17th European Symposium on Computer Aided Process Engineering – ESCAPE17
V. Plesu and P.S. Agachi (Editors)
© 2007 Elsevier B.V. All rights reserved.

A New De Novo Approach for Optimizing Peptides that Inhibit HIV-1 Entry

Ho Ki Fung,[a] Christodoulos A. Floudas,[a] Martin S. Taylor,[b] and Robert F. Siliciano[b]

[a]*Department of Chemical Engineering, Princeton University, Princeton, NJ 08540, USA, floudas@titan.princeton.edu*
[b]*School of Medicine, Johns Hopkins University, Baltimore, MD21205*

Abstract

A new de novo protein design framework and its application to the redesign of an HIV-1 entry peptide inhibitor is presented.

Keywords

Peptide and protein design and discovery; Drug design; In silico sequence selection; structure prediction; de novo protein design; optimization

1. Introduction

Like most surface glycoproteins of enveloped viruses, the human immunodeficiency virus type 1 (HIV-1) envelop glycoprotein, which consists of two subunits gp120 and gp41, plays a vital role in the attachment, fusion, and entry events of host cell infection. Gp120 determines viral tropism by binding to the target cell receptor CD4 and other chemokine receptors (CCR5 or CXCR4 or both). This leads to conformational change in gp41 and the subsequent exposure of the fusion peptide, which fuses the viral and host cell membranes [1-3].
Treatment of AIDS was traditionally based on nucleoside analog reverse transcriptase and protease inhibitors, which exhibited problems of high cost, metabolic side-effects in patients, and drug resistance [2]. However, recently an

anti-HIV drug appeared in the market which functions by a different mechanism. It is a linear 36-residue peptide called enfuvirtide (or the commercial name Fuzeon) marketed jointly by Roche and Trimeris in 2003. It inhibits HIV-1 gp41 and prevents viral entry into the host cell.
The objective of our work is to de novo design an HIV-1 gp41 inhibitor that is even shorter than Fuzeon. At the outset, through literature search we found that the Kim's group had performed experiments on some potent short constrained inhibitors that bind to the hydrophobic pocket of gp41 [4]. Out of six peptides tested, they found the best crosslinked 14-residue inhibitor, C14linkmid, to have an IC_{50} value of 35 μM for cell-cell fusion. Most importantly, the crystal structure of the bound complex was elucidated already, which provides an excellent design template for us to initiate the design.

2. Our de novo protein design framework

Our two-stage de novo protein design framework not only selects and ranks amino acid sequences for a particular fold using a novel integer linear programming (ILP) model, but also validates the specificity to the fold for these sequences based on the full-atomistic forcefield AMBER [5]. The two stages are outlined as below:

2.1. Stage one: in silico sequence selection

The ILP model we use for sequence selection into a single template structure, which is the most computationally efficient one among 13 equivalent formulations we studied, takes the form:

$$\min_{y_i^j, y_k^l} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=i+1}^{n} \sum_{l=1}^{m_k} E_{ik}^{jl}(x_i, x_k) w_{ik}^{jl}$$

$$\text{s.t.} \quad \sum_{j=1}^{m_i} y_i^j = 1 \quad \forall i$$

$$\sum_{j=1}^{m_i} w_{ik}^{jl} = y_k^l \quad \forall i, k > i, l \qquad (1)$$

$$\sum_{l=1}^{m_k} w_{ik}^{jl} = y_i^j \quad \forall i, k > i, j$$

$$y_i^j, y_k^l, w_{ik}^{jl} = 0 - 1 \quad \forall i, j, k > i, l$$

Set $i = 1,...,n$ defines the number of residue positions along the backbone. At each position *i* there can be a set of mutations represented by $j\{i\} = 1,...,m_i$, where for the general case $m_i = 20 \; \forall i$. The equivalent sets $k \equiv i$ and $l \equiv j$ are

defined, and $k > i$ is required to represent all unique pairwise interactions. Binary variables y_i^j and y_k^l are introduced to indicate the possible mutations at a given position. Specifically, variable y_i^j (y_k^l) will be one if position i (k) is occupied by amino acid j (l), and zero otherwise. The composition constraints require that there is exactly one type of amino acid at each position. The pairwise energy interaction parameters E_{ik}^{jl} were empirically derived by solving a linear programming parameter estimation problem, which restricts the low energy high resolution decoys for a large training set of proteins to be ranked energetically less favorable than their native conformations [8].

