NIH Public Access
Author Manuscript
Curr Drug Targets. Author manuscript; available in PMC 2010 June 30.

Published in final edited form as:
Curr Drug Targets. 2010 March ; 11(3): 264–278.

Computational Methods for *De novo* Protein Design and its Applications to the Human Immunodeficiency Virus 1, Purine Nucleoside Phosphorylase, Ubiquitin Specific Protease 7, and Histone Demethylases

M.L. Bellows and **C.A. Floudas***
Department of Chemical Engineering, Princeton University, Princeton, NJ, USA

Abstract

This paper provides an overview of computational *de novo* protein design methods, highlighting recent advances and successes. Four protein systems are described that are important targets for drug design: human immunodeficiency virus 1, purine nucleoside phosphorylase, ubiquitin specific protease 7, and histone demethylases. Target areas for drug design for each protein are described, along with known inhibitors, focusing on peptidic inhibitors, but also describing some small-molecule inhibitors. Computational design methods that have been employed in elucidating these inhibitors for each protein are outlined, along with steps that can be taken in order to apply computational protein design to a system that has mainly used experimental methods to date.

Keywords

Computational protein design; human immunodeficiency virus 1; purine nucleoside phosphorylase; ubiquitin specific protease 7; histone demethylases

1 INTRODUCTION

De novo protein design, also referred to as the inverse protein folding problem, is the determination of an amino acid sequence, or set of sequences, that will fold into a given 3-dimensional (3D) protein template, which may be fixed or flexible. While the protein folding problem seeks to identify the one lowest energy conformation for a given amino acid sequence, the protein design problem exhibits degeneracy, in that many amino acid sequences fold into a given template, with the different sequences giving different properties (activity, specificity) to the protein. It therefore has a wide range of applications, from improved design of inhibitors and new sequences with increased stability to the design of catalytic sites of enzymes and drug discovery [1–3].

Until recently, protein design consisted primarily of experimental techniques such as rational design, mutagenesis, and directed evolution. Although these methods produce good results, they are restrictive because of the limited sequence search space (estimated to be only 10^3 – 10^6). Computational methods, on the other hand, can increase this search space to 10^{128}, making computational protein design more popular.

© 2010 Bentham Science Publishers Ltd.

*Address correspondence to this author at the Department of Chemical Engineering, Princeton University, Princeton, NJ 08544, USA; Tel: (609) 258-4595; Fax: (609) 258-0211; floudas@princeton.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Several successes in protein design include increasing the stability and specificity of a target protein [4–6] to locking proteins into useful conformations [7]. Computational methods aid the protein design process by determining folding kinetics [4,8] and protein-ligand interactions [9]. They help with protein docking [10–12] and assist peptide and protein drug discovery [13–15].

Despite these successes, there are limitations. Currently, it is very difficult to design a protein consisting of 100 or more amino acids. If one assumes an average of 100 rotamers for all 20 amino acids at each position, this problem reaches a complexity of $100^{100} = 10^{200}$. Coupled with the NP-hard nature [16,17] of the problem, designing larger proteins (> 100 amino acids) proves a great challenge. In addition to improving the computational efficiency of *de novo* design algorithms, another challenge is to incorporate true backbone flexibility. These two challenges are interrelated, as incorporating backbone flexibility increases the computational complexity of an algorithm.

The next few sections outline the methodologies and recent advances in computational protein design, using both fixed and flexible backbone templates and describing both deterministic methods and stochastic methods.

2 COMPUTATIONAL METHODS

The various computational methods employed for *de novo* protein design belong to two classes: those that use fixed backbone templates and those that use flexible backbone templates. A fixed backbone template consists of fixed backbone atom coordinates and fixed rotamer conformations. This was first proposed by Ponder and Richards [18]. This is usually the case when only an X-ray crystal structure of the design template is known. Flexible backbone templates, on the other hand, are more true to nature, as protein structures are inherently flexible. Flexible templates can be a set of fixed backbone atom coordinates, such as the set of structure models obtained from NMR structure determination. Instead of a set of fixed atoms coordinates, the backbone atoms can take on a range of values between specified bounds. The rotamers can also consist of a set of discrete rotamers for each residue or the rotamer angles can be allowed to vary between a specified range.

2.1 Fixed Backbone Templates

2.1.1 Deterministic Methods—Deterministic algorithms include those that use (a) dead end elimination (DEE) methods, (b) self-consistent mean field (SCMF) methods, (c) power law (PL) methods or (d) those that utilize quadratic assignment-like models coupled with deterministic global optimization. The deterministic methods (a), (b), and (c) use a discrete set of rotamers, which are used for tractability of the search problem, while methods (d) can use either a discrete or a continuous set of rotamers.

DEE methods historically use fixed-backbone templates and a discrete set of rotamers [19–23]. DEE works by systematically eliminating rotamers that cannot be part of the sequence with the lowest free energy. The energy function used in DEE is a combination of individual terms (rotamer to template) and pairwise terms (rotamer to rotamer) and is given by Eq. 1, where $E(i_a)$ is the rotamer-template energy for rotamer i_a of amino acid i. Likewise, $E(i_a,j_b)$ is the pairwise rotamer-rotamer energy between rotamer i_a of amino acid i and rotamer j_b of amino acid j. Finally, N is the total number of amino acid positions.

$$E=\sum_{i=1}^{N} E(i_a)+\sum_{i=1}^{N-1}\sum_{j>i}^{N} E(i_a, j_b) \tag{1}$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The general idea behind DEE is that if the pairwise energy, $E(i_a,j_b)$, between rotamer i_a and j_b is greater than that between i_c and j_b, for all j_b in some set $\{B\}$, then the rotamer i_a can be removed, or pruned, from the set because it cannot be part of the global energy minimum conformation. This relationship is expressed by Eq. 2. This equation can be made more tractable by the bounds implied by Eq. 2, given in Eq. 3.

$$E(i_a)+\sum_{j\neq i}^{N}E(i_a,j_b)>E(i_c)+\sum_{j\neq i}^{N}E(i_c,j_b),\quad \forall\{B\} \tag{2}$$

$$E(i_a)+\sum_{j\neq i}^{N}\min_{b}E(i_a,j_b)>E(i_c)+\sum_{j\neq i}^{N}\max_{b}E(i_c,j_b) \tag{3}$$

Recent advances to the DEE algorithm allow faster computation of larger protein design problems. BroMAP (branch-and-bound rotamer optimization using MAP estimation) expands smaller search trees while performing moderate computation in each node, reducing the total run time [24]. The BroMAP method generates lower bounds using approximate maximum-a-posteriori (MAP) estimation. Another advance, called MinDEE (for minimized-DEE) not only prunes rigid rotamers as described by the original DEE algorithm, but guarantees that rotamers belonging to the energy minimized global minimum energy conformation will not be pruned [25]. This is important because conformations that survive the original DEE algorithm are often subjected to further rotamer energy minimizations. When this occurs, the combined process is heuristic and the provable guarantee is lost. Other modifications to DEE include X-DEE (extended DEE), which gives gap-free lists of low-energy states for a given energy range and was applied to the determination of protonation states of a protein [26], and type-variant DEE, which can be used in multistate protein design [27]. Further descriptions of DEE modifications and successes can be found in Fung *et al*. [3].

SCMF methods involve predicting the values of elements of a conformational matrix $P(i,a)$ to determine the probability that amino acid i will adopt a rotamer conformation a. Koehl and Delarue [28] were among the first to describe such a method in protein design. This iterative process begins with an initial guess for the conformational matrix, giving equal probability to all rotamers (Eq. 4).

$$P(i,a)=\frac{1}{A},\quad a=1,2,\ldots,A \tag{4}$$

The energy, based upon the mean field potential $E(i,a)$, is then calculated, which is dependent upon the conformational matrix.

$$E(i,a)=U(x_{i_a})+U(x_{i_a},x_0)+\sum_{j=1,j\neq i}^{N}\sum_{b=1}^{B}P(j,b)\,U\left(x_{i_a},x_{j_b}\right) \tag{5}$$

In equation 5, x_0 is the coordinates of atoms in the fixed template, while x_{ia} and x_{jb} are the coordinates of atoms of position i with rotamer conformation a and position j with rotamer conformation b, respectively. The potential energy, U, can described by classical Lennard-Jones potential. Once the energy has been calculated, the conformational matrix is then updated (Eq. 6) and this is repeated until convergence is obtained. It should be noted, however, that

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

although this method is deterministic, it does not guarantee convergence to the global minimum solution, as DEE does.

$$P(i,a) = \frac{e^{\frac{E(i,a)}{RT}}}{\sum_{a=1}^{A} e^{\frac{E(i,a)}{RT}}} \qquad (6)$$

The SCMF method was extended by the Saven group [29–32], introducing information theory concepts and formulating it as an optimization problem. They maximized the sequence entropy subject to composition and mean-field energy constraints. Bhattacherjee and Biswas [33] recently developed a SCMF based theory in order to consider the effect of correlated mutations by evaluating site-specific amino acid pair probabilities in a library of sequences.

The power law (PL) method is based on the similarity of a topology optimization problem and a protein design problem [34]. Instead of solving the topology optimization problem in discrete space, however, it is solved in continuous space in order to reduce the intensive computation required. The problem introduced by this method is the artifact of obtaining non-discrete optimal solutions. This is solved by introducing solid isotropic material with penalization (SIMP). The rigidity of a material is expressed as an artificial power law in continuous space. Using this formulation, the interaction energy between the known amino acid at position *j* and the unknown amino acid at position *i* is given by Eq. 7.

$$E_{ij}\left(x_j\right) = -E_{ij}^{0} x_i^{p} \qquad (7)$$

The energy value between the amino acids at sites *i* and *j* is given by E^0_{ij}. This is also referred to as the stiffness of the material, in terms of the SIMP formulation. The state of the undetermined amino acid at position *i* is represented by the variable x^p_i. Equation 7 is used as the energy function in the quadratic programming problem given by Eq. 8, where B_{ij} is constructed in order to obtain equality constraints that limit the selection of one amino acid per position, *N* is the number of residue positions, *m* is the number of amino acid types available, and Q_{ij} contains the interaction energy between residues.

$$\begin{aligned} &\min \quad x_i^p Q_{ij} x_j^p \\ &\text{subject to} \quad \sum_{j=1}^{Nm} B_{ij} x_j = c_i \quad \text{for} \quad i=1,2,\ldots,N+m \\ &\qquad 0 \le x_i \le 1 \quad \text{for} \quad i=1,2,\ldots,Nm \end{aligned} \qquad (8)$$

Klepeis *et al*. [5,6] proposed a novel approach to *de novo* protein design using a quadratic assignment-like global optimization model to select sequences. The original integer linear programming (ILP) model uses a single fixed backbone template structure. Further advances by Fung *et al*. [16,35,36], incorporating backbone flexibility into the model, are presented in subsequent sections.

2.1.2 Stochastic Methods—Because the protein design problem is NP-hard [16,17], deterministic methods may reach a limit and stochastic methods are employed to find locally

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

optimal solutions. These methods include genetic algorithms, Monte Carlo sampling, and combinatorial library methods.

Monte Carlo (MC) methods work by performing a mutation of a residue at a certain position and calculating the energies of the sequence before and after the mutation. The mutation will be accepted if it lowers the energy, or, if the energy is higher, it will accept the mutation based upon the Metropolis criterion [37] (Eq. 9).

$$P_{accept}=min\left\{1, e^{-\beta\Delta E}\right\}, \quad \beta=\frac{1}{kT} \tag{9}$$

In Eq. 9, P_{accept} is the probability that a move will be accepted, ΔE is the change in energy before and after adding a mutation, T is the temperature, and k is the Boltzmann constant. Rosetta Design [38] is a MC based algorithm. This algorithm has options for both fixed-backbone template design and flexible-backbone template design (considering an ensemble of fixed backbones), however, only local minima are obtained, not necessarily the global minimum. Two complementary Markov chain Monte Carlo (MCMC) protein design methods were developed by Thomas *et al*. [39]. One uses constrained shuffling which generates a small set of high-likelihood new sequences. The other uses component sampling, which generates all the sequences that meet given constraints. Zou and Saven [40] combine MC methods with SCMF methods to bias the sequence selection using predetermined sequence identity probabilities. The probabilities are determined using SCMF methods, while the sequence selection is performed using MC methods. Successes of MC methods consist of new sequences for the fixed backbones of the β1 domain of protein G, λ repressor, and sperm whale myoglobin [41,42], and the redesign of the src SH3 domain, the λ repressor, U_1A, protein L, tenascin, procarboxypeptidase, acylphosphatase, S6, and FKBP12 [4].

