Journal of Computer-Aided Molecular Design, **14:** 705–718, 2000.
KLUWER/ESCOM

Computational design of D-peptide inhibitors of hepatitis delta antigen dimerization

Carl D. Elkin[a,b], Harmon J. Zuccola[b], James M. Hogle[a,b] & Diane Joseph-McCarthy[c,*]
[a]*Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, MA 02139, U.S.A.;* [b]*Department of Biological Chemistry & Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, U.S.A.;* [c]*Biological Chemistry Department, Wyeth Research, 87 CambridgePark Dr., Cambridge, MA 02140, U.S.A.*

Received 23 August 1999; Accepted 28 April 2000

Key words: coiled-coil formation, D-peptide ligands, multiple copy simultaneous search (MCSS), structure-based drug design

Summary

Hepatitis delta virus (HDV) encodes a single polypeptide called hepatitis delta antigen (DAg). Dimerization of DAg is required for viral replication. The structure of the dimerization region, residues 12 to 60, consists of an anti-parallel coiled coil [Zuccola et al., Structure, 6 (1998) 821]. Multiple Copy Simultaneous Searches (MCSS) of the hydrophobic core region formed by the bend in the helix of one monomer of this structure were carried out for many diverse functional groups. Six critical interaction sites were identified. The Protein Data Bank was searched for backbone templates to use in the subsequent design process by matching to these sites. A 14 residue helix expected to bind to the D-isomer of the target structure was selected as the template. Over 200 000 mutant sequences of this peptide were generated based on the MCSS results. A secondary structure prediction algorithm was used to screen all sequences, and in general only those that were predicted to be highly helical were retained. Approximately 100 of these 14-mers were model built as D-peptides and docked with the L-isomer of the target monomer. Based on calculated interaction energies, predicted helicity, and intrahelical salt bridge patterns, a small number of peptides were selected as the most promising candidates. The ligand design approach presented here is the computational analogue of mirror image phage display. The results have been used to characterize the interactions responsible for formation of this model anti-parallel coiled coil and to suggest potential ligands to disrupt it.

Introduction

Hepatitis Delta virus (HDV) is a 36 nm virus particle consisting of an envelope and an internal nucleocapsid [1]. The envelope is made up of the surface antigen of Hepatitis B virus (HBV), while the nucleocapsid is composed of the HDV-encoded Delta Antigen (DAg), and a single-stranded circular RNA. Because HBV proteins are incorporated into the HDV envelope, HDV depends on the presence of HBV for transmission and is therefore considered a defective, satellite virus of HBV. However, HDV can replicate independently once HDV RNA is introduced into cells. Coinfection with HDV and HBV is associated with an unusually high incidence of fulminating acute hepatitis that is often fatal.

Dimerization and subsequent oligomerization of DAg is known to be required for transactivation of viral replication, encapsidation of the viral RNA and recruitment of HBV surface antigen [2, 3]. Residues 12 to 60 of DAg contain canonical heptad repeats and have been implicated in *in vitro* dimerization. Site-specific mutants with various substitutions at the position of two of the Leu residues in this dimerization region either reduced or prevented dimerization and therefore *trans* activation of HDV RNA replication [2]. The structure of the peptide consisting of residues 12 to 60 of DAg has been solved by X-ray diffraction methods to 1.8 Å resolution [4]. The

*To whom correspondence should be addressed. E-mail: djoseph@genetics.com

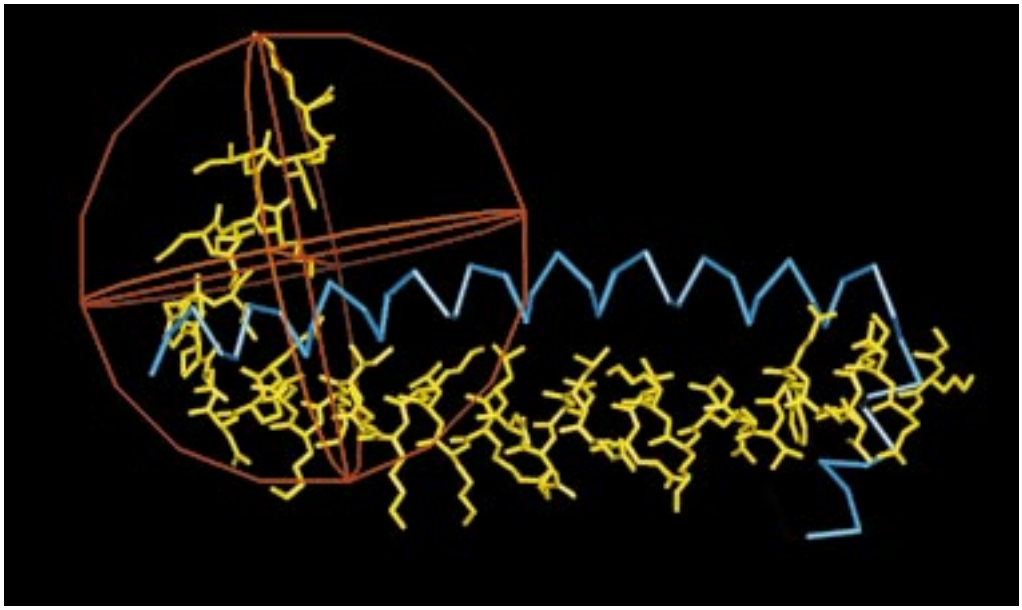

Figure 1. Structure of the dimer formed by a peptide corresponding to residues 12 to 60 of DAg. All non-hydrogen atoms of the target monomer are shown in yellow, an α-carbon trace of the complementary monomer is in blue, and the red sphere denotes the MCSS distribution region. The sequence for residues 12 to 60 is as follows: GREDILEQW VSGRKKLEEL ERDLRKLKKK IKKLEEDNPW LGNIKGIIGK.

monomer is L-shaped consisting of a long N-terminal helix and a short C-terminal helix separated by a bend. The N-terminal helices associate to form a dimeric anti-parallel coiled-coil approximately 60 Å long (Figure 1) [4]. The unusual anti-parallel orientation of this coiled-coil is stabilized by extensive hydrophobic interactions between residues in the bend region of one monomer with residues near the N-terminus of the long helix of the complementary monomer.

Transfecting cells with a plasmid containing the N-terminal third of DAg has been shown to inhibit viral replication [2, 5], suggesting that small molecules capable of disrupting dimerization and subsequent oligomerization could be viable as antivirals. It should be noted that the disruption of the dimer poses a significant challenge since the coiled-coil region has been shown to be very stable (its melting point is approximately 80 °C [6]). The recent demonstration, however, that peptides corresponding to a coiled-coil region (a 30-mer) and a helical region (a 35-mer) of HIV-1 gp41, respectively, exhibit antiviral activity in patients [7–9], together with the crucial role of the DAg as a transactivator of replication early in infection when levels of DAg are low, provide cogent arguments for the viability of this approach.

The overall objective of the present study is to design ligands that will disrupt dimerization by binding to the hydrophobic pocket formed by one monomer (the target monomer), thereby preventing binding of the other monomer (the complementary monomer). The design of short D-peptides was undertaken primarily because peptides are relatively easy to synthesize, and because D-peptides are not as readily degraded *in vivo* and are less likely to be immunogenic [10, 11]. Peptidyl inhibitors could later serve as templates for the design of small molecule peptidomimetics.

