Biophysical Journal, Volume 98

Supporting Material

The Effect of End Constraints on Protein Loop Kinematics

Steven Hayward and Akio Kitao

METHODS

Constraint surface exploration

Being a linear approximation, Equation (6) is only valid for small changes in torsion angles. What if one wants to determine structures far from the starting structure that still satisfy the fixed end groups constraint?
One can use the following simple algorithm:

$n = 0$
$\mathbf{Y}(\boldsymbol{\tau}(n))\boldsymbol{\delta\tau}^0(n) = \mathbf{0}$
$n \rightarrow n+1$
$\boldsymbol{\tau}(n) \rightarrow \boldsymbol{\tau}(n) + \boldsymbol{\delta\tau}^0(n)\boldsymbol{\lambda}(n)\Delta s$

(Algorithm A1)

where $\boldsymbol{\delta\tau}^0(\mathrm{n})$ is the matrix of null space vectors at the nth iteration step, $\boldsymbol{\lambda}(\mathrm{n})$ is a normalised column vector (the ith row corresponding to the *i*th null space vector) that determines the local direction of movement on the constraint surface, and Δs is the step size. In principle, systematic variation of $\boldsymbol{\lambda}(\mathrm{n})$ should allow for exploration of the whole constraint surface. This will be demonstrated for a pentapeptide in Results.

Steepest descent and gradient ascent on the constraint surface

Steepest descent or gradient ascent can be performed to minimize or maximize any function of the torsions $\Phi(\boldsymbol{\tau})$ keeping to the constraint surface:

$n = 0$

$\mathbf{Y}(\boldsymbol{\tau}(n))\boldsymbol{\delta\tau}^0(n) = \mathbf{0}$

$n \to n+1$ (Algorithm A2)

$\boldsymbol{\tau}(n) \to \boldsymbol{\tau}(n) + \sum_j \Delta\Phi\left(\boldsymbol{\tau}(n) + \boldsymbol{\delta\tau}_j^0(n)\Delta s\right)\boldsymbol{\delta\tau}_j^0(n)$

where $\Delta\Phi\left(\boldsymbol{\tau}(n) + \boldsymbol{\delta\tau}_j^0(n)\Delta s\right) = \Phi\left(\boldsymbol{\tau}(n) + \boldsymbol{\delta\tau}_j^0(n)\Delta s\right) - \Phi(\boldsymbol{\tau}(n))$.

Torsion angle targeting

There are occasions when it is desirable to be able to change selected torsion angles in a segment with constrained end groups to specified values. The method, which is also a steepest descent algorithm, has the current point on the constraint surface move iteratively in the direction of the projection onto the null space (which defines the hyperplane tangential to the curved surface of constraint) of the vector joining the target point and the current point in the subspace defined by the set of torsion angles that are targeted. If $\boldsymbol{\tau}_{targ}$ is a column vector of same length as $\boldsymbol{\tau}(\mathrm{n})$ with components that are the target values or otherwise zeros, then:

$\mathrm{n} = 0$

$\mathbf{Y}(\boldsymbol{\tau}(\mathrm{n}))\boldsymbol{\delta\tau}^0(\mathrm{n}) = \mathbf{0}$

$\boldsymbol{\Delta T}(\mathrm{n}) = \boldsymbol{\tau}_{targ} - \boldsymbol{\tau}_{\mathrm{zeros}}(\mathrm{n})$ (Algorithm A3)

$\mathrm{n} \to \mathrm{n}+1$

$\boldsymbol{\tau}(\mathrm{n}) \to \boldsymbol{\tau}(\mathrm{n}) + \boldsymbol{\delta\tau}^0(\mathrm{n})\boldsymbol{\delta\tau}^0(\mathrm{n})^{\mathrm{t}}\boldsymbol{\Delta T}(\mathrm{n})\Delta\mathrm{s}$

