

Thème 5 : Cinétique et repliement des protéines

19 juin 2015

Prédiction de la vitesse de repliement des protéines à partir du contact géométrique et de la séquence d'acides aminés

Zheng Ouyang, Jie Liang

Résumé

La vitesse de repliement $\ln k_f$ des protéines varie selon 8 ordres de magnitude : il existe une corrélation entre $\ln k_f$ la topologie de la protéine repliée. Si la structure native est connue, alors $\ln k_f$ peut être prédite grâce à une corrélation avec la localité des contacts dans la protéine. Une des mesures qui permet cela, donnée par les auteurs est le nombre de contacts géométriques N_α . Le nombre de contacts géométriques est non seulement utile pour prédire le $\ln k_f$ des repliements rapides à deux états mais aussi pour les repliements plus complexes à états multiples. La vitesse de repliement des protéines peut aussi être prédite par la séquence d'acides aminés sans connaissance de la topologie native.

Introduction

Les protéines à repliement rapide ont tendance à avoir plus de structures locale en hélice α alors que les protéines lentes à se replier ont plus de feuillet β . La première valeur inventée pour prédire $\ln k_f$ est l'**ordre de contact relatif moyen** (*RCO*, *average relative contact order*) qui mesure la fraction relative de contacts locaux contre les contacts non-locaux non-covalents. Plusieurs variations sont arrivées par la suite : long-range ordre (*LRO*), total contact distance (*TCD*), chain topology parameter (*CTP*), la **distance effective des protéine** L_{eff} ou L . Un dernier critère est l'**ordre de contact absolu** (*ACO*, *absolute contact order*) qui fonctionne bien aussi pour les repliements plus complexes. Tous les résultats obtenus avec ces valeurs demandent une connaissance minimale de la structure de la protéine, comme la classe structurale ou les prédictions prioritaires de la structure secondaire. Le but des auteurs est de développer une **méthode générale qui peut prédire la vitesse de repliement des protéines de différentes classes uniquement grâce à la séquence d'acides aminés**. Les coefficients de corrélation obtenus pour N_α (cf. table 3) suggèrent que toutes les protéines de toutes les classes, qu'elles soient à repliement à deux états ou plus, doivent se replier selon un unique mécanisme pour lequel les déterminants les plus importants de la vitesse de repliement sont l'emboîtement spatial (spatial packing) et les interactions en "fermeture éclair" (zippering interactions).

Matériels et méthodes

Modèle et données

Set de données L'article se base sur les vitesses de repliement de 80 protéines, 45 à deux états, 35 à états multiples. Parmi celles-ci, il y a des protéines α , β et $\alpha\beta$. Pour les protéines à états multiples, ils ont utilisé la vitesse la plus faible car les vitesses plus rapides sont dues à des pièges cinétiques : la vitesse la plus faible correspond à l'apparition de la protéine native et est donc comparable avec les vitesses des protéines à deux états. Les tables 1 et 2 montrent les vitesses de repliements obtenues expérimentalement. On voit qu'elles varient selon 8 ordres de grandeur.

Table 1. *The set of 45 two-state proteins*

PDB ID	Length	N_a	$\ln k_f$	Structural class
1L2Y	20	2	12.40	α
1VII	36	7	11.51	α
2PDD	43	10	9.69	α
1PRB	53	10	12.90	α
1BAS	53	4	5.91	α
1IDY	54	2	8.73	α
1FEX	59	11	8.19	α
1BDD	60	8	11.69	α
2A3D	73	7	12.7	α
1IMQ	86	39	7.28	α
1LMB	87	29	8.50	α
1ENH	54	4	10.53	α
1PGB	16	9	12.0	β
1PIN	32	19	9.37	β
1EOM	37	22	8.85	β
1EOL	37	19	10.37	β
1K9Q	40	22	8.37	β
1FMK	57	43	4.05	β
1SHG	57	45	2.10	β
1NYF	58	43	4.54	β
1PKS	76	51	-1.06	β
1C8C	64	39	6.95	β
1PSE	69	43	1.17	β
1C9O	66	47	7.20	β
1G6P	66	50	6.30	β
1CSP	67	48	6.54	β
1MJC	69	47	5.23	β
2AIT	74	61	4.21	β
1K8M	87	59	-0.71	β
1TEN	89	75	1.06	β
1FNF	90	76	-0.92	β
1WIT	93	80	0.41	β
1QTU	115	70	-0.36	β
1DIV	56	25	6.61	$\alpha\beta$
2PTL	62	36	4.10	$\alpha\beta$
2CI2	65	35	3.87	$\alpha\beta$
1RFA	78	49	7.0	$\alpha\beta$
2HQL	72	51	0.18	$\alpha\beta$
1HDN	85	51	2.69	$\alpha\beta$
1URN	96	49	5.76	$\alpha\beta$
2ACY	98	61	0.84	$\alpha\beta$
1APS	98	64	-1.47	$\alpha\beta$
1DIV	93	52	0.0	$\alpha\beta$