The computational strategy that we describe is the analogue of experimental mirror image phage display [12]. The approach is used to design a series of helical D-peptides as potential inhibitors of DAg oligomerization. The first step in the process involved the use of the MCSS method [13, 14] to determine optimal locations of various functional groups in the hydrophobic pocket centered at the bend in the dimerization domain. Analysis of the resulting functional group maps led to the identification of six critical interaction sites. Four of these interaction sites are satisfied when the complementary monomer binds; additional functional groups which can preferentially satisfy these four sites as well as ones which can interact at the other two sites were identified and incorporated into the designed ligands. The Protein Data Bank (PDB) was searched for suitable structural templates that matched up to four of these interaction sites. Of approximately 100 search

hits, a 14-residue helical structure expected to bind to the mirror image of the target was selected as the scaffold for subsequent design work. Over 200 000 sequences were generated based on the MCSS results and starting from the scaffold sequence. These sequences were then screened using a neural network secondary structure prediction algorithm [15] and only those that were predicted to be highly helical were retained. About 100 of the remaining sequences were then modeled as D-peptides and energy minimized with the fixed target structure. Subsequent modifications were made to a number of these peptides to optimize intrahelical salt bridge formation on the noninteracting side of the helix and to potentially reduce the cost of synthesis. Based on calculated interaction energies, predicted helicities, and expected intrahelical salt bridge patterns, a small number of candidate D-peptides have been selected for synthesis. Assays to measure their interaction with the target and inhibition of DAg dimerization are currently under development.

Materials and methods

Preparation of targets

Monomer B of the crystal structure of residues 12 to 60 of DAg [4] (PDB1A92) was taken as the target structure. The N-terminus was acetylated and the C-terminus was N-methylated. An all-hydrogen target was prepared using the hbuild command in CHARMM [16] with standard PARAM22 parameters [17]. A polar-hydrogen target was prepared using standard PARAM19 parameters [18].

MCSS calculations

The MCSS method [13, 14] determines energetically favorable positions for functional groups on the surface of a protein with known three-dimensional structure. N copies of a given functional group (where N is typically a few thousand) are randomly placed in the binding-site sphere or box and subjected to simultaneous energy minimization. Each copy of the group interacts with the full force field of the target structure but the group copies do not interact with each other. If two copies of a group minimize to the same position then one of them is discarded. After functional group maps have been calculated for a diverse set of groups, other computational methods can be used to link the minima together to form candidate molecules whose functional groups are optimally positioned to bind in the target site [19–21].

Table 1. Functional groups

Abbreviation	Description	Energy cutoff[a]
ARGR	Arginine side chain	−75
ASNR	Asparagine side chain	−20
ASPR	Aspartic acid side chain	−20
CHXR	Methyl cyclohexane	−4
CYSR	Cysteine side chain	−5
ETHR	Ethanol	−10
GLNR	Glutamine side chain	−25
GLUR	Glutamic acid side chain	−40
HISR	Histidine side chain	−10
ILER	Isoleucine side chain	−4
LEUR	Leucine side chain	−4
LYSR	Lysine side chain	−50
MAMM	Methyl ammonium	−40
MEOH	Methanol (serine side chain)	−20
METR	Methionine side chain	−7
NITR	Acetonitrile	−20
NMAC	N-methyl acetamide	−15
PHEN	Phenol	−20
PHER	Phenylalanine side chain	−5
PRPN	Propane (valine side chain)	−4
THIZ	Thiazole	−19
THRR	Threonine side chain	−20
TRPR	Tryptophan side chain	−10
WATR	Water	−10

[a]In kcal/mol.

For the present application, a 16 Å radius sphere centered on atom CG2 of Val 21 of the complementary monomer was specified for the initial distribution of group copies; this sphere encompasses the hydrophobic core region formed by the bend in the target monomer (Figure 1). Functional group maps were calculated for the groups listed in Table 1 using both the polar representation (the extended atom polar hydrogen representation for the group and the target) and the hybrid representation (the polar hydrogen representation for the group and the all hydrogen representation for the target). (A detailed comparison of the resulting polar- and hybrid-representation minima will be published elsewhere.) Both representations gave very similar results, overall. Since the interaction energy provides a relative ranking of the minima in the binding site, a cutoff energy was chosen for each group and only minima with energy below the cutoff were considered. For each hybrid-representation map, a histogram was constructed of the number of minima per

interaction energy bin. Each histogram indicated a set of minima with significantly lower interaction energies than others of that type. For each group, the energy cutoff was taken as the lower of the solvation free energy for the group or the highest energy of the minima in this set on the histogram. To summarize the MCSS results, minima were clustered according to the interactions that they make with the target structure in order to identify six critical interaction sites.

Identification of clusters

Clusters of minima were identified using an algorithm that initially assigns each group minimum to its own cluster. Then starting with the first minimum, the center-of-mass (CM) distance is calculated between that group copy and all others. If the distance is less than a specified cutoff then the minima are considered to be in the same cluster. This is done successively for all other minima. When the distance between minima in two different clusters is less than a specified cutoff, the two clusters are combined into one cluster. The CM cutoff was chosen to be 3 Å unless otherwise stated.

Database searches

The Sequence Database Search program in Quanta [22] was used to search a database for peptide structures suitable for use as templates or scaffolds in the design process; the database consisted of the first 300 protein structures in the 1993 PDB. The search criteria specified the critical interaction sites to match (defined as atom positions) as well as the distance and allowed length of the chain between sites, but not the relative orientation of the sites. Therefore about half the hits retrieved would be expected to bind to the actual target and half to the mirror image of the target (Figure 2). The strictest search constraints that yielded hits specified an overlap with four of the critical interaction sites. The search that resulted in the selection of the template described below retrieved all segments of structures with residues that overlapped sites 1, 2, 3 and 5; the CB atom of each site was within 1.0 Å of the CB atom of the matched protein residue and matched residues were between 4 and 8 Å apart and separated sequentially by no more than 6 residues. Specifically, hydrophobic residues were allowed at site 1, Trp at site 2, positive residues at site 3, and any residue at site 5. Residues 314 to 327 of lactate dehydrogenase (PDB accession #PDB2LDX) [23], a helical segment located on the exterior of the protein, were selected as the scaffold. This 14-residue helix would be expected to bind to the mirror image of the target, which is equivalent to the D-isomer of this peptide binding to the actual target.

Mutant sequence generation

A C program was written to automatically generate all possible mutant sequences given a starting sequence and a list of allowed mutations at each position. The starting template sequence, residues 314 to 327 of PDB2LDX, was used. The allowed mutations were based upon the MCSS results and the initial docking of the D-isomer of the template structure with the target. Residues were chosen for mutation because their side chains could potentially overlap with one of the sites and the selected residues were mutated to those amino acids with corresponding low energy MCSS minima at nearby positions. The resulting sequences were submitted directly as input to a secondary structure prediction program.

Secondary structure prediction

The neural network-based Secondary Structure/Class Prediction program [15] was used to predict secondary structure. While this neural network was trained on a database of protein structures and not peptides, it seems likely that a short sequence predicted to be highly helical will be at least temporarily a helix in solution. The probability that a peptide is helical does not depend on whether it is composed of D- or L-amino acids.

D-peptide structure generation

For approximately 100 of the sequences predicted to be the most helical, model structures were built using Quanta [22]. The mutant structures were generated by positioning each mutated side chain so that it overlapped as closely as possible with the corresponding side chain in the starting scaffold structure (residues 314 to 327 of PDB2LDX as docked to interact with the D-isomer of the target). If this was not possible the side chain was placed in its most common rotamer [24]. The D-isomer of the peptide was then generated by reflecting the L-peptide through the $x = 0$ plane. At this point, when possible, side chain torsion angles were manually adjusted to position each side chain so that it overlapped with clusters of low energy MCSS minima for similar functional groups.