Genetic algorithms, in comparison, generate a number of random amino acid sequences and exchange them for a fixed template [43–47]. This method originated in genetics and evolution [48]. High-energy sequences are eliminated, while low energy ones form hybrids. The algorithm terminates when the solution converges. Genetic algorithms have been used to design ligands for prolyl oligopeptidase, p53, and DNA gyrase [49] and to design short peptides with high stability that resemble the antibody epitopes of thrombin and blood coagulation factor VIII [50].

Finally, combinatorial library methods attempt to maximize the entropy subject to a set of constraints [30,32,51–57].

2.2 Flexible Backbone Templates

2.2.1 Deterministic Methods—Georgiev and Donald [58] incorporated backbone flexibility into the DEE method. As previously mentioned, the original DEE algorithm does not take backbone flexibility into account. Modifications to DEE have attempted to do so, but the modified DEE algorithms don't have the provable guarantee of eliminating rotamers that are not part of the global minimum energy conformation. Georgiev and Donald [58] provided modifications in which this provable guarantee is kept intact and allows for backbone flexibility. Equation 3 is expressed equivalently as

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\begin{aligned} & E_{\Theta}(i_a) + \sum_{j \neq i}^{N} \min_{b} E_{\Theta}(i_a, j_b) \\ - \quad & E_t - \sum_{j \neq i}^{N} \max_{b} E_{\Delta}(j_b) - \sum_{j \neq i}^{N} \sum_{k \neq i, k > j}^{N} \max_{b,d} E_{\Delta}(j_b, k_d) \\ > \quad & E_{\Phi}(i_c) + \sum_{j \neq i}^{N} \max_{b} E_{\Phi}(i_c, j_b), \quad \forall \{B\}, \{D\}, \end{aligned} \tag{10}$$

where E_t represents the difference between lower and upper bounds on the template energy. E_t is the interval of possible template energies. $E_{\Theta}(i_a)$ represents a lower bound on the sum of the energy interactions between the atoms of rotamer i_a and the energy interactions between the atoms of rotamer i_a and the template atoms. Similarly, $E_{\Theta}(i_a, j_b)$ is defined as the lower bound pairwise energy between rotamers i_a and j_b. $E_{\Theta}(i_a)$ and $E_{\Theta}(i_c, j_b)$ are defined as the maximum bounds, similar to the descriptions for E_{Θ}. Finally, the interval terms, E_{Δ}, are defined as

$$\begin{aligned} E_{\Delta}(i_a) &= E_{\Phi}(i_a) - E_{\Theta}(i_a) \\ E_{\Delta}(i_a, j_b) &= E_{\Phi}(i_a, j_b) - E_{\Theta}(i_a, j_b). \end{aligned} \tag{11}$$

What Eq. 10 essentially does is, given a compact space of backbone conformations, compare a lower bound on the energy achievable when amino acid i has a rotamer identity of i_a to an upper bound on the energy achievable with a competing rotamer identity say i_e, of the same amino acid. Georgiev *et al.* [59] also introduced backbone flexibility into the DEE algorithm by incorporating back rub motions. Back rub motions are side-chain coupled local backbone motion and is commonly observed in proteins.

A novel modeling and computational framework introduces a quadratic assignment-like global optimization model [5,6,16,35,36,60] to select sequences. It incorporates backbone flexibility by describing the template in terms of distance bins, with each bin given the same energy. A single backbone template or a set of multiple templates obtained from NMR experiments, molecular dynamics simulations, or docking calculations can be used. From this, a rank ordered list of sequences is obtained. Equation 12 shows the basic mixed-integer linear programming (MILP) model developed by Fung *et al*. [16].

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

$$\begin{aligned} \text{subject to} \quad & \sum_{j=1}^{m_i} y_i^j = 1 \quad \forall \quad i \\ & \sum_{j=1}^{m_i} w_{ik}^{jl} = y_k^l \quad \forall \quad i, k>i, l \\ & \sum_{l=1}^{m_k} w_{ik}^{jl} = y_i^j \quad \forall \quad i, k>i, j \\ y_i^j, \quad & y_k^l, w_{ik}^{jl} \in \{0, 1\} \quad \forall \quad i, j, k>i, l \end{aligned} \tag{12}$$

The model minimizes the sum of all pairwise energy interactions. The energy term, $E^{jl}{}_{ik}$, is dependent upon the distance between residue i and residue k. The distances can either be the C^{α}–C^{α} distances or the centroid-centroid distances. Backbone flexibility is introduced in two

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

ways, the first is by discretizing the energy function into bins based upon the distance between a pair of residues. Therefore, a pair of residues may take on a range of distances that fall within a particular distance bin and still retain the same energy value. The second way flexibility is introduced is by expanding the model depicted in Eq. 12 to allow for the distances spanned by two resides to fall across multiple bins. This is achieved in two ways: by introducing a weighted average energy term or by introducing a binary variable that selects a particular distance bin for a pair of residues depending upon all the bins the pair spans. Details of these model formulations can be found in Fung *et al*. [36].

The sequences selected using the MILP models are subsequently re-ranked in a validation stage that consists of fold specificity calculations, approximate binding affinity calculations, or both. Both validation methods incorporate true backbone flexibility. A fold specificity for each sequence is calculated by generating an ensemble of structures of the mutant sequence and of the native sequence. The ensemble incorporates backbone flexibility by allowing a continuous range between upper and lower bounds on the C^{α}-C^{α} distances and ϕ and ψ angles. The fold specificity is determined according to Eq. 13, where $\beta = 1/k_B T$.

$$f_{specificity} = \frac{\sum_{i \in novel} \exp^{-\beta E_i}}{\sum_{i \in native} \exp^{-\beta E_i}} \tag{13}$$

The set *novel* contains the ensemble of structures of the mutant sequence, while the set *native* contains the ensemble of structures of the native sequence. The fold specificity is a measure of how likely a mutant sequence will fold into the native structure. A higher fold specificity indicates a greater likelihood. Fold specificities are calculated based on (a) the full atomistic forcefield either through the ASTRO-FOLD approach [61–70] and deterministic global optimization [71–77] or (b) the AMBER forcefield *via* a novel NMR structure refinement method [3,35].

The second validation method is most applicable when designing proteins or peptides to bind to another target protein. This method re-ranks the sequences from the selection stage by calculating an approximate binding affinity (Eq. 14).

$$K* = \frac{q_{PL}}{q_P q_L} \tag{14}$$

The approximate binding affinity (K*) is a ratio of the partition functions of the complex (*PL*), target protein (*P*), and peptide or ligand (*L*). The partition functions are calculated using rotamerically-based conformation ensembles for each species. Backbone flexibility is introduced in this validation method by the way in which the ensembles are constructed, allowing the peptide ensemble to take on a number of discrete backbone and rotamer configurations and allowing the complex ensemble to take on a number of discrete peptide backbone and rotamer configurations and docked configurations. Details of the approximate binding affinity calculation can be found in Bellows *et al*. [78].

In addition to the two methods mentioned above, groups have applied the methodologies pertaining to fixed backbone templates to multiple discrete templates in order to add some flexibility to their design. Su and Mayo [79] and Ross *et al.* [80] used the DEE algorithm with fixed backbone template for each of their backbone templates in order to generate several sets of perturbed backbones from the native structure of their design. Kono and Saven [30] used

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

their SCMF method on a set of fixed backbone templates in order to design sequences that would be robust to distance changes in the template.

2.2.2 Stochastic Methods—Monte Carlo simulations have been used to generate multiple discrete templates based upon an initial fixed-backbone template [7,81,82]. Using this, the Pande group [81,82] used genetic algorithms with the discrete templates to design sequences that exhibited higher diversity than the corresponding natural sequence alignments. Kraemer-Pecore *et al*. [7] also used genetic algorithms with multiple discrete templates to identity a sequence that folded into the WW domain.

Another example of using genetic algorithms with Monte Carlo simulations involves iterating between sequence space and structure space. Desjarlais and Handel [83] generated multiple discrete backbone templates and used a genetic algorithm to exchange not only rotamers but also backbone torsional information in recombination. The backbone structures were then refined using Mote Carlo. Using this method, they were able to design three new core variants of the protein 434 cro.

Monte Carlo methods have also been used by Kuhlman *et al*. [84] and Saunders and Baker [85] where the method again uses a set of discrete backbone templates. A Monte Carlo search gives the lowest energy sequence for each template, then performs structure prediction of the sequences to allow for shifts in structure space, and continues to iterate between sequence space and structure space for a desired number of iterations. They have successfully used this method to design a new sequence for Top 7, a 93 residue α/β protein with a novel fold.

Finally, Harbury *et al*. [86–88] have used a continuum backbone template with discrete rotamers to design a family of α-helical bundle proteins with a right-handed superhelical twist. Their method works by creating a continuum template using backbone parameterization and then performing a sequence search from rotamer libraries. Backbone movements are introduced by treating the parameters as variables during the sequence search for energy minimization.

3 APPLICATION DOMAINS

The computational protein design methods described can be applied to a number of systems. Four systems are outlined below. All involve using *de novo* protein design methods to design inhibitors that target the human immunodeficiency virus 1, purine nucleoside phosphorylase, ubiquitin specific protease 7, and histone demethylases. Some systems have a long history of using computational *de novo* design to discover inhibitors, while others have been treated with only experimental techniques. The following sections review the work that has been done in each of the four areas with regard to target identification, inhibitor design and discovery, and what computational methods, if any, have been applied to the system.

3.1 Human Immunodeficiency Virus 1

The human immunodeficiency virus 1 (HIV-1) is the virus that causes AIDS, the acquired immunodeficiency syndrome. While there is no vaccine for HIV-1, there are treatments for those who have been infected. Called antiretroviral treatment, it reduces the risks associated with HIV if begun right after infection. The treatment was first developed in 1995 and consists of a cocktail of at least three drugs. The cocktail usually includes two nucleoside analogue reverse transcriptase inhibitors and either a protease inhibitor or a non-nucleoside reverse transcriptase inhibitor [89]. These drugs work by targeting different aspects of the HIV replication process, leading to a rise in CD4 T cells (the target of HIV). However this is then followed by an increase in drug-resistant variants of HIV.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Recently, new studies have focused on inhibiting HIV-1 entry into cells using fusion inhibitors. This is promising because the inhibitors can potentially target the numerous intermediates that are formed during viral entry into the cell [90]. HIV-1 infects cells using a multistep process. The process begins when gp160, an envelope glycoprotein, proteolytically cleaves into a surface subunit (gp120) and a transmembrane subunit (gp41) [91]. This is done by a cellular convertase [92]. In the next step, gp120 binds to the host cell receptor CD4, leading to a conformational change in gp120 and causing the extension of the V3 loop of gp120. This extension binds to chemokine receptor CCR5 or CXCR4 or both. The binding of the chemokine receptor to the V3 loop leads to the exposure of gp41. gp41 is divided into three domains: extracellular domain, transmembrane domain, and cytoplasm domain [93]. The extracellular domain can be further divided into four functional domains: fusion peptide, N-terminal heptad repeat (NHR or HR1), C-terminal heptad repeat (CHR or HR2), and a tryptophan rich region [93]. Both NHR and CHR contain a number of 4-3 heptad repeats, and these repeats generally form coiled-coil structures [94]. After gp41 has been exposed, the NHR and CHR come together to form a 6-helix bundle. Once the 6-helix bundle is formed, the viral and cell membranes fuse together and HIV has successfully infected the host cell.

