Nucleic Acids Research, 2005, Vol. 33, Web Server issue ***W347–W351***
doi:10.1093/nar/gki411

EVAcon: a protein contact prediction evaluation service

Osvaldo Graña, Volker A. Eyrich[1], Florencio Pazos, Burkhard Rost[1] and Alfonso Valencia*

Protein Design Group, Centro Nacional de Biotecnologia (CNB-CSIC), Cantoblanco, 28049 Madrid, Spain and [1]CUBIC, Department of Biochemistry and Molecular Biophysics, Columbia University, 650 West 168th Street BB217, New York, NY 10032, USA

Received February 14, 2005; Revised and Accepted March 21, 2005

ABSTRACT

Here we introduce EVAcon, an automated web service that evaluates the performance of contact prediction servers. Currently, EVAcon is monitoring nine servers, four of which are specialized in contact prediction and five are general structure prediction servers. Results are compared for all newly determined experimental structures deposited into PDB (~5–50 per week). EVAcon allows for a precise comparison of the results based on a system of common protein subsets and the commonly accepted evaluation criteria that are also used in the corresponding category of the CASP assessment. EVAcon is a new service added to the functionality of the EVA system for the continuous evaluation of protein structure prediction servers. The new service is accesible from any of the three EVA mirrors: PDG (CNB-CSIC, Madrid) (http://www.pdg.cnb.uam.es/eva/con/index.html); CUBIC (Columbia University, NYC) (http://cubic.bioc.columbia.edu/eva/con/index.html); and Sali Lab (UCSF, San Francisco) (http://eva.compbio.ucsf.edu/~eva/con/index.html).

INTRODUCTION

Contacts between amino acid side chains are fundamental during protein folding, especially those contacts between pairs of residues that preserve the native state of the protein. These contacts are often well separated in the sequence but close in the protein 3D structure.

New methods to predict contacts, based on different approaches, are continuously appearing (1–6). The importance of contact prediction has been pointed out in international experiments, such as the Critical Assessment of protein Structure Prediction [CASP, (7)], where contact prediction is evaluated within the *ab initio* category, and the Critical Assessment of Fully Automated protein Structure Prediction [CAFASP, (8)] where these kind of predictions are evaluated as well.

These experiments revealed that contact prediction techniques are still not of sufficient quality to allow them to be used as the only source of information in protein structure prediction, even though, as demonstrated in the comparison carried out in the context of the CAFASP3 challenge, they can predict contacts of a similar quality to other *ab initio* methods, i.e. fragments-based methods, such as Rosetta (9), threading methods, such as GenTHREADER (10) and FUGUE (11), and comparative modelling servers, such as Pcomb (12).

The need to monitor progress in this field motivated us to develop an automatic contact prediction evaluation service that will allow the scientific community to compare results of these kind of predictions in a systematic way. In particular we are interested in establishing the limits reachable by contact prediction methods in comparison with other prediction approaches.

EVAcon evaluates raw contact predictions produced by the contact specialists and also evaluates, in terms of implicitly predicted 3D contacts, methods for the full reconstruction of protein models (i.e. comparative modelling and threading/fold recognition servers). The results of EVAcon allow a direct comparison between contact predictions produced by the specialists and predictions extrapolated from the 3D models produced by protein structure prediction servers. Overall, the results are potentially informative about the ability of the contact specialists to produce 3D restraints, something that could be used by other *ab initio* methods to produce protein models, or that could be used as additional information to select models generated by other prediction methods (13–15). EVAcon realises the continuous evaluation of protein structure prediction servers, using the facilities of the EVA system for assessing the servers, retrieving the new structures from PDB and building homogeneous datasets for testing the methods (16,17).