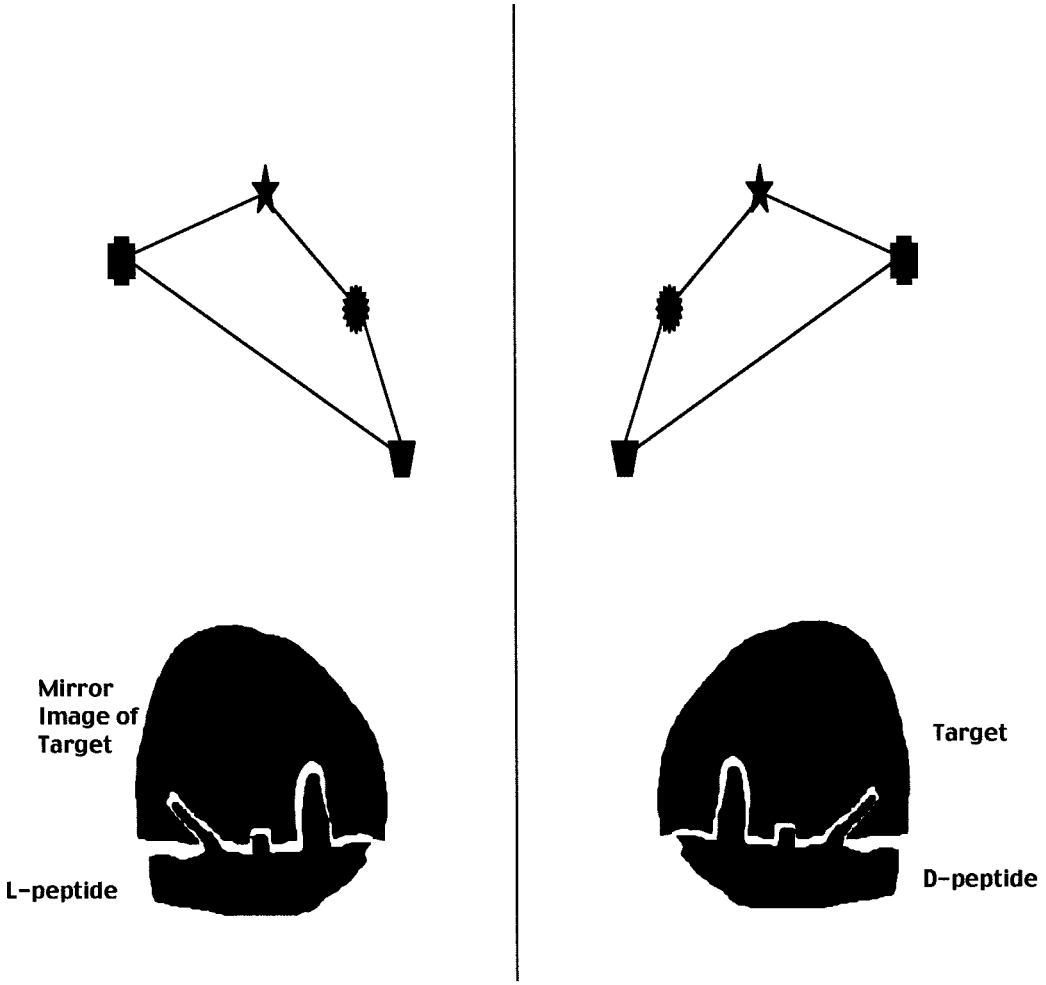

Figure 2. A schematic representation of four of the critical interaction sites and their mirror image. The search program was given information about which residue types were allowed at each site (represented by different shaped polygons) and the distances between sites (represented by the connecting lines). A particular structure as well as its mirror image satisfy the same set of constraints. Below the sites, an L-peptide is docked to the mirror image of the target, and after reflection the corresponding D-peptide is docked to the target.

Engineering amphipathicity and salt bridges

Additionally, residues on the solvent-exposed side of the helix (1, 4, 5, 8, 9, and 12) were constrained to be hydrophilic while those on the inner surface, expected to bind to the target (2, 3, 6, 7, 10, 11, 13, and 14), were hydrophobic unless the MCSS results dictated otherwise. Given the importance of the periodicity of polar and nonpolar amino acids in determining the secondary structure of self-assembling oligomeric peptides [25], this hydrophilicity pattern requirement is expected to make the peptides more helical. Salt bridges between i and i + 3 residues and between i and i + 4 residues have also been shown to stabilize small peptides in a helical conformation [26]. In addition, a negatively charged residue at the N-terminus of a helix further increases its stability by counteracting part of the helix dipole [26, 27]. Therefore residues on the solvent-exposed outer surface of the helix (1, 4, 5, 8, 9, and 12) were mutated to engineer salt bridges to stabilize the helix. There were four possible patterns of salt bridges: *4-4* (residues 1:5, 8:12), *4-4-3* (1:5, 4:8, 9:12), *3-4-4* (1:4, 5:9, 8:12), and *3-3-3* (1:4, 5:8, 9:12) (see Figure 3). With the *4-4-3* and *3-4-4* patterns, interactions may occur between neighboring salt bridges and this could have a destabilizing effect on the helix structure; e.g., in the *4-4-3* case, a Glu 4:Lys 5 interaction seems to weaken both the Glu 1: Lys 5 and the Glu 4: Lys 8 salt bridges

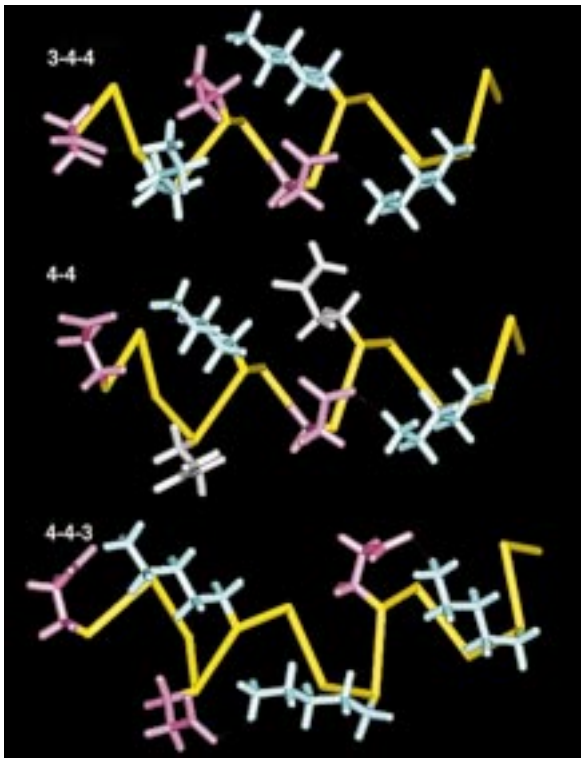

Figure 3. The three salt bridge schemes employed are shown; *3-4-4* between residues 1:4, 5:9, and 8:12, *4-4* between 1:5 and 8:12, *4-4-3* between 1:5, 4:8, and 9:12. Lys residues are in blue, Glu in pink, and Gln in gray.

and distort the helix structure somewhat. With the *4-4* pattern, residues 4 and 9 (which are unspecified by the MCSS results or the salt bridge requirements) were only allowed to be Gln, because in general Gln has the highest helical propensity [28]. Since i: i + 4 salt bridges have been shown to be more effective at stabilizing helices than i: i + 3 salt bridges [26], the *3-3-3* salt bridge scheme was not used.