3.1.1 Targets—There are two main targets of HIV-1 for fusion inhibition: gp41 and gp120. gp41 is targeted in order to prevent the formation of the 6-helix bundle, thereby preventing membrane fusion. gp120 is targeted in order to prevent the binding of the chemokine receptor. This prevents the exposure of gp41, thereby stopping membrane fusion. Targeting gp120 is more upstream in the infection process than targeting gp41.

A number of research groups have elucidated the 3D structure of gp41, which is necessary for any computational protein design. Weissenhorn *et al*. [95], Chan *et al*. [96], and Tan *et al*. [97] used X-ray crystallography to determine the structure of gp41, confirming that it forms a coiled coil structure and consists of a 6-helix bundle. Zhou *et al*. [98] have also reported the structure of gp41 in complex with a small molecule inhibitor, providing further information as to the precise location of the binding site of gp41. The small molecule inhibitor binds to the hydrophobic pocket of the NHR of gp41. This prevents the formation of the interface between the NHR and CHR. The identification of a hydrophobic pocket of the NHR of gp41 as a potential drug target is further supported by Chan *et al*. [99]. They show that the inhibitory activity of the peptides they examined depend upon their ability to bind to the hydrophobic, coiled-coil cavity. Sia *et al*. [100] elucidated the crystal structure of a complex of gp41 with a short constrained C-peptide, which also binds to the hydrophobic core of gp41 NHR. In addition, Balogh *et al*. [101] presented a novel method that gives explicit structural constraints on a small molecule inhibitor bound to the hydrophobic pocket of gp41 for NMR. The functional role of the coiled-coil domain of gp41 was explored by Wild *et al*. [102]. They compared the effects of amino acid substitutions in the coiled-coil domain on both structure and function and found a correlation between the destabilizing effects of the substitutions on the structure and virus entry.

The 3D structure of gp120, especially the V3 loop, has been more difficult to obtain compared to gp41. This is because of the high flexibility of the V3 loop, and the fact that gp120 needs to bind to CD4 in order to induce the conformational change that extends the V3 loop. Vranken *et al*. [103] determined the structure of the V3 loop using NMR. This provided 20 backbone structures, giving a flexible backbone template that can be used for computational protein design. Recently, the crystal structure of gp120 bound to CD4 and CCR5 was determined by Huang *et al*. [104]. This structure shows gp120 with the V3 loop extension and the chemokine receptor bound to the V3 loop. This provides the location of the binding site for CD4 and for the CCR5 chemokine receptor on the V3 loop. Tan and Rader [105] further analyzed the binding regions of gp120 using a flexibility analysis of all known gp120 structures. They identified two rigid regions, which may serve as stable targets for drug design.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Other studies have also attempted to determine conserved regions of gp120 and gp41, which can be targets for broadly neutralizing antibodies [106–111].

3.1.2 Inhibitors—

There are two FDA (U. S. Food and Drug Administration) fusion inhibitors on the market. The first, Fuzeon (enfuvirtide), was developed by Roche and Trimeris and has been in use since 2003 [90,112]. The second, Selzentry (maraviroc), was developed by Pfizer and approved in 2007 [113].

Fuzeon is used in patients that have developed a drug resistance to standard antiretroviral treatments. Its sequence is based on the sequence of gp41 CHR and targets gp41 NHR. Selzentry differs from Fuzeon in that it is a small molecule and targets gp120. Selzentry binds to the chemokine receptor CCR5. One issue with Selzentry is that it is only effective in patients with HIV using the CCR5 receptor. The drug has limited activity in patients with HIV using the CXCR4 receptor or both receptors.

In addition to the two FDA approved fusion inhibitors, there are many more that have been published in the open literature. Table **1** provides the names, sequences, IC_{50} values, and references for a number of HIV fusion inhibitors.

A few of the peptides contain D-amino acids, which are resistant to proteolytic digestion, thereby increasing their half-life *in vivo* [114,115,120]. Recent work has also focused on finding peptidic inhibitors of enfuvirtide-resistant HIV strains [117,121–124], which have developed due to the increased use of Fuzeon. The mechanisms of action of second-generation fusion inhibitors are being investigated [94] as well as the minimal sequence and minimal interface for fusion inhibitors [125,126]. Liang [127] provides a review of inhibitors targeting CXCR4, one of the chemokine receptors that bind to gp120. In addition, various small molecules are being investigated to find new HIV-1 inhibitors that may have increased bioavailability and reduced cost of production over peptidic inhibitors [93,101,127–131].

3.1.3 Computational Methods—

A number of computational methods have been used to design novel HIV-1 fusion inhibitors. Singh *et al*. [132] used homology modeling and docking studies to investigate the interactions between the chemokine receptor CXCR4 and flavonoids. They report that the compounds they examined may become important new antiviral drugs. Berchanski and Lapidot [133] designed new conjugates of neomycin with two arginine peptide chains to target CXCR4 using a multistep docking procedure [134]. Imai *et al*. [135] used their computer program ANTIS to design inhibitors that target CCR5. The Jiang group [136–139] used a virtual screening method based on computer modeling to identify gp41 small-molecule inhibitors. They screened 20,000 compounds and found that one of the compounds, ADS-J1, had an IC_{50} in the low micromolar range. Recently, Bellows *et al*. [140] used their *de novo* design framework to discover novel peptide inhibitors of gp41. The peptides are 12 amino acids in length and the most promising one showed an IC_{50} of 31μM in inhibition studies of HIV-1 entry into host cells.

In addition to targeting gp41 and gp120, computational methods have elucidated inhibitors that target the protein disulfide isomerase (PDI) of HIV-1 [141].

3.2 Purine Nucleoside Phosphorylase

Purine nucleoside phosphorylase (PNP) catalyzes purine metabolism, converting adenosine into adenine, inosine into hypoxanthine, and guanosine into guanine. It was discovered that people who are deficient in PNP have an immunodeficiency problem. Altered pathways of purine metabolism lead to T-cell deficiency. This is because deoxyguanosine is not properly metabolized and builds up in the blood. It is then transported and phosphorylated by T-cell deoxynucleoside kinases, forming harmful elevated levels of dGTP in the T-cells [142].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

This lead to the realization that inhibition of PNP could regulate undesirable T-cell proliferation [143]. Diseases such as rheumatoid arthritis, psoriasis, inflammatory bowl disorders, and multiple sclerosis are caused by inappropriate activation of T-cells by self-antigens and could be targets for PNP inhibitors [142].

3.2.1 Targets—A number of groups have determined the crystal structure of PNP in complex with various ligands. Canduri *et al*. [144] presented a number of crystal structures of PNP in complex with ligands. They determined the structure of human PNP in complex with guanosine at 2.80 Å resolution, with 3'-deoxyguanosine at 2.86 Å resolution, and 8-azaguanine at 2.85 Å resolution. Canduri *et al*. [145] have also determined the crystal structure of human PNP in complex with inosine and 2',3'-dideoxyinosine at 2.8 Å resolution. Finally, Canduri *et al*. [146] reported the crystal structure of human PNP in complex with hypoxanthine and sulfate ion at 2.6 Å resolution. These crystal structures provide information about the location of the binding site of PNP. The ligands described in the aforementioned complexes are the natural ligands that bind during purine metabolism.

In contrast to the structures cited above, dos Santos *et al*. [147] have reported the crystallographic structure of human PNP in complex with acyclovir. Acyclovir is a potent inhibitor of replicant herpes simplex virus and also inhibits PNP, but with lower activity. This provides a structural template of an inhibitor bound to PNP. Silva *et al*. [148] described the kinetics and determined the crystal structure of human PNP in complex with a synthetic substrate, 7-methyl-6-thioguanosine.

Recent work has also been done to determine the active sites of PNP and the binding mechanisms of the ligands. Deng *et al.* [149] used NMR data to determine active site contacts in human PNP. Wielgus-Kutrowska *et al*. [150] studied the binding of 2-amino-9-[2-(phosphonome-thoxy) ethyl]-6-sulfanylpurine (PME-6-thio-Gua) to PNP from Cellulomanoas sp. They determined that the binding of the ligand to PNP is a one-step process.

Other groups have determined the structures of PNP from a number of other species. Caceres *et al*. [151] determined a first time model of PNP from Streptococcus agalactiae using molecular modeling and dynamics simulations. Schnick *et al*. [152] and Shi *et al*. [153] both report a crystal structure of PNP from Plasmodium falciparum in complex with various ligands and inhibitors. Others include models of PNP from calf spleen [154] and Listeria monocytogenes [155]. These models and structures may provide further insight to the binding site and mechanisms of human PNP.

3.2.2 Inhibitors—There are a number of inhibitors of PNP reported in the open literature, however all consist of small-molecule inhibitors. There are no peptidic inhibitors, making this an interesting challenge for *de novo* protein design. Because there are no peptidic inhibitors, there is no complex of PNP with a peptide, therefore no design template. One can, however, attempt to construct a template using the information of where and how the small-molecule inhibitors bind to PNP.

One of the earliest inhibitors of PNP was 9-(3,4-dioxopentyl)hypoxanthine [156]. It was found to be 25 times more effective in inhibiting PNP than butanedione. Inhibitors with equilibrium dissociation constants in the picomolar range were designed based on the transition-state structure of PNP [157]. More recent studies have found novel multi-substrate inhibitors [158], transition state analogue inhibitors [142,159], and analogues of forodesine HCl as inhibitors [160]. Clinch *et al*. [161] developed two potent inhibitors of human PNP that are currently in clinical trials and which were used as templates for further inhibitor design. In addition, Todorova and Schwarz [162] have investigated the effect of the phosphate substrate on inhibitor binding to human PNP. They determined the thermodynamics of three inhibitors

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of PNP (acyclovir, ganciclovir, and 9-benzylguanine) as a function of phosphate ion concentration.

3.2.3 Computational Methods—A number of computational methods have been used to aid in the design and development of small-molecule inhibitors of PNP, but none have been used to develop peptidic inhibitors. DeWitte and Shakhnovich [163] developed a *de novo* design method called SMoG (Small Molecule Growth), which is used for lead design and is based on a simple model for protein-ligand interactions. They applied their method to PNP, comparing the IC_{50} of known PNP ligands to the SMoG energy score. Shimada *et al*. [164] used SMoG to generate ligands complexed to actual protein structures and then used these models to construct databases from which knowledge-based protein-ligand potentials were derived. Macedo Timmers *et al*. [165] developed a set of empirical scoring functions that have the potential to evaluate binding affinities and docking results. These scoring functions were applied PNP in an effort to guide virtual screening. Another method applied to PNP relies on the use of drug-binding databases for virtual screening initiatives [166].

3.3 Ubiquitin Specific Protease 7

Ubiquitin specific protease 7 (USP7), also known as Herpesvirus-Associated USP (HAUSP) is a deubiquitinating enzyme that cleaves ubiquitin from its substrates. Ubiquitination is a prominent step in the regulation of many cellular processes. The major function of ubiquitination is to target proteins for degradation [167]. The catalytic core of USP7 consists of three domains: Fingers, Thumb, and Palm domains. In addition, the N-terminal domain of UPS7 consists of a TRAF-like MATH domain. The catalytic cleft is located between the Palm and the Thumb domains.

USP7 binds to p53, which contributes to its deubiquitination and stabilization. In addition, over expression of USP7 leads to p53-dependent cell growth repression and apoptosis. This indicates that USP7 may act as a tumor suppressor [168]. Furthermore, USP7 contains an additional binding site that can be targeted with inhibitors such as ubiquitin aldehyde (Ubal), a potent and irreversible inhibitor of many deubiquitinating enzymes, to regulate deubiquitinating activity [168]. This is important as deviations in the ubiquitin pathway can lead to a number of clinical disorders such as neurodegenerative disorders [169], genetic diseases [170], and immune surveillance/viral pathogenesis [170]. USP7 is also known to interact with a number of viruses, such as the Herpes simplex virus and the Epstein-Barr virus, making it a potential target for anti-viral treatment [168].