Table 2. *The set of 35 multistate proteins*

PDB ID	Length	N_α	$\ln k_f$	Structural class
2ABD	86	33	6.48	α
2CRO	65	26	5.35	α
1UZC	69	17	8.68	α
1CEI	85	22	5.8	α
1BRS	90	43	3.37	α
2ASE	156	85	3.50	α
1TTT	89	72	3.6	β
1FNF	93	75	5.48	β
1HNG	96	73	1.8	β
1ADW	123	84	0.64	β
1EAL	127	80	1.3	β
1HFC	131	94	3.4	β
1OPA	133	97	1.4	β
1HCD	118	78	1.1	β
1BEB	156	103	-2.20	β
1B9C	224	171	-2.76	β
1IIB	151	105	-4.01	β
1PGB	56	33	6.40	$\alpha\beta$
1UBQ	76	49	5.90	$\alpha\beta$
1GXT	89	53	4.39	$\alpha\beta$
1SCE	97	22	4.17	$\alpha\beta$
1HMK	121	57	2.79	$\alpha\beta$
3CHY	128	62	1.0	$\alpha\beta$
1HEL	129	69	1.25	$\alpha\beta$
1DK7	146	97	0.83	$\alpha\beta$
1JOO	149	75	0.30	$\alpha\beta$
2RN2	155	85	1.41	$\alpha\beta$
1RA9	159	96	-2.46	$\alpha\beta$
1PHP	175	97	2.30	$\alpha\beta$
1PHP	219	121	-3.44	$\alpha\beta$
2BLM	260	136	-1.24	$\alpha\beta$
1QOP	268	132	-2.5	$\alpha\beta$
1QOP	392	218	-6.9	$\alpha\beta$
1BTA	89	48	1.11	$\alpha\beta$
1L63	162	64	4.10	$\alpha\beta$

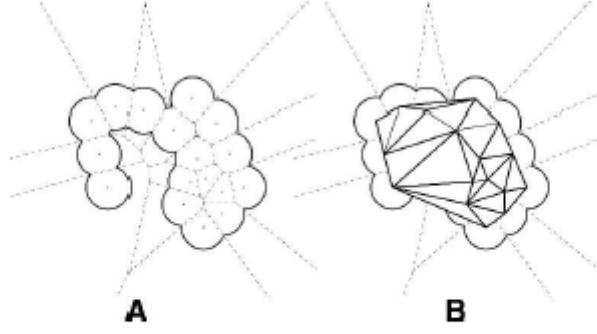


Figure 1. Voronoi diagram of a simple 2D molecule. (A) The molecule is formed by disks of uniform size. The dashed lines represent the Voronoi diagram, in which each region contains one atom. (B) The Delaunay edges of the molecule.

Définition du contact géométrique La plupart du temps, un contact est déclaré si deux résidus sont à une distance limite spécifique. Ce genre de définition peut inclure des paires qui n'ont pas d'interaction stérique. Le contact géométrique réfère à un contact non-local, bien stabilisé et s'appuie sur le critère de Voronoi (méthode statistique). Sur la figure 1, chaque cellule de Voronoi contient un seul atome et chaque point à l'intérieur de cette cellule est plus proche de cet atome que de n'importe quel autre. Une cellule de Voronoi est défini par ses arêtes qui sont les bisectrices perpendiculaires des segments reliant deux atomes. Ces segments sont les appelés arêtes de Delaunay. Les résidus i et j forme un contact géométrique s'ils sont connecté par une arête de Delaunay et que leur arête de Voronoi correspondante intersectionne le corps de la protéine. Ils ont fait le choix que les résidus devaient être distant d'au moins 4 résidus et que leur distance spatiale ne pouvait pas être plus grande que $6,5\text{\AA}$. N_α est donc le nombre total de résidus dans la protéine ayant ce type de contact.

$$RCO = \frac{1}{L * N} \sum^N \Delta S_{i,j}$$

$$ACO = \frac{1}{N} \sum^N \Delta S_{i,j}$$

où N = le nombre total de contact

$\Delta S_{i,j}$ = la séquence séparant les résidus i et j

L = le nombre total de résidus

RCO calcule les importances relatives des contact locaux et distants tandis que ACO est la séquence de séparation moyenne, non normalisée par la longueur de la chaîne. C'est la valeur N qui prendra la valeur N_{alpha} ou une autre valeur (basée seulement sur la distance maximale de $6,5\text{\AA}$ par exemple).

Résultats et discussions

La définition de contact géométrique basée sur le critère de Voronoi améliore la corrélation avec vitesse de repliement

Dans la table 3 plus la valeur est négative, plus la corrélation est significative.

Comparaison de la définition de contact géométrique et définition basée sur la distance

La mesure basée sur la distance déclare deux résidus en contact si leurs atomes sont séparés par $6,5\text{\AA}$. Avec la mesure géométrique, ~ 5200 résidus sont en contact, avec l'autre, ~ 8400 : il y a un donc $\sim 40\%$

Table 3. *Correlation coefficients of structure-derived parameters with protein folding rates*

Parameter	Correlation coefficient		
	Two-state	Multistate	All
RCO	-0.53	0.06	-0.15
ACO	-0.83	-0.64	-0.76
L	-0.72	-0.79	-0.72
$L^{1/2}$	-0.72	-0.83	-0.78
$L^{2/3}$	-0.72	-0.82	-0.76
$L^{3/5}$	-0.72	-0.82	-0.77
$\ln(L)$	-0.69	-0.84	-0.79
N_α	-0.86	-0.86	-0.83

Here RCO is the contact order, ACO the absolute contact order, L is chain length, N_α is the alpha contact number.