where $\boldsymbol{\tau}_{zeros}(n)$ is the same as $\boldsymbol{\tau}(n)$ except the same set of components are set to zero as for $\boldsymbol{\tau}_{targ}$, i.e. the components corresponding to the torsions that are not targeted. Targeting can be stopped when targets have been achieved, i.e. when $\boldsymbol{\Delta T}(n) = \mathbf{0}$. However, in some cases it may not be possible to achieve target values which will occur when $\boldsymbol{\delta\tau}^0(\mathrm{n})^{\mathrm{t}}\boldsymbol{\Delta T}(\mathrm{n}) = \mathbf{0}$, i.e. when the vector joining the target and the current point in

the subspace is orthogonal to the tangential hyperplane. Thus during a targeting run both $\|\mathbf{\Delta T}(n)\|$ and $\left\|\mathbf{\delta\tau}^0(\mathrm{n})^{\mathrm{t}}\mathbf{\Delta T}(\mathrm{n})\right\|$ are monitored. Targeting was terminated when $\|\mathbf{\Delta T}(\mathrm{n})\|$ fell below 0.1 or $\left\|\mathbf{\delta\tau}^0(\mathrm{n})^{\mathrm{t}}\mathbf{\Delta T}(\mathrm{n})\right\|$ fell below 0.001. As constraining particular torsions is simply a matter of removing them from the calculation by eliminating the corresponding columns of $\mathbf{Y}(\boldsymbol{\tau})$, it is a simple matter to simultaneously target selected torsions whilst constraining others.

RESULTS

Exploring conformations on the constraint surface for short polypeptides

Segments shorter than a pentapeptide

Searching for lower rank structures in tri-and tetra- peptides

By variation of $\boldsymbol{\tau}$ it is possible to find structures where for $\mathrm{N_{res}} = 4$, $\mathrm{rank}(\mathbf{Y}(\boldsymbol{\tau})) = 5$ providing a single null space vector. This was achieved by using a simplex search method implemented in the MATLAB function `fminsearch` to search for $\boldsymbol{\tau}$ such that in a singular value decomposition of $\mathbf{Y}(\boldsymbol{\tau})$, one of the singular values becomes equal to zero. The search was started from a large number of randomly generated structures. It appears there are a vast number of such structures covering the whole Ramachandran plot when plotting ϕ, ψ angles for individual residues. Movement along the null space vector for each of these structures shows a wide variety of small movements, but these movements are in directions that lead immediately to full rank structures ($\mathrm{rank}(\mathbf{Y}(\boldsymbol{\tau})) = 6$, for which the only solution is $\mathbf{\delta\tau}^0 = \mathbf{0}$, i.e. they become dynamically trapped straight away).

It is even possible to find structures for $\mathrm{N_{res}} = 3$, where $\mathrm{rank}(\mathbf{Y}(\boldsymbol{\tau})) = 3$, again giving a single null space vector. Starting from various extended conformations the resulting structures all had a central residue with Ramachandran coordinates (180,180) which means the backbone atoms between C_1^α and C_3^α lie in the same plane. The allowed small movement is a crankshaft like rotation of the peptide units (26-29) either side of C_2^α which rotate in opposite directions. These effect a small translation of C_2^α. Only a small rotation is allowed as any movement results in full

rank (rank$(\mathbf{Y}(\boldsymbol{\tau}))=4$ for which the only solution is $\boldsymbol{\delta\tau}^0=\mathbf{0}$ i.e. they become dynamically trapped straight away). Repeating the procedure from various α-helix conformations, all resulting structures have a central residue with Ramachandran coordinates (0,0).

Pentapeptides

With one angle constrained

Using the iterative procedure described in Algorithm(A1), structures far from the starting structure can be reached. In the case of a pentapeptide where the ϕ_2 is constrained, e.g. the second residue mimics a proline, there are 7 free torsions and solving Equation (6) gives a single null space vector. Algorithm(A1) with $\boldsymbol{\lambda}(\mathrm{n})$ =1 and $\Delta s = 0.1$ was applied to α-helix, extended starting structures and various loop structures (see legend to Figure 1S for details). This resulted in cycling behaviour. Figure 1S shows the ϕ,ψ angles for residues 2, 3 and 4 in a Ramachandran plot. As one can see residues 3 and 4 trace closed cycles, whilst residue 2 moves up and down the vertical straight line. This means that the state point moves on a 1-dimensional closed loop in torsion angle space and that torsion angle changes along the segment are all perfectly correlated. It is noticeable that the extended structure is able to change its conformation much more than the α-helix. This is generally true when each of the ϕ angles of residues 2-5 is constrained. Residues 2 and 3 in the extended structure case remain in the allowed region for all amino acids residue types in the Ramachandran plot. For residue 4 regions visited are low energy only for a glycine residue. Movies showing the movement corresponding to one cycle for both the α-helix and extended starting structures are available at Supporting Material. Starting from a neighbouring structure gives a neighbouring trajectory in most cases (see "Cusps on Surface" in main paper).

