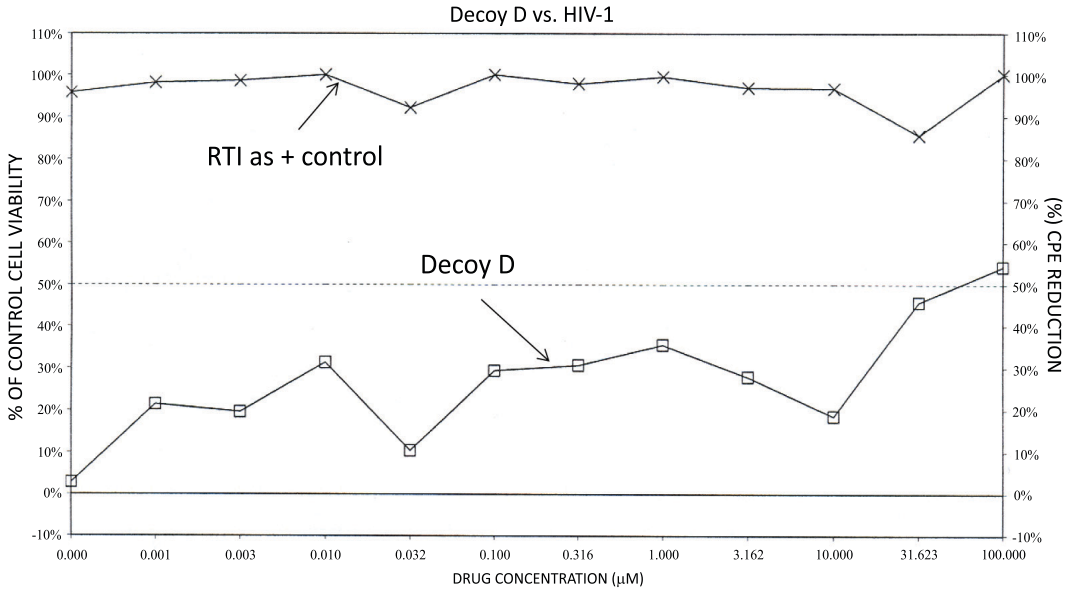

Fig. (6). HIV-1 challenge test results for Decoy D. Upper curve is the control data using AZT. HIV Challenge Tests for Decoy D were conducted by Southern Research Institute.

obtained based on only the first three of the four criteria given above at a 20 percent tolerance level:

A peptide (Decoy A): EEDQY

B peptide (Decoy B): YDEDY

C peptide (Decoy C): DWHDDQ

D peptide (Decoy D): DWWYW

Further ranking of the peptides based on binding strengths can also be done using the open source all-atom static energetic mapping algorithm [6]. However, due to the small number of peptides and the small number of residues, all four of the peptides were synthesized and tested. We note that currently we have been able to greatly expand the number of ligands to hundreds of thousands using multiprocessors, increased the total simulation time per residue, employ all four of the bond constraints in the peptide search, and reduce the tolerance levels to 10-15 percent. These improvements generally result in 8 to 12 residue segments and stronger binding affinities as compared to the initial study and results given here.

The four D-peptides listed above were synthesized (Sigma Genosys) and sent to Southern Research Institute for HIV-1 Cytoprotection Assays using CEM-SS cells and HIV − 1RF (Project 12209.01). Only one of the ligands (Decoy D) demonstrated IC50 inhibition as shown in Fig. (**6**).

Fig. (**7**) shows the computer generated Decoy D peptide attached in one of the helical groves of the C-heptad helix as predicted from the attached residues and peptide bond constraint criteria. As shown in Fig. (**8**), we conducted further longer time dynamic simulations with the Decoy D peptide on the target surface in order to verify its superior binding strength and stability compared to the other three peptides. Decoy A, B, and C peptides exhibited less overall interaction energies compared to Decoy D and had less overall interactions with helical grooves of GP41.