Table 4. *Comparing distance-based and geometry-based definitions of contacts, for correlating with folding rates*

Parameter	Correlation coefficient	
	Distance contact	Geometric contact
RCO	-0.11	-0.15
ACO	-0.77	-0.77
N_α	-0.81	-0.83

The geometric definition gives a small improvement.

des paires qui sont non nécessaires ou qui dégrade la corrélation.

Aussi, N_α est un meilleur prédicteur de $\ln k_f$ que la longueur de la chaîne bien que que $\ln k_f$ et L sont bien corrélés ($R = 0,91$) (cf. figure 3)

Différents contacts géométriques contribuent différemment à la vitesse de repliement

Chaque type de résidu i à une contribution w_i :

$$\ln k_f = a + n_\alpha * w$$

où n_α *lpha* = vecteur à 20 dimensions enregistrant le nombre de chaque type de résidu jouant dans les contacts géométriques

w = vecteur à 20 dimensions des contributions

A ce stade, w doit encore être déterminé. En utilisant le set de 80 protéines, ils obtiennent les contributions optimales et la constante a . Celle-ci minimise l'erreur résiduelle du $\ln k_f$ prédit avec les valeurs obtenues expérimentalement par une mesure de distance Euclidienne. La table 5 montre que la valine,

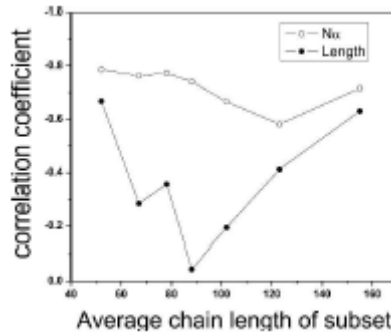


Table 5. The weight parameters for the different residue types in determining protein folding rates

Residue	w	Residue	w
Glu	0.451	HIS	-0.111
Phe	0.181	THR	-0.153
Pro	0.112	GLN	-0.154
Arg	0.106	SER	-0.171
Ala	0.091	ASP	-0.179
Cys	0.069	LEU	-0.259
Gly	0.040	TYR	-0.319
Asn	0.032	TRP	-0.387
Lys	-0.043	ILE	-0.400
Met	-0.055	VAL	-0.612

w is the relative weight of each residue type, given that it is in geometric contact.

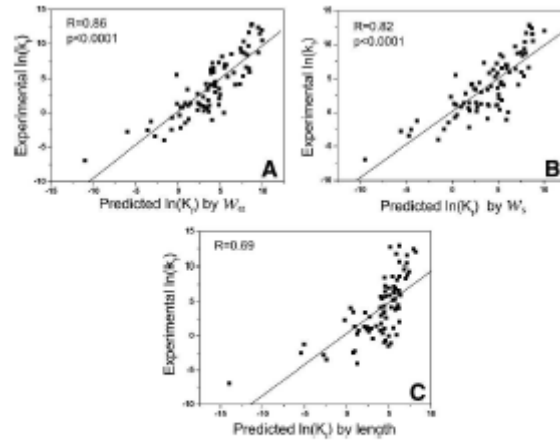


Figure 5. Scatter plots of the predicted and experimentally measured values of $\ln k_f$ in jackknife leave-one-out tests: (A) using weighted geometric contact number; (B) using sequence information only, and (C) using chain length.

l'isoleucine, le tryptophane et la tyrosine ralentissent le repliement tandis que la glutamine et la phénylalanine l'accélèrent.

Limite supérieure de repliement des protéines

En considérant que la plus petite protéine fait 20 résidus et que le résidu le plus rapide à se lier est la glutamine, on arrive à ceci : $\ln k_f = 10,29 + 20 * 0,451 = 19,3$, soit environ 4 nsec.

Vitesse de repliement et structure avec un alphabet réduit d'acides aminés

Pour éviter de devoir faire trop d'essais, ils utilisent un alphabet à 2 résidus, plus exactement 2 classes contenant les résidus. Ces classes ont été choisies après différents essais. Ici, $n_\alpha = (n_1, n_2)$ et $w = (w_1, w_2)$. Cela permet encore d'augmenter le coefficient de corrélation ($R = -0,87$) (cf. figure 5). Ils ne précisent pas comment les contributions sont calculées (somme, moyenne...?).

Prediction de la vitesse de repliement d'après la séquence

Comme les protéines sont repliées de manière compacte, ils font l'hypothèse que tous les résidus d'un même type ont la même probabilité d'avoir un contact géométrique. Avec cette hypothèse, la vitesse de

Table 6. Predicting protein folding rates using reduced alphabets of amino acids

Method	Amino acid groups	Weights	R
w_α	+(A C E F M N R G H K L P T)	0.015	-0.86
	-(D I Q S V W Y)	-0.324	
w_x	+(A C E F M N R)	0.999	-0.82
	-(D I Q S V W Y G H K L P T)	-0.534	

Here, w_α is the vector of weights for geometric contacts, w_x is the vector of weights for sequence, R is the correlation coefficient of leave-one-out jackknife tests; "+" denotes the group of residues which facilitate protein folding, and "-" denotes the group of residues which may retard folding.

repliement peut-être déterminée uniquement d'après la séquence et l'abilité de chaque acide aminé à former un contact géométrique.