Interaction energy calculation

Each mutant D-peptide structure was energy minimized in the presence of the fixed target and in vacuum. Ligand-target (LT) and ligand-ligand (LL) interaction and internal ligand energy terms were calculated (L_0). Minimizations consisted of 200 steps of Steepest Descent followed by up to 3500 steps of Adopted Basis Newton-Raphson until a gradient tolerance of 0.01 kcal/mol · Å was reached. All energy calculations were carried out using CHARMM with a constant dielectric of 4.0, the shift truncation function for the electrostatic energy terms, and switch for the van der Waals with a nonbond list cutoff of 13.0 Å (and ctonnb 10.0 Å and ctofnb 12.0 Å) [16]. Test calculations confirmed that the D- and L-isomers of a given peptide are isoenergetic using the standard CHARMM PARAM22 force field [17].

After minimization with the target, each peptide-target complex was examined and overlayed with MCSS maps for groups similar to the peptide side chains. Occasionally side chain torsion angles were manually adjusted to move a side chain so that it overlapped clusters of corresponding minima. Whenever any side chains were moved, the new structure was minimized as described above and the lowest energy final structure was accepted. In all cases, the all-atom root-mean-square difference (RMSD) between the starting peptide structure and the structure after minimization in the presence of the target was less than 1.1 Å. Several of the peptides were subsequently modified to ensure optimal intrahelical salt bridge formation on the non-interacting side of the helix, and again minimization was repeated. Normalized interaction energies (LT + LL − L_0) were tabulated, by adding an approximate ligand strain term (LL − L_0) to the ligand-target interaction term. Peptides were selected as candidate ligands, based on the normalized interaction energies, predicted helicity, and intrahelical salt bridge patterns.

Results

Detailed analysis of MCSS minima

MCSS maps were generated for polar, charged, aromatic, and hydrophobic functional groups (see Table 1 for a list). In general, most of the low energy minima are located in or nearby the hydrophobic pocket, often nearby charged groups and in small clusters (ranging in size from one to ten minima). As a representative example, the functional group map for TRPR is shown in Figure 4. To test the use of the methodology with this system, the calculated functional group minima were compared to the positions of the side chains of the complementary monomer that interact with the target monomer. For nine of these ten side chains, the MCSS calculations identified minima for the corresponding functional group as well as chemically similar groups at the side chain location; at the location of the tenth, Arg 13 of the complementary monomer, minima were found for hydrophobic residues. However, it is the aliphatic part of the Arg side chain that

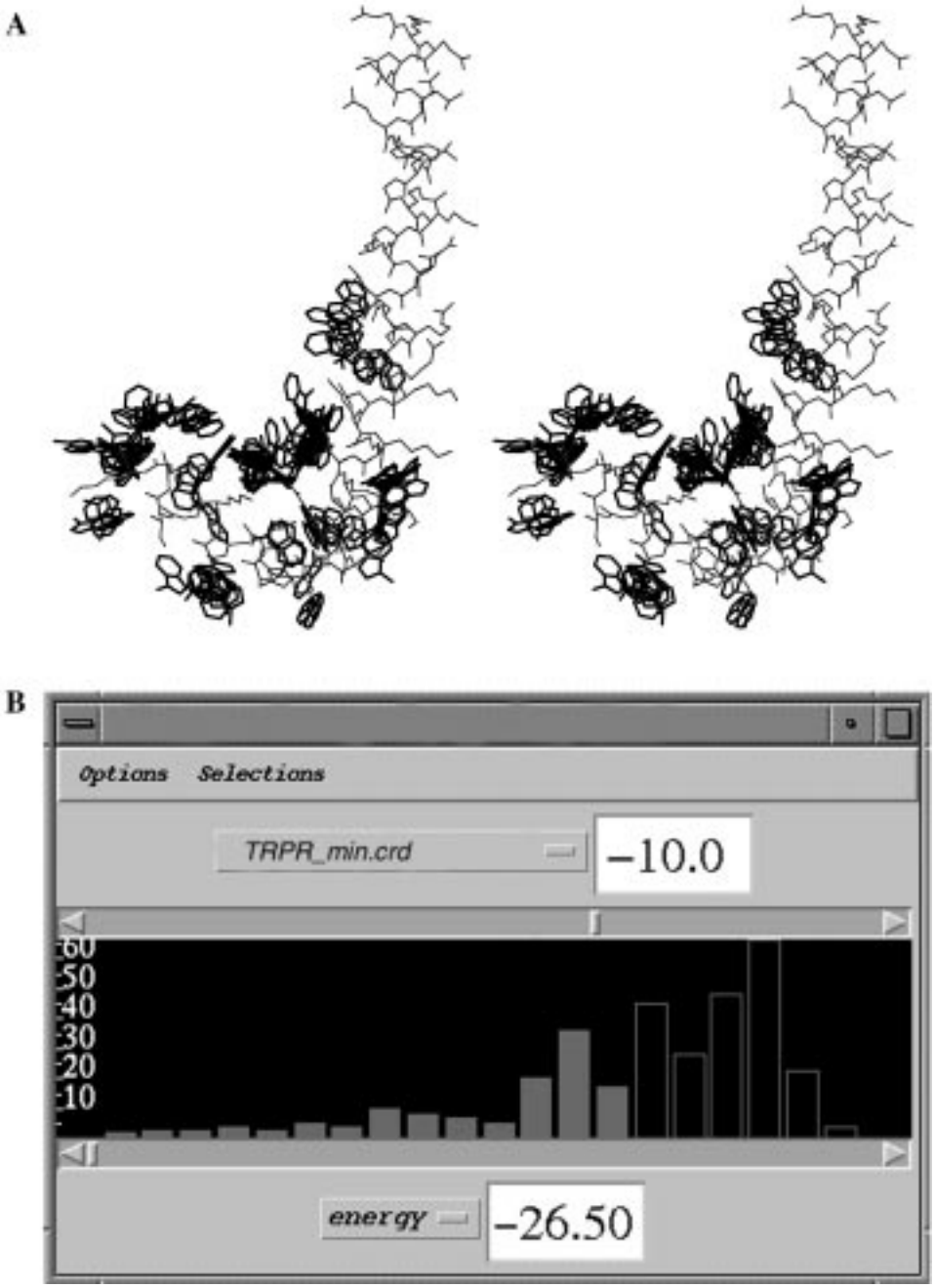

Figure 4. TRPR minima. In (A) a stereoplot of the target monomer is shown in thin black lines with the 113 lowest energy hybrid hydrogen TRPR minima in thick black lines ($E < -10$ kcal/mol). In (B) a histogram of the TRPR minima interaction energies is shown. The solid gray bars correspond to minima with $-25.8 < E < -10$ kcal/mol and the unfilled bars to energies greater than -10 kcal/mol.