Systematic search method

The systematic methods involved iterating along the second null space vector (so with $\boldsymbol{\lambda}(n)=(0 \quad 1)^t$), as determined directly by the MATLAB function `null`. After a large number of steps, a single iteration step along the first null space vector was made (so with $\boldsymbol{\lambda}(n)=(1 \quad 0)^t$). This process was repeated a large number of times. In order to be sure that the whole surface was covered a second method of exploration was performed. In this method iteration steps were performed along the null space vector determined by evaluating $\boldsymbol{\lambda}(\mathrm{n})$ at each step such that the τ_k angle did not change. This was achieved by setting $\boldsymbol{\lambda}(n)$ equal to the normalised vector of $\left(1 \quad -\frac{\delta\tau^0_{k1}(n)}{\delta\tau^0_{k2}(n)}\right)^t$, where $\delta\tau^0_{kj}(n)$ is the change in

τ_k in the j-th null space vector at step *n*. As the component of $\left(\boldsymbol{\delta\tau}_1^0(n) \quad \boldsymbol{\delta\tau}_2^0(n)\right)\begin{pmatrix} 1 \\ -\dfrac{\delta\tau_{k1}^0(n)}{\delta\tau_{k2}^0(n)} \end{pmatrix}^t$ corresponding to τ_k is 0, this means τ_k does not change. After a large number of steps a single iteration step was performed along the null space vector determined by evaluating $\boldsymbol{\lambda}(\mathrm{n})$ such that the τ_k angle would increase or decrease. This was achieved by setting $\boldsymbol{\lambda}(n)$ equal to the normalised vector of $\pm\begin{pmatrix} 1 \\ \dfrac{1-\delta\tau_{k1}^0(n)}{\delta\tau_{k2}^0(n)} \end{pmatrix}^t$. As the component of $\left(\boldsymbol{\delta\tau}_1^0(n) \quad \boldsymbol{\delta\tau}_2^0(n)\right)\begin{pmatrix} 1 \\ \dfrac{1-\delta\tau_{k1}^0(n)}{\delta\tau_{k2}^0(n)} \end{pmatrix}^t$ corresponding to τ_k is 1 this means τ_k is forced to change. This process was repeated a large number of times. Thus the whole surface was explored by performing cycles such as those shown in Figure S1 over all allowed τ_k angles. This produces different traces to those produced by the first method.

LADH loop definition variation

The loop 290-301 was varied ±1 residue at either end (but not simultaneously; 291-301 was not included as ϕ_{291} is one of the targeted torsion angles). The distance between the C^{γ}'s of Pro296 and Leu57 which starts at 4.3 Å is given after targeting in Table SIII below. Our result which states that without the constraints on $\psi_{294}, \phi_{295}, \psi_{295}, \phi_{296}$ the block to domain closure is not removed, is unaltered by this slight variation in brace residue definition. In all cases targets were achieved except for segment 290-300 where targets were nearly achieved ($\left\|\boldsymbol{\Delta\mathbf{T}}(\mathrm{n})\right\| = 3.91°$).

Table SI

Parameter values used

Parameter	Value
C-N distance	1.33 Å
N-C^{α} distance	1.47 Å
C^{α}-C distance	1.53 Å
C-O distance	1.23 Å
C–N–C^{α} bond angle	122°
N–C^{α}–C bond angle	109.5°
C^{α} –C–N bond angle	115°
C^{α} –C–O bond angle	121°
Peptide-bond torsion	180°

Table SII

ϕ, ψ angles of decapeptide structures derived from α-helix and extended structures for which the correlation coefficients between ψ_1 and ϕ_{10} is +1 or -1.