Les propensités de contact géométrique ont été estimées sur base d'un set de 1670 protéines non redondantes avec < 25% d'identité de séquence. Ces propensités définissent le vecteur à 20 dimensions p :

$$\ln k_f = a + n_2 * P(p \circ w)$$

où n_2 = est le vecteur à deux dimensions de l'alphabet réduit

P = dénote la projection du produit des vecteurs à 20 dimensions sur l'espace à deux dimensions

$$P(p \circ w) = \left\{ \sum_{i \in A_1} p_i * w_i, \sum_{i \in A_2} p_i * w_i \right\} \quad (1)$$

Le modèle résultant est le suivant : $\ln k_f = 9,334 + n_2 * w_s$. Voir w_s dans la table 6.

Conclusions

La quantité N_{alpha} est hautement anti-corrélée avec la vitesse de repliement. Elle donne une meilleure corrélation que les quantités RCO, ACO et L. Elle n'est pas biaisée par les boucles courtes tant que 2 pairs sont séparés d'au moins 3 résidus tandis que les mesures RCO et ACO fonctionnent mieux avec des contacts distants. La base physique de ces corrélations peuvent être dues aux mécanismes de zipping et d'assemblage. Les contacts entre les résidus qui sont largement séparés sont plus difficiles car leur recherche de conformation est plus coûteuse en terme d'entropie et le repliement semble se former par mécanisme local de fermeture éclair.

Les différents acides aminés ont des propensités différents pour la vitesse de repliement. Le repliement est ralenti par Val, Ile, Trp et Tyr et accéléré par Glu et Phe.

Les états de transition et la signification de la valeur Φ dans la cinétique de repliement des protéines

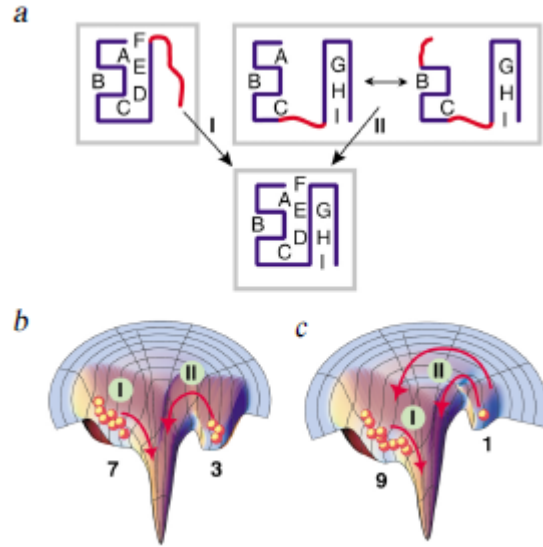
Oskhan, Bahar et Dill

Φ comme la règle cinétique de la "nativité" (nativeness) de la transition d'état

La valeur Φ (Φ -value) permet d'étudier l'étape limitante dans la vitesse de repliement des protéines. La valeur Φ est le changement dans l'énergie libre de l'état de transition, induit par une mutation, divisé par le changement d'énergie libre à l'équilibre du repliement.

La valeur Φ est traditionnellement vue comme une règle cinétique, montrant à quel point l'état de transition d'un monomère ressemble à l'état natif. Sa valeur est typiquement entre 0 et 1. A $\Phi = 0$,

Fig. 2 A negative Φ -value is defined as an acceleration of folding following a destabilizing mutation. Destabilizing (increasing the energy) of a step in the slow flow channel (II) rechannels the flow into the faster channel (I). **a.** Rate limiting conformational steps along channels I and II. **b.** For wild type protein, greater population and faster flow is along route I. **c.** Mutant protein has a destabilized energy well (right side), forcing more population of channel I. Hypothetical populations of 7:3 shifted to 9:1 illustrate a Φ -value of approximately -0.30.



la structure est dépliée au site de la mutation autant qu'à l'état dénaturé; à $\Phi = 1$, la structure est repliée, au site de mutation, autant que dans la protéine native. Des valeurs fractionnelles indiquent que la structure est présente mais faible. Des interprétations classiques ne peuvent donc pas expliquer des valeurs négatives ou supérieures à 1. Pourtant ces valeurs existent.

La nativité de l'état de transition est exprimé par la coordonnée ξ , où $\xi = 0$ réfère à la conformation dénaturée et $\xi = 1$ à la conformation native. Pour toutes les quantités à une dimension, comme une coordonnée de nativité, il y a la notion d'entre-deux. Ainsi toutes les conformations de la protéine peuvent être mises le long de la coordonnée d'entre D (dénaturé) et N (natif) ($\xi_D < \xi < \xi_N$). Cela veut dire que la conformation doit être en les états dénaturé natif. Par-là, $\Phi < 0$ signifierait que la conformation est plus dénaturée que l'état dénaturé et inversement pour $\Phi = 0$. Il existe pourtant une base physique à ces valeurs non classiques qui peut être expliquée par un modèle de réseau exact. Alors que les valeurs classiques se retrouvent pour des systèmes ayant une simple coordonnée de réaction, les valeurs non classiques arrivent pour des courants parallèles et couplés, comme dans un paysage énergétique en entonnoir (cf. figure 2).