Table 2. Distribution of MCSS minima

Group	Cluster number	Cluster size	Energy range	Nearest neighbors
Minima for large polar groups				
NMAC	1	2	−54.8, −53.7	Glu 31, Lys 38
	2	9	−51.7, −29.5	Glu 45, Lys 55
	3	1	−49.2	Lys 26, Asp 33
ASNR	1	7	−55.4, −42.8	Glu 45, Lys 55
	2	6	−52.7, −47.4	Glu 31, Lys 38
	3	1	−48.0	Lys 26, Asp 33
GLNR	1	20	−57.9, −30.8	Glu 45, Lys 55
	2	4	−55.8, −52.5	Glu 31, Lys 38
	3	1	−50.4	Lys 26, Asp 33
Minima for small polar groups				
WATR	1	2	−46.9, −46.7	Glu 45, Lys 55
	2	1	−40.3	Lys 26
	3	2	−34.2, −33.3	Glu 31
	4	1	−31.8	Lys 39, Lys 36
ETHR	1	1	−25.2	Lys 38
	2	1	−24.4	Lys 26
	3	2	−23.4, −23.2	Glu 45, Lys 55
MEOH	1	6	−41.9, −39.4	Glu 45, Lys 55
	2	2	−31.2, −30.8	Lys 38
	3	2	−31.2, −30.8	Lys 36
CYSR	1	1	−18.3	Glu 45, Lys 55
	2	1	−16.1	Lys 38
	3	2	−14.9, −14.0	Lys 26
THRR	1	8	−42.2, −23.0	Glu 45, Lys 55
	2	7	−37.9, −20.6	Glu 31
	3	5	−31.5, −27.2	Asp 33
METR	1	7	−11.8, −8.1	Glu 45, Lys 55
	2	3	−11.3, −7.8	Lys 38
	3	11	−8.7, −7.2	Lys 42
	4	8	−8.1, −7.1	Lys 43
	5	10	−7.9, −7.1	Lys 40
NITR	1	2	−32.1, −31.6	Lys 55
	2	1	−29.5	Lys 38
	3	1	−26.1	Lys 55
Minima for charged groups				
ARGR	1	2	−99.7, −99.3	Glu 45, Lys 55
	2	18	−84.5, −46.7	Lys 38, Ile 41
	3	9	−72.5, −65.9	Glu 46, Asp 47
	4	1	−71.9	Glu 31
LYSR	1	9	−113.0, −107.1	Glu 45, Lys 55
	2	1	−94.6	Glu 31
	3	28	−93.3, −89.1	Glu 45
MAMM	1	6	−111.3, −104.1	Glu 46
	2	2	−91.3, −90.9	Glu 45
	3	2	−79.5, −78.5	Asp 47
ASPR	1	2	−87.2, −86.2	Lys 40
	2	2	−84.3, −83.8	Lys 38

Table 2. (continued)

Group	Cluster number	Cluster size	Energy range	Nearest neighbors
ASPR	3	1	−83.5	Lys 26
	4	2	−83.0, −82.2	Lys 55
	5	6	−82.7, −82.5	Lys 60
	6	1	−72.0	Lys 43
GLUR	1	4	−89.1, −87.6	Lys 40
	2	4	−85.9, −85.3	Lys 38
	3	1	−85.0	Lys 26
	4	3	−84.6, −84.0	Lys 60
	5	5	−84.5, −83.7	Lys 55
	6	4	−84.0, −83.8	Lys 43
Minima for aromatic groups				
PHEN	1	9	−41.8, −23.5	Leu 51, Gly 52
	2	3	−40.6, −38.1	Ile 41, Lys 55
PHER	1	12	−8.8, −5.3	Pro 49
	2	2	−8.0, −5.6	Leu 51, Gly 52
	3	4	−7.6, −5.8	Leu 37
	4	11	−7.5, −5.7	Ile 41, Lys 55
	5	13	−7.5, −5.1	Leu 44
TRPR	1	2	−23.9, −20.6	Lys 38
	2	17	−23.0, −16.9	Ile 41, Leu 51, Ile 54
	3	2	−21.1, −17.0	Pro 49
	4	8	−19.8, −10.2	Lys 43, Asp 47
	5	4	−18.4, −15.3	Leu 30
HISR	1	1	−40.6	Lys 38, Leu 34
	2	9	−38.4, −13.0	Asp 33
	3	6	−37.9, −23.5	Leu 34, Lys 55
THIZ	1	10	−27.5, −21.7	Lys 55
	2	2	−26.0, −24.2	Glu 31
Minima for hydrophobic groups				
CHXR	1	11	−7.9, −4.6	Pro 49
	2	47	−7.3, −4.0	Ile 41, Lys 55
	3	20	−7.1, −5.4	Ile 41, Lys 38
	4	4	−6.9, −6.1	Lys 39, Lys 40
	5	8	−6.8, −5.6	Leu 37
PRPN	1	8	−5.4, −4.2	Glu 45, Pro 49
	2	6	−5.2, −4.6	Ile 41
	3	10	−5.2, −4.1	Leu 51, Ile 58
	4	4	−5.0, −4.7	Leu 37
	5	4	−4.8, −4.6	Leu 34
	6	4	−4.7, −4.3	Lys 43
	7	4	−4.7, −4.2	Lys 40
ILER	1	9	−6.5, −4.1	Ile 41
	2	7	−6.0, −4.4	Leu 37
	3	12	−5.6, −4.0	Lys 40
	4	16	−5.6, −4.4	Ile 41
LEUR	1	9	−5.8, −4.9	Ile 41
	2	7	−5.7, −4.6	Lys 40
	3	5	−5.7, −5.1	Leu 37
	4	4	−5.6, −4.9	Ile 41

Table 3. Sequences generated[a]

Sequence position	Allowed amino acids	Sites potentially occupied
1	E	
2	FMRTYW	1
3	EFDKMNQRW	1
4	EKQ	
5	EKQ	
6	ALHI	4
7	DEHIKLNRSTV	4
8	EKQ	4
9	EKQ	5
10	EFIMWY	2
11	DEFMQSVWY	2
12	K	
13	AFHKRSY	3
14	KR	3

[a]Not all combinations of the allowed amino acids at each position were considered. Often the selection of an amino acid for one position limits the amino acids allowed at another position. For example, all peptides with a K at position 5 were required to have an E at position 4 to ensure an acceptable salt bridge pattern. The total number of sequences generated was about 200 000.

interacts with the target monomer; the guanidinium group of Arg 13 points away from the dimer interface towards solvent. The MCSS method thus correctly identified the residues known from experiment to interact directly with the target. A description of the distribution of all calculated minima is given in Table 2. The number of minima per cluster generally increases with the size of the functional group, especially for the lower energy clusters that tend to be smaller. (Positively charged minima are sometimes located nearby positively charged residues, however the orientations of the minima are such that the like charges are never pointing directly towards each other.) While all energy values and minima counts given are for the hybrid hydrogen sets unless otherwise stated, the trends are the same with the polar hydrogen minima.

Identification of critical sites

The functional group maps were summarized by partitioning the binding region into six critical interaction sites. Clusters of low energy minima were identified at each site. The designed peptide ligands were required to optimally fill as many of these six interaction sites as possible to optimize their overall binding affinity for the target. The locations of the sites are shown in Figure 5. A brief description of the types of minima found at each site follows.

Site 1 is defined by the position of Arg 13 of the complementary monomer, specifically the aliphatic part of the side chain. Hydrophobic groups, such as CHXR, PHER and ILER have large numbers of minima with relatively low energies here. Similarly, the hydrophobic portions of the various types of minima at this position interact with Trp 50 of the target monomer. The charged groups ARGR, ASPR, GLUR and LYSR have no minima at this site.

Site 2 is defined by the position of Trp 20 in the complementary monomer. It is located at the center of the hydrophobic pocket formed by the bend in the helix of the target monomer, and therefore optimally filling this site is crucial. A functional group at this site can potentially interact with a number of hydrophobic target residues (e.g., Ile 41, Leu 51, Ile 54, and Ile 58). As expected, all hydrophobic groups have low energy minima at this site. The larger aromatic groups such as TRPR fill the pocket better. The inclusion at this position of unnatural amino acids that are larger than TRPR (e.g., naphthalene or anthracine) may be considered in future studies.

Site 3 is defined by Arg 24 of the complementary monomer, which is nearby Ile 41 and Glu 45 of the target monomer. While this arginine interacts with the neighboring Glu 28 of the same monomer in the X-ray structure, a group at this position can potentially interact with Glu 45 of the target monomer instead. Thus, to facilitate interaction with Glu 45, this site should be occupied by a positively charged group. ARGR has many low energy minima at this location.

