



EMBL

Reflect και OnTheFly Αυτόματος Εμπλουτισμός Ιστοσελίδων και Εγγράφων

Ευάγγελος Παφίλης¹, Γεώργιος Παυλόπουλος¹, Heiko Horn¹, Seán I. O'Donoghue¹, Lars J. Jensen¹, Michael Kuhn¹, Nigel P. Brown¹, Sean Hooper² and Reinhard Schneider¹
{pafilis, pavlopou, heiko.horn, odonoghue, lars.jensen, mkuhn, nigel.brown, reinhard.schneider}@embl.de, shooper@lbl.gov
¹EMBL, Heidelberg, Germany
²Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, USA.

Reflect: Αυτόματη Αναγνώριση και Εμπλουτισμός Ονομάτων Βιοχημικών Οντοτήτων σε Ιστοσελίδες

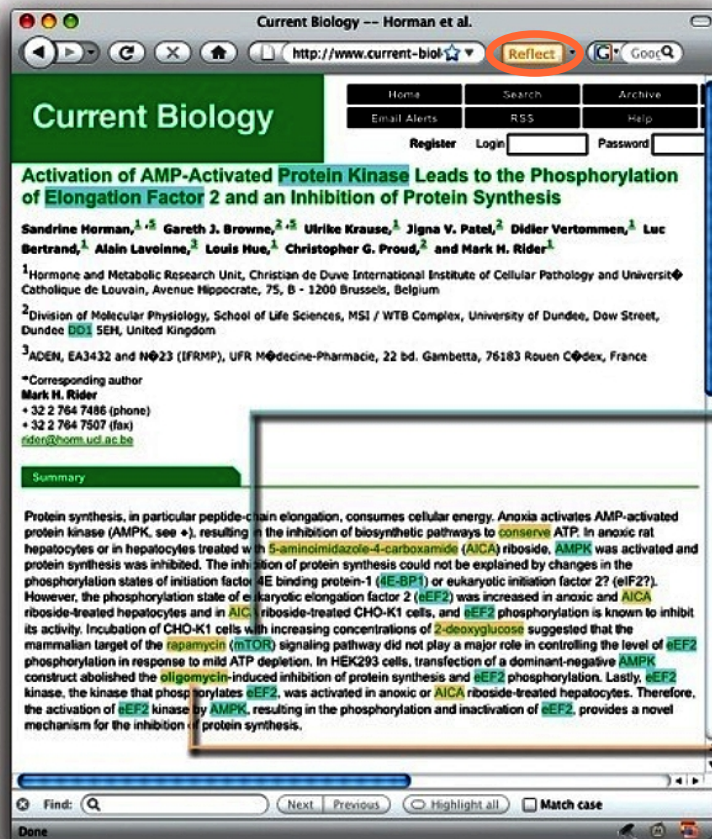
Reflect Server

18 GB RAM

Πρωτεΐνες (1,5 εκατομύρια)
Χημικές Ενώσεις (4,3 εκατομύρια)
373 οργανισμοί

- Επέκταση πλοηγτών
- Οποιαδήποτε ιστοσελίδα
- ~1" για ένα άρθρο 5 σελίδων (ADSL)
- 86% precision, 86 % recall (Biocreative I, yeast test corpus [1])

<http://reflect.ws>
contact@reflect.ws



Identifier: ENSP00000354587
Synonyms: FKBP12-rapamycin complex-associated protein; FK506-binding protein
Domains: Domains, Sequence, Structure, Locus, Literature
Sequence: APYRAVLAHQDLFSLAQGCTDKAPDLLEAELTAPAGESTSMK

Structure, **Interaction partners**, **Subcellular location**, **Organism**

oligomycin
Chemical: C1=CC=C(C=C1)C2=CC=CC=C2C3=CC=CC=C3C4=CC=CC=C4C5=CC=CC=C5C6=CC=CC=C6C7=CC=CC=C7C8=CC=CC=C8C9=CC=CC=C9C10=CC=CC=C10C11=CC=CC=C11C12=CC=CC=C12C13=CC=CC=C13C14=CC=CC=C14C15=CC=CC=C15C16=CC=CC=C16C17=CC=CC=C17C18=CC=CC=C18C19=CC=CC=C19C20=CC=CC=C20C21=CC=CC=C21C22=CC=CC=C22C23=CC=CC=C23C24=CC=CC=C24C25=CC=CC=C25C26=CC=CC=C26C27=CC=CC=C27C28=CC=CC=C28C29=CC=CC=C29C30=CC=CC=C30C31=CC=CC=C31C32=CC=CC=C32C33=CC=CC=C33C34=CC=CC=C34C35=CC=CC=C35C36=CC=CC=C36C37=CC=CC=C37C38=CC=CC=C38C39=CC=CC=C39C40=CC=CC=C40C41=CC=CC=C41C42=CC=CC=C42C43=CC=CC=C43C44=CC=CC=C44C45=CC=CC=C45C46=CC=CC=C46C47=CC=CC=C47C48=CC=CC=C48C49=CC=CC=C49C50=CC=CC=C50C51=CC=CC=C51C52=CC=CC=C52C53=CC=CC=C53C54=CC=CC=C54C55=CC=CC=C55C56=CC=CC=C56C57=CC=CC=C57C58=CC=CC=C58C59=CC=CC=C59C60=CC=CC=C60C61=CC=CC=C61C62=CC=CC=C62C63=CC=CC=C63C64=CC=CC=C64C65=CC=CC=C65C66=CC=CC=C66C67=CC=CC=C67C68=CC=CC=C68C69=CC=CC=C69C70=CC=CC=C70C71=CC=CC=C71C72=CC=CC=C72C73=CC=CC=C73C74=CC=CC=C74C75=CC=CC=C75C76=CC=CC=C76C77=CC=CC=C77C78=CC=CC=C78C79=CC=CC=C79C80=CC=CC=C80C81=CC=CC=C81C82=CC=CC=C82C83=CC=CC=C83C84=CC=CC=C84C85=CC=CC=C85C86=CC=CC=C86C87=CC=CC=C87C88=CC=CC=C88C89=CC=CC=C89C90=CC=CC=C90C91=CC=CC=C91C92=CC=CC=C92C93=CC=CC=C93C94=CC=CC=C94C95=CC=CC=C95C96=CC=CC=C96C97=CC=CC=C97C98=CC=CC=C98C99=CC=CC=C99C100=CC=CC=C100

Ένα online άρθρο επεξεργασμένο από τη Reflect και περιλήψεις για πρωτεΐνες και χημικές ενώσεις. Στο κείμενο τα ονόματα των πρωτεϊνών έχουν επισημανθεί με μπλε, ενώ οι χημικές ενώσεις με πορτοκαλί. Η επεξεργασία του κειμένου απαιτεί μοναχά ένα κλικ στο κουμπί Reflect και διαρκεί περίπου ένα δευτερόλεπτο. Κάνοντας κλικ σε οποιαδήποτε από τα επισημασμένα ονόματα εμφανίζεται η αντίστοιχη περίληψη (εικόνα: Dr Sean I O' Donoghue).

OnTheFly: Αυτόματος Εμπλουτισμός Εγγράφων και Δημιουργία Δικτύων Βιολογικής Γνώσης

- Εμπλουτισμός εγγράφων PDF, Word, Excel και απλών αρχείων κειμένου
- Αυτόματη μετατροπή σε HTML
- Εύκολη χρήση: υποστήριξη για Drag n' Drop
- ~15" για ένα PDF 15 σελίδων
- Δημιουργία δικτύων συσχετίσεων (associations) σύμφωνα με τη βάση STITCH [2] για τις πρωτεΐνες και τις χημικές ενώσεις που αναφέρονται στο έγγραφο
- Δυνατότητα επιλογής πολλαπλών εγγράφων και δημιουργίας κοινού δικτύου

<http://onthefly.embl.de>
pavlopou@embl.de, pafilis@embl.de

A. Το OnTheFly υποστηρίζει τη λειτουργικότητα Drag n' Drop B, C: εικόνες από την εμπλουτισμένη HTML έκδοση ενός πλήρους άρθρου, αρχικά διαθέσιμου ως PDF. D. Το δίκτυο αλληλεπιδράσεων σύμφωνα με τη βάση STITCH για τις πρωτεΐνες του γραφήματος C (υποδεικνύονται με ελαφρύ μπλε).

A: Screenshot of the OnTheFly interface showing a file upload area and a list of files. A green arrow points to the 'Reflect' button.

B: Screenshot of a scientific article with protein names highlighted in blue and chemical structures in orange.

C: Screenshot of a network graph showing interactions between proteins (YMR13C, YDL176W, YLR077W, YDR255C, YMR15C, YDL160W, YLR077W, YDR255C).

D: Screenshot of a network graph showing interactions between proteins and chemical structures.

Επόμενα Βήματα

- Δυνατότητα διαδραστικής (interactive) επεξεργασίας των λεξικών και των περιλήψεων στα πρότυπα της Wikipedia [3]
- Πιο πολλά λεξικά / πλούσιες περιλήψεις
 - οργανισμοί
 - μεταβολικά μονοπάτια
 - ασθένειες

Βιβλιογραφία

Η πληροφορία για τις πρωτεΐνες και τις χημικές ενώσεις προέρχεται από τις βάσεις δεδομένων STRING (<http://string.embl.de>), STITCH (<http://stitch.embl.de>), SMART (<http://smart.embl.de>) και PDBSum (www.ebi.ac.uk/pdbsum).

- [1] Biocreative: <http://biocreative.sourceforge.net>
- [2] STITCH: <http://stitch.embl.de>
- [3] Wikipedia: <http://www.wikipedia.org>