Les routes de repliement de protéines modèles

Sur la figure 1, la structure native émerge systématiquement avec un seul temps de relaxation, à travers de multiples routes microscopiques couplées. La structure native de la protéine a un modèle de réseau à deux dimensions équivalent à un feuillet β à deux brins anti-parallèles adjacent à une hélice α . Le processus de repliement implique deux routes principales. Dans le canal dominant I, le repliement le plus rapide avec le meilleur flux, l'étape limitante dans la vitesse est la conformation où il n'y a pas encore le second brin (boîte grise de la figure 1). Le long du canal II, plus lent et avec un flux plus faible, l'étape limitante possède le feuillet mais une hélice partielle.

La base physique pour les valeurs Φ non classiques

Une valeur négative de Φ signifierait qu'une mutation accélérerait le repliement. En déstabilisant le contact G, c'est-à-dire en diminuant son pouvoir attracteur (Φ négatif), le feuillet β est déstabilisé, le courant est ralenti dans le canal II et la populations d'hélices augmentent dans le canal I (cf. table 1 et figure 2).

Un Φ positif a une explication inverse : la déstabilisation d'un contact dans le canal rapide redirige le courant vers le canal lent. En conclusion, les valeurs non classiques déterminent des routes parallèles de

Fig. 1 Folding kinetics of a lattice model chain. The top row shows the high energy, high entropy conformations at the top of the funnel. Lower rows represent lower energy, lower entropy conformations deeper on the landscape. The bottom is the native (N) conformation. Channel I is the fast route, and channel II is slower. The boxed nonnative structures are the rate limiting steps for these channels, found from landscape mapping (see text).

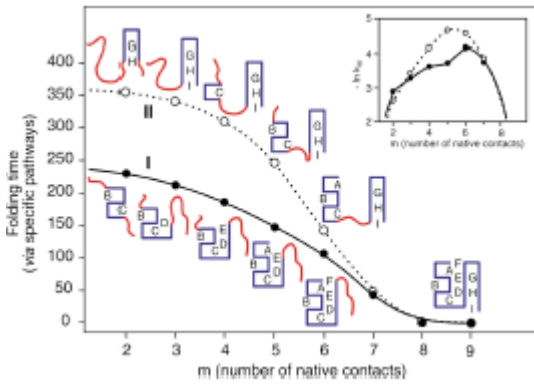
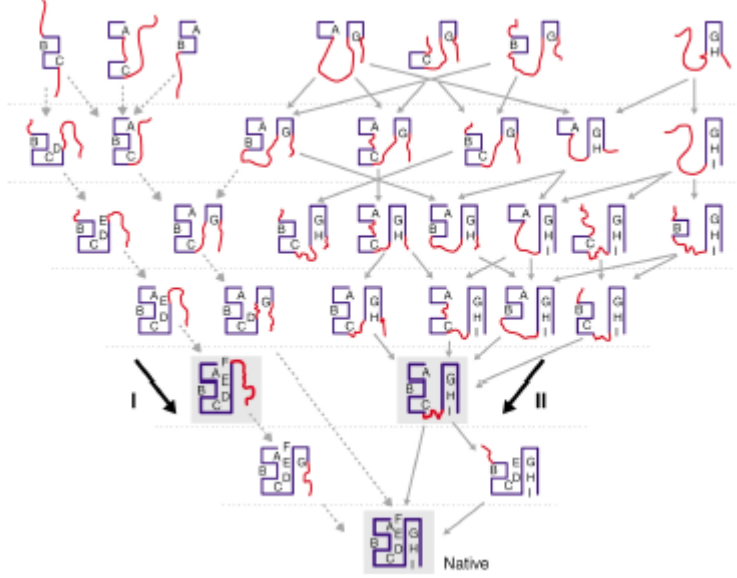


Fig. 3 Landscape mapping. A chain is put into a specific conformation with m native contacts before time $t = 0$. Folding is initiated at $t = 0$. The passage time required for each single step transition between macroconformations is computed. The y-axis shows the cumulative time to reach the native state, which is the sum of times for the individual steps. Performing the helix (I) speeds up folding more than performing the β -sheet (II). Inset: The logarithms of the individual step rates k_{ij} map the relative energies on the landscape. For example, the barrier along the β -sheet route is higher than for the helix route.

repliement et non des artéfacts expérimentaux.

Les macroroutes ont des barrières, les microroutes n'en n'ont pas

Pour le 16-mères, le nombre de conformations possible est de ~ 800.000 . Une microroute est une trajectoire particulière entre deux conformations, i et j , avec une seule constante de vitesse k_{ij} . Une macroroute est l'ensemble des conformations ayant un set particulier de contacts. Il y a ~ 270 macroroutes. Une macroroute entre 2 macroconformations, m et n est la collection de toutes les routes d'une macroconformation à une autre.

