Journal of Computer-Aided Molecular Design, 11 (1997) 539–546.
KLUWER/ESCOM

J-CAMD 417

Design of stapled DNA-minor-groove-binding molecules with a mutable atom simulated annealing method

Wynn L. Walker[a], Mary L. Kopka[b], Richard E. Dickerson[b] and David S. Goodsell[c,*]

[a]*Department of Biomathematics and* [b]*Molecular Biology Institute, University of California, Los Angeles, CA 90024, U.S.A.*
[c]*Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, U.S.A.*

Received 30 December 1996
Accepted 23 June 1997

Keywords: Rational drug design; DNA-minor-groove-binding molecules

Summary

We report the design of optimal linker geometries for the synthesis of stapled DNA-minor-groove-binding molecules. Netropsin, distamycin, and lexitropsins bind side-by-side to mixed-sequence DNA and offer an opportunity for the design of sequence-reading molecules. Stapled molecules, with two molecules covalently linked side-by-side, provide entropic gains and restrain the position of one molecule relative to its neighbor. Using a free-atom simulated annealing technique combined with a discrete mutable atom definition, optimal lengths and atomic composition for covalent linkages are determined, and a novel hydrogen bond 'zipper' is proposed to phase two molecules accurately side-by-side.

Introduction

The design of sequence-specific DNA-binding molecules remains an elusive goal. Natural molecules with narrow specificity abound: for instance, netropsin and distamycin bind tightly and specifically to regions with four and five successive AT base pairs, and anthramycin specifically alkylates guanine. A current goal is to use these existing molecules, finely honed by evolution, as leads for the design of custom molecules tailored to bind to a sequence of pharmaceutical interest.

Netropsin, distamycin, and similar molecules bind singly, with high affinity, in the minor groove of DNA at A/T-rich sequences. The structural basis for this recognition is well characterized, being due to the narrow, deep, negatively charged nature of the minor groove in A/T sequences [1,2]. The molecules bind at the center of a narrow, deep minor groove, and are thus unable to distinguish AT base pairs from TA base pairs, only to distinguish AT or TA from CG or GC. When binding singly to a DNA double helix, these molecules recognize an entire base pair, not individual bases on each strand.

Netropsin-type molecules also bind to mixed-sequence DNA in a 2:1 mode [3], with two molecules bound side-by-side in a widened minor groove. The structural basis for this mode has been studied at atomic resolution in two crystallographic systems: distamycin A bound to d(I-C-I-C-I-C-I-C) and two DNA-RNA chimers of this sequence [4,5], and an imidazole-substituted analogue of netropsin (a lexitropsin) bound to d(C-A-T-G-G-C-C-A-T-G) [6]. Because the molecules bind essentially to individual strands, it is now possible to break the AT versus TA and CG versus GC ambiguity and design true sequence-reading molecules.

Covalent linkage of the two molecules provides an entropic bonus to side-by-side inhibitor binding, by prefitting two molecules in the 2:1 mode. Linked molecules have been synthesized and tested. Two approaches have been reported. The first approach creates a hairpin molecule, with two molecules bound end-to-end, which then fold to adopt the side-by-side structure. There remains some ambiguity in binding: these molecules can also bind in an extended conformation in a narrower minor groove. Dervan and co-workers [7] report success in tailoring these linkers to favor one binding mode over the other, using a shorter linker to force the end-to-end binding mode and a longer linker to allow side-by-side binding. This was successful enough to allow the authors to syn-

*To whom correspondence should be addressed.

thesize a molecule that adopts both modes, with one end forming a hairpin and the other end binding as a single chain.

The second approach is to staple two molecules together at some intermediate point. Stapled molecules may bind in two ways: in the side-by-side mode, or with one half unbound, dangling free in the surrounding solvent. Chen and Lown [8] have reported a theoretical study proposing that this latter mode may offer a way to cross-link two separate DNA double helices. Mrksich and Dervan [9] have reported the synthesis of a series of stapled pyridine-2-carboxamidenetropsins, linked by aliphatic chains through to the central rings of each molecule. Binding results yielded the preference of C4>C3>C5>C6 (where C4 denotes an aliphatic linker of four carbon atoms connecting the pyrrole ring nitrogen atoms on each inhibitor), with a range of about a factor of 2 in binding strength from best to worst, and an increase of a factor of 10 over the unstapled molecule. Chen and Lown [10] report a series of stapled distamycins, also linked by aliphatic chains at the central rings. They report a gradual increase in side-by-side binding strength from C4 to C7 to poly(dA–dT)-poly(dA–dT), with the longest linker providing a 1000× increase in binding strength.