3.3.1 Targets—USP7 regulates p53 and is bound by at least two viral proteins (ICP0 from the Herpes simplex virus and EBNA1 from the Epstein-Barr virus). ICP0 binds to a domain between amino acids 599–801, while EBNA1 and p53 bind to a domain between amino acids 62–205 [171,172]. Holowaty and Frappier [171] showed that binding of EBNA1 to USP7 disrupts the USP7-p53 interaction.

The first crystal structure of USP7 consists of the catalytic core (amino acids 205–560: Fingers, Thumb, and Palm domains) in isolation and in complex with ubiquitin aldehyde (Ubal) [167]. Interestingly, upon Ubal binding, USP7 undergoes a dramatic conformational change in the active site, with some loops moving 5 – 8 Å. The second binding site, located in the TRAF-like domain, binds p53 and MDM2 in a mutually exclusive manner, with MDM2 having the higher binding affinity [173]. Monoubiquitylation of the tumor suppressor p53 mediated by MDM2 promotes its mitochondrial apoptosis, however, the apoptotically active non-ubiquitylated p53 can also be generated *via* the p53-USP7 complex [174]. Hu *et al*. [173] determined the crystal structure of the TRAF-like domain of UPS7 in complex with p53 and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

the TRAF-like domain of USP7 in complex with MDM2. Furthermore, the crystal structure of the p53 binding domain has also been elucidated in isolation and bound to EBNA1 [175].

Conflicting studies show that binding of p53 to USP7 either promotes the deubiquitination and subsequent stabilization of p53 [176,177] or that disruption of USP7 stabilizes p53 [178, 179]. It is thought that MDM2, rather than p53, is the preferred substrate of USP7, given by the higher binding affinity of MDM2 over p53 [173]. Hence, it may be that USP7 deubiquitinates multiple targets, MDM2 and p53 being two of them, and that the net deubiquitination of the targets determines the stability of p53. USP7 antagonists have the potential to stabilize p53 (which may be sufficient to cause tumor regression) in tumors that over express MDM2 by promoting the degradation of MDM2 [168].

The crystal structures of USP7 in isolation and bound to its substrates provide a number of possible templates for *de novo* protein design. We have recently verified these multiple binding sites by applying the binding site detection algorithm ConCavity [180]. ConCavity combines evolutionary sequence conservation estimates with structure-based surface pocket prediction in identifying binding sites. Three main binding sites were identified, the Ubal binding site, the p53/MDM2 binding site, and the catalytic cleft. This is consistent with the complexes that have been reported.

3.3.2 Inhibitors—Colland *et al*. [181] recently identified the first lead-like inhibitor of USP7. Using high-throughput screening, they discovered HBX 41,108, a small-molecule that inhibits USP7 deubiquitination. HBX 41,108 was shown to inhibit USP7 in an uncompetitive way, interacting with the enzyme-substrate complex rather than competing with substrate binding. HBX 41,108 has an IC_{50} of 0.42 μM and treatment with the small molecule stabilized p53, activated the transcription of a p53 target gene, and inhibited cancer cell growth. It also induced p53-dependent apoptosis in cancer cells containing the wild-type p53.

It is interesting that the small-molecule inhibitor interacts with the enzyme-substrate complex, rather than with just the enzyme itself. Molecular docking studies were carried out by Colland *et al*. [181] in order to understand the structural basis of HBX 41,108 binding to USP7. HBX 41,108 was docked to the ubiquitin-bound form of USP7. The inhibitor interacts with Asp295, Val256, Phe283, Trp285, His294, Leu299, and Val302, which form a hydrophobic groove. A crystal or NMR structure of this enzyme-substrate-inhibitor complex confirming these interactions would provide a very good design template for protein design, allowing the design of peptidic inhibitors of USP7. However, peptidic inhibitors could also be designed to inhibit the interaction of the substrate with USP7, using one of the known crystal structures mentioned previously, and targeting the ubiquitin binding site or the TRAF-like MATH domain of USP7.

3.3.3 Computational Methods—While computational methods have not been utilized to design novel inhibitors of USP7, various *in silico* methods have been used to examine the MDM2-mediated p53 ubiquitination network [182], develop a pharmacophore model for inhibitors of ubiquitin isopeptidases, and identify USP7 as a tumor necrosis factor receptor-associated factor domain-containing protein [183].

3.4 Histone Demethylases

Demethylases are a recently identified class of proteins that reverse Lys and Arg methylation. The first histone demethylase was identified by Shi *et al*. [184]. Histones are part of chromatin and aid in the packaging of DNA into the chromosomes [185]. DNA winds around histones like spools, forming nucleosomes [186]. Histones affect gene regulation by chemical modifications along the N-terminal histone tail that protrudes from the histone core around which the DNA is wound. These chemical modifications consist of phosphorylation, acetylation, methylation, adenoribosylation, ubiquitylation, and suoylation. While

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

phosphorylation and acetylation have long been known to be reversible processes (mediated by kinases and phosphatases for phosphorylation and acetyltransferases and deacetylases for actylation), methylation was long thought to be an irreversible process. This changed with the discovery of two histone demethylases: lysine-specific demethylase-1 (LSD1) [184] and JmjC domain-containing demethylase-1 (JHDM1) [187]. These two enzymes are capable of demethylating histone tails, making histone methylation a dynamic process, such as phosphorylation and acetylation. Since then, other demethylases have been identified, expanding the class of JmjC domain-containing demethylases to include JHDM3A/JMJD2A, GASC1/JMJDC, and JHDM2A [188]. LSD1 is limited to the demethylation of dimethylated and monomethylated lysine. In contrast, JmjC domain-containing demethylases have been shown to demthylate trimethylated lysine [189–191].

Since histone methylation is linked to a number of important biological processes (stem cell maintenance and differentiation, X inactivation, and DNA damage response [185,192]), it is expected that histone demethylation is also involved in these processes. Demethylases can act as transcriptional repressors due to the fact that methylation of histone H3 activates gene expression. This implies that inhibitors of demethylases can be used as therapeutic treatments for diseases that involve improper changes in gene expression, such as cancer [193,194].

3.4.1 Targets—A number of crystal and solution structures of LSD1 have been determined. The earliest is a solution structure of the SWIRM domain of LSD1 [195], which is comprised of 6 compact α helices. Stavropoulos *et al*. [186] presented the crystal structure of LSD1 at 2.9 Å resolution, finding that LSD1 has a highly symmetric, closely packed domain structure, from which a long helical `tower' domain extends. Yang *et al*. [196], Yang *et al*. [197], and Forneris *et al*. [198] determined the crystal structure of LSD1 in complex with CoREST, a co-repressor that, when bound to LSD1, enables demethylation of histone H3 Lys4 (H3-K4) within nucleosomes. Other crystal structures of LSD1 include the structure of LSD1 at a resolution at 2.8 Å [199], LSD1 bound to an H3 peptide [200], and LSD1 in complex with tranylcypromine, a known monoamine oxidase inhibitor [201].

The most prominent member of the JmjC domain-containing demethylase family is JMJD2A. Chen *et al*. [202] first reported the crystal structure of JMJD2A with and without α-ketoglutarate in the presence of Fe(II). The structure of the core domain consists of the JmjN domain, the JmjC domain, the C-terminal domain, and a zinc-finger motif. Following this, Chen *et al*. [203] determined the structures of the catalytic core of JMJD2A in complex with methylated H3-K36 peptide substrates in the presence of Fe(II). In addition, structures of JMJD2A-Ni(II)-Zn(II) bound to tri-, di- and monomethyl forms of H3-K9 and the di- and trimethyl forms of H3-K36 have been reported [204,205]. JMJD2A has also been found to bind another histone, H4. Lee *et al*. [206] presented the crystal structures JMJD2A in complex with the trimethyl form of H3-K4 and JMJD2A in complex with the trimethyl form of H4-K20, showing that the two histones adopt radically different binding modes. Other JmjC domain-containing demethylase structures include the DNA binding domain of the H3-K4 demethylase RBP2 [207], the DNA binding domain of JARID1C [208], and solution structure of JARID1A [209].

Other groups have described the binding of the LSD1/CoREST1/HDAC co-repressor complex to SUMO-2/3 [210], the binding of HIF-1 α to JMJD1A and JMJD2B [211], and the use of LSD1 as a potential therapeutic target [212,213].

3.4.2 Inhibitors—A recent review by Mai and Altucci [214] focuses on the chemical aspect and potential application in cancer therapy of molecules that inhibit not only histone demethylases but also DNA methyltransferases, histone deacetylases, histone acetyltransferases, and histone methyltransferases. Culhane *et al*. [215] identified two peptidic

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

inhibitors of LSD1. The first peptide (**1**) exhibited time-dependent inhibition of LSD1. While the second peptide (**2**) did not, it was still moderately potent against LSD1 with an IC_{50} of 15.6 ± 1.7 μM. Table **2** provides the sequences of the two peptides. Schmidt and McCafferty [188] have studied monoamine oxidase inhibitors (MAOIs) as potential inhibitors of LSD1. They showed that *trans*-2-phenylcyclopropylamine (2-PCPA) is the most potent MAOI and that it is a time-dependent, mechanism-based irreversible inhibitor of LSD1 with an IC_{50} of 20.7 ± 2.1 μM. In addition, 2-PCPA has been used as template to develop other mechanism-based inactivators of LSD1 [216].

Rose *et al*. [217] identified inhibitor scaffolds for JMJD2 histone demethylases. The best inhibitor (**6a**) had an IC_{50} of 1.4 μM. The seven highest-ranked inhibitors based upon enzyme-inhibitor interactions from this study (**6a**, **5a**, **7**, **3**, **8**, **1c**, **1a**) are shown in Table **2**. Recently, Hamada *et al*. [218] designed JHDM inhibitors based upon the crystal structure of JMJD2A in complex with NGO and histone trimethylated lysine peptide. They tested four compounds and all showed inhibitory activity against JMJD2 proteins. Other known inhibitors of JHDM include NGO, succinic acid, and 2,4-lutidinic acid.

3.4.3 Computational Methods—In the literature examined to date, the design of demethylase inhibitors has been done experimentally, rather than computationally. New small-molecule inhibitors were designed based upon scaffolds of known demethylase inhibitors. There are many known crystal and solution structures of demethylases in complex with histones and various inhibitors, which can be used as templates for computational protein design. Even though there are no structures of peptidic inhibitors, the knowledge that can be gleaned from the small-molecule complexes can provide a starting point for computational design.

4 SUMMARY

A number of computational protein design methods were described and recent advances outlined. These methods included both deterministic (such as DEE, SCMF, and quadratic assignment-like models) and stochastic (such as Monte Carlo, genetic algorithms, and combinatorial library approaches) methods, and showed varying incorporation of backbone flexibility (from rigid backbones and discrete rotamers, to continuous backbones with continuous rotamer angles). A number of successes using the various computational methods were highlighted. Reviews on protein design by Floudas [1] and Floudas *et al*. [2] cite a number of successes in the area of protein design, from the design of active sites on α-lytic protease to improve specificity toward the substrate [219] to the design of inhibitors for complement component C3 [5,6]. Despite these successes, there still remain many challenges for protein design. These include improved methodologies, such as improved sequence selection using flexible-backbone templates and simultaneous sequence and structure selection with flexible templates, and new design applications, such as design of inhibitors or peptides, among others.

Four application domains for computational *de novo* protein design were described, outlining potential drug targets. Peptides can be designed to inhibit HIV-1, PNP, USP7, and histone demethylases. To date, there have been a number of successes with peptidic inhibitors of HIV-1 and histone demethylases, however inhibitors of PNP and USP7 consist only of small-molecule inhibitors. All four systems have great promise for computational protein design, using the information of both peptidic and small-molecule inhibitors bound to the target protein as design templates.

Acknowledgments

CAF gratefully acknowledges financial support from the National Science Foundation, the National Institutes of Health (R01 GM52032, R24 GM069736) and the US Environmental Protection Agency, EPA (R 832721-010). Although the research described in the article has been funded in part by the U.S. Environmental Protection Agency's STAR program

through grant (R 832721-010), it has not been subjected to any EPA review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

REFERENCES

[1]. Floudas CA. Research challenges, opportunities and synergism in systems engineering and computational biology. AIChE J 2005;51:1872–84.