Les microroutes n'ont pas de barrières et suivent un paysage énergétique en entonnoir, les macroroutes ont des barrières cinétiques. Cela provient de la balance de deux effets : (1) la vélocité des changements de conformations (le long des microroutes) augmentent de manière monotone en descendant dans le paysage et (2) le nombre de microroutes diminue en descendant dans le paysage. Cela mène à des étapes en goulot (bottleneck) le long des macroroutes. Sur la figure 3, le temps maximum par étape identifie les macroétapes limitantes : ce sont les étapes $m=6$ et $m=5$ pour les canaux I et II respectivement. Cette barrière résulte d'une propriété du paysage, pas des microroutes. Elle est due à la réduction du nombre de routes le long de l'entonnoir énergétique à cause d'un problème énergétique le long d'une microroute.

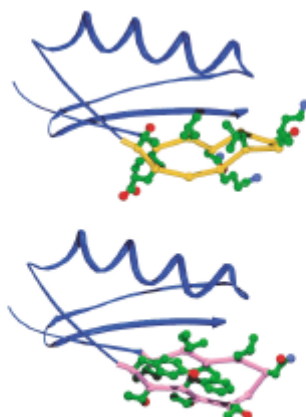


Fig. 1 Protein G (ref. 10; top) and a model of the design variant Nug2. The side chains for residues 6–16 are shown.

Le repliement implique un "nucleated zipping"

Il y a une grande hétérogénéité de vitesses si la chaîne part d'une conformation partiellement formée. Souvent, l'état natif est atteint plus rapidement en partant d'une conformation à deux contacts que d'une conformation à 6 contacts. Cette hétérogénéité est due au fait que ces conformations à 6 contacts ne sont pas sur une macroroute qui permet d'arriver à la conformation native : elles doivent remonter dans le paysage énergétique car c'est un cul-de-sac (cf. figure 2).

Il y a des noyaux de repliement, c'est-à-dire que certains contacts permettent une accessibilité cinétique à l'état natif. Les événements de repliement sont des fermetures éclairs (zippers) : les contacts les plus locaux se forment en premiers, les structures secondaires forment les contacts non locaux ensuite.

Finalement, la valeur Φ n'est pas une règle cinétique pour la progression vers l'état natif : elle est corrélée avec le changement de vitesse dû à des mutations.

La valeur absolue de $|\Phi|$ définit les contacts clés ("gatekeeper" contacts) : c'est le degré avec lequel un contact contrôle les processus vers l'état natif. Donc des sites avec $|\Phi| \gg 0$ sont des contacts clés, ceux avec $|\Phi| \sim 0$ ont peu de contrôle sur le courant.

Redesign de la voie de repliement du protéine par des méthodes de calcul

Nauli, Kuhlman et Baker

Résumé

Ils ont utilisé une stratégie basée sur des méthodes de calcul pour switcher la voie de repliement de la protéine G. Celle-ci implique normalement la formation du second et non du premier β -turn lors de l'étape limitante. Les conformations du squelette et les séquences d'acides aminés maximisant la densité d'interactions dans le premier β -haupin ont été identifiées et deux variants contenant 11 acides aminés ont été conçus. Ceux-ci sont plus stables que les protéines natives et la vitesse de repliement est 100x plus rapide.

Introduction

Les voies principales de repliement maximisent la formation des interactions natives favorables et minimisent la perte d'entropie configurationnelle de la chaîne. Cela suggère qu'il est possible d'altérer le mécanisme de repliement en changeant la stabilité intrinsèque. Il est plus facile d'étudier cela avec

Table 1 Sequences of designed proteins ¹	
WT	6-ILNCKTLKGET-16
NuG1	6-FIVIGDRVVVV-16
NuG2	6-VIVLNGTTFTY-16

¹Residues 1-5 and 17-57 are identical to wild type protein G. Turn residues are shown in bold. All variants of protein G were made and purified using published methods⁹. The identity of the proteins were verified by mass spectrometry.

des protéines ayant une certaine symétrie car il y a beaucoup de chemins possibles avec des barrières d'énergie libre proches en magnitude.

Les domaines IgG des protéines G et L consistent tous les deux en une hélice α collée à un feuillet bêta à 4 brins formés par deux β hairpins symétriquement opposés. Malgré la similarité dans la structure et la fonction, il n'y a pas de similarité de séquence. La symétrie des 2 protéines est cassée durant le repliement. Dans la protéine G, le second β -turn est présent à l'étape limitante et pas le premier, c'est l'inverse pour la protéine L.

Hypothèse de l'article : le repliement de cette classe de protéines implique la formation du β -hairpin ayant la plus faible énergie libre à l'étape limitante. Pour tester l'hypothèse, ils ont switché la voie de repliement de G vers celle de L (par ordinateur) en diminuant la stabilité intrinsèque du premier hairpin et en augmentant celle du second.