*To whom correspondence should be addressed. Tel: +34 91 585 4570; Fax: +34 91 585 4506; Email: valencia@cnb.uam.es

METHOD OVERVIEW

Analysis of contacts

A set of estimations are applied for the evaluation of the contact predictions following the standards commonly accepted in the field and used for the evaluations of CASP and CAFASP (see http://predictioncenter.llnl.gov/ for the description of the evaluation). Two residues are considered to be in contact when the separation between their C-beta atoms (C-alpha for Gly) is $\leqslant$8 Å. There is also the possibility of re-evaluating the predictions with a variable distance threshold (which can be chosen by the user) for the definition of contact. For 3D models without side chains, the positions of the potential side chains are generated with MaxSprout (18).

Predictions from contact specialists are usually received as a list of predicted contacts sorted by a reliability value. We evaluate different levels of prediction, by selecting a number of predicted contact pairs proportional to the protein length. For 3D prediction servers that do not have values of reliability associated to each pairwise distance, we sample an equivalent number of predicted pairs. Evaluations are carried out for a number of pairs corresponding to: $L/10$, $L/5$, $L/2$, L and 2 times the protein length (L).

The predictions are evaluated in terms of accuracy (Acc), improvement over random (Imp), coverage (Cov) and Xd.

$$\mathrm{Acc} = \frac{\text{number of correctly predicted contacts}}{\text{all contacts predicted}}$$

$$\mathrm{Imp} = \frac{\mathrm{Acc}}{\mathrm{AccRand}}$$

The accuracy of a random prediction (AccRand) is obtained assuming that all the possible pairs of the experimental structure in the corresponding ranges of sequence distance separation are potential contacts (C).

$$\mathrm{AccRand} = \frac{\text{number of observed contacts}}{\mathrm{C}}$$

$$\mathrm{Cov} = \frac{\text{number of correctly predicted contacts}}{\text{number of observed contact}}$$

$$\mathrm{Xd} = \sum_{i=1}^{i=15} \frac{(\mathrm{Pip} - \mathrm{Pia})}{(\mathrm{di} \times 15)}$$

Xd is the weighted harmonic average difference between the distance distribution of the predicted contacts and the all-pairs distance distribution.

There are 15 distance bins covering the range from 0 to 60 Å. The sum runs for all the distance bins. di is the distance representing each bin, the upper limit (normalized to 60). Pip is the percentage of predicted pairs whose distance is included in the i bin. Pia is the same for all the pairs. Defined in this way, Xd > 0 indicates the positive cases where the population of predicted contacts' distances is shifted to lower distances (19). For a random prediction, Xd will have a value around zero.

The predictions are also evaluated for several values of sequence deviation from the predicted residue (delta, 14), giving the percentage of correctly predicted contacts within delta residues of the experimental contact. We apply delta values from 0 to 5. This means that a predicted contact between two residues i and j is considered correct if there is a contact observed between any (i − delta, i + delta) and any (j − delta, j + delta).

All these measures are calculated for three different ranges of contacts: short range contacts (at least 6 residues of sequence separation between the residues of the pair), medium range contacts (at least 12 residues of sequence separation between the residues of the pair) and long range contacts (at least 24 residues of sequence separation between the residues of the pair).

Servers participating

A set of prediction servers have been evaluated in EVAcon during the last 30 weeks (Table 1). We would like to invite any other automated servers to join the EVA evaluation system.

Viewing results

The current interface provides an easy navigation for target structures and prediction methods (Figure 1):

(i) EVAcon provides a fast graphical view with directly accessible results that give an overview of the performance of the servers. This initial view is composed of static pages that are regenerated every week. These pages include results (with graphs, tables and raw files for Acc, Imp, Cov and Xd) for each individual server, a summary of the mean accuracy over the weeks of the current year and also a comparison of the accuracies of the servers for the set of all the proteins during the current week.

(ii) The dynamic section provides an in-depth view of the results, suitable for method developers. A query-based system allows a flexible display of results. For example, it makes it possible to define subsets of proteins of interest common to all methods, and to visualize the comparative performances of the servers on these proteins. The results are presented as graphs, tables and raw files.

(iii) EVAcon includes an additional facility for the direct evaluation of contact predictions that is intended for developers