Our goal in the work reported here is to optimize the phasing of the stapled molecule to DNA in a manner that maximizes its sequence specificity, by designing linkers that optimize the longitudinal positioning of one molecule relative to its neighbor. In addition, linkers of minimal length are sought, to reduce entropic costs. Previously, we described the design of linked molecules based on the binding of natural antibiotics arranged head-to-tail along the minor groove of DNA, using an exhaustive search method to evaluate linker geometries [11]. Here, we describe the design of molecules based on side-by-side binding of netropsin-like molecules in the wider minor groove of mixed-sequence regions, using an improved linker design method based on simulated annealing and a discrete mutable atom definition.

Methods

Design proceeds in two steps. First, the binding of two individual molecules in the minor groove of DNA is modelled. In previous work [11], models were obtained by overlapping crystallographic coordinates; here, the side-by-side molecules to be linked are present in a single crystallographic experiment. Second, two bonds are chosen for linkage, and the shortest and most rigid linker is designed.

Design of linkers

A number of techniques for the design of linkers have been reported, typically for use in *de novo* molecular design (reviewed in Ref. 12). For instance, Lewis [13,14] described a method for 'growing' molecular chains along a molecular graph embedded between the two target molecules, and many programs, such as LUDI [15] and CAVEAT [16], search fragment libraries for appropriate linker geometries. These techniques allow great breadth in the molecules that are designed, which is ideal when faced with an empty active site to fill. Our current problem, however, has narrower goals: the design of linkers that are of minimal size and flexibility, and that are synthetically accessible. Thus, we have developed a mutable atom simulated annealing technique that effectively and rapidly samples this narrower range of linker geometries. In each simulation, the ideal conformation and atom type composition of a linker of given length are determined. Atom types are determined by allowing each position to mutate during the simulation, similar to techniques reported for use in *de novo* molecular design [17,18].

Simulated annealing begins by placing the linker atoms on a line between the positions to be linked. These atoms then perform a random walk in space throughout the simulation. An energy is evaluated at each new linker conformation, which is used to decide probabilistically whether the step will be accepted or rejected. As the temperature is reduced over the simulation, the linker freezes into an energetically optimal conformation.

To improve the performance of the simulated annealing protocol, we applied an anisotropic random walk method, where the magnitudes of the steps taken at each time point are varied throughout the simulation to reflect the topography of the search space. We applied the method of Vanderbilt and Louie [19], which periodically analyzes the successful excursions of the random walk and uses this information to forecast an appropriate size and direction for future steps. The simulation is broken into cycles of M accepted random walk steps. At the end of each cycle (l), the first (A) and second (B) moments of each variable (X) over the previous M steps are computed:

$$A_i^{(l)} = \frac{\sum_{m=1}^{M} X_i^{(m,l)}}{M}$$

$$B_{ij}^{(l)} = \frac{\sum_{m=1}^{M} \left(X_i^{(m,l)} - A_i^{(l)}\right)\left(X_j^{(m,l)} - A_j^{(l)}\right)}{M}$$

where i and j vary over the x, y, and z coordinates of each linker atom. In our simulations, cycles of $M = 75n$ steps, where n is the number of atoms in the linker, provided enough points for an adequate representation of the local search space. The covariance matrix (C) for the next cycle is defined as a linear function of the previous covariance matrix and the current second moment matrix:

$$C^{(l)} = \frac{\alpha\chi B^{(l)}}{\beta M} + (1-\alpha)C^{(l)}$$

where α is a damping constant controlling the rate at

which information from B is folded into C; β scales B such that, if no steps are rejected, $B^{(1)}/\beta M$ will provide an unbiased estimate of the covariance matrix for individual walk increments; and χ is a growth factor greater than 1, chosen to ensure that the random walk continually explores enough of the search space. If the simulation were a free random walk, it would cover $\chi^{1/2}$ times as much ground in each successive cycle. Values of $\beta = 0.11$ and $\chi = 3$ were taken from the original report, and $\alpha = 0.7$ was determined experimentally in the current system. Displacement vectors for the next cycle are chosen to be elements of the hypercube with mean zero and covariance C, using a Gaussian random number generator.

The anisotropic simulated annealing method provides substantial improvement in performance. Test simulations on an eight-atom linker were used to optimize parameters for both isotropic and anisotropic methods. The isotropic method required an average of 640 000 function evaluations to find the low-energy conformation one half of the time, whereas the anisotropic method required an average of 19 000 function evaluations for similar performance. The added cost of overhead to calculate the covariance matrices resulted in a 5–10-fold overall reduction in computation time in these tests.

Force field

The energy is evaluated with a typical molecular mechanics force field, which includes a penalty for deviation from ideal covalent geometry and terms for the non-bonded interaction energy. Penalties for deviations in bond lengths and bond angles are evaluated using a harmonic potential around the ideal values, with functional forms and values identical to those used in the program AMBER [20]. The energies of non-bonded interactions between linker and the linked molecule and between linker and DNA are calculated using a grid-based approach, with the program AutoGrid [21]. The energy includes a traditional 6-12 potential for dispersion/repulsion interactions, and a directional 10-12 potential for hydrogen bond interactions (for functional forms and parameters, see Ref. 22). Torsional restraints were implemented and tested, but not used in the current study, due to the extended nature of these short linkers.