Site 4 is defined by a HISR minimum ($E = -21.2$ kcal/mol), which interacts with the hydrocarbon chain of Lys 40 of the target monomer. Minima with relatively low energies are found here for cyclic functional groups, such as HISR, THIZ, PHEN. This site, which is occupied by Gln 19 in the complementary monomer, could also be occupied by a five-membered or six-membered aromatic ring or by other polar groups.

Site 5 is defined by the location of Leu 17 of the complementary monomer, and interacts with Leu 51 and Ile 54 of the target monomer. Based on functional group maps for NITR, this site could be occupied by a polar side chain with a partial negative charge. Many minima nearby this site have relatively high energies, indicating that this site is not as well defined as the others and corresponds to a shallower and probably broader potential energy well.

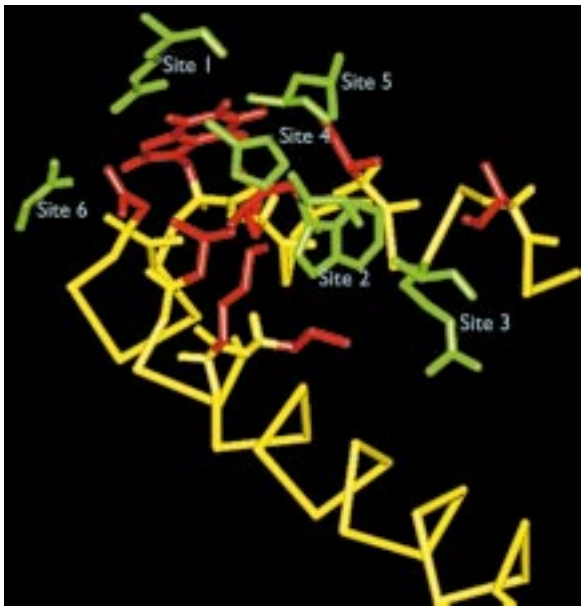

Figure 5. The six critical interaction sites identified by MCSS are shown in green. The target backbone is in yellow and its side chains that interact with the complementary monomer and fall within the MCSS region are drawn in red.

Site 6 is defined by an NMAC minimum (E = −15.5 kcal/mol), located near Asn 48, Pro 49 and Trp 50 of the target monomer. The best minima at this site, in terms of their relative energies, were generally for polar groups including NMAC and CYSR; the site could be occupied by one of these groups or a closely related group. The complementary monomer does not place a side chain at this site and therefore does not make an analogous interaction.

Candidate D-peptide ligands

A 14 residue D-peptide helix was selected as the backbone scaffold, as described in the Methods section. Over 200 000 sequences were generated starting from the scaffold sequence and based on the MCSS results and subject to the additional constraints that an appropriate pattern of hydrophobicity /hydrophilicity and of intrahelical salt bridges must be satisfied. Residues on the inner surface of the peptide whose side chains could potentially occupy one of the six critical interaction sites were mutated to satisfy the MCSS-derived requirements of that site. The amino acids allowed at those positions are summarized in Table 3. Increasing the helicity of the peptide by engineering amphipathicity and salt bridges on the solvent-exposed side of the helix is expected to stabilize the more favored conformation of the peptide for binding and to decrease the entropic cost of binding the peptide to the target. Only natural amino acids were incorporated, so that the sequences could be screened using a secondary structure prediction algorithm [29]. Approximately 100 of the sequences predicted to be most helical were then model built as D-peptides and energy minimized in the presence of the fixed target monomer.

For each model built D-peptide, an interaction energy with the target structure was calculated (LT + LL − L_0). These interaction energies ranged in energy from −103.9 to −48.4 kcal/mol (Table 4). The van der Waals (VDW) and electrostatic (ELEC) components of these energies were examined; other studies have indicated that the VDW component can sometimes be a reasonable predictor of binding affinity [30–35]. The VDW component of the interaction energies ranged from −49.7 to −21.2 kcal/mol, while the ELEC ranged from −60.4 to −10.8 kcal/mol.

Peptides 1, 4 and 7 from Table 5 (listed in bold print) represent a diverse set that would be predicted by different criteria to bind well. Peptide 1 has the lowest interaction energy as well as the lowest VDW component, although it is predicted to be less helical than most of the other top 20 peptides (Table 4). A model of peptide 1 interacting with the target monomer is shown in Figure 6. Peptide 4 has a low interaction energy and is predicted to be more helical than nearly all of the other 200 000 sequences studied. Peptide 7 has a low (favorable) interaction energy, a very low VDW component, and is predicted to be highly helical. The three peptides (see Figure 3) are expected to adopt different intrahelical salt bridge patterns (peptide1: *4-4-3*, peptide 4: *3-4-4*, and peptide 7: *4-4*). Peptides 1, 4, and 7 have been synthesized and are 12%, 12%, and 24% helical, respectively, by circular dichroism spectroscopy with significantly increased helical content (up to 77%) as trifluoroethanol is added.

Discussion

A novel computational ligand design scheme has been developed that is analogous to the experimental mirror image phage display technique [12]. It involves the use of the MCSS algorithm to optimally place functional groups in the binding region, database searches of the PDB to select a template structure expected to bind to the mirror image of the target, and the generation of mutant sequences that incorporate the MCSS results and are engineered to be helical. The approach has

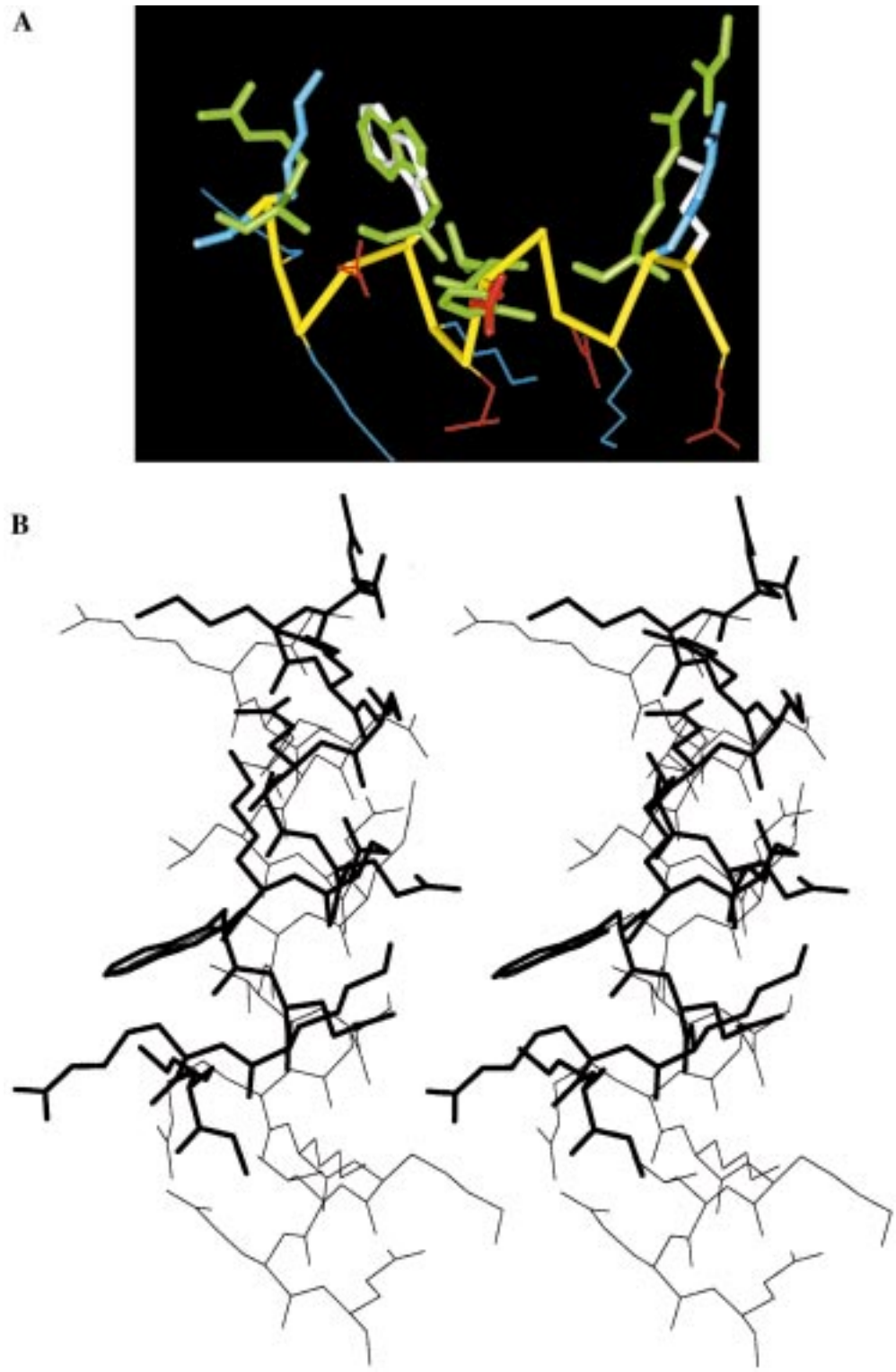

Figure 6. A model of one of the candidate peptides (EMRKEAEEKWEKRK) is shown in (A) overlapping the MCSS sites. The backbone is in yellow, positively charged side chains in blue, negatively charged side chains in red, and neutral side chains in white. The MCSS critical interaction sites are shown in green. In (B) a stereoplot of the same candidate peptide is shown in thick black lines overlapping with the complementary monomer in thin black lines, demonstrating that it should not be possible for the peptide and complementary monomer to bind simultaneously.

Table 4. Comparison of interaction energies of the candidate peptides

#	Sequence	Interaction energy[a]	Total bonded	VDW	Electrostatic	Salt bridge[b]
1	**EWKEKADKEWEKRK**	**−103.9**	**6.2**	**−49.7**	**−60.4**	**4-4-3**
2	EWKEKADKEWEKRR	−103.1	5.3	−49.3	−59.2	4-4-3
3	EWKQKADEQWEKRR	−91.8	5.9	−40.5	−57.2	4-4
4	**EMRKEAEEKWEKRR**	**−90.5**	**5.6**	**−42.4**	**−53.7**	**3-4-4**
5	EMREKAEKEWEKRR	−88.9	7.6	−38.8	−57.8	4-4-3
6	ERRKEAEEKYWKRK	−88.2	6.7	−47.5	−47.4	3-4-4
7	**ERRQKAEEQYWKRK**	**−86.0**	**3.8**	**−46.4**	**−43.5**	**4-4**
8	ERRQKAEEQYWKRR	−82.0	6.0	−46.9	−41.1	4-4
9	ERREKAEKEYWKRK	−78.9	7.7	−45.2	−41.4	4-4-3
10	EYRQKATEQWEKRR	−75.1	6.3	−40.8	−40.6	4-4
11	EMRQKAEEQMWKRR	−71.6	12.3	−45.4	−38.6	4-4
12	EMEQKAEEQMWKRR	−71.4	8.4	−41.9	−37.9	4-4
13	EWKKEADEKWEKRR	−71.3	8.7	−34.4	−45.6	3-4-4
14	EMEQKAEEQMWKRR	−70.4	9.5	−41.4	−38.5	4-4
15	EMRQKAEEQWEKRR	−68.5	1.4	−36.9	−33.0	4-4
16	EYEQKAEEQMWKRR	−66.0	11.4	−42.3	−35.1	4-4
17	EMRKEAEEKWEKRK	−64.2	12.6	−44.1	−32.8	3-4-4
18	ETEQKANEQHVKRR	−63.9	2.9	−35.1	−31.6	4-4
19	EMRQKAEEQWWKRR	−59.1	9.5	−44.9	−23.7	4-4
20	EWRQKAEEQYEKRK	−55.2	5.3	−44.7	−15.8	4-4
21	EAEQAAAEAAAKAR[c]	−49.7	2.6	−23.3	−29.0	4-4
22	GREDILEQWVSGRK[d]	−52.1	3.2	−42.4	−13.0	–

[a]The peptide was minimized in the presence of the target and its energy when minimized in a vacuum was subtracted to yield the interaction energy.
[b]4-4 indicates salt bridges between residues 1:5 and 8:12, 4-4-3 between residues 1:5, 4:8, and 9:12, and 3-4-4 between residues 1:4, 5:9 and 8:12.
[c]Peptide 21, which is all alanine residues except for the salt-bridges, is included as a control.
[d]Peptide 22 consists of residues 12 to 25 of the complementary monomer from the crystal structure. This interaction energy (−52.1 kcal/mol) is included as a further control.

been applied to design candidate D-peptide ligands to bind to one monomer of HDAg in the hydrophobic core of the dimerization region, and thereby block dimerization. Fourteen-residue D-peptides of considerable diversity have been designed. Each of the most promising candidate peptide ligands has a favorable interaction energy with the target monomer, is predicted to be helical, is expected to make intrahelical salt bridges that stabilize the structure, and has the appropriate pattern of alternating hydrophobic and hydrophilic residues required for coiled-coil formation upon binding. A similar computational scheme could be used to design non-helical D-peptide ligands.

Each of the selected peptides is predicted to occupy four of the six critical interaction sites described. The candidate ligands are fairly similar to the complementary monomer, which is not surprising since the complementary monomer also satisfies four of the six critical interaction sites. In the designed peptides, however, the side chains are positioned differently and make different interactions because they are in a left-handed helix compared to the right-handed helix of the complementary monomer. In each, a positively charged residue at position 3 is predicted to form a salt bridge with Asp 47, a negatively charged residue at position 7 with Lys 40, and a positively charged residue at position 14 with Glu 34, respectively, of the target monomer. The salt bridge between residue 7 and Lys 40 is expected to shield site 2, at the center of the hydrophobic pocket, from solvent. Since the backbone of each peptide is very close to site 4, it was not possible to incorporate a five or six-member ring here and maintain the expected helical structure of the peptide. On each of the candidate peptides, only three or four residues are identical to the original scaffold. Develop-

Table 5. Predicted helicity[a] at each position of the candidate[b] sequences

Peptide number	Residue Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	**11.1**	**28.4**	**26.5**	**64.2**	**57.0**	**44.3**	**17.1**	**46.4**	**74.6**	**78.1**	**67.5**	**49.1**	**17.1**	**2.9**
2	11.1	28.4	26.3	68.8	64.2	47.3	24.5	40.3	68.8	78.1	69.8	67.9	25.0	5.5
3	13.7	48.5	56.7	92.1	95.3	95.6	82.6	93.2	96.6	94.9	96.4	82.5	37.7	8.4
4	**18.3**	**70.1**	**83.8**	**96.3**	**99.2**	**99.2**	**96.6**	**96.6**	**94.4**	**94.9**	**94.4**	**68.3**	**38.2**	**7.9**
5	26.6	71.7	77.8	96.6	99.2	99.0	93.7	96.3	94.0	94.9	94.0	83.1	37.9	7.9
6	6.5	11.1	20.9	53.9	96.6	98.3	99.4	95.7	96.0	96.0	90.3	67.3	17.8	4.1
7	**8.5**	**27.7**	**46.4**	**83.2**	**95.4**	**98.3**	**100.0**	**100.0**	**95.7**	**95.7**	**94.1**	**81.8**	**40.1**	**10.0**
8	8.5	23.5	44.3	83.2	95.4	98.3	100.0	100.0	95.7	95.7	94.1	86.0	38.2	7.8
9	3.8	20.3	26.3	83.5	95.3	95.3	94.9	93.2	95.7	93.2	94.4	76.8	25.9	8.4
10	25.2	70.4	79.6	94.9	100	95.7	94.4	94.4	89.8	96.6	92.7	83.1	50.4	8.4
11	19.8	88.3	92.7	98.5	98.8	99.2	98.3	98.3	94.6	96.3	96.4	81.8	44.1	7.8
12	18.3	82.1	89.9	95.7	99.2	96.6	98.3	98.3	93.2	9.06	96.4	81.8	42.4	7.8
13	13.2	22.8	38.2	81.9	81.9	68.2	48.6	84.4	83.8	84.4	80.7	65.0	21.1	4.1
14	18.3	79.5	93.0	98.5	98.8	99.2	98.3	98.3	93.2	99.4	96.4	81.8	44.1	10.0
15	19.8	89.8	91.3	95.7	98.3	96.6	99.2	98.3	94.6	96.3	96.4	81.8	51.5	10.0
16	16.1	73.1	77.0	96.3	99.2	99.2	98.3	99.2	93.2	95.7	94.6	86.0	44.1	10.0
17	18.1	70.1	82.3	96.3	99.2	99.2	96.6	96.6	96.4	93.2	94.4	76.8	37.9	7.9
18	19.8	49.2	55.1	86.8	95.6	96.6	87.8	83.1	82.6	82.6	82.6	60.0	24.8	5.5
19	18.8	89.8	92.1	95.7	98.8	96.6	99.2	98.3	93.2	96.3	96.4	81.8	42.4	7.8
20	13.7	61.7	79.6	94.9	98.5	96.6	96.6	98.5	93.2	96.3	94.0	73.2	49.1	10.0
21	29.5	98.3	95.7	95.7	98.8	97.4	98.0	98.3	96.0	98.3	95.3	91.2	81.0	18.6

[a]The probability that residue *m* of peptide *n* will be helical (as predicted by the Secondary Structure Prediction Program [15]) is given in column *m*, row *n*.
[b]See Table 4 for the sequences.

ment of an assay to measure inhibition of dimerization is underway.

Future work will involve the design of combinatorial libraries of candidate D-peptide ligands based upon the results presented here. These libraries may include both natural and unnatural amino acids. The binding mode of peptides that are shown experimentally to interact with DAg will be confirmed by NMR spectroscopy. Ultimately, based on our D-peptide results, we would like to design small molecule peptidomimetics that bind in the hydrophobic core formed by the bend in the helix of one monomer of DAg and thereby block dimerization. In parallel, the computational approach presented here is being applied to design D-peptide ligands to inhibit formation of higher order DAg oligomers which is mediated by residues 50–60.

Acknowledgements

This work was supported in part by a Giovanni Armenise-Harvard Foundation fellowship to D.J.-M. and NIH grant AI32480 to J.M.H.

References

1. Fields, B.N., Knipe, D.M. and Howley, P.M. (Eds.), Fields Virology, 3rd ed., Lippincott-Raven Publishers, Philadelphia, PA, 1996.
2. Xia, Y. and Lai, M.M.C., J. Virol., 66 (1992) 6641.
3. Wang, J. and Lemon, S., J. Virol., 67 (1993) 446.
4. Zuccola, H., Lemon, S., Erickson, B. and Hogle, J.M., Structure, 6 (1998) 821.
5. Chao, M., Hsieh, S.Y. and Taylor, J., J. Virol., 64 (1990) 5066.
6. Rozzelle, J., Wang, J., Wagner, D., Erickson, B. and Lemon, S., Proc. Natl. Acad. Sci. USA, 92 (1995) 382.
7. Wild, C., Oas, T., McDanal, C., Bolognesi, D. and Matthews, T., Proc. Natl. Acad. Sci. USA, 89 (1992) 10537.
8. Wild, C., Shugars, D., Greenwell, T., McDanal, C. and Matthews, T., Proc. Natl. Acad. Sci. USA, 91 (1994) 9770.
9. Chen, C., Matthew, T., McDanal, C., Bolognesi, D. and Greenberg, M., J. Virol., 69 (1995) 3771.
10. Milton, R., Milton, S. and Kent, S., Science, 256 (1993) 1445.
11. Guichard, G., Benkirane, N., Zeder-Lutz, G., Van Regenmortel, M., Briand, J. and Muller, S., Proc. Natl. Acad. Sci. USA, 91 (1994) 9765.
12. Schumacher, T., Mayr, L., Minor, D., Milhollen, M., Burgess, M. and Kim, P., Science, 271 (1996) 1854.
13. Miranker, A. and Karplus, M., Proteins, 11 (1991) 29.
14. Evensen, E., Joseph-McCarthy, D. and Karplus, M., MCSS version 2.1 (1997).

15. Chandonia, J. and Karplus, M., Secondary structure/class prediction program (1997).
16. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M., J. Comput. Chem., 4 (1983) 187.
17. MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. and Karplus, M., J. Phys. Chem. B, 102 (1998) 3586.
18. Neria, E., Fischer, S. and Karplus, M., J. Chem. Phys., 105 (1996) 1902.
19. Caflisch, A., Miranker, A. and Karplus, M., J. Med. Chem., 36 (1993) 2142.
20. Eisen, M., Wiley, D., Karplus, M. and Hubbard, R., Proteins, 19 (1994) 199.
21. Miranker, A. and Karplus, M., Proteins, 23 (1995) 472.
22. Quanta, MSI (1992).
23. Hogrefe, H., Griffith, J., Rossman, M. and Goldberg, E., J. Biol. Chem., 262 (1987) 13155.
24. Ponder, J.W and Richards, F.M., J. Mol. Biol., 193 (1987) 775.
25. Xiong, H., Buckwalter, B.L., Shieh, H.-M. and Hecht, M.H., Proc. Natl. Acad. Sci. USA, 92 (1995) 6349.
26. Marqusee, S. and Baldwin, R., Proc. Natl. Acad. Sci. USA, 84 (1987) 8898.
27. Sindelar, C., Hendsch, Z. and Tidor, B., Protein Sci., 7 (1998) 1898.
28. Chou, P. and Fasman, G., Biochemistry, 13 (1974) 211.
29. Chandonia, J.M. and Karplus, M., Protein Sci., 4 (1995) 275.
30. Grootenhuis, P.D.J. and Van Helden, S.P., Comput. Approaches Supramol. Chem., (1994) 137.
31. Grootenhuis, P.D.J. and van Galen, P.J.M., Acta Crystallogr., D51 (1995) 560.
32. Holloway, M.K., Wai, J.M., Halgren, T.A., Fitzgerald, P.M.D., Vacca, J.P., Dorsey, B.D., Levin, R.B., Thompson, W.J., Chen, L.J., deSolms, S.J., Gaffin, N., Ghosh, A.K., Giuliani, E.A., Graham, S.L., Guare, J.P., Hungate, R.W., Lyle, T.A., Sanders, W.M., Tucker, T.J., Wiggins, M., Wiscount, C.M., Woltersdorf, O.W., Young, S.D., Darke, P.L. and Zugay, J.A., J. Med. Chem., 38 (1995) 305.
33. Joseph-McCarthy, D., Hogle, J.M. and Karplus, M., Proteins, 29 (1997) 32.
34. Kurinov, I.V. and Harrison, R.W., Nat. Struct. Biol., 1 (1994) 735.
35. Caflisch, A. and Karplus, M., Perspect. Drug Discov. Design, 3 (1995) 51.