[2]. Floudas CA, Fung HK, McAllister SR, Mönnigmann M, Rajgaria R. Advances in protein structure prediction and *de novo* protein design: a review. Chem Eng Sci 2006;61:966–88.

[3]. Fung HK, Welsh WJ, Floudas CA. Computational *de novo* peptide and protein design: rigid templates versus flexible templates. Ind Engin Chem Res 2008;47:993–1001.

[4]. Dantas G, Kuhlman B, Callender D, Wong M, Baker D. A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. J Mol Biol 2003;332:449–60. [PubMed: 12948494]

[5]. Klepeis JL, Floudas CA, Morikis D, et al. Integrated structural, computational and experimental approach for lead optimization: design of compstatin variants with improved activity. J Am Chem Soc 2003;125:8422–23. [PubMed: 12848533]

[6]. Klepeis JL, Floudas CA, Morikis D, Tsokos CG, Lambris JD. Design of peptide analogs with improved activity using a novel *de novo* protein design approach. Ind Engin Chem Res 2004;43:3817–26.

[7]. Kraemer-Pecore CM, Lecomte JT, Desjarlais JR. A *de novo* redesign of the WW Domain. Protein Sci 2003;12:2194–2205. [PubMed: 14500877]

[8]. Gordon D, Marshall S, Mayo S. Energy Functions for Protein Design. Curr Opin Struct Biol 1999;9:509–13. [PubMed: 10449371]

[9]. Lilien RH, Stevens BW, Anderson AC, Donald BR. A novel ensemble-based scoring and search algorithm for protein redesign and its application to modify the substrate specificity of the gramicidin synthetase a phenylalanine adenylation enzyme. J Comput Biol 2005;12:740–61. [PubMed: 16108714]

[10]. Daily MD, Masica D, Sivasubramanian A, Somarouthu S, Gray JJ. CAPRI Rounds 3–5 reveal promising successes and future challenges for rosettadock. Proteins: Structure, Function, and Genetics 2005;60:181–6.

[11]. Gray JJ, Moughon S, Wang C, et al. Protein- protein docking with simultaneous optimization of rigid-body displacement and side chain conformations. J Mol Biol 2003;331:281–99. [PubMed: 12875852]

[12]. Gray JJ, Moughon SE, Kortemme T, et al. Protein-protein docking predictions for the CAPRI experiment. Proteins: Structure, Function, and Genetics 2003;52:118–22.

[13]. Jacobson M, Sali A. Comparative protein structure modeling and its applications to drug discovery. Annu Rep Med Chem 2004;39:259–76.

[14]. Petricoin EF, Rajapaske V, Herman EH, et al. Toxicoproteomics: serum proteomic pattern diagnostics for early detection of drug induced cardiac toxicities and cardioprotection. Toxicol Pathol 2004;32:122–30. [PubMed: 15209412]

[15]. Steiner S, Witzmann FA. Proteomics: applications and opportunities in preclinical drug development. Electrophoresis 2000;21:2099–2104. [PubMed: 10892720]

[16]. Fung HK, Rao S, Floudas CA, Prokopyev O, Pardalos PM, Rendl F. Computational comparison studies of quadratic assignment like formulations for the *in silico* sequence selection problem in *de novo* protein design. J Comb Optim 2005;10:41–60.

[17]. Pierce NA, Winfree E. Protein design is NP-hard. Protein Engin 2002;15:779–82.

[18]. Ponder JW, Richards FM. Tertiary templates for proteins. J Mol Biol 1987;193:775–91. [PubMed: 2441069]

[19]. Desmet J, Maeyer MD, Hazes B, Lasters I. The dead-end elimination theorem and its use in side-chain positioning. Nature 1992;356:539–42.

[20]. Gordon B, Hom G, Mayo S, Pierce N. Exact rotamer optimization for protein design. J Comput Chem 2003;24:232–43. [PubMed: 12497602]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

[21]. Looger LL, Hellinga HW. Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: Implications for protein design and structural genomics. J Mol Biol 2001;307:429–45. [PubMed: 11243829]

[22]. Pierce N, Spriet J, Desmet J, Mayo S. Conformational splitting: a more powerful criterion for dead-end elimination. J Comput Chem 2000;21:999–1009.

[23]. Voigt C, SL Mayo FA, Wang ZG. Computational method to reduce the search space for directed protein evolution. Proc Natl Acad Sci USA 2001;98:3778–83. [PubMed: 11274394]

[24]. Hong EJ, Lippow SM, Tidor B, Lozano-Perez T. Rotamer optimization for protein design through MAP estimation and problem-size reduction. J Comput Chem 2009;30:1923–45. [PubMed: 19123203]

[25]. Georgiev I, Lilien RH, Donald BR. The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. J Comput Chem 2008;29:1527–42. [PubMed: 18293294]

[26]. Kloppmann E, Ullmann GM, Becker T. An extended dead-end elimination algorithm to determine gap-free lists of low energy states. J Comput Chem 2007;28:2325–35. [PubMed: 17471458]

[27]. Yanover C, Fromer M, Shifman JM. Dead-end elimination for multistate protein design. J Comput Chem 2007;28:2122–9. [PubMed: 17471460]

[28]. Koehl P, Delarue M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. J Mol Biol 1994;239:249–75. [PubMed: 8196057]

[29]. Fu X, Kono H, Saven JG. Probabilistic approach to the design of symmetric protein quaternary structures. Protein Engin 2003;16:971–7.

[30]. Kono H, Saven J. Statistical theory of protein combinatorial libraries: packing interactions, backbone flexibility, and the sequence variability of a main-chain structure. J Mol Biol 2001;306:607–28. [PubMed: 11178917]

[31]. Saven JG, Wolynes PG. Statistical mechanics of the combinatorial synthesis and analysis of folding macromolecules. J Phys Chem B Condens Matter Mater Surf Interfaces Biophys 1997;101:8375–89.

[32]. Zou JM, Saven J. Statistical theory of combinatorial libraries of folding proteins: energetic discrimination of a target structure. J Mol Biol 2000;296:281–94. [PubMed: 10656832]

[33]. Bhattacherjee A, Biswas P. Statistical theory of protein sequence design by random mutation. J Phys Chem B Condens Matter Mater Surf Interfaces Biophys 2009;113:5520–7. [PubMed: 19323540]

[34]. Koh SK, Liu G, Zhu WH. A continuous protein design model using artificial power law in topology optimization. J Mech Des 2009;131:041001–14.

[35]. Fung HK, Floudas CA, Taylor MS, Zhang L, Morikis D. Toward full-sequence *de novo* protein design with flexible templates for human beta-defensin-2. Biophys J 2008;94:584–99. [PubMed: 17827237]

[36]. Fung HK, Taylor MS, Floudas CA. Novel formulations for the sequence selection problem in *de novo* protein design with flexible templates. Optimization Methods and Software 2007;22:51–71.

[37]. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087–92.

[38]. Liu Y, Kuhlman B. Rosetta design server for protein design. Nucleic Acids Res 2006;34(Web Server Issue):W235–8. [PubMed: 16845000]

[39]. Thomas J, Ramakrishnan N, Bailey-Kellogg C. Protein design by sampling an undirected graphical model of residue constraints. IEEE-ACM Trans Comput Biol 2009;6:506–16.

[40]. Zou JM, Saven JG. Using self-consistent fields to bias Monte Carlo methods with applications to designing and sampling protein sequences. J Chem Phys 2003;118:3843–54.

[41]. Koehl P, Levitt M. *De novo* protein design. I. In search of stability and specificity. J Mol Biol 1999;293:1161–81. [PubMed: 10547293]

[42]. Koehl P, Levitt M. *De novo* protein design. II. plasticity in sequence space. J Mol Biol 1999;293:1183–93. [PubMed: 10547294]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

[43]. Desjarlais J, Handel T. *De novo* design of the hydrophobic cores of proteins. Protein Sci 1995;4:2006–18. [PubMed: 8535237]

[44]. Hecht D, Fogel GB. Novel *in silico* approach to drug discovery *via* computational intelligence. J Chem Inf Model 2009;49:1105–21. [PubMed: 19348414]

[45]. Holland, JH. Adaptation in Natural and Artificial Systems. The MIT Press; Boston: 1993.

[46]. Lazar GA, Desjarlais JR, Handel TM. *De novo* design of the hydrophobic core of ubiquitin. Protein Sci 1997;6:1167–78. [PubMed: 9194177]

[47]. Scott LPB, Chahine J, Ruggiero JR. Using genetic algorithm to design protein sequence. Appl Math Comput 2008;200:1–9.

[48]. Tuffery P, Etchebest C, Hazout S, Lavery R. A new approach to the rapid determination of protein side chain conformations. J Biomol Struct Dyn 1991;8:1267–89. [PubMed: 1892586]

[49]. Belda I, Madurga S, Llora X, et al. ENPDA: an evolutionary structure-based *de novo* peptide design algorithm. J Comput Aided Mol Des 2005;19:585–601. [PubMed: 16267689]

[50]. Hohm T, Limbourg P, Hoffmann D. Multiobjective evolutionary method for the design of peptidic mimotopes. J Comput Biol 2006;13:113–25. [PubMed: 16472025]

[51]. Moore GL, Maranas CD. Identifying residue-residue clashes in protein hybrids by using a second-order mean-field approach. Proc Natl Acad Sci USA 2003;100:5091–6. [PubMed: 12700353]

[52]. Moore GL, Maranas CD. Computational challenges in combinatorial library design for protein engineering. AIChE J 2004;50:262–72.

[53]. Park S, Yang X, Saven JG. Advances in computational protein design. Curr Opin Struct Biol 2004;14:487–94. [PubMed: 15313244]

[54]. Saraf MC, Gupta A, Maranas CD. Design of combinatorial protein libraries of optimal size. Proteins Struct Funct Bioinformatics 2005;60:769–77.

[55]. Saraf MC, Horswill AR, Benkovic SJ, Maranas CD. FamClash: a method for ranking the activity of engineered enzymes. Proc Natl Acad Sci USA 2004;101:4142–7. [PubMed: 14981242]

[56]. Saraf MC, Moore GL, Maranas CD. Using multiple sequence correlation analysis to characterize functionally important protein regions. Protein Engin 2003;16:397–406.

[57]. Zhang NG, Zeng C. Reference energy extremal optimization: A Stochastic search algorithm applied to computational protein design. J Comput Chem 2008;29:1762–71. [PubMed: 18351599]

[58]. Georgiev I, Donald BR. Dead-end elimination with backbone flexibility. Bioinformatics 2007;23:i185–94. [PubMed: 17646295]

[59]. Georgiev I, Keedy D, Richardson JS, Richardson DC, Donald BR. Algorithm for backrub motions in protein design. Bioinformatics 2008;24:I196–I204. [PubMed: 18586714]

[60]. Taylor MS, Fung HK, Rajgaria R, Filizola M, Weinstein H, Floudas CA. Mutations affecting the oligomerization interface of g-protein coupled receptors revealed by a novel *de novo* protein design framework. Biophys J 2008;94:2470–81. [PubMed: 18178645]

[61]. Klepeis JL, Floudas CA. Free energy calculations for peptides *via* deterministic global optimization. J Chem Phys 1999;110:7491–512.

[62]. Klepeis JL, Floudas CA. Ab initio prediction of helical segments in polypeptides. J Comput Chem 2002;23:245–66. [PubMed: 11924737]

[63]. Klepeis JL, Floudas CA. Ab initio tertiary structure prediction of proteins. Journal of Global Optimization 2003;25:113–40.

[64]. Klepeis JL, Floudas CA. ASTRO-FOLD: a combinatorial and global optimization framework for ab initio prediction of three-dimensional structures of proteins from the amino acid sequence. Biophys J 2003;85:2119–46. [PubMed: 14507680]

[65]. Klepeis JL, Floudas CA. Prediction of _-Sheet Topology and Disulfide Bridges in Polypeptides. J Comput Chem 2003;24:191–208. [PubMed: 12497599]

[66]. Klepeis JL, Floudas CA, Morikis D, Lambris J. Predicting peptide structures using NMR data and deterministic global optimization. J Comput Chem 1999;20:1354–70.