Mutable atoms

To design linkers incorporating heteroatoms, a hybrid discrete/continuous simulated annealing method was used to determine the optimal combination of atom types in each length of linker. A random combination of atom types is chosen to begin the simulation, and the conformation is optimized by the continuous anisotropic random walk method described above. The final configuration is then fixed, and the atom types of each position are optimized by a discrete isotropic random walk through combinations of atom type. Continuous phases (optimizing conformation) and discrete phases (optimizing atom composition) are then alternately repeated a fixed number of times.

The discrete phase currently allows linker atoms to be carbon, nitrogen, or oxygen. Nitrogen is able to donate a hydrogen bond (the amine hydrogen is implicitly assumed) and oxygen is able to accept a hydrogen bond. At a given step of the simulation, an atom is chosen randomly and mutated. The linker is then tested for chemical feasibility, rejecting atomic combinations that would not be stable or that would not fit into the synthetic scheme. Currently, we enforce a single requirement: that all heteroatoms be separated by at least two carbon atoms in viable linkers. If a linker is judged to be viable, the energy is evaluated and the simulation continues. If unviable, a second mutation is made and tested for viability. Mutation continues until a viable linker is obtained. This technique is necessary to ensure that all possible atomic combinations are accessible to the search algorithm, as there may be combinations that are not accessible by a single point mutation from a given starting combination. By allowing successive mutations of unviable linkers until a viable one is found, we improve the connectivity of the space of possible atomic linkers being searched by the algorithm.

For the systems under study here, with three to six atoms exploring three atom types each, three alternating discrete/continuous cycles were sufficient to find an ideal solution. For the longer linkers, this provides a consider-

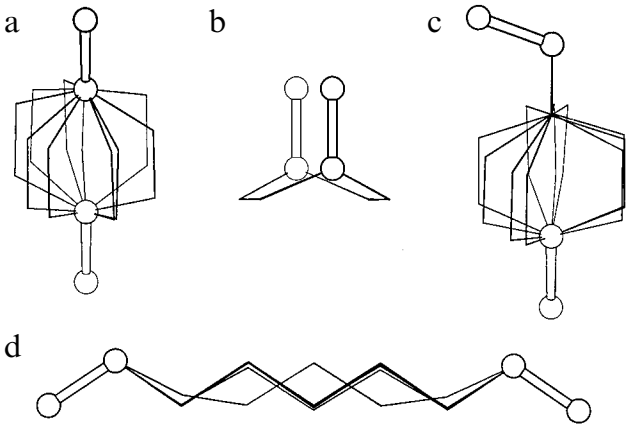

Fig. 1. Results of test systems. The two bonds to be connected are in ball-and-stick representation, and the simulated linkages are shown in bonds. (a) Test case 1: energies ranged from 0.0003 to 0.0008 kcal/mol. (b) Test case 2: energies ranged from 0.0007 to 0.0027 kcal/mol. (c) Test case 3: energies ranged from 0.0033 to 0.0060 kcal/mol. (d) Test case 4: energies ranged from 0.016 to 0.046 kcal/mol for the nine successful solutions; the one outlier had an energy of 5.09 kcal/mol. Simulations began at a temperature $k_BT = 100.0$ kcal/mol, and were reduced by a factor of 0.92 for 100 cycles, each with 3000 individual steps in conformation. Isotropic step sizes were used throughout. Each experiment is composed of 10 of these simulations, yielding 10 separate final linker conformations for each test case.

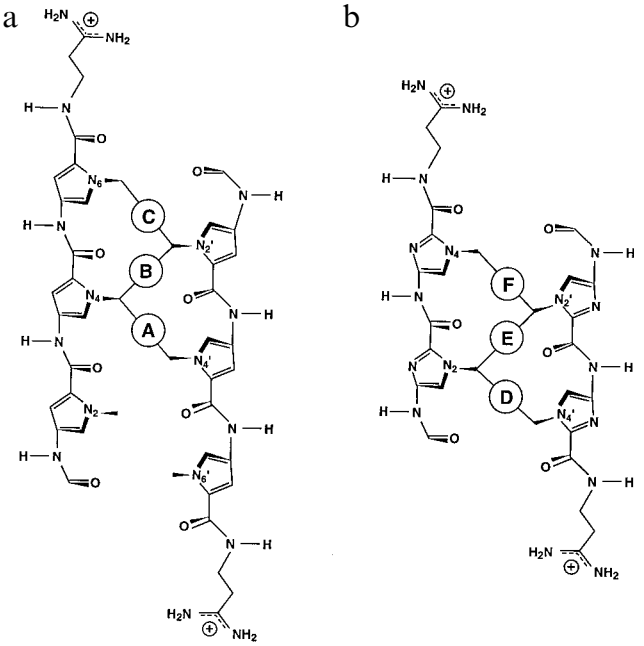

Fig. 2. Stapled linkages. (a) The three crystallographically unique sites for the linkage of distamycin, as bound to d(I-C-I-C-I-C-I-C), labelled A, B, and C. Site A lies on a crystallographic twofold axis. (b) The three linkages tested for diimidazole lexitropsin bound to d(C-A-T-G-G-C-C-A-T-C-G), labelled D, E, and F.