0.1 Construire des nouveaux hairpins

La conformation du squelette du premier β -hairpin a été redesigné pour supprimer l'angle ϕ positif des brins β et pour inclure des *beta*-turn plus favorable de type I' ou II' plutôt que du type natif I. Les type I' et II' sont trouvés dans les *beta*-hairpins tandis que leurs images miroirs, type I et II, connectent souvent des brins β dépareillés. En greffant le nouveau β -turn dans la protéine G, ~ 300 conformations alternatives du squelette ont été créées. Les séquences avec une faible énergie libre ont été trouvées en cherchant après les conformations compactes et enterrant bien les résidus hydrophobes. Les séquences alternatives sont considérées pour les résidus 6-16, les autres résidus restent dans leur conformation native. De-là sortent NuG2 et (tournant de type I') et NuG1 (tournant de type II') (cf. table 1). Ils ont tous les deux plus d'acides aminés hydrophobes que la protéine sauvage.

Augmenter la stabilité dans le redesign des protéines

Le spectre des deux protéines sont presque identiques à celui du type sauvage (cf. figure 2a) et des données de RX montre la structure de NuG2 diffère complètement de la structure sauvage à l'endroit du tournant. Des expériences de dépliement à l'équilibre montrent les protéines redesignées possèdent des transitions de dépliement coopératifs typiques de la protéine native et ont été stabilisées de manière importante (cf. figure 2b). Les protéines redesignées se replient 100x plus rapidement que la protéine native. A leur connaissance, ce sont les premiers à obtenir une protéine qui se replie plus rapidement que la protéine native par design ordinateur.

En parallèle, pour augmenter la stabilité du hairpin, ils ont muté l'Asp 46 en Ala pour réduire la stabilité du second β -hairpin (mutation D46A).

Le premier hairpin se replie maintenant avant le second

Des mutations dans les deux β -turns ont l'effet opposé sur les vitesses de repliement et dépliement NuG1/D46A comparé à l'effet des mêmes mutations dans la protéine native. Une mutation dans le second tournant, D46A, ralentit le repliement de la protéine native et accélère le dépliement de NuG1/D46A. Une mutation dans le premier tournant accélère le dépliement de G/D46A et ralentit le repliement de NuG1/D46A. Ces changements sont reflétés par les valeurs Φ (cf. table 2 et figure 4).

Les mutations menant à NuG2/D46A indiquent aussi un mécanisme de switch du repliement (cf. table 2 et figure 4b). Comme avec NuG1/D46A, la mutation T46A accélère la vitesse de dépliement et

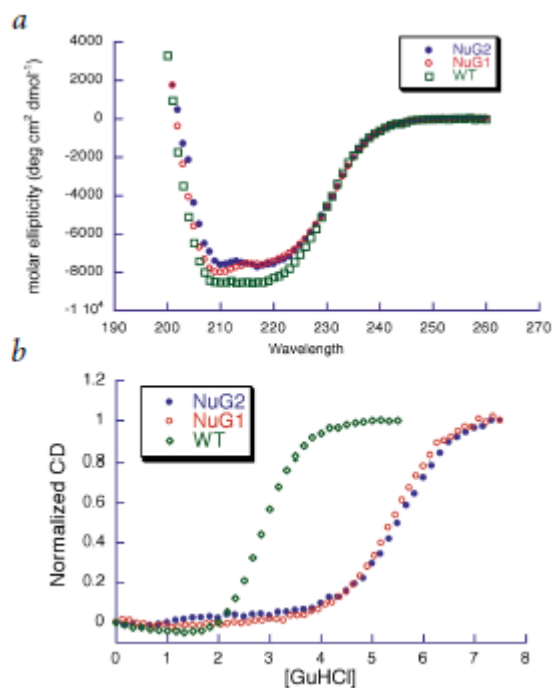


Fig. 2 CD spectra and stabilities of protein G and design variants. **a**, CD spectra of wild type (squares), NuG1 (open circles) and NuG2 (closed circles). **b**, GuHCl-induced denaturation of wild type (diamonds), NuG1 (open circles) and NuG2 (closed circles) followed by CD at 220 nm. Protein concentrations were 7 μ M in 50 mM sodium phosphate, pH 6.0, at 295 K.

Table 2 Thermodynamic and kinetic parameters of wild type and redesigned proteins ¹						
Stabilities ²	50 °C			25 °C		
	Method I	Method I	Method II	k _f ^{1.5M} (s ⁻¹)	k _f ^{2.5M} (s ⁻¹)	k _u ^{0M} (s ⁻¹)
WT	-2.8	-4.5	-4.8	9.6	0.89	6.2
NuG1	-6.1	-7.8	-8.9	1,140	170	0.70
NuG2	-6.4	-7.6	-9.1	1,050	160	0.61
ϕ -values of wild type and redesigned proteins						
	k _f ^{1.5M}	k _f ^{2.5M}	k _u ^{0M}	ϕ ^{1.5M}	ϕ ^{2.5M}	
WT	9.6	0.89	6.2	—	—	
T11A/WT	12.5	1.4	19	-0.30	-0.58	
T49A/WT	3.0	0.22	5.4	1.1	1.1	
NuG1/D46A	200	30	4.5	—	—	
NuG1/D46A/G10A	52	10	7.4	0.73	0.68	
NuG1/D46A/D11A	71	13	5.1	0.89	0.86	
NuG1/D46A/T49A	86	21	24	0.33	0.18	
NuG2/D46A	110	19	2.3	—	—	
NuG2/D46A/N10A	57	9.2	1.6	2.1	2.0	
NuG2/D46A/G11A	17	2.5	1.4	1.4	1.4	
NuG2/D46A/T49A	55	12	8.6	0.35	0.26	

¹Thermodynamic and kinetic parameters were calculated as detailed in the Methods section.