[67]. Klepeis JL, Pieja MT, Floudas CA. A new class of hybrid global optimization algorithms for peptide structure prediction: alternating hybrids and applications for met-enkephalin and melittin. Biophys J 2003;84:869–82. [PubMed: 12547770]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

[68]. Klepeis JL, Pieja MT, Floudas CA. A new class of hybrid global optimization algorithms for peptide structure prediction: integrated hybrids. Comp Phys Commun 2003;151:121–40.

[69]. Klepeis JL, Wei Y, Hecht MH, Floudas CA. Ab initio prediction of the 3-dimensional structure of a *de novo* designed protein: a double blind case study. Proteins 2005;58:560–70. [PubMed: 15609306]

[70]. McAllister SR, Mickus BE, Klepeis JL, Floudas CA. A novel approach for alpha-helical topology prediction in globular proteins: generation of interhelical restraints. Proteins 2006;65:930–52. [PubMed: 17029234]

[71]. Adjiman C, Androulakis I, Floudas CA. A global optimization method, _bb, for general twice-differentiable constrained nlps - ii. implementation and computational results. Comp Chem Eng 1998;22:1159–79.

[72]. Adjiman CS, Dallwig S, Floudas CA. A global optimization method, alpha BB, for general twice-differentiable constrained NLPs - I. Theoretical advances. Comp Chem Eng 1998;22:1137–58.

[73]. Androulakis I, Maranas C, Floudas CA. Alpha BB: a global optimization method for general constrained nonconvex problems. J Global Optimi 1995;7:337–63.

[74]. Androulakis I, Maranas C, Floudas CA. Prediction of Oligopeptide Conformations *via* Deterministic Global Optimization. J Global Optimi 1997;11:1–34.

[75]. Esposito WR, Floudas CA. Global optimization in parameter estimation of nonlinear algebraic models *via* the error-in-variables approach. Ind Engin Chem Res 1998;37:1841–58.

[76]. Esposito WR, Floudas CA. Deterministic global optimization in nonlinear optimal control problems. J Global Optim 2000;17:97–126.

[77]. Floudas CA, Pardalos PM. State-of-the-Art in global optimization – computational methods and applications - preface. J Global Optimi 1995;7:113.

[78]. Bellows ML, Fung HK, Floudas CA, Lopez de Victoria A, Morikis D. New compstatin variants through two *de novo* protein design frameworks. Biophys J. 2010 accepted for publication.

[79]. Su A, Mayo SL. Coupling backbone flexibility and amino acid sequence selection in protein design. Protein Sci 1997;6:1701–7. [PubMed: 9260282]

[80]. Ross SA, Sarisky CA, Su A, Mayo SL. Designed protein G core variants fold to native- like structures: sequence selection by orbin tolerates variation in backbone specification. Protein Sci 2001;10:450–4. [PubMed: 11266631]

[81]. Larson SM, England JL, Desjarlais JR, Pande VS. Thoroughly sampling sequence space: large-scale protein design of structural ensembles. Protein Sci 2002;11:2804–13. [PubMed: 12441379]

[82]. Larson SM, Garg A, Desjarlais JR, Pande VS. Increased detection of structural templates using alignments of designed sequences. Proteins 2003;51:390–6. [PubMed: 12696050]

[83]. Desjarlais JR, Handel TM. Side chain and backbone flexibility in protein core design. J Mol Biol 1999;290:305–18. [PubMed: 10388574]

[84]. Kuhlman B, Dantas G, Ireton GC, g Verani, Stoddard B, Baker D. Design of a novel globular protein fold with atomic-level accuracy. Science 2003;302:1364–8. [PubMed: 14631033]

[85]. Saunders CT, Baker D. Recapitulation of protein family divergence using flexible backbone protein design. J Mol Biol 2005;346:631–44. [PubMed: 15670610]

[86]. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS. High-resolution protein design with backbone freedom. Science 1998;282:1462–7. [PubMed: 9822371]

[87]. Harbury PB, Tidor B, Alber T, Kim PS. Repacking protein cores with backbone freedom: structure prediction for coiled coils. Proc Natl Acad Sci USA 1995;92:8408–12. [PubMed: 7667303]

[88]. Plecs JJ, Harbury PB, Kim PS, Alber T. Structural test of the parameterized-backbone method for protein design. J Mol Biol 2004;342:289–97. [PubMed: 15313624]

[89]. Charles, A.; Janeway, J.; Travers, P. Immunobiology: the immune system in health and disease. 3rd ed.. Current Biology Ltd.; Garland Publishing Inc.; 1997.

[90]. Baldwin CE, Sanders RW, Berkhout B. Inhibiting HIV-1 entry with fusion inhibitors. Curr Med Chem Sep;2003 10:1633–42. [PubMed: 12871113]

[91]. Chan DC, Kim PS. HIV entry and its inhibition. Cell 1998;93:681–4. [PubMed: 9630213]

[92]. Luciw, PA. Fields Virology. Lippincott-Raven Publishers; 1996.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

[93]. Liu SW, Wu SG, Jiang SB. HIV entry inhibitors targeting gp41: From polypeptides to small-molecule compounds. Curr Pharm Des 2007;13:143–62. [PubMed: 17269924]

[94]. Qi Z, Shi WG, Xue N, et al. Rationally designed anti-HIV peptides containing multifunctional domains as molecule probes for studying the mechanisms of action of the first and second generation HIV fusion inhibitors. J Biol Chem 2008;283:30376–84. [PubMed: 18662985]

[95]. Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC. Atomic structure of the ectodomain from HIV-1 gp41. Nature 1997;387:426–30. [PubMed: 9163431]

[96]. Chan DC, Fass D, Berger JM, Kim PS. Core structure of gp41 from the HIV envelope glycoprotein. Cell 1997;89:263–73. [PubMed: 9108481]

[97]. Tan KM, Liu JH, Wang JH, Shen S, Lu M. Atomic structure of a thermostable subdomain of HIV-1 gp41. Proc Natl Acad Sci USA 1997;94:12303–8. [PubMed: 9356444]

[98]. Zhou GF, Ferrer M, Chopra R, et al. The structure of an HIV-1 specific cell entry inhibitor in complex with the HIV-1 gp41 trimeric core. Bioorg Med Chem 2000;8:2219–27. [PubMed: 11026535]

[99]. Chan DC, Chutkowski CT, Kim PS. Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. Proc Natl Acad Sci USA 1998;95:15613–7. [PubMed: 9861018]

[100]. Sia SK, Carr PA, Cochran AG, Malashkevich VN, Kim PS. Short Constrained Peptides That Inhibit HIV-1 Entry. Proc Natl Acad Sci 2002;99:14664–9. [PubMed: 12417739]

[101]. Balogh E, Wu D, Zhou GY, Gochin M. NMR second site screening for structure determination of ligands bound in the hydrophobic pocket of HIV-1 gp41. J Am Chem Soc 2009;131:2821–3. [PubMed: 19206471]

[102]. Wild C, Dubay JW, Greenwell T, et al. Propensity for a leucine zipper-like domain of 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;91:12676–80. [PubMed: 7809100]

[103]. Vranken WF, Budeskinsky M, Fant F, Boulez K, Borremans FA. The complete consensus V3 loop peptide of the envelope protein gp120 of HIV-1 shows pronounced helical character n solution. FASEB Lett 1995;374:117–21.

[104]. Huang C, Lam SN, Acharya P, et al. Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. Science 2007;317:1930–4. [PubMed: 17901336]

[105]. Tan HP, Rader AJ. Identification of putative, stable binding regions through flexibility analysis of HIV-1 gp120. Proteins Struct Funct Bioinformatics 2009;74:881–94.

[106]. Li J, Chen X, Huang JH, Jiang SB, Chen YH. Identification of critical antibody-binding sites in the HIV-1 gp41 six-helix bundle core as potential targets for HIV-1 fusion inhibitors. Immunobiology 2009;214:51–60. [PubMed: 19159827]

[107]. Liu J, Deng YQ, Dey AK, Moore JP, Lu M. Structure of the HIV-1 gp41 membrane- proximal ectodomain region in a putative prefusion conformation. Biochemistry 2009;48:2915–23. [PubMed: 19226163]

[108]. Maria Ferraz R, Rodriguez-Carmona E, Ferrer-Miralles N, Meyerhans A, Villaverde A. Screening HIV-1 antigenic peptides as receptors for antibodies and CD4 in allosteric nanosensors. J Mol Recognit 2009;22:255–60. [PubMed: 19177495]

[109]. Pantophlet R, Burton DR. GP120: Target for neutralizing HIV-1 antibodies. Annu Rev Immunol 2006;24:739–69. [PubMed: 16551265]

[110]. Pinter A. Roles of HIV-1 Env variable regions in viral neutralization and vaccine development. Curr HIV Res 2007;5:542–53. [PubMed: 18045110]

[111]. Zhou TQ, Xu L, Dey B, et al. Structural definition of a conserved neutralization epitope on HIV-1 gp120. Nature 2007;445:732–7. [PubMed: 17301785]

[112]. Lalezari JP, Henry K, O'Hearn M, et al. Enfuvirtide, an HIV-1 fusion inhibitor, for drug-resistant HIV infection in North and South America. N Engl J Med 2003;348:2175–85. [PubMed: 12637625]

[113]. Dorr P, Westby M, Dobbs S, et al. Maravoric (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. Antimicrob Agents Chemother 2005;49:4721–32. [PubMed: 16251317]

[114]. Welch BD, VanDemark AP, Heroux A, Hill CP, Kay MS. Potent D-peptide inhibitors of HIV-1 entry. Proc Natl Acad Sci USA 2007;104:16828–33. [PubMed: 17942675]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

[115]. Gaston F, Granados GC, Madurga S, et al. Development and characterization of peptidic fusion inhibitors derived from HIV-1 gp41 with partial D-amino acid substitutions. Chem Med Chem 2009;4:570–81. [PubMed: 19206122]

[116]. Stoddart CA, Nault G, Galkina SA, et al. Albumin-conjugated C34 peptide HIV-1 fusion inhibitor equipotent to C34 and T-20 *in vitro* with sustained activity in SCID-HU THY/LIV mice. J Biol Chem 2008;283:34045–52. [PubMed: 18809675]

[117]. He YX, Xiao YH, Song HF, et al. Design and evaluation of sifuvirtide, a novel HIV-1 fusion inhibitor. J Biol Chem 2008;283:11126–34. [PubMed: 18303020]

[118]. Wang RR, Yang LM, Wang YH, et al. Sifuvirtide, a potent HIV fusion inhibitor peptide. Biochem Biophys Res Commun 2009;382:540–4. [PubMed: 19289098]

[119]. Yu H, Tudor D, Alfsen A, Labrosse B, Clavel F, Bomsel M. Peptide P5 (residues 628–683), comprising the entire membrane proximal region of HIV-1 gp41 and its calcium-binding site, is a potent inhibitor of HIV-1 infection. Retrovirology 2008;5:93. [PubMed: 18925934]

[120]. Pritsker M, Jones P, Blumenthal R, Shai YC. A synthetic all D-amino acid peptide corresponding to the N-terminal sequence of HIV-1 gp41 recognizes the wild-type fusion peptide in the membrane and inhibits HIV-1 envelope glycoprotein-mediated cell fusion. Proc Natl Acad Sci USA 1998;95:7287–92. [PubMed: 9636141]

[121]. Dwyer JJ, Wilson KL, Davison DK, et al. Design of helical, oligomeric HIV-1 fusion inhibitor peptides with potent activity against enfuvirtide resistant virus. Proc Natl Acad Sci USA 2007;104:12772–7. [PubMed: 17640899]

[122]. Izumi K, Kodama E, Shimura K, et al. Design of peptide based inhibitors for human immunodeficiency virus type 1 strains resistant to T-20. J Biol Chem 2009;284:4914–20. [PubMed: 19073606]

[123]. Naito T, Izumi K, Kodama E, et al. SC29EK, a Peptide fusion inhibitor with enhanced alpha-helicity, inhibits replication of human immunodeficiency virus type 1 mutants resistant enfuvirtide. Antimicrob Agents Chemother 2009;53:1013–8. [PubMed: 19114674]

[124]. Nishikawa H, Nakamura S, Kodama E, et al. Electrostatically constrained alpha-helical peptide inhibits replication of HIV-1 resistant to enfuvirtide. Intl J Biochem Cell Biol 2009;41:891–9.