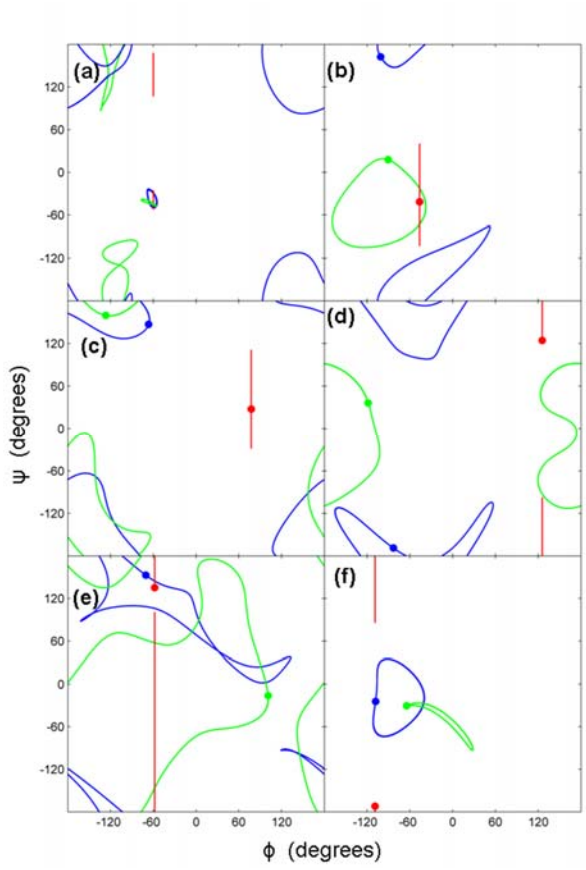
	Res N^{O}:	1	2	3	4	5	6	7	8	9	10	corr($\delta\psi_1$ $\delta\phi_{10}$)	$\delta\phi_{10} / \delta\psi_1$
α	ϕ	—	-31.84	-52.67	-79.18	-63.87	-56.46	-80.65	-57.22	-63.58	-56.09	1.0	0.8
	ψ	-33.31	-45.77	-38.63	-30.62	-48.57	-33.71	-34.43	-48.34	-87.87	—		
α	ϕ	—	-10.50	-64.74	-62.81	-96.23	-66.03	-61.69	-95.76	-67.34	-76.25	-1.0	-1.6
	ψ	-44.41	-35.12	-58.39	-31.51	-34.19	-58.32	-32.37	-33.46	-10.85	—		
β	ϕ	—	-90.49	-128.37	-104.50	-114.75	-110.55	-140.97	-148.97	-163.26	-130.61	1.0	1.0
	ψ	101.3	154.19	134.45	134.14	123.82	148.09	154.15	166.65	111.45	—		
β	ϕ	—	-99.916	166.65	171.89	-75.27	153.25	-143.21	-131.36	9.30	-64.88	-1.0	-3.1
	ψ	94.33	-171.78	-177.65	91.78	-175.58	127.47	-168.20	37.47	-118.07	—		