²In kcal mol⁻¹.

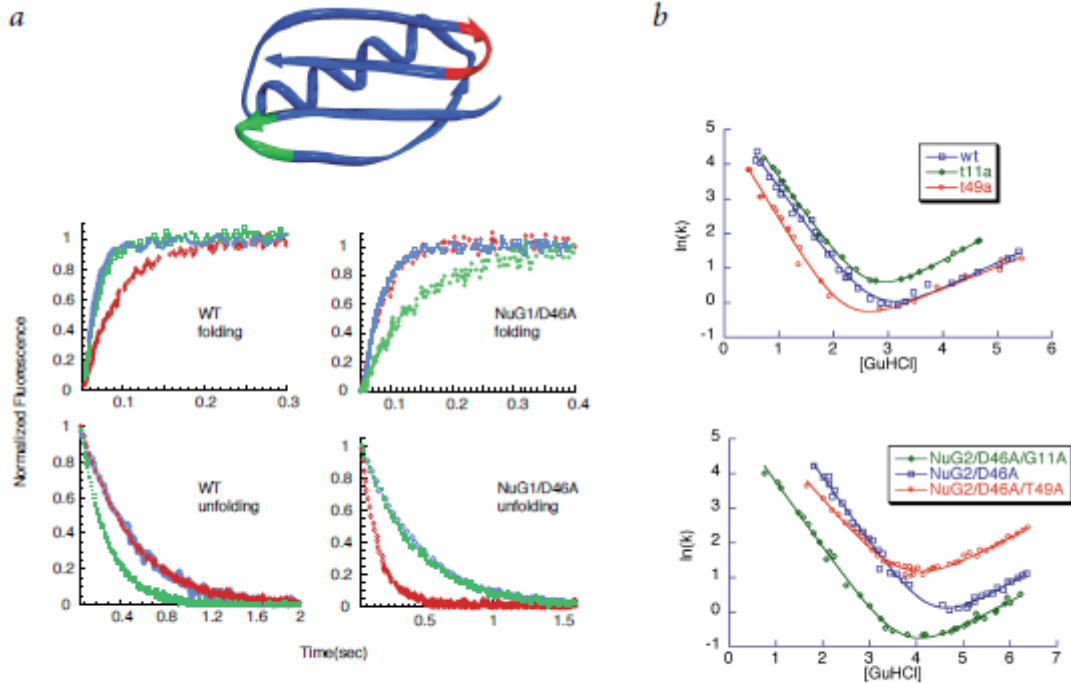


Fig. 4 The folding mechanism of NuG1/D46A is switched. **a**, Folding and unfolding kinetic traces for the parent protein (blue), a destabilizing second-hairpin mutant (red) and a destabilizing first-hairpin mutant (green). In the WT (wild type) panels, the symbols correspond to WT (blue), T49A (red) and T11A (green). In the NuG1/D46A panels, the symbols correspond to NuG1/D46A (blue), NuG1/D46A/T49A (red) and NuG1/D46A/G10A (green). Final protein concentrations were at least 7 μ M in 50 mM sodium phosphate, pH 6.0, at 295 K (ref. 11). For WT, the final GuHCl concentrations for folding and unfolding were 0.4M and 4M, and for NuG1 they were 2.5M and 6M. **b**, Plots of $\ln(k)$ versus GuHCl concentration of point mutants based on WT (top) and NuG2/D46A (bottom). The coloring scheme is the same as in (a).

la mutation dans le premier tournant la ralentit. La différence entre NuG1/D46A et NuG2/D46A est que dans le second cas, la mutation dans le premier tournant ralentit la vitesse de dépliement en plus que celle de repliement.

Conclusions

Les résultats montrent que la stabilité de la protéine peut être améliorée en optimisant la conformation du squelette et la séquence d'acide aminé du β -hairpin en utilisant une approche informatique. Cela est possible en construisant un coeur de protéines hydrophobes sans altérer la conformation du squelette.

La topologie de l'état natif est un déterminant de la vitesse de repliement de la protéine.

Dans le cas où il existe plusieurs chemins pour arriver à la conformation native, la route impliquant la formation de sous-structures avec les interactions les plus favorables est sélectionnée.

Fig. 3 The redesigned proteins fold faster than wild type protein G. Plots of $\ln(k)$ versus GuHCl concentration for WT (green), NuG1 (red) and NuG2 (blue) protein G. The rate constant, k , is derived from single exponential fits to stopped-flow fluorescence experiments and is equal to the folding rate constant, k_f , plus the unfolding rate constant, k_u . The relaxation rate constants at low concentrations of GuHCl (the left branch of the v-curve) are dominated by k_f , whereas at high concentrations of denaturant (the right branch of the v-curve) k_u dominates.

