

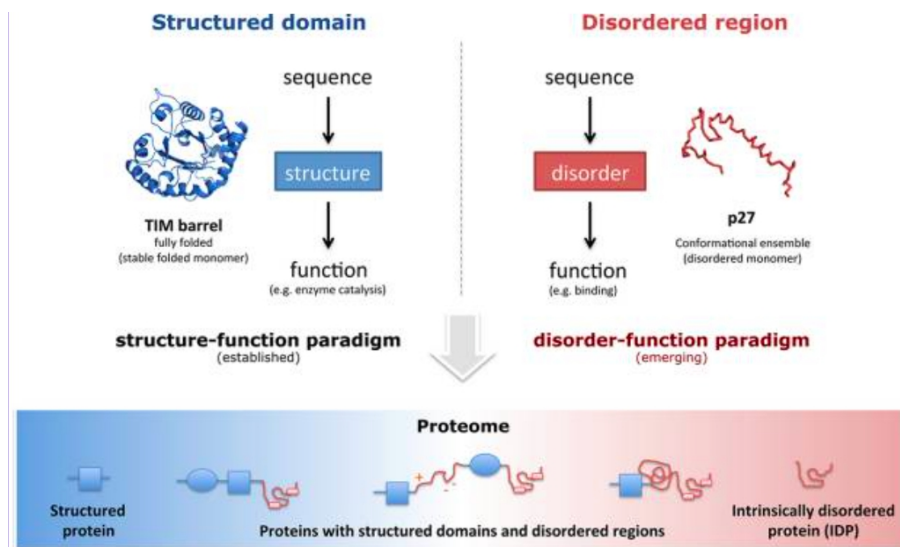
Résumé Thème 3 Biophysic II

REVIEW : « Classification of intrinsically disordered regions and proteins. » 2014

1 Introduction

1) Les segments de protéines non caractérisées sont une source de nouveauté fonctionnelle

Fonction potentielle d'une protéine vient souvent de l'analyse computationnelle de leur séquence. Caractérisation des segments de protéines non annotées et non caractérisés devrait conduire à la découverte de nouvelles fonctions et fournir des indications importantes sur les processus biologiques existants.



Domaines structurés et régions intrinsèquement désordonnées (IDRs) = 2 classes fondamentales des briques fonctionnelles des protéines.

Synergie entre ces 2 classes augmente polyvalence fonctionnelle des protéines.

2) Le modèle **structure-fonction** améliore la prédiction de la fonction

Fonction de la protéine dépend essentiellement de la structure bien définie et en 3D de la chaîne polypeptidique.

Ce modèle est basé sur :

- concepts expliquant spécificité des enzymes
- structure de protéines repliées déterminés par rayons X

La fonction peut être déduite de la séquence et de sa structure.

La similitude structurelle entre les protéines peut révéler des relations évolutives éloignées qui ne sont pas facilement détectables par des méthodes basées sur la séquence.

3) La **classification** facilite d'avantage la prédiction de la fonction

Domaines structurés peuvent avoir fonctions moléculaires distinctes.

Les séquences qui composent domaines structurés peuvent être organisées dans des familles de séquences homologues.

Systèmes de classification : Pfam / SCOP

4) Des régions et des protéines intrinsèquement désordonnées

Régions intrinsèquement désordonnées (IDR) :

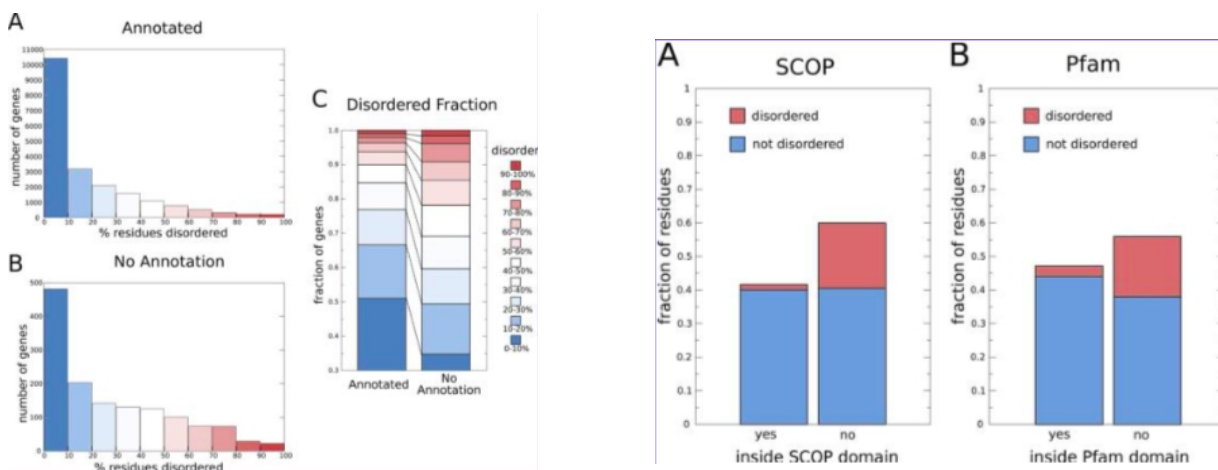
- Nombreuses protéines doivent adopter une structure bien définie pour remplir leur fonction MAIS

grande partie de ces protéines, constituées de segments polypeptidiques incapable de former une structure 3D définie, sont quand même fonctionnelles.

- manque d'aa hydrophobes encombrants → pas de noyau hydrophobe bien organisé (constitue domaine structuré)
- participe activement à diverses fonctions médiées par protéines.
- identification des régions homologues plus difficile

Protéine intrinsèquement désordonnée (IDP) : protéines avec des séquences totalement désordonnées qui n'adoptent pas de structure tertiaire.

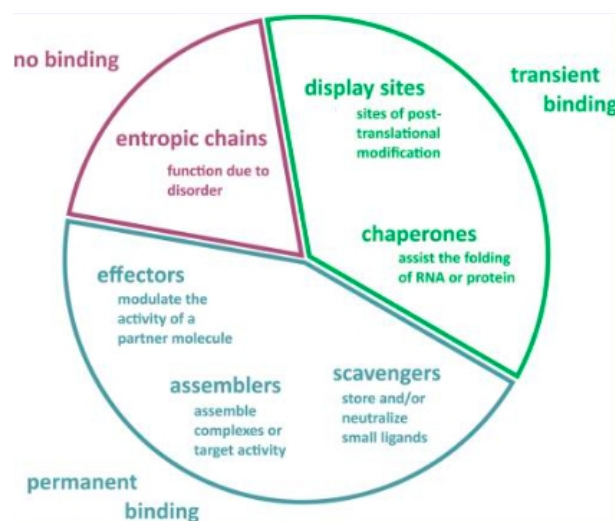
5) Le besoin de classification applicable aux régions et protéines intrinsèquement désordonnées
séquences de protéines **modulaires** = régions structurées + désordonnées



2 Fonction

Une protéine unique peut être constituée de plusieurs régions désordonnées qui appartiennent à différentes catégories fonctionnelles.

6 classes de fonctions différentes des régions désordonnées: entropic chains, display sites, chaperones, effectors, assemblers, and scavengers. (voir figure)



1) **Entropic chains** : remplissent des fonctions qui bénéficient directement de leur désordre conformationnelle → ils fonctionnent sans jamais devenir structuré.

« flexible linker » : permettent mouvement de domaines positionnés sur deux extrémités du segment de liaison l'un par rapport à l'autre

« spacers » : régulent les distances entre les domaines

La flexibilité est une caractéristique fonctionnelle qui doit être maintenue.

Une autre sous-catégorie de chaînes entropiques : ressorts entropiques.

2) **Display sites** : sites de modifications post-traductionnelles (PTM) affectent stabilité, rotation, potentiel d'interaction, et localisation des protéines dans la cellule. La flexibilité conformationnelle des régions de protéines désordonnées comme les « display sites » offre des avantages sur les régions structurées :

- facilite dépôt de PTM en permettant l'interaction transitoire mais spécifique avec les sites catalytiques des enzymes de modification
- permet accès facile et reconnaissance de PTM au sein de l'IDR par des protéines effectrices qui interviennent dans les résultats en aval lors de la liaison.

3) **Chaperones** : protéines qui assistent les molécules d'ARN et de protéines pour atteindre leur état fonctionnel replié. La polyvalence des segments désordonnés semble bien adaptée pour la fonction de chaperon.

4) **Effectors** : interagissent avec d'autres protéines et modifient leur activité. Lors de la liaison à leur partenaires d'interactions IDRs subissent souvent une transition « disorder-to-order » (ou « coupled folding and binding »). IDR peut également affecter l'activité d'autres parties au sein de la même protéine, soit par interactions concurrentielles soit par modulation allostérique.

Les régions auto-inhibitrices sont généralement enrichies pour le désordre intrinsèque et ont souvent des structures différentes dans les états fonctionnellement actifs et inhibiteurs de la protéine.

Couplage allostérique ne nécessite pas toujours une route structurelle bien définie pour se propager à travers la protéine mais peut aussi être déterminée par la stabilité des conformations individuelles de la protéine qui changent lors de la liaison à leur partenaires d'interactions.

5) **Assemblers** : Assembleurs désordonnés rassemblent plusieurs partenaires de liaison pour favoriser la formation de complexes protéiques d'ordre supérieur. La présence de différentes régions fonctionnelles dans les segments désordonnés permet liaison et rassemblement des différents partenaires.

-> mortier structurel (rassembler les protéines en stabilisant les complexes qu'elles forment)

-> échafaudage (squelette pour l'ensemble régulé spatio-temporelle des différents partenaires de signalisation)

Régions d'échafaudage : un des plus hauts degrés de désordre parmi les catégories fonctionnelles

6) **Scavengers** : stockent et neutralisent petits ligands.

3 Caractéristiques fonctionnelles

1) **Motifs linéaires (3-10 aa)**

majorité des motifs linéaires ont été trouvées comme enrichies en IDR.

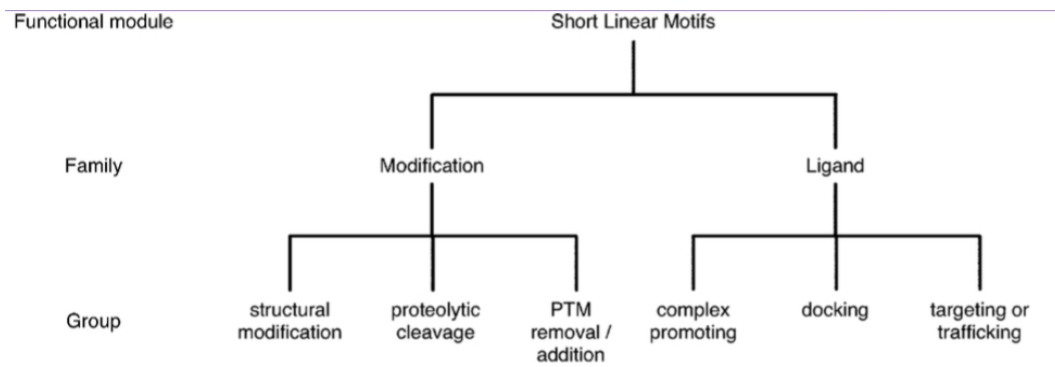
En régulant les interactions de faible affinité, ces motifs de séquences courtes peuvent cibler des protéines à une localisation subcellulaire particulière, recruter des enzymes qui modifient l'état chimique du motif par des modifications post-traductionnelles (PTM), contrôler la stabilité d'une protéine, et promouvoir le recrutement de facteurs de liaison pour faciliter la formation du complexe.

Motifs linéaires sont aidés par la nature flexible des régions désordonnées qui les entourent

-> se lient principalement sur les surfaces des domaines globulaires

Une conséquence de ces propriétés est que les virus et les bactéries pathogènes ont évolué pour imiter ces motifs linéaires, ce qui leur permet de manipuler la régulation de processus cellulaires.

- ceux qui agissent comme sites de modification
- ceux qui agissent comme des ligands



Ils ont tendance à agir comme des commutateurs moléculaires pour 2 raisons:

- faible affinité en raison de la surface de liaison limitée
- faible encombrement permet aux motifs de se produire plusieurs fois dans la même protéine, ce qui favorise les interactions d'avidité élevée et le recrutement de plusieurs facteurs.

2) Caractéristiques de reconnaissance moléculaire (MoRFs) (10-70 aa)

- favorisent les interactions protéine-protéine spécifiques (transitions « disorder-to-order »)
- peuvent servir de points de contact initial pour les événements d'interaction

3) Domaines intrinsèquement désordonnés (IDDs)

- variété de fonctions mais le plus souvent impliqués dans liaison à ADN, ARN et protéines.
- Certaines familles de domaine exige seulement présence d'un désordre dans leur voisinage pour le fonctionner, tandis que d'autres semblent compter sur la présence de régions désordonnées dans des endroits spécifiques par rapport au début ou à la fin du domaine des protéines.

4) Continuum des caractéristiques fonctionnelles

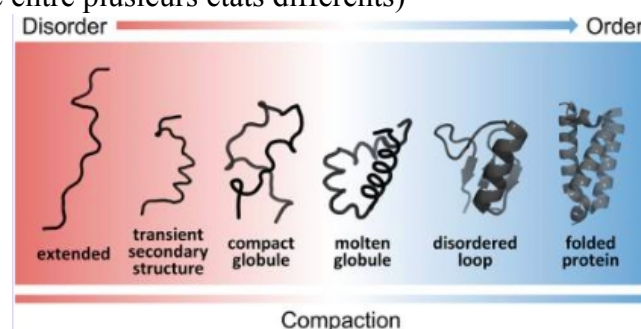
Longueur souvent utilisée pour distinguer les différents types de modules de liaison désordonnés. Ces 3 caractéristiques fonctionnelles sont différents états dans le même continuum de mécanismes impliquant des régions désordonnées contraignantes.

4 Structure

-> systèmes de classification qui sont fondés sur les caractéristiques structurales des protéines désordonnées.

1) Continuum structurelle

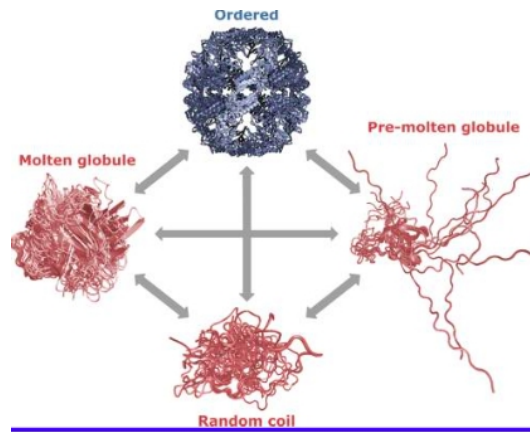
continuum de conformation : allant de entièrement structuré à complètement désordonné. (fluctuation stochastique entre plusieurs états différents)



2) Ensembles conformationnels

Les populations de conformations individuelles au sein de l'ensemble varient dans des conditions différentes.

3) **Modèle de protéine quartet** -> propose que la fonction des protéines peut provenir de quatre types d'états conformationnels et les transitions entre elles: « random coil, pre-molten globule, molten globule, and folded »



4) Structure supertertiaire

=> Etat de la variabilité conformationnelle au sein d'une protéine se situe entre la structure tertiaire des protéines et la structure quaternaire des assemblages multiprotéiques.

5 Séquence

=> systèmes de classification en fonction de la séquence de IDR

- relations dans l'ensemble structure-séquence
- saveurs des pronostics
- espace de la complexité séquence-désordre
- classées sur la base du degré global de désordre
- longueur des régions désordonnées
- position des régions désordonnées
- répétitions en tandem

6 Interaction des protéines

-> utilisation d'une combinaison de 2 mécanismes de liaison : la sélection conformationnelle et le repliement induit.

- **complexes flous** :

-> dans complexes de désordre statique : régions désordonnées peuvent adopter plusieurs conformations bien définies dans le complexe

-> dans désordre dynamique : oscillations entre différents états d'un ensemble à l'état lié.

- **plasticité de liaison** : une autre catégorisation des IDRs basée sur plasticité des liaisons

7 Evolution

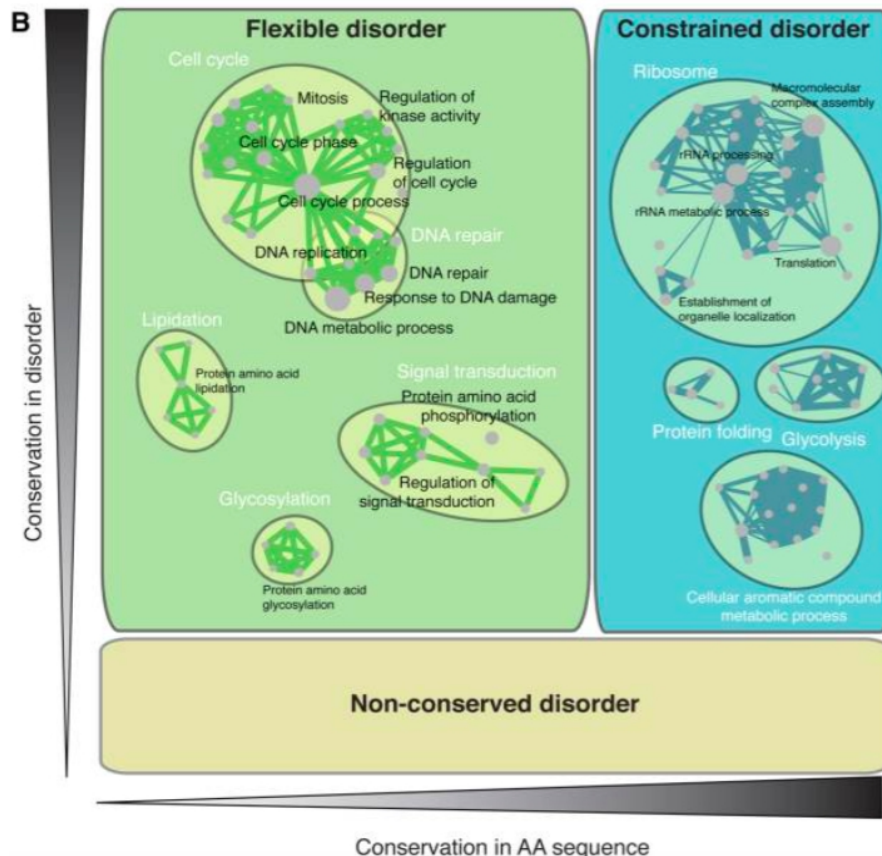
Régions désordonnées évoluent généralement plus rapidement que domaines structurés.

-> classifications de l'évolution des régions des protéines désordonnées.

IDR avec des fonctions et des propriétés similaires ont tendance à avoir des caractéristiques d'évolution similaires.

1) Conservation de la séquence : varie en fonction des fonctions spécifiques

- les régions désordonnées qui contiennent Slims constituent désordre flexible afin de fournir la flexibilité pour permettre l'accès au motif linéaire par des protéines qui les unissent en tant que ligands ou d'introduire des modifications post-traductionnelles.



2) Lignage et spécificité des espèces 3) Histoire évolutionnaire et mécanisme de l'expansion répétée

8 Régulation

disponibilité modifiée des IDPs est associée à des maladies (cancer, maladies neurodégénératives)

La régulation de la disponibilité des IDPs garantit qu'ils sont disponibles dans les dosages appropriés et pour la bonne quantité de temps,

=> rôles fonctionnels possibles de protéines avec IDR basé sur propriétés régulatrices cellulaires :

1) motifs d'expression (abondance de la transcription) 2) épissage alternatif 3) cinétique de dégradation 4) processus post-traductionnel et sécrétion

9 Propriétés biophysiques

Beaucoup de travaux biophysique ont été réalisés sur le désordre structural de protéines en utilisant une variété de techniques expérimentales.

- 1) solubilité : distribution des charges et composition de la séquence
- 2) la possibilité de subir des transitions de phase
- 3) rôle de la biominéralisation.

10 Discussion

->améliorer la couverture de l'annotation en prédisant la fonction de protéines nouvelles et indéfinies basées sur des informations communiquées par les protéines les mieux étudiées.

->combinaisons de plusieurs systèmes de classification existants pourraient atteindre prédiction de la fonction de meilleure qualité pour les IDRs.

1) les méthodes actuelles de prédiction de la fonction des IDRs et IDPs :

- approches basées sur des motifs linéaires
- approches basées sur des sites PTM
- approches basées sur caractéristiques de reconnaissance moléculaire
- approches basées sur les domaines intrinsèquement désordonnés

Article : « Prediction of protein binding in disordered proteins » 2009

Introduction

Explication sur pourquoi beaucoup de protéines ne sont pas bien structurées et présentent des régions désordonnées et ont pourtant fonctions biologiques.

IUP = Intrinsically unstructured proteins

IDP = Intrinsically disordered proteins

De nombreuses protéines désordonnées fonctionnent en se liant spécifiquement à d'autres protéines, ADN ou ARN :

=> processus de couplage « folding and binding » (transition de l'état désordonné à un état plus ordonné avec des éléments structuraux secondaires et tertiaires stables)

La flexibilité inhérente des protéines désordonnées offre plusieurs avantages dans la liaison.

Méthode : Dans cet article ils utilisent une méthode générale pour identifier les régions de liaison spécifiques subissant une transition « disorder-to-order »

-> basée sur la méthode générale de prédiction de désordre = IUPred (basé sur l'hypothèse que les protéines désordonnées ont une composition spécifique d'acides aminés qui ne permet pas la formation d'une structure bien définie stable.)

-> procédé utilise des potentiels statistiques qui peuvent être utilisés pour calculer l'énergie d'interaction par paires à partir de coordonnées connues.

-> les résidus désordonnés peuvent être prédits en ayant des énergies par paire estimées défavorables.

-> l'estimation de l'énergie pour chaque résidu est basée sur le type d'acide aminé, et la composition d'acides aminés de sa région séquentielle et IUPred tient compte du fait que la tendance au désordre des résidus peut être modulée par leur environnement.

Ici, l'environnement est pris en compte simplement au niveau de la composition d'acides aminés.

=> c'est la méthode ANCHOR : ce modèle simple capture la propriété essentielle des régions de liaison désordonnées et permet leur prédiction robuste.

Résultats

Schéma de l'algorithme :

la caractéristique essentielle de ces régions de liaison : comportement différent suivant si isolées ou liées à leur protéine partenaire.

Etat libre -> se comportent comme des protéines désordonnées (ensemble structural très flexible).

Etat lié -> adoptent une conformation rigide, similaire à des régions des structures globulaires.

Prédiction basée sur trois propriétés :

1 un résidu donné appartient à une longue région désordonnée + filtre les domaines globulaires.

2 état isolé : résidu ne parvient pas à former suffisamment de contacts favorables avec ses propres voisins pour se replier sinon adopterait toute seule une structure bien définie.

3 vérifie la faisabilité qu'un résidu donné peut former suffisamment d'interactions favorables avec les protéines globulaires lors de la liaison -> garantit qu'il y a un gain d'énergie par interaction avec les régions globulaires.

Paramètres ont été déterminés pour minimiser la différence entre les énergies estimées et les énergies calculées à partir des structures connues de l'ensemble de données de protéines globulaires.

Prédiction de site de liaison désordonné basée sur trois scores différents :

- mesurer la tendance du voisinage d'un acide aminé à être désordonné

- estimer l'énergie d'interaction par paires du résidu donné peut gagner en formant contacts intrachânes.

- l'énergie par paires que le résidu peut obtenir par interaction avec une protéine globulaire est approximée en utilisant la composition moyenne d'aa des protéines globulaires.

Net avantage de cette approche car réduit considérablement la dépendance au jeu de données des complexes désordonnés -> laisse seulement 5 paramètres à être optimisés sur ce petit jeu de données.

La construction de la **méthode de prévision ANCHOR** démontré sur le domaine N-terminal de la protéine p53 humaine connu pour être capable de se lier à 3 différentes protéines globulaires.

(*Figure 1 p.5*) :

-> montre les 3 sites de liaison du N-terminal + calcul des 3 scores cités plus haut

-> Attention : le désordre n'est pas une condition stricte de succès de prédiction d'un site de liaison désordonné

Test de l'algorithme :

ensemble d'apprentissage positif = courts complexes désordonnés

3 ensembles d'apprentissage négatifs = protéines globulaires, interfaces ordonnées et protéines désordonnées

Division des données en 3 jeux de paramètres (testé 1 contre 2) -> ne diffèrent pas donc méthode robuste (*Table 2 et 3 p.6*) . Un site de liaison a été considéré trouvée si au moins cinq de ses acides aminés sont classés correctement.

Autre manière de décrire algorithme :

- « ROC curves » = TPR vs FPR (*Figure 2 p.6*)

- Aire sous la courbe (AUC)

Puisque les régions interactives d'une protéine désordonnée et d'une ordonnée sont intrinsèquement différentes on s'attend à ce que le prédicteur ne reconnaisse que les sites de liaison des protéines désordonnées qui interagissent avec des protéines globulaires, mais ne font pas partie des protéines globulaires elles-mêmes.

=> algorithme spécifique à des sites de liaison désordonnés par opposition à des protéines globulaires, l'interface entre les protéines globulaires ou des protéines désordonnées en général.

Evaluation du prédicteur (basé sur acide aminé)

Il y a des différences entre la composition des sites de liaison désordonnés et la composition en acides aminés de l'un des ensembles négatifs : protéines globulaires, interfaces ordonnées et protéines désordonnées (*Figure 3 p.7*)

Structures secondaires et efficacité de ANCHOR

On considère les 3 types de structures 2ndaires : helix/extended/coil

La teneur en structure secondaire dans ce type d'interactions penche beaucoup vers conformation « coil ».

=> pas de différence significative dans l'efficacité de la prédiction sur les différents éléments de structure secondaire.

Test sur les longues régions désordonnées

forte préférence pour prédire plusieurs régions interagissant pour les chaînes plus longues.

On trouve courts sites de liaison désordonné dans son ensemble ou manque complètement les sites de liaison. (*Figure 5A p.9*)

répartition de la fraction des acides aminés identifiés avec succès lors de prédiction dans les deux types de sites de liaison. (*Figure 5B p.9*)

prédiction de ANCHOR pour protéine p27 humaine (*Figure 6 p.10*)

Protéine du syndrome de Wiskott-Aldrich (WASp)

cas supplémentaire montrant la sortie de prédiction pour une séquence de protéine complète.

prédiction de ANCHOR capte correctement les sites de liaison désordonnée de WASp (*Figure 7 p.11*)

Analyses protéomiques complètes

aperçu de l'évolution des sites de liaison désordonnés, le prédicteur a été exécuté sur les 736 protéomes complets -> tendance à l'augmentation de :

- la quantité de protéines désordonnées
- la fraction d'aa désordonnés prévu pour être dans les sites de liaison désordonnés quand la complexité de l'organisme augmente. (*Figure 8 p.11*)

Distribution des longueurs de sites de liaison et de régions désordonnés, prédites par ANCHOR, dans protéomes complets. (*Figure 9 p.12*)

Discussion => résumé de tous les points de la partie résultats (p.13/14/15)

Régions en transition « disorder-to-order » pendant la liaison sont des éléments essentiels dans le processus de reconnaissance moléculaire impliquant des protéines désordonnées.

Le succès de ANCHOR confirme hypothèse que, malgré ces différences les régions de liaison désordonnées ont une propriété commune qui les prédisposent pour le couplage « folding and binding ».