[125]. Nishikawa H, Oishi S, Fujita M, et al. Identification of minimal sequence for HIV-1 fusion inhibitors. Bioorg Med Chem 2008;16:9184–7. [PubMed: 18819810]

[126]. Oishi S, Ito S, Nishikawa H, et al. Design of a novel HIV-1 fusion inhibitor that displays a minimal interface for binding affinity. J Med Chem 2008;51:388–91. [PubMed: 18197613]

[127]. Liang XY. CXCR4, inhibitors and mechanisms of action. Chemical Biol Drug Des 2008;72:97–110.

[128]. Li G, Haney KM, Kellogg GE, Zhang Y. Comparative docking study of anibamine as the first natural product CCR5 antagonist in CCR5 homology models. J Chem Inf Model 2009;49:120–32. [PubMed: 19166361]

[129]. Liu K, Lu H, Hou L, et al. Design, synthesis, and biological evaluation of N-carboxyphenylpyrrole derivatives as potent HIV fusion inhibitors targeting gp41. J Med Chem 2008;51:7843–54. [PubMed: 19053778]

[130]. Lu RJ, Tucker JA, Zinevitch T, et al. Design and synthesis of human immunodeficiency virus entry inhibitors: sulfonamide as an isostere for the alpha-ketoamide group. J Med Chem 2007;50:6535–44. [PubMed: 18052117]

[131]. Madani N, Schon A, Princiotto AM, et al. Small- molecule CD4 mimics interact with a highly conserved pocket on HIV-1 gp120. Structure 2008;16:1689–1701. [PubMed: 19000821]

[132]. Singh S, Malik BK, Sharma DK. Targeting HIV-1 through molecular modeling and docking studies of CXCR4: Leads for therapeutic development. Chem Biol Drug Des 2007;69:191–203. [PubMed: 17441905]

[133]. Berchanski A, Lapidot A. Computer-based design of novel HIV-1 entry inhibitors: neomycin conjugated to arginine peptides at two specific sites. J Mol Model 2009;15:281–94. [PubMed: 19057930]

[134]. Berchanski A, Lapidot A. Prediction of HIV-1 entry inhibitors neomycin-arginine conjugates interaction with the CD4-gp120 binding site by molecular modeling and multistep docking procedure. Biochim Biophys Acta 2007;1768:2107–19. [PubMed: 17560540]

[135]. Imai M, Baranyi L, Okada N, Okada H. Inhibition of HIV-1 infection by synthetic peptides derived CCR5 fragments. Biochem Biophys Res Commun 2007;353:851–6. [PubMed: 17210123]

[136]. Jiang S, Lin K, Zhang L, Debnath AK. A screening assay for antiviral compounds targeted to the HIV-1 gp41 core structure using a conformation-specific monoclonal antibody. J Virol Methods 1999;80:85–96. [PubMed: 10403680]

[137]. Liu S, Boyer-Chatenet L, Lu H, Jiang S. Rapid and automated fluorescence-linked immunosorbent assay for high throughput screening of HIV-1 fusion inhibitors targeting gp41. J Biomol Screen 2003;8:685–93. [PubMed: 14711394]

[138]. Liu S, Zhao Q, Jiang S. Determination of the HIV-1 gp41 postfusion conformation modeled by synthetic peptides: applicable for identification of the HIV-1 fusion inhibitors. Peptides 2003;24:1301–13.

[139]. Liu SW, Jiang SB. High throughput screening and characterization of HIV-1 entry inhibitors targeting gp41: theories and techniques. Curr Pharm Des 2004;10:1827–43. [PubMed: 15180543]

[140]. Bellows, ML.; Fung, HK.; Floudas, CA., et al. Discovery of entry inhibitors for HIV-1 vis a novel *de novo* protein design framework. 2009. in preparation

[141]. Gowthaman U, Jayakanthan M, Sundar D. Molecular docking studies of dithionitrobenzoic acid and its related compounds to protein disulfide isomerase: computational screening of inhibitors to HIV-1 entry. BMC Bioinformatics 2008;9:S14. [PubMed: 19091013]

[142]. Schramm VL. Development of transition state analogues of purine nucleoside phosphorylase as anti-T-cell agents. Biochim Biophys Acta 2002;1587:107–17. [PubMed: 12084452]

[143]. Boehncke WH, Gilbertsen RB, Hemmer J, Sterry W. Evidence for a pathway independent from 2'-deoxyguanosine and reversible by il-2 by which purine nucleoside phosphorylase inhibitors block t-cell proliferation. Scand J Immunol 1994;39:327–32. [PubMed: 8128192]

[144]. Canduri F, Silva RG, dos Santos DM, et al. Structure of human PNP complexed with ligands. Acta Crystallogr D Biol Crystallogr 2005;61:856–62. [PubMed: 15983407]

[145]. Canduri F, dos Santos DM, Silva RG, et al. Structures of human purine nucleoside phosphorylase complexed with inosine and ddI. Biochem Biophys Res Commun 2004;313:907–14. [PubMed: 14706628]

[146]. Canduri F, Fadel V, Dias MVB, et al. Crystal structure of human PNP complexed with hypoxanthine and sulfate ion. Biochem Biophys Res Commun 2005;326:335–8. [PubMed: 15582582]

[147]. dos Santos DM, Canduri F, Pereira JH, et al. Crystal structure of human purine nucleoside phosphorylase complexed with acyclovir. Biochem Biophys Res Commun 2003;308:553–9. [PubMed: 12914786]

[148]. Silva RG, Pereira JH, Canduri F, de Azevedo WF Jr, Basso LA, Santos DS. Kinetics and crystal structure of human purine nucleoside phosphorylase in complex with 7-methyl-6- thioguanosine. Arch Biochem Biophys 2005;442:49–58. [PubMed: 16154528]

[149]. Deng H, Cahill SM, Abad JL, et al. Active site contacts in the purine nucleoside phosphorylase - hypoxanthine complex by NMR and ab initio calculations. Biochemistry 2004;43:15966–74. [PubMed: 15595853]

[150]. Wielgus-Kutrowska B, Antosiewicz JM, Dlugosz M, Holy A, Bzowska A. Towards the mechanism of trimeric purine nucleoside phosphorylases: stopped-flow studies of binding of multisubstrate analogue inhibitor - 2-amino-9-[2-(phosphonomethoxy)ethyl]-6-sulfanylpurine. Biophys Chem 2007;125:260–8. [PubMed: 16989940]

[151]. Caceres RA, Timmers LFS, Dias R, Basso LA, Santos DS, de Azevedo WF. Molecular modeling and dynamics simulations of PNP from Streptococcus agalactiae. Bioorg Med Chem 2008;16:4984–3. [PubMed: 18381241]

[152]. Schnick C, Robien MA, Brzozowski AM, et al. Structures of Plasmodium falciparum purine nucleoside phosphorylase complexed with sulfate and its natural substrate inosine. Acta Crystallogr D Biol Crystallogr 2005;61:1245–54. [PubMed: 16131758]

[153]. Shi WX, Ting LM, Kicska GA, et al. Plasmodium falciparum purine nucleoside phosphorylase - Crystal structures, immucillin inhibitors, and dual catalytic function. J Biol Chem 2004;279:18103–6. [PubMed: 14982926]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

[154]. Luic M, Koellner G, Yokomatsu T, Shibuya S, Bzowska A. Calf spleen purine-nucleoside phosphorylase: crystal structure of the binary complex with a potent multisubstrate analogue inhibitor. Acta Crystallogr D Biol Crystallogr 2004;60:1417–24. [PubMed: 15272165]

[155]. Timmers LFSM, Caceres RA, Basso LA, Santos DS, De Azevedo WF. Structural bioinformatics study of PNP from Listeria monocytogenes. Protein Pept Lett 2008;15:843–9. [PubMed: 18855758]

[156]. Salamone SJ, Jordan F. Synthesis of 9-(3,4-dioxopentyl) hypoxanthine, the 1st arginine directed purine derivative: An irreversible inactivator for purine nucleoside phosphorylase. Biochemistry 1982;21:6383–8. [PubMed: 6817790]

[157]. Miles RW, Tyler PC, Furneaux RH, Bagdassarian CK, Schramm VL. One-third-the-sites transition-state inhibitors for purine nucleoside phosphorylase. Biochem 1998;37:8615–21. [PubMed: 9628722]

[158]. Toms AV, Wang WR, Li YB, Ganem B, Ealick SE. Novel multisubstrate inhibitors of mammalian purine nucleoside phosphorylase. Acta Crystallogr D Biol Crystallogr 2005;61:1449–58. [PubMed: 16239721]

[159]. Lewandowicz A, Tyler PC, Evans GB, Furneaux RH, Schramm VL. Achieving the ultimate physiological goal in transition state analogue inhibitors for purine nucleoside phosphorylase. J Biol Chem 2003;278:31465–8. [PubMed: 12842889]

[160]. Kamath VP, Xue J, Juarez-Brambila JJ. Synthesis of analogs of forodesine HCl, a human purine nucleoside phosphorylase inhibitor-Part II. Bioorg Med Chem 2009;19:2627–9.

[161]. Clinch K, Evans GB, Frohlich RFG, et al. Third- Generation Immucillins: Syntheses and Bioactivities of Acyclic Immucillin Inhibitors of Human Purine Nucleoside Phosphorylase. J Med Chem 2009;52:1126–43. [PubMed: 19170524]

[162]. Todorova NA, Schwarz FP. Effect of the phosphate substrate on drug-inhibitor binding to human purine nucleoside phosphorylase. Arch Biochem Biophys 2008;480:122–31. [PubMed: 18938130]

[163]. DeWitte RS, Shakhnovich EI. SMoG: *de novo* design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. J Am Chem Soc 1996;118:11733–44.

[164]. Shimada J, Ishchenko AV, Shakhnovich EI. Analysis of knowledge-based protein-ligand potentials using a self-consistent method. Protein Sci 2000;9:765–75. [PubMed: 10794420]

[165]. Macedo Timmers LFS, Caceras RA, et al. Structural studies of human purine nucleoside phosphorylase: towards a new specific empirical scoring function. Arch Biochem Biophys 2008;479:28–38. [PubMed: 18790691]

[166]. Macedo Timmers LFS, Pauli I, Caceras RA, de Azevedo WF Jr. Drug-Binding Databases. Curr Drug Targ 2009;9:1092–9.