Table SIII

Variation in loop definition for LADH

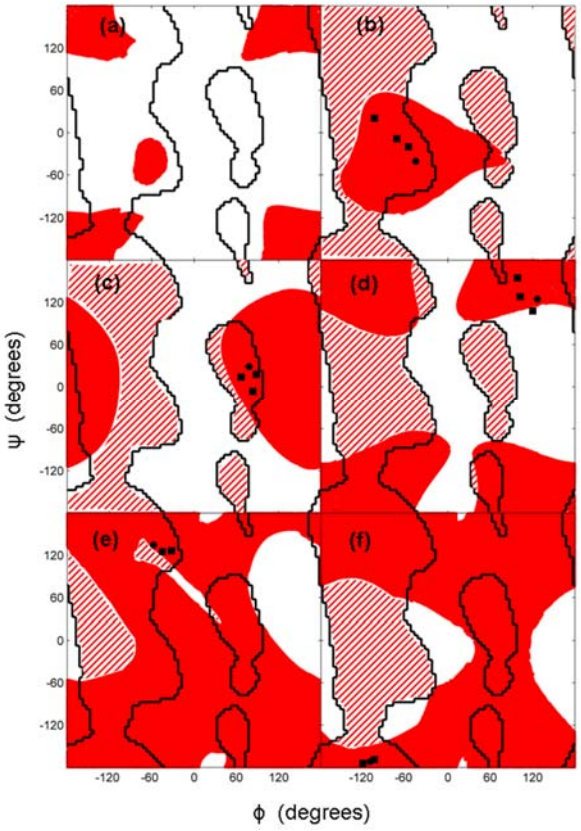
Segment	**$\psi_{294}, \phi_{295}, \psi_{295}, \phi_{296}$ constrained, Pro296 C$^{\gamma}$ atom-Leu57 C$^{\gamma}$ atom distance after targeting**	**$\psi_{294}, \phi_{295}, \psi_{295}, \phi_{296}$ unconstrained, Pro296 C$^{\gamma}$ atom-Leu57 C$^{\gamma}$ atom distance after targeting**
290-301	11.8 Å	6.9 Å
290-302	11.0 Å	6.8 Å
290-300	14.7 Å	7.3 Å
289-301	12.6 Å	8.6 Å

Figure S1



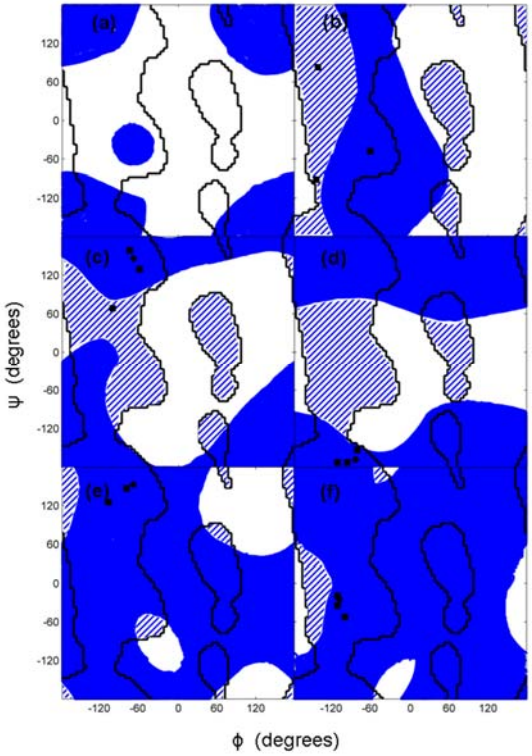
Starting from a pentapeptide structures with the ϕ of the residue 2 constrained the torsion angles were changed according to Algorithm A1. The null space comprised a single vector. The ϕ,ψ traces for residues 2 (red), residue 3 (green) and residue 4 (blue) are shown in a Ramachandran plot. The dots indicate the starting values (omitted in (a) for clarity). **(a)** Two sets of traces are shown: one set starting from an extended structure with the following ϕ,ψ angles $(-123,136)_1$, $(-60,136)_2$, $(-123,136)_3$, $(-123,136)_4$, $(-123,136)_5$; the other starting from an α-helix with the following ϕ,ψ angles $(-57,-47)_1$,$(-60,-47)_2$, $(-57,-47)_3$,$(-57,-47)_4$, $(-57,-47)_5$. Movies of the structures performing one cycle are available in Supporting Material. **(b)** Starting from loop α-α 1.1.5 (using Oliva et al. (25) classification code, PDB code: 1ECA, segment 49-53A) **(c)** α-β 1.2.5 (PDB code: 5P21, segment 137-141A) **(d)** β-α 3.1.1 (PDB code: 2TMD, segment 395-399A) **(e)** β-β link 2.1.1 (PDB code: 1EFT, segment 248-252A) **(f)** β-β hairpin 2.3.2 (PDB code: 1HOE, segment 16-20A).

Figure S2



See legend to Figure 2 in main paper.

Figure S3



See legend to Figure 2 in main paper.