able reduction in computation time: each mutable atom experiment performed here required 3–5 times the computation needed to evaluate a single linker, whereas to test exhaustively all atom combinations, given the rules for viability described above, requires seven individual experiments for three-atom linkers and 37 experiments for seven-atom linkers.

Results and Discussion

Test cases

We considered several test systems, designed to yield ideal solutions, to test the basic free-atom simulated annealing method. The computed solutions are presented in Fig. 1. The first case is composed of two colinear bonds spaced such that two intervening aliphatic carbon atoms, forming three bonds, would form an ideal link. As expected, a family of rotationally symmetric solutions was obtained (Fig. 1a) with very low scores. The second case is composed of two parallel, non-colinear bonds, again placed to be ideally linked by two atoms. Two families of solutions were found, corresponding to the two symmetrically identical positions of the linker (Fig. 1b). The third case, with two coplanar bonds to be linked, requires a longer aliphatic linker of three intervening atoms. As in the first test case, a family of rotationally symmetric solutions was obtained for a section of this linker that is flanked by colinear bonds (Fig. 1c). The final test case required a six-atom linker in an all-trans conformation. Nine out of 10 simulations found the proper conformation, one simulation froze into a 'corkscrew' conformation of higher energy (Fig. 1d).

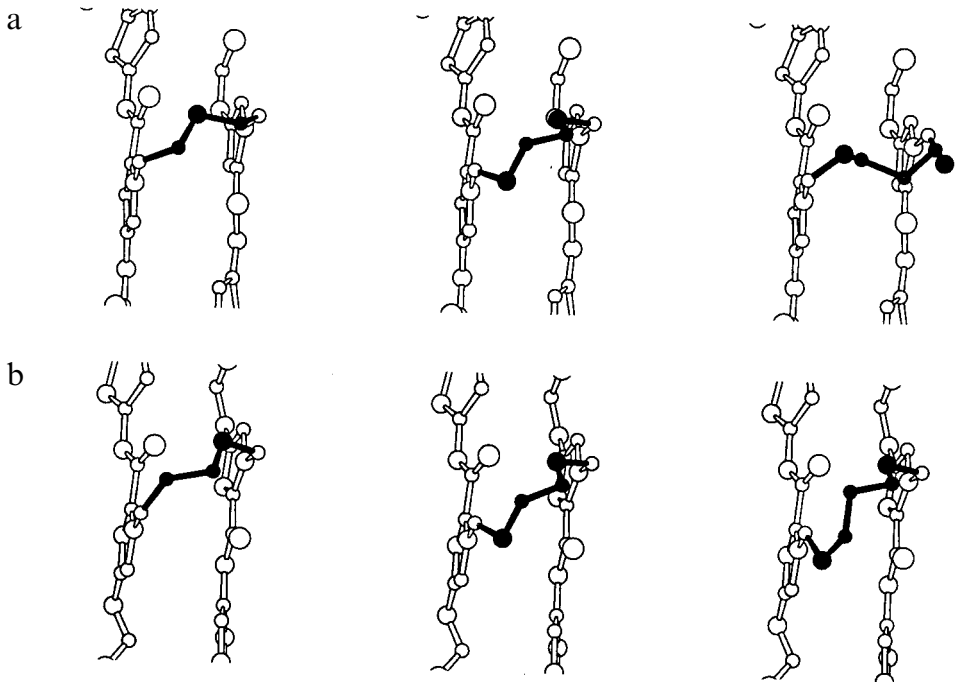

Fig. 3. Selected results of stapled linkages. (a) Site B in the d(I-C-I-C-I-C-I-C):distamycin complex linked by (from left to right) three-, four-, and five-atom linkers. (b) Site D in the d(C-A-T-G-G-C-C-A-T-C-G):lexitropsin complex linked by three-, four-, and five-atom linkers. Oxygen atoms are the largest, nitrogen atoms intermediate, and carbon atoms the smallest. Linker atoms are filled, and crystallographically determined atoms are clear.