[167]. Hu M, Li P, Li M, et al. Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. Cell 2002;111:1041–54. [PubMed: 12507430]

[168]. Daviet L, Colland F. Targeting ubiquitin specific proteases for drug discovery. Biochimie 2008;90:270–83. [PubMed: 17961905]

[169]. Chung KKK, Dawson VL, Dawson TM. The role of the ubiquitin-proteasomal pathway in Parkinson's disease and other neurodegenerative disorders. Trends Neurosci 2001;24:S7–14. [PubMed: 11881748]

[170]. Schwartz AL, Ciechanover A. The ubiquitin-proteasome pathway and pathogenesis of human diseases. Annu Rev Med 1999;50:57–74. [PubMed: 10073263]

[171]. Holowaty MN, Frappier L. HAUSP/USP7 as an Epstein-Barr virus target. Biochem Soc Trans 2004;32:731–2. [PubMed: 15494000]

[172]. Holowaty MN, Sheng Y, guyen T, Arrowsmith C, Frappier L. Protein interaction domains of the Ubiquitin-specific protease, USP7/HAUSP. J Biol Chem 2003;278:47753–61. [PubMed: 14506283]

[173]. Hu M, Gu L, Li M, Jeffrey PD, Gu W, Shi Y. Structural basis of competitive recognition of p53 and MDM2 by HAUSP/USP7: Implications for the regulation of the p53-MDM2 pathway. PLoS Biol 2006;4:e27. [PubMed: 16402859]

[174]. Marchenko ND, Wolff S, Erster S, Becker K, Moll UM. Monoubiquitylation promotes mitochondrial p53 translocation. EMBO J 2007;26:923–34. [PubMed: 17268548]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

[175]. Saridakis V, Sheng Y, Sarkari F, et al. Structure of the p53 binding domain of HAUSP/USP7 bound to epstein-barr nuclear antigen 1: implications for EBV-mediated immortalization. Mol Cell 2005;18:25–36. [PubMed: 15808506]

[176]. Li M, Chen D, Shiloh A, et al. Deubiquitination of p53 by HAUSP is an important pathway for p53 stabilization. Nature 2002;416:648–653. [PubMed: 11923872]

[177]. Xirodimas DP, Saville MK, Bourdon JC, Hay RT, Lane DP. MDM2-mediated NEDD8 conjugation of p53 inhibits its transcriptional activity. Cell 2004;118:83–97. [PubMed: 15242646]

[178]. Cummins JM, Rago C, Kohli M, Kinzler KW, Lengauer C, Vogelstein B. Tumor suppression: disruption of HAUSP gene stabilizes p53. Nature 2004;428:1–2. [PubMed: 15058298]

[179]. Cummins JM, Vogelstein B. HAUSP is required for p53 destabilization. Cell Cycle 2004;3:689–92. [PubMed: 15118411]

[180]. Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA. Predicting Protein Ligand Binding Sites by Combining Evolutionary Sequence Conservation and 3D Structure 2009;5 (12):e1000585.

[181]. Colland F, Formstecher E, Jacq X, et al. Small molecule inhibitor of USP7/HAUSP ubiquitin protease stabilizes and activates p53 in cells. Mol Cancer Ther 2009;8:2286–95. [PubMed: 19671755]

[182]. Brazhnik P, Kohn KW. HAUSP-regulated switch from auto- to p53 ubiquitination by Mdm2 (*in silico* discovery). Math Biosci 2007;210:60–77. [PubMed: 17585950]

[183]. Zapata JM, Pawlowski K, Haas E, Ware CF, Godzik A, Reed JC. A diverse family of proteins containing tumor necrosis factor receptor-associated factor domains. J Biol Chem 2001;276:24242–52. [PubMed: 11279055]

[184]. Shi YJ, Lan F, Matson C, et al. Histone demethylation mediated by the nuclear arnine oxidase homolog LSD1. Cell 2004;119:941–53. [PubMed: 15620353]

[185]. Shi Y, Whetstine JR. Dynamic Regulation of Histone Lysine Methylation by Demethylases. Mol Cell 2007;25:1–14. [PubMed: 17218267]

[186]. Stavropoulos P, Blobel G, Hoelz A. Crystal structure and mechanism of human lysine specific demethylase-1. Nat Struct Mol Biol 2006;13:626–32. [PubMed: 16799558]

[187]. Tsukada Y, Fang J, Erdjument-Bromage H, et al. Histone demethylation by a family of JmjC domain-containing proteins. Nature 2006;439:811–6. [PubMed: 16362057]

[188]. Schmidt DMZ, McCafferty DG. *Trans*-2-phenylcyclopropylamine is a mechanism-based inactivator of the histone demethylase LSD1. Biochemistry 2007;46:4408–16. [PubMed: 17367163]

[189]. Cloos PAC, Christensen J, Agger K, et al. The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. Nature 2006;442:307–11. [PubMed: 16732293]

[190]. Klose RJ, Yamane K, Bae Y, et al. The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. Nature 2006;442:312–6. [PubMed: 16732292]

[191]. Whetstine JR, Nottke A, Lan F, et al. Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. Cell 2006;125:467–81. [PubMed: 16603238]

[192]. Cloos PAC, Christensen J, Agger K, Helin K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes Dev 2008;22:1115–40. [PubMed: 18451103]

[193]. Shi Y. Histone lysine demethylases: emerging roles in development, physiology and disease. Nat Rev Genet 2007;8:829–33. [PubMed: 17909537]

[194]. Szewczuk LM, Culhane JC, Yang MJ, Majumdar A, Yu HT, Cole PA. Mechanistic analysis of a suicide inactivator of histone demethylase LSD1. Biochemistry 2007;46:6892–902. [PubMed: 17511474]

[195]. Tochio N, Umehara T, Koshiba S, et al. Solution structure of the SWIRM domain of human histone demethylase LSD1. Structure 2006;14:457–68. [PubMed: 16531230]

[196]. Yang M, Gocke CB, Luo X, et al. Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. Mol Cell 2006;23:377–87. [PubMed: 16885027]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

[197]. Yang MJ, Culhane JC, Szewczuk LM, et al. Structural basis for the inhibition of the LSD1 histone demethylase by the antidepressant *trans*-2-phenylcyclopropylamine. Biochemistry 2007;46:8058–65. [PubMed: 17569509]

[198]. Forneris F, Binda C, Adamo A, Battaglioli E, Mattevi A. Structural basis of LSD1 CoREST selectivity in histone H3 recognition. J Biol Chem 2007;282:20070–4. [PubMed: 17537733]

[199]. Chen Y, Yang Y, Want F, et al. Crystal structure of human histone lysine-specific demethylase 1 (LSD1). Proc Natl Acad Sci 2006;103:13956–61. [PubMed: 16956976]

[200]. Yang M, Culhane JC, Szewczuk LM, et al. Structural basis of histone demethylation by LSD1 revealed by suicide inactivation. Nat Struct Mol Biol 2007;14:535–9. [PubMed: 17529991]

[201]. Mimasu S, Sengoku T, Fukuzawa S, Umehara T, Yokoyama S. Crystal structure of histone demethylase LSD1 and tranylcypromine at 2.25 angstrom. Biochem Biophys Res Commun 2008;366:15–22. [PubMed: 18039463]

[202]. Chen ZZ, Zang JY, Whetstine J, et al. Structural insights into histone demethylation by JMJD2 family members. Cell 2006;125:691–702. [PubMed: 16677698]

[203]. Chen Z, Zang J, Kappler J, et al. Structural basis of the recognition of a methylated histone tail by JMJD2A. Proc Natl Acad Sci 2007;104:10818–23. [PubMed: 17567753]

[204]. Couture JF, Collazo E, Ortiz-Tello PA, Brunzelle JS, Trievel RC. Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. Nat Struct Mol Biol 2007;14:689–95. [PubMed: 17589523]

[205]. Ng SS, Kavanagh KL, McDonough MA, et al. Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. Nature 2007;448:87–91. [PubMed: 17589501]

[206]. Lee J, Thompson JR, Botuyan MV, Mer G. Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. Nat Struct Mol Biol 2008;15:109–11. [PubMed: 18084306]

[207]. Tu S, Teng YC, Yuan C, et al. The ARID domain of the H3K4 demethylase RBP2 binds to a DNA CCGCCC motif. Nat Struct Mol Biol 2008;15:419–21. [PubMed: 18270511]

[208]. Koehler C, Bishop S, Dowler EF, et al. Backbone and side chain 1H, 13C and 15N resonance assignments of the Bright/ARID domain from the human JARID1C (SMCX) protein. Biomol NMR Assign 2008;2:9–11. [PubMed: 19636912]

[209]. Wang GG, Song J, Wang Z, et al. Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. Nature 2009;459:847–51. [PubMed: 19430464]

[210]. Ouyang J, Shi Y, Valin A, Xuan Y, Gill G. Direct binding of CoREST1 to SUMO-2/3 contributes to gene-specific repression by the LSD1/CoREST1/HDAC complex. Mol Cell 2009;34:145–54. [PubMed: 19394292]

[211]. Beyer S, Kristensen MM, Jensen KS, Johansen JV, Staller P. The histone demethylases JMJD1A and JMJD2B are transcriptional targets of hypoxia-inducible factor HIF. J Biol Chem 2008;283:36542–2. [PubMed: 18984585]

[212]. Lee MG, Wynder C, Schmidt DM, McCafferty DG, Shiekhattar R. Histone H3 lysine 4 demethylation is a target of nonselective antidepressive medications. Chem Biol 2006;13:563–7. [PubMed: 16793513]

[213]. Stavropoulos P, Hoelz A. Lysine-specific demethylase 1 as a potential therapeutic target. Expert Opin Ther Targets 2007;11:809–20. [PubMed: 17504018]

[214]. Mai A, Altucci L. Epi-drugs to fight cancer: from chemistry to cancer treatment, the road ahead. Intl J Biochem Cell Biol 2009;41:199–213.

[215]. Culhane JC, Szewczuk LM, Liu X, Da GP, Marmorstein R, Cole PA. A mechanism-based inactivator for histone demethylase LSD1. J Am Chem Soc 2006;128:4536–7. [PubMed: 16594666]

[216]. Gooden DM, Schmidt DMZ, Pollock JA, Kabadi AM, McCafferty DG. Facile synthesis of substituted *trans*-2-arylcyclopropylamine inhibitors of the human histone demethylase LSD1 and monoamine oxidases A and B. Bioorg Med Chem 2008;18:3047–51.

[217]. Rose NR, Ng SS, Mecinovic J, et al. Inhibitor scaffolds for 2-oxoglutarate-dependent histone lysine demethylases. J Med Chem 2008;51:7053–6. [PubMed: 18942826]

[218]. Hamada S, Kim TD, Suzuki T, et al. Synthesis and activity of N-oxalylglycine and its derivatives as Jumonji C-domain-containing histone lysine demethylase inhibitors. Bioorg Med Chem 2009;19:2852–5.

[219]. Wilson C, Mace J, Agard D. Computational method for the design of enzymes with altered substrate specificity. J Mol Biol 1991;220:495–506. [PubMed: 1856870]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

HIV Fusion Inhibitors. Amino Acid letter “B” Denotes the Unnatural Amino Acid α- Aminoisobutyric Acid (Aib). Cross-Linked Amino Acids are Denoted by the ^ Character. Upper-Case Letters Indicate L-Amino Acids while Lower-Case Letters Indicate D-Amino Acids

Name	Sequence	Length	IC_{50}(μM)	Refs.
$(PIE7)_3$	PEG-(KGACDYPEWQWLCAA)3	45	0.00025	[114]
C34M2	WMEWDREINNYTSLIHS LIEESQNQQeKNEkELL	34	0.00031	[115]
Selzentry	Small molecule ligand	n/a	0.00040	[113]
C34	WMEWDREINNYTSLIHSLIEESQNQQEKNEQELL	34	0.0006	[116]
Fuzeon	YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWF	36	0.0037	[114]
Sifuvirtide	SWETWEREIENYTRQIYRILEESQEQQDRNERDLLE	36	0.00394	[117,118]
P5	WMEWDREINNYTSLIHSL IEESQNQQEKNEQELLEL DKWASLWNWFNITNWLWYIK	56	0.06	[119]
PIE7	KGACDYPEWQWLCAA	15	0.62	[114]
C14linkmid	MTWQ^EWDREIQ^NYT	14	35	[100]
C14Aib	MTWBEWDREIBNYT	14	144	[100]
C14wt	MTWMEWDREINNYT	14	>500	[100]

Table 2

Histone Demethylase Inhibitors. The IC_{50} of 1 was Not Available. Peptides 1 and 2 are Modified at the N-Terminus. See Culhane *et al.* [215] for Details

Name	Sequence	Length	IC_{50} (μM)	Ref.
6a	small molecule ligand	n/a	1.4	[217]
7	small molecule ligand	n/a	6.6	[217]
1	QTARKSTGGKAPRKQLA	17		[194,215]
2	QTARKSTGGKAPRKQLA	17	15.6	[215]
2-PCPA	small molecule ligand	n/a	20.7	[188,197]
8	small molecule ligand	n/a	27	[217]
5a	small molecule ligand	n/a	28	[217]
1a	small molecule ligand	n/a	78	[217]
1c	small molecule ligand	n/a	100	[217]
3	small molecule ligand	n/a	540	[217]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript