

PROGETTO FIRB 2001-2003
Bioinformatica per la Genomica e la Proteomica
III Rendicontazione

Stato di avanzamento del lavoro scientifico in percentuale del lavoro previsto

L'unita' di Bologna e' impegnata nei seguenti workpackages: n.1 e n. 4.

- Relativamente a WP1 lo stato di avanzamento e' 100% del lavoro previsto
- Relativamente a WP4 lo stato di avanzamento e' 100% del lavoro previsto

Problemi

Non si registrano particolari problemi

Risultati conseguiti nell'annualità

WP4.-Attività 1 Modellizzazione su larga scala di proteine globulari e di membrana. Nell'ambito del progetto si sono ultimate le attività come di seguito riassunte.

E' proseguita l'attività di messa in funzione e installazione su server dei predittori sviluppati durante l'attività FIRB 2001-2003 per renderli fruibili alla comunità scientifica. Come descritto precedentemente i predittori sviluppati e basati su sistemi di hidden Markov model e reti neurali sono di estrema utilità nell'ambito dell'analisi di genomi e proteomi, per la loro documentata efficienza e capacità discriminante. I predittori sono stati quindi unificati in un unico web-server (Trample) per il loro utilizzo on-line per l'analisi di sequenze (<http://gpcr.biocomp.unibo.it>).

Il server accessibile al sito www.biocomp.unibo.it contiene ora i seguenti predittori:

Per la predizione di modelli topologici di proteine di membrana (beta strand transmembrana (Porine)):

- HMM-B2TMR basato su Hidden Markov Models
- B2TMR basato su reti neurali

Per la predizione di modelli topologici di proteine di membrana di tipo all-alfa:

- Ensemble
- HTMR
- Psi Kyte-Doolittle
- Kyte-Doolittle
- PRODIV_
- TMHMM_0.91
- MEMSAT
- TMHMM_2.0

Sono inoltre presenti alcuni predittori per altre caratteristiche strutturali come:

- La predizione di struttura secondaria
- La predizione di peptide segnali (SPEPLIP)
- Un predittore per ritrovare data un sequenza remoti omologhi strutturali (BLINK)

Una delle caratteristiche piu' importanti di TRAMPLE e' il fatto che possiede un completo sistema di gestione dei risultati o "content management system". Ciascun utente possiede quindi un proprio spazio in cui puo' immagazzinare, accedere e annotare (anche con documenti) i risultati delle predizione fatte utilizzando una semplice interfaccia web. In questo modo si migliora significativamente l'usabilita' delle predizioni rispetto ai sistemi tradizionali di predizione via web. I risultati delle predizioni sono generati dinamicamente come pagine web.

L'header ha un menu' che consente la navigazione dell'application server e l'accesso ad un diario personale delle predizioni. Esiste poi sulla sinistra della pagina un menu pull-down che permette all'utente di salvare i risultati in cartelle proprie, che sono divise in Current, Temp e Backup.

Sulla sommita' della pagina viene riportato graficamente un plot che mostra l'affidabilita' delle predizioni in termini di differenze di probabilità.

Esiste poi la possibilita' di generare una vista ingrandita di una porzione della predizione per mezzo in un apposito sistema di zooming, con un semplice click del mouse. Tutte le predizioni sono

comunque mostrate in forma tabellare a colori sul fondo della pagina. Inoltre e' possibile esportare le predizioni scaricando un file in formato XML oppure in forma tabulare compatta in csv (Comma Separated Values). Per quel che riguarda il Content management system l'architettura di TRAMPLE e' basata sul concetto di "job session" o session. Ciascuna session contiene le sequenze utilizzate dall'utente, i risultati dei calcoli e le eventuali annotazioni dell'utente. L'utente dopo ciascuna predizione, che può essere anche multipla, può quindi salvare, cancellare o annotare i propri esperimenti computazionali e classificarli in una dimensione temporale per le diverse sessioni, cosa che e' preclusa nelle classiche applicazioni web. Una libreria JavaScript (<http://kupu.oscom.org/>) e' inclusa in TRAMPLE per fornire ulteriori funzionalita' di editing. Il diario, può essere utilizzato ed aggiornato, utilizzando il menu sulla sommita' della pagina.

Lo screening su larga scala di genomi è stato sperimentato sul proteoma di Escherichia coli e sulla sua variante patogena e le predizioni su tutte le catene è disponibile al sito www.biocomp.unibo.it/Firb2001-03.

E' stato poi implementato un tool per identificare possibili epitomi conformazionali sulla superficie di proteine di struttura nota. Questo strumento è in grado, data la struttura di una proteina e una lunghezza attesa dell'epitopo di elencare tutti i peptici, con un numero prescelto di mismatch, in grado di mimare la conformazione locale della proteina. Le sequenze risultanti possono essere salvati in formato adatto ad essere ricercato per similarità da programmi tipo BLAST.

WP1. Attività 2 Aggiornamento e mantenimento di banche dati sviluppate dai proponenti il progetto e integrazione di queste ultime con altre banche dati mondiali e con pacchetti per l'analisi.

In questo ambito si sono sviluppate ed aggiornate le seguenti basi banche dati:

- Una banca dati per le superfici di interazione tra proteine disponibile al sito www.biocomp.unibo.it, sempre inserita nell'ambiente di TRAMPLE. In questa banca dati per ogni proteina del PDB è visibile la superficie eventualmente in contatto con un'altra proteina corredata dei dati di predizioni e di affidabilità della predizione
- Per affrontare il problema della duplicazione di sforzi e dell'uso inefficiente delle risorse nel campo della predizione di strutture proteiche abbiamo realizzato una risorsa pubblica il "Protein Model Database" (PMDB). Questa banca dati permette di depositare e recuperare modelli pubblicati in letteratura insieme ad eventuali validazioni sperimentali. Al momento contiene più di 74,000 modelli per circa 240 proteine ed e' reperibile al sito del CASPUR.

Tipologia del risultato	SI/NO	Nr. Attività	Workpackage s	Descrizione
Nuove idee, nuove conoscenze, nuovi modelli interpretativi di fenomeni complessi;	SI	1	4	Si sono perfezionati alcuni predittori di proteine di membrana e globulari
Realizzazione di nuova strumentazione scientifica e/o di dispositivi avanzati;	SI	1	4	Si e' installato un web server TRAMPLE per un utilizzo continuo ed interattivo sei software sviluppati per la predizione della struttura proteiche Si è implementato un sistema per la identificazione di epitopi conformazionali Si è anche implementata una banca dati, con annessi tools do ricerca, di modelli tridimensionali di proteine
Messa in opera di metodologie scientifiche avanzate;				

Realizzazione di prototipi;	SI	1	4	La suite di programmi installati lavora su un computer parallelo di tipo Beowulf già a disposizione presso il Gruppo di Biocomputing della Università di Bologna
Sintesi di nuove molecole e/o di materiali artificiali;				
Proposta di nuove tecnologie;				
Contributo all'innovazione della produzione di beni di servizio;	Si	1	4	L'accesso al servizio rende disponibile una lista di proteine per il genoma di escherichia coli annotate su basi strutturali
Sviluppo di software innovativo;	Si	1	4	Integrazione in un application server e di un content management system di nuovi metodi predittivi basati su machine learning
Instaurazione di rapporti di lavoro con giovani ricercatori	SI	1	6	Training ed impiego di 6 giovani ricercatori, presso il Dipartimento di Biologia Unibo, ed il Dipartimento di Biochimica, Uni La Sapienza, Roma.
Altri risultati e/o precisazioni				

Modalità con la quale si ritiene di documentare i risultati.

Modalità	SI/NO	Nr. Attività	Workpackages	Descrizione
Pubblicazioni Scientifiche	Si	1	4	Vedi elenco allegato
Pubblicazioni su supporto informatico (CD, Web, etc.)	Si	1	4	Vedi www.biocomp.unibo.it
Edizioni critiche, lessici, liste di frequenza, etc..				
Rapporti tecnici e/o progetti				
Brevetti;				
Comunicazioni a congressi nazionali;	Si	1	4	Vedi elenco allegato
Comunicazioni a congressi internazionali;	Si	1	4	Vedi elenco allegato
Diffusione dei risultati sul piano informativo;				Sito web www.biocomp.unibo.it

Diffusione dei risultati sul piano divulgativo;				Bologna Fiere 28 Febbraio-1marzo 2005
---	--	--	--	---------------------------------------

Pubblicazioni Scientifiche 2005 pertinenti al progetto:

- Capriotti E, Fariselli P, Casadio R -I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure- Nucl Acids Res 33:W303-W305 (2005)
- Fariselli P, Finelli M, Rossi I, Amico M, Zauli A, Martelli PL, Casadio R -TRAMPLE: the transmembrane protein labelling environment- Nucl Acids Res 33:W198-W201 (2005)
- Fariselli P, Martelli PL, Casadio R -A new decoding algorithm for hidden Markov models improves the prediction of the topology of all-beta membrane proteins - BMC Bioinformatics 6,Suppl 4:S12 (2005)
- Becker L, Bannwarth M, Meisinger C, Hill K, Model K, Krimmer T, Casadio R, Truscott KN, Schulz GE, Pfanner N, Wagner R -Preprotein Translocase of the Outer Mitochondrial Membrane: Reconstituted Tom40 Forms a Characteristic TOM Pore- J Mol Biol 353(5): 1011-20 (2005)
- Morozzo della Rocca B, Miniero DV, Tasco G, Dolce V, Falconi M, Ludovico A, Cappello AR, Sanchez P, Stipani I, Casadio R, Desideri A, Palmieri F -Substrate-induced conformational changes of the mitochondrial oxoglutarate carrier: a spectroscopic and molecular modelling study- Mol Membr Biol (Cover Figure) 22(5):443-52 (2005)
- Capriotti E, Fariselli P, Calabrese R, Casadio R -Predicting protein stability changes from sequence with Support Vector Machines- Bioinformatics 21 Suppl 2: ii54-ii58 (2005)
- Pellegrini-Calace, M. and Tramontano, A. Identification of a novel putative mitogen-activated kinase cascade on human chromosome 21 by computational approaches. Bioinformatics, in press
- Castrignano', T., D'Onorio De Meo, P., Cozzetto, D., Talamo, I. and Tramontano, A. 2006. The PMDB Protein Model Database Nuclei Acids Research, 34, 306-309
- Brenner, S., Tramontano, A. 2005. Sequences and topology: a decade of genomes, Curr. Opin. Struct. Biol., 15(3) 245-247.
- Raimondo, D., Giorgetti, A., Miele, AE., Tramontano, A. 2005. Evaluating the usefulness of protein structure models for molecular replacement, Bioinformatics, 21(S2) 72-76.
- Moulton, J., Fidelis, K., Rost, B., and Hubbard, T. and Tramontano, A. 2005. Critical Assessment of Methods of Protein Structure Prediction (CASP) - Round VI, Proteins, S7, 3-7.
- Soro, S., Tramontano, A. 2005. Function Prediction at CASP6, Proteins, S7, 214-224.
- Francisci, S., De Luca, C. Oliva, R., Morea, V., Tramontano, A. and Frontali, L., 2005. Aminoacylation and conformational properties of yeast mitochondrial mutants with respiratory deficiency, RNA, 11: 914-927.

- Cozzetto, D., Di Matteo, A. and Tramontano, A. 2005 Ten years of prediction. and counting. FEBS Journal, 272, 881-882.

Partecipazione a congressi nazionali 2005:

- Casadio R. -Un sistema sw di annotazione automatica di genomi- Research to Business, Bologna, 28/02/2005-1/03/2005, pag 43
- Casadio R., Fariselli P., Martelli PL. -How many membrane proteins in the human genome?- BITS 2005, Milano, 17-19/3/2005, (Eds. Milanese et al.), pag.20
- Fariselli P, Martelli PL, Casadio R. -The posterior Viterbi: a new decoding algorithm for Hidden Markov Models-, BITS 2005, Milano, 17-19/3/2005, (Eds. Milanese et al.), pag.37 (Oral Presentation)

Partecipazioni a congressi Internazionali 2005

- Casadio R. -Biomolecular recognition and proteomics- Biocomplexity: a paradigm useful for other disciplines? International Italo-Canadian Workshop organized by the Universities of York (Toronto, Canada) and Bologna (Italia), Dipartimento di Patologia Sperimentale, Bologna 13-15/07/2005 (Invited talk)
- Capriotti, E. Fariselli, P. Calabrese, R. Casadio, R. - Predicting protein stability changes from sequence with Support Vector Machines- 4th European Conference on Computational Biology, ECCB/jbi, Madrid (Spagna) 28/09-1/10 2005 pag.54
- Calabrese, R. Capriotti, E. Fariselli, P. Casadio, R. -Prediction of the insurgence of human genetic diseases due to single point protein mutation- 4th European Conference on Computational Biology, ECCB/jbi, Madrid (Spagna) 28/09-1/10 2005 pag.155
- Raimondo D., Giorgetti A., Cozzetto D., and Tramontano A. ***Quality and effectiveness of protein structure comparative models*** Second International Symposium on Mathematical and Computational Biology / Fifth Brazilian Symposium on Mathematical and Computational Biology. Dec. 03–Dec 08, Petropolis, Brazil.
- Raimondo D., Giorgetti A., and Tramontano A. Evaluating the usefulness of protein structure models for molecular replacement, Bioinformatics. ECCB 2005, Madrid, Spain