TABLE 1
SIMULATION OF STAPLED LINKAGES[a]

Site[b]	Linker	E_{bond}	E_{angle}	E_{int}	E_{tot}
A. d(I-C-I-C-I-C-I-C):distamycin					
A	-C-C-C-C-C-	2.18	5.03	5.83	13.04
	-C-N-C-C-N-	5.08	4.50	–6.56	3.02
	-C-C-C-C-C-C-	0.09	3.30	–0.02	3.37
	-N-C-C-C-C-N-	0.54	1.69	–16.84	–14.61
B	-C-C-C-	0.37	11.95	1.90	14.22
	-C-N-C-	0.35	10.49	–5.99	4.84
	-C-C-C-C-	0.19	3.93	–1.28	2.84
	-N-C-C-N-	0.19	3.36	–16.14	–12.60
	-C-C-C-C-C-	0.19	2.53	–1.88	0.85
	-N-C-C-C-N-	0.09	4.39	–26.01	–21.53
C	-C-C-C-C-	11.43	8.45	8.00	27.88
	-C-N-C-C-	15.76	10.86	2.66	29.26[c]
	-C-C-C-C-C-	0.21	1.22	–2.13	–0.70
	-C-N-C-C-N-	0.64	0.77	–10.24	–8.83
B. d(C-A-T-G-G-C-C-A-T-G):lexitropsin					
D	-C-C-C-C-C-	5.55	6.44	8.17	20.17
	-C-N-C-C-N-	4.73	4.83	2.81	12.36
	-C-C-C-C-C-C-	2.10	3.21	2.35	7.67
	-N-C-C-C-C-N-	1.86	0.87	–13.52	–10.79
E	-C-C-C-	1.33	26.94	8.21	36.49
	-C-C-N-	1.99	20.98	4.73	27.71
	-C-C-C-C-	1.64	7.99	4.26	13.89
	-N-C-C-N-	2.33	9.65	–11.24	0.75
	-C-C-C-C-C-	1.11	3.95	–0.66	4.40
	-N-C-C-C-N-	0.81	1.59	–18.18	–15.78
F	-C-C-C-C-C-	2.11	4.07	2.93	9.11
	-N-C-C-C-N-	1.53	1.11	–9.32	–6.67
	-C-C-C-C-C-C-	1.41	3.62	–3.14	1.88
	-N-C-C-C-C-N-	1.42	0.33	–16.14	–14.36

Linker	Total energy (kcal/mol)					
	Site A	Site C	Site D	Site F	Site B	Site E
C. Summary of stapled linkage results						
-C-C-C-C-C-	13.04	–0.70	20.17	9.11		
-C-N-C-C-N-	3.02	–8.83	12.36			
-N-C-C-C-N-				–6.67		
-C-C-C-C-C-C-	3.37		7.67	1.88		
-N-C-C-C-C-N-	–14.61		–10.79	–14.36		
-C-C-C-					14.22	36.49
-C-N-C-					4.84	
-C-C-N-						27.71
-C-C-C-C-					2.84	13.89
-N-C-C-N-					–12.60	0.75
-C-C-C-C-C-					0.85	4.40
-N-C-C-C-N-					–21.53	–15.78

[a] For aliphatic linkers, simulations began at a temperature $k_BT = 75.0$ kcal/mol, and were reduced by a factor of 0.75 for 30 cycles, each with 45n individual steps in conformation, where n is the number of atoms in the linker that are free to move. Anisotropic step sizes were updated after each cycle. A typical experiment is composed of 50 of these simulations, yielding 50 separate final linker conformations. The conformation of lowest total energy is given here. For the mutable atom experiments, three rounds of alternating continuous conformation search and discrete atom type search were performed. Each round contained 75n steps in 50 cycles of temperature reduction. The best solution is included for each site and linker length. All combinations of five bond linkers at sites B and E were tested individually to confirm adequate coverage by the mutable atom algorithm (data not shown).

[b] Site: site of linkage as shown in Fig. 4; E_{bond}: bond deformation energy; E_{angle}: angle deformation energy; E_{int}: interaction energy from linker to DNA and polyamide; E_{tot}: total energy. All values are in kcal/mol.

[c] The aliphatic linker was found to have the lowest energy by the mutable atoms technique. This is the linker of lowest energy that includes at least one heteroatom.

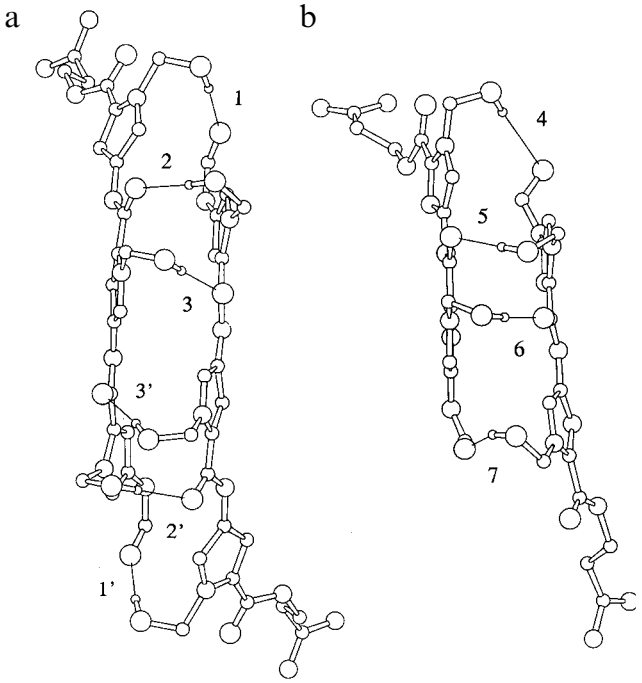

Fig. 4. Simulation of hydroxyl zippers. (a) Hydroxyls added to distamycin in the d(I-C-I-C-I-C-I-C) complex. Hydroxyls 1, 2, and 3 are identical by symmetry with hydroxyls 1', 2', and 3', respectively. (b) Hydroxyls added to lexitropsin in the d(C-A-T-G-G-C-C-A-T-C-G) complex. Oxygen atoms are the largest, followed by nitrogen, then oxygen, and hydrogen is the smallest. Hydrogen bonds are drawn as thin lines.

Stapling two molecules side-by-side

Stapled molecules were designed based on the positions of two separate molecules bound side-by-side to a single DNA double helix. Two 2:1 complexes solved to atomic level by X-ray crystallography were available at the time of this study: distamycin A bound to the oligonucleotide d(I-C-I-C-I-C-I-C) [4] and a diimidazole lexitropsin bound to the oligonucleotide d(C-A-T-G-G-C-C-A-T-G) [6]. These two structures are remarkably similar: the relative placement of one molecule next to its neighbor and the placement of the two molecules relative to the DNA base planes is structurally close. Given this high structural similarity, we assume that the structure of the free molecules will provide an adequate model for the binding of an ideally stapled molecule. We focussed on linkages through the pyrrole (or imidazole) *N*-methyl group, since they have been shown to be synthetically tractable [10, 23].

Figure 2 shows the possible sites of linkage for these two crystallographic structures examined in the current study. In both cases, the molecules bind antiparallel, placing the charged tails in maximally distant positions. The molecules are staggered such that the formyl and amide groups are stacked on the pyrrole and imidazole rings. In the distamycin structure, the oligonucleotide is centered on a crystallographic twofold axis, so only one half of the structure is unique. Three linkage positions are structurally unique: A – N4 to N4', crossing the crystallographic twofold axis; B – N4 to N2'; and C – N6 to N2'. (In what follows, only the letter designations (A, B, etc.) will be used for clarity.) In the diimidazole lexitropsin structure, the two molecules are crystallographically unique, so three linkage positions may be tested: D – N2 to N4'; E – N2 to N2'; and F – N4 to N2'.

Note that the linkers fall into two structural classes: A, C, D, and F are structurally similar, as are B and E. The results of linkage experiments are included in Table 1, and pictured in Fig. 3. The total interaction energies included in Table 1 show an appreciable range of values for a given linker in each of the two structural classes, where ideally we would expect the values to be similar. This is due to the rigid model employed in the simulation: the crystallographic structures are used as reported, without the ability to make small compensatory motions to idealize the bonding geometry between linker and inhibitor during the simulation. The bond length and bond angle potentials are sensitive to deviations on the order of 0.1–0.2 Å, so small variations in the crystallographic structure can have large effects in energy. We are currently exploring methods of softening potentials to model small compensatory motions, without making the model overly permissive for unrealistic geometries.

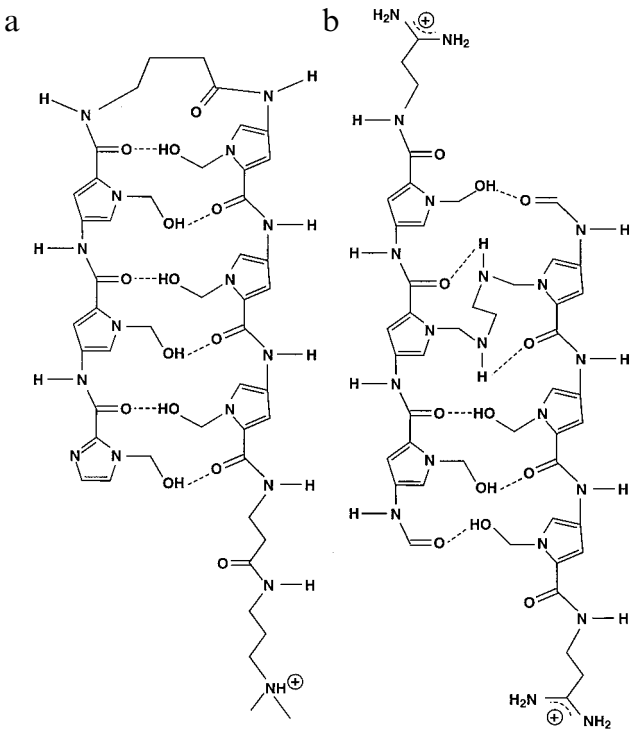

Fig. 5. Suggestions for synthesis. (a) The hairpin molecule ImPyPy-γ-PyPyPy-β-Dp, which binds specifically to the sequence 5'-TGTTA-3' [24], is modified by the addition of hydroxyl groups to reduce the possibility of extended binding modes. (b) Distamycin is stapled and zippered to form a stable, side-by-side structure.

TABLE 2
SIMULATION OF HYDROXYL LINKAGES[a]

Site	E_{bond}	E_{angle}	E_{hb}	E_{tot} [b]
d(I-C-I-C-I-C-I-C):distamycin				
1	0.00	0.06	–5.00	–4.94
2	0.00	0.04	–5.00	–4.95
3	0.00	0.14	–4.98	–4.83
d(C-A-T-G-G-C-C-A-T-C-G):lexitropsin				
4	0.00	0.02	–0.33	–0.31
5	0.00	0.02	–5.00	–4.99
6	0.00	0.20	–4.99	–4.79
7	0.01	1.47	–4.97	–3.48

[a] Simulations began at a temperature $k_BT = 75.0$ kcal/mol, and were reduced by a factor of 0.75 for 50 cycles, each containing 150 individual steps in conformation. Anisotropic step sizes were updated after each cycle. Each experiment is composed of 10 of these simulations, yielding 10 similar final linker conformations. The conformation of lowest total energy is given here.
[b] See footnote b of Table 1.

Using purely aliphatic linkers, sites B and E may be linked by three atoms with appreciable distortion, or by four atoms with no distortion. Three-atom linkers give unfavorable energies in both crystallographic structures, and for optimal fit require that the rings tilt towards one another by about 0.5 Å – a difficult adjustment because the molecules are in tight steric contact along their entire length. Four-atom linkers bridge the two rings with better geometry, with smaller angular distortions of the linker. The second class of linkage requires a longer chain of five or six aliphatic carbon atoms: site C shows good geometry with a five-atom linker, but sites A, D, and F require the longer six-atom linker.

The aliphatic linkers show significant steric clash with the neighboring carbonyl oxygens on the two molecules, resulting in disappointing energies. This may be alleviated by replacing key carbon atoms in the linker with nitrogen atoms, reducing the steric overlap and allowing the formation of hydrogen bonds. Optimal sites for these nitrogen atoms were obtained with the mutable atom technique. In all cases, the technique placed nitrogen atoms at the ends of the linkers, in close contact with neighboring carbonyl oxygens. The energetic penalties for bond and angle deformations changed little, but the interaction energies improved substantially, replacing an unfavorable carbon–oxygen contact with a favorable amine to oxygen hydrogen bond. The best suggestion for the shorter linkages at sites B and E is a four-atom linkage with the structure -N-C-C-N- and the longer distance at sites A, C, D, and F may be bridged with a small repositioning of rings by a five-atom linker -N-C-C-C-N-, and with no distortion by the longer linker -N-C-C-C-C-N-.

Linking molecules with a hydrogen bond zipper

A second and quite different method of holding two inhibitors together side-by-side is provided by what we have termed 'zippers': a row of hydrogen bonds between carbonyl oxygens on one molecule and hydroxymethyl groups introduced into the rings of the other molecule, as shown in Figs. 4 and 5. In both the crystal structures, the molecules bind side-by-side such that each aromatic ring is stacked against the peptide plane of its neighbor, not as one might expect, ring upon ring. Thus, the ideal stapled linkage would connect the ring methyl position with its closest peptide carbonyl oxygen, a span of about three bonds. A covalent linkage between these sites may not be synthetically tractable, but the hydroxyl hydrogen bond proposed here might offer an attractive alternative.

For the calculation, the hydrogen bond from hydroxyl hydrogen to carbonyl oxygen was treated as a covalent bond, with appropriately modified restraints. The hydrogen bond energy was evaluated with a typical 10-12 formulation identical to that used for the non-bonded interaction. The angle across the hydrogen atom, subtended by the hydroxyl H-O bond and the hydrogen bond, was given an equilibrium value of 180° and a weak harmonic potential with coefficient 5 kcal/mol Å. The angle across the carbonyl oxygen, subtended by the C=O bond and the hydrogen bond, was not restrained, with zero energy at all angles. The results are shown in Table 2 and Fig. 4.

Hydroxyls in these positions are perfectly placed to form hydrogen bonds with the neighboring carbonyl oxygens. All adopt nearly ideal bond and angle geometry (due to the looser restraints on the hydrogen bond geometry) and all but one form nearly ideal hydrogen bonds. The one outlier is number 4, at the end of the lexitropsin complex, but a facile rotation of the terminal formyl group will bring this bond into ideal geometry, similar to the formyl group of the neighboring inhibitor.

Two strategies for linking molecules side-by-side

This paper presents two improvements on reported stapled minor-groove-binding molecules. The first is the suggestion for optimal stapled linkages. Our best suggestion is the use of the four-atom linkage -N-C-C-N- between pyrrole or imidazole ring methyl positions. This short linkage will favor binding in a single mode, whereas a longer linkage of five or six atoms might allow the slippage of one half of the molecule by one position. The amine groups add additional strength to the linkage by forming hydrogen bonds with the neighboring carbonyl oxygens.

The second suggestion is for the addition of hydroxyls to the remaining pyrrole or imidazole methyls. These hydroxyls are perfectly placed to form hydrogen bonds with carbonyl oxygens on the opposite molecule, forming a hydrogen bond zipper. For molecules synthesized as a hairpin, these hydroxyls would favor binding in the hairpin mode over binding in extended modes (Fig. 5a). For stapled molecules, we might expect added stability due to the presence of these interactions (Fig. 5b).

Acknowledgements

The authors wish to thank Elliot Landaw for helpful comments. This work was supported in part by GM-31299, and NIH Training Grant GM08185 and a fellowship from the Program in Mathematics and Molecular Biology at the University of California, Berkeley, supported by NSF Grant DMS-9406348. This is manuscript 10251-MB of the Scripps Research Institute.

References

1 Kopka, M.L., Yoon, C., Goodsell, D.S., Pjura, P. and Dickerson, R.E., Proc. Natl. Acad. Sci. USA, 82 (1985) 1376.
2 Kopka, M.L., Yoon, C., Goodsell, D.S., Pjura, P. and Dickerson, R.E., J. Mol. Biol., 183 (1985) 553.
3 Pelton, J.G. and Wemmer, D.E., Proc. Natl. Acad. Sci. USA, 86 (1989) 5723.
4 Chen, X., Ramakrishnan, B., Rao, S.T. and Sundaralingam, M., Nat. Struct. Biol., 1 (1994) 169.
5 Chen, X., Ramakrishnan, B., Rao, S.T. and Sundaralingam, M., Nat. Struct. Biol., 2 (1995) 733.
6 Kopka, M.L., Goodsell, D.S., Han, G.W., Chiu, T.K., Lown, J.W. and Dickerson, R.E., Structure, 5 (1997) 1033.
7 Trauger, J.W., Baird, E.E. and Dervan, P.B., Chem. Biol., 3 (1996) 369.
8 Chen, Y.-H. and Lown, J.W., Biophys. J., 68 (1995) 2041.
9 Mrksich, M. and Dervan, P.B., J. Am. Chem. Soc., 115 (1993) 9892.
10 Chen, Y.-H. and Lown, J.W., J. Am. Chem. Soc., 116 (1994) 6995.
11 Walker, W.L., Kopka, M.L., Filipowsky, M.E., Dickerson, R.E. and Goodsell, D.S., Biopolymers, 35 (1995) 543.
12 Rosenfeld, R., Vajda, S. and DeLisi, C., Annu. Rev. Biophys. Biomol. Struct., 24 (1995) 677.
13 Lewis, R.A., Roe, D.C., Huang, C., Ferrin, T.E., Langridge, R. and Kuntz, I.D., J. Mol. Graph., 10 (1992) 66.
14 Lewis, R.A., J. Comput.-Aided Mol. Design, 4 (1990) 205.
15 Böhm, H.-J., J. Comput.-Aided Mol. Design, 6 (1992) 61.
16 Lauri, G. and Bartlett, P.A., J. Comput.-Aided Mol. Design, 8 (1994) 51.
17 Miranker, A. and Karplus, M., Proteins, 23 (1995) 472.
18 Gehlhaar, D.K., Moerder, K.E., Zichi, D., Sherman, C.J., Ogden, R.C. and Freer, S.T., J. Med. Chem., 38 (1995) 466.
19 Vanderbilt, D. and Louie, S.G., J. Comput. Phys., 56 (1984) 259.
20 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
21 Goodsell, D.S. and Olson, A.J., Proteins, 8 (1990) 195.
22 Morris, G.M., Goodsell, D.S., Huey, R. and Olson, A.J., J. Comput.-Aided Mol. Design, 10 (1996) 293.
23 Mrksich, M. and Dervan, P.B., J. Am. Chem. Soc., 116 (1994) 3663.
24 Mrksich, M., Parks, M.E. and Dervan, P.B., J. Am. Chem. Soc., 116 (1994) 7983.