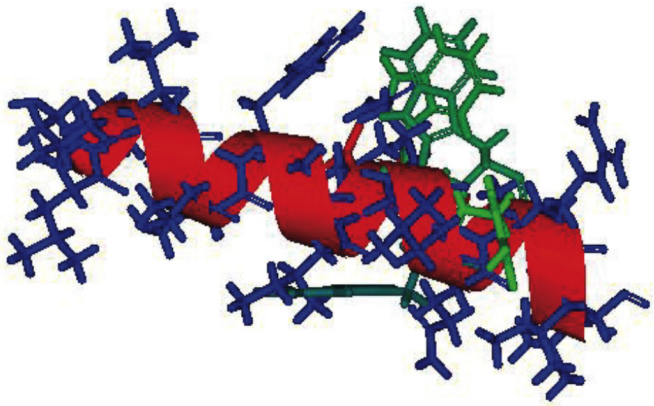

Fig. (7). Computer Synthesized Peptide (Decoy D; in green) binding in a helical groove of GP41 (red and blue) from HIV-1.

We note that a static energetic mapping of the binding strengths and subsequent ranking of the peptides are now formally available [6], which can provide a rapid ranking method. However, longer time (millisecond) dynamic algorithms may be necessary due to natural conformational changes associated with the nascent peptide and protein target. Formal off-rate screening is also a challenge due to the long simulation times required and alternative methods are necessary, such as temperature programmed desorption, which are under current study. In that regard, we note that the algorithm given here does not provide equilibrium binding data. This initial study points to the importance of carrying out longer time dynamic simulations with nascent peptides prior to any experimental testing in order to optimize the process of fragment peptide discovery; this is in addition to the number of ligands, constraints, and tolerance issues noted above. If possible, the addition of in-*vitro* kinetic binding studies, not done for the four peptides above, is also important in viral applications given the complexities and costs of cell challenge experimental methods [6].

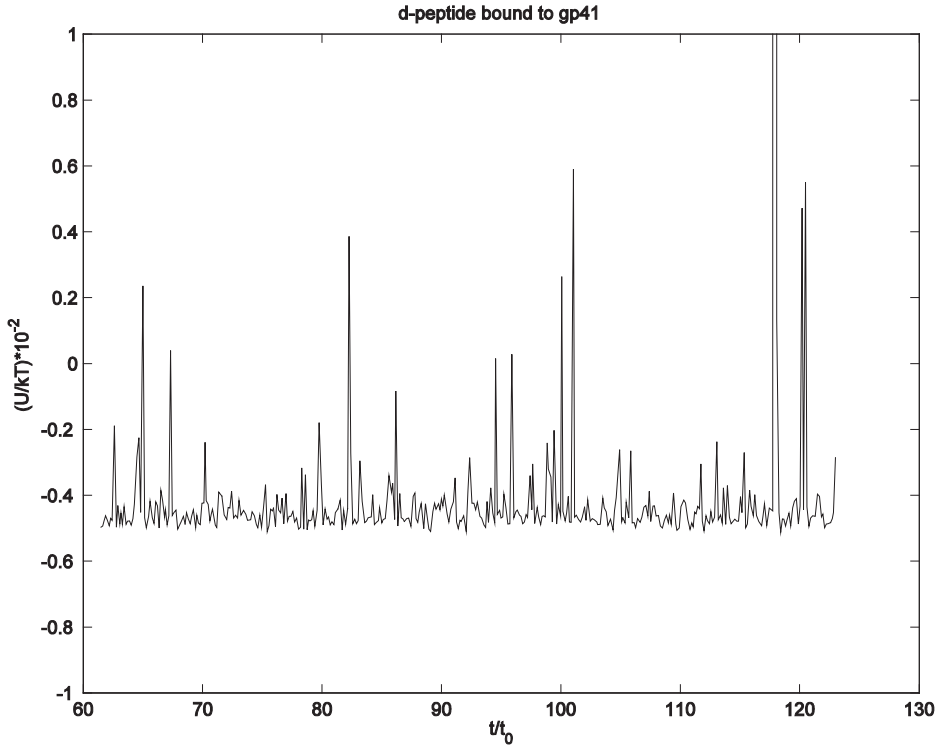

Fig. (8). Total interaction energy, U, time profile for the attached state of Decoy D in helical groove. Note that this is a typical dynamic energetic signature for attached ligands. The total interaction energy is scaled by kT, where k is Boltzmann's constant and T is absolute temperature (371.15K) and time is scaled by to, 10^{-11} seconds.

CONCLUSION

We have demonstrated the ability to conduct an *ab initio* design of peptide inhibitory molecules to an envelope glycoprotein target molecule GP41 of HIV-1 using a real-time fragment discovery algorithm. By passing on the order of one hundred thousand amino acid residues in their natural protein state over the target surface and then using general bond constraint criteria for peptides, we were able to obtain several potential peptide candidates in a docked state for further study and testing. Our approach includes the development of potentially multi-use rapid viral detection and therapeutic agents based on *ab initio* computationally generated small peptides with molecular weights less than 1000 Da. In general, peptides can be designed specifically from any validated viral protein target molecule by this method. Since our approach heavily relies on real-time computational methods, there is a potential to significantly reduce the time necessary for the identification of efficacious candidate compounds and also potentially respond in a very rapid manner to newly mutated forms [15]. Clearly, however, more bench studies are needed in order to validate the fragment discovery method beyond the limited results for HIV-1 GP41 inhibition given here, including other viral and non-viral protein targets.

The challenges of this approach include the use of a greater number of residues or fragments; longer simulation times, tighter bond constraints on the discovery of nascent peptides, and longer time dynamic simulations of nascent peptides in the bound state of the target for the purposes of ranking and partial validation. The addition of flexibility of nascent peptides and target proteins, not considered here, would also be desirable using either implicit solvent or molecular dynamics methods [9], albeit at the expense of longer computational times.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

The author would like to acknowledge the Southern Research Institute for carrying out the HIV-1 cell challenge assays.

REFERENCES

[1] McGregor, D.P. *Curr. Opin. Pharmacol.*, **2008**, *8*, 616-619.
[2] Chan, D.C. Kim, P.S. HIV entry and its inhibition. *Cell,* **1998**, *93,* 681-684.
[3] Chan, D.C.; Fass, D.; Berger, J.M.; Kim, P.S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell,* **1997**, *89*, 263-273.
[4] Debnath, A.K. Progress in identifying peptides and small molecule inhibitor targeted to gp41 of HIV-1. *Exp. Opin. Investigl. Drugs*, **2006**, *15,* 465-478.
[5] Wild, C.; Dubay, J.W.; Greenwell, T. Propensity for a leucine zipper-like domain of the human immunodeficiency virus type 1 gp41 to form oligomers correlates with a role in virus-induced fusion rather than assembly of the glycoprotein complex. *Proc. Natl. Acad. Sci. USA,* **1994**, *91*, 12676-12680.
[6] Krall, A.; Brunn, J.; Kankanala, S.; Peters, M.H. A simple contact mapping algorithm for identifying potential peptide mimetics in

protein-protein interaction partners. Proteins. Structure, 17 Functions. *Bioinformatics*, **2014**, *82,* 2253-2262.

[7] Peters, M.H. The Smoluchowski diffusion equation for structured macromolecules near structured surfaces. *J. Chem. Phys*., **2000**, *112,* 5488-5498.

[8] Zhang, Y.; Peters, M.H.; Li, Y. Nonequilibrium multiple time scale dynamic simulation of receptorligand interactions in structured protein systems proteins: Structure, Function. *Genetics,* **2003**; *52*: 339-348.

[9] Peters, M.H. Langevin dynamics for the transport of flexible biological macromolecules in confined geometries. *J. Chem. Phys.,* **2011**; *134*, 025105-025116.

[10] Duan, Y.; Wu, C.; Chowdhury, S.; Lee, MC.; Xiong,G.; Zhang, W. A point charge force field for molecular mechanics simulations of proteins based on condensed phase quantum mechanical calculations. *J. Comput. Chem*., **2003**, *24*(16), 1999-2012.

[11] Ramstein, J.; Lavery, R. Energetic coupling between DNA bending and base pair opening. *Proc. Natl. Acad. Sci. U.S.A*., **1988,** *85*,7231-7235.

[12] Smith, P.E.; Pettitt, B.M. Modeling solvent in biomolecular systems. *J. Phys. Chem.,* **1994**, *98*, 9700-9711.

[13] Hasel, W.; Hendrickson, T.F.; Clark Still, W. A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahed. Comput. Methodol*., **1988**; *1*, 103-116.

[14] Gropp, W.; Lusk, E.; Skjellum, A. Using MPI. The MIT Press, Cambridge MA, **1999**.

[15] Davey, S. State of the world's vaccines and immunization. The World Health Organization, **2006**.