2.2. Stage two: approximate method for fold validation

Driven by the full atomistic forcefield AMBER [5], simulated annealing calculations are performed for an ensemble of several hundred random structures generated for each sequence from stage one using CYANA 2.1 [9,10] within the upper and lower bounds on C^{α}-C^{α} distances and dihedral angles input by the user. This feature allows our framework to observe true backbone flexibility [11]. The TINKER package [12] is subsequently used for local energy minimization of these conformers. A fold specificity factor is finally computed for each sequence using the following equation:

$$f_{specificity} = \frac{\sum_{i\in \text{ new sequence conformers}} \exp(-\beta E_i)}{\sum_{i\in \text{ native sequence conformers}} \exp(-\beta E_i)} \tag{2}$$

3. The de novo design

3.1. Design template

The crystal structure of C14linkmid bound to the hydrophobic core of gp41 (PDB code: 1GZL), as elucidated by [4] at a resolution of 1.9Å, was shown in Figure 1. Only chain A and chain C in the PDB file are shown in the diagram and used for the design. Both chains exist in helical form in the complex. The crosslink, made by diaminoalkane, is between position 629 and position 636, and is supposed to constrain the C-peptide helix to reduce its entropy loss upon binding [4]. Energy minimization is driven by the high resolution centroid-centroid forcefield [8]. In the second stage, the bounds on the angles and distances input into the CYANA 2.1 package were $\pm 10^{o}$ around the template and ±10% of those in the template respectively.

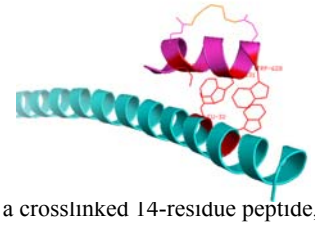

Fig. 1 Crystal structure of a crosslinked 14-residue peptide, C14linkmid (pink), bound to the hydrophobic core of gp41(cyan) [7]. This provides the template for the de novo design of the gp41 inhibitor.

3.2. Mutation set

While positions 629 and 636 are fixed at their native GLN to preserve the diaminoalkane crosslinker, other positions are varied with the mutation set selected to preserve the nature of the native residue (see Table 1).

Table 1. Mutations set of *in silico* sequence selection for the redesign of an HIV-1 entry inhibitor.

Positions	Native residue	Allowed mutations
628	W	A,I,L,M,F,Y,W,V
629	Q	Q
630	E	R,N,D,Q,E,G,H,K,S,T
631	W	A,I,L,M,F,Y,W,V
632	D	R,N,D,Q,E,G,H,K,S,T
633	R	R,N,D,Q,E,G,H,K,S,T,C
634	E	R,N,D,Q,E,G,H,K,S,T
635	I	A,I,L,M,F,Y,W,V
636	Q	Q
637	N	R,N,D,Q,E,G,H,K,S,T
638	Y	A,I,L,M,F,Y,W,V
639	T	R,N,D,Q,E,G,H,K,S,T

3.3. Biological constraints

Two case studies, which differ by the charge restricted on the segment from position 630 to position 635, were performed. One fixes the charge to be the same as native and the other allows the charge to vary between ±1 of native. These constraints were implemented in the form of linear biological constraints. In each case study 500 sequences were generated in the sequence selection stage, and their fold specificities were confirmed using CYANA 2.1 and TINKER. The requisite biological constraints are:

$$\sum_i y_i^{Arg} + \sum_i y_i^{Lys} - \sum_i y_i^{Asp} - \sum_i y_i^{Glu} = -2 \quad \forall 630 \le i \le 635 \tag{3}$$

and

$$-3 \le \sum_i y_i^{Arg} + \sum_i y_i^{Lys} - \sum_i y_i^{Asp} - \sum_i y_i^{Glu} \le -1 \quad \forall 630 \le i \le 635 \tag{4}$$

respectively. In addition, an upper bound of 5 is imposed on the total number of mutations, which translates into the equation:

$$\sum_{i=1}^{n} \sum_{j=1}^{m_i} y_i^j \le 5 \quad \forall i, j \notin \text{native residues} \tag{5}$$

3.4. Results

The top 10 sequences out of all those from the sequence selection stage ranked according to fold specificity are listed in Table 2. Results from the two case studies observe roughly the same pattern of: -W^{628}-Q^{629}-(D/E)^{630}-W^{631}-(D/R)^{632}-(R/N)^{633}-(E/N/D)^{634}-(W/Y)^{635}-Q^{636} (R/Q/N)^{637}-(Y/W/L)^{638}-R^{639}-. High degree of consistency exists for the preferences at position 628, position 631, and position 639: the first two positions do not prefer to be mutated under the conditions we imposed, and the third one strongly prefers ARG. It is interesting to note that the batch of sequences with their charge on [630,635] fixed at native in the sequence selection model actually performed better in the fold specificity stage than their counterparts which are allowed to vary between ±1 of the native charge on [630,635]. This can be seen by noticing that the same sequence with the highest fold specificity for the native charge ±1 batch only ranks second in the native charge batch.

Table 2. Top 10 sequences ranked according to fold specificity for the HIV-1 gp41 inhibitor.

native charge on [630,635]												
Fold specificity rank	Positions											
	628	629	630	631	632	633	634	635	636	637	638	639
	W	Q	E	W	D	R	E	I	Q	N	Y	T
1	W	Q	**D**	W	D	R	E	**W**	Q	**R**	Y	**R**
2	W	Q	E	W	**R**	**D**	E	**W**	Q	**R**	Y	**R**
3	W	Q	E	W	D	R	**D**	**Y**	Q	**R**	**W**	**R**
4	W	Q	**D**	W	D	R	**D**	**Y**	Q	N	**W**	**R**
5	W	Q	E	W	**R**	**E**	E	**W**	Q	**R**	Y	**R**
6	W	Q	**N**	W	D	**N**	E	**W**	Q	**R**	Y	**R**
7	W	Q	**Q**	W	D	**N**	E	**W**	Q	**R**	Y	**R**
8	W	Q	E	W	D	R	E	**W**	Q	**Q**	**L**	**R**
9	W	Q	**D**	W	D	R	E	**W**	Q	**Q**	Y	**R**
10	W	Q	**D**	W	D	R	E	**W**	Q	**Q**	**L**	**R**
native charge ±1 on [630,635]												

1	W	Q	E	W	**R**	**D**	E	**W**	Q	**R**	Y	**R**
2	W	Q	**D**	W	D	R	**N**	**W**	Q	N	**L**	**R**
3	W	Q	**D**	W	D	R	**N**	**L**	Q	N	**W**	**R**
4	W	Q	E	W	**R**	**N**	E	**W**	Q	**R**	Y	**R**
5	W	Q	**D**	W	D	R	**N**	**Y**	Q	N	**W**	**R**
6	W	Q	E	W	**R**	**Q**	E	**W**	Q	**R**	Y	**R**
7	W	Q	**D**	W	D	R	**N**	**W**	Q	**Q**	Y	**R**
8	W	Q	E	W	**E**	R	**N**	**Y**	Q	N	**W**	**R**
9	W	Q	**D**	W	**Q**	R	E	**W**	Q	**Q**	Y	**R**
10	W	Q	E	W	D	R	**N**	**Y**	Q	**Q**	**W**	**R**

4. Conclusions

In this paper, we predicted active analogs for an anti HIV-1 entry peptide inhibitor using our novel framework for de novo protein design.

Acknowledgements

CAF acknowledges financial support from the National Science Foundation, the National Institutes of Health and the US Environmental Protection Agency (R01 GM52032, R24 GM069736, GAD R 832721-010). This work has not been reviewed by and does not represent the opinions of USEPA.

References

1. D.C. Chan, D. Fass, J.M. Berger, and P.S. Kim, Cell, 89 (1997) 263
2. P.A. Galanakis, G.A. Spyroulias, A. Rizos, P. Samolis, and E. Krambovitis, Curr. Med. Chem., 12 (2005) 1551
3. C. Huang, M. Tang, M. Zhang, S. Majeed, E. Montabana, R. L. Stanfield, D.S. Dimitrov, B. Korber, J. Sodroski, I.A. Wilson, R. Wyatt, and P.D. Kwong, Science, 310 (2005) 1025
4. S.K. Sia, P.A. Carr, A.G. Cochran, V.N. Malashkevich, and P.S. Kim, PNAS, 99 (2002) 14664
5. W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman, J. Am. Chem. Soc., 117 (1995) 5179
6. H.K. Fung, S. Rao, C.A. Floudas, O. Prokopyev, P.M. Pardalos, and F. Rendl, J. Comb. Optim., 10 (2005) 41
7. H.K. Fung, M.S. Taylor, and C.A. Floudas, Optim. Methods & Software, 22 (2007) 51
8. R. Rajgaria, S.R. McAllister, and C.A. Floudas, Proteins, 65 (2006) 726
9. P. Guntert, C. Mumenthaler, and K. Wuthrich, J. Mol. Bio., 273 (1997) 283
10. P. Guntert, J. Mol. Bio., 278 (2004) 353
11. C.A. Floudas, AICHE J., 51 (2005) 1872
12. J. Ponder, TINKER, software tools for molecular design. Washington University School of Medicine. St. Louis, MO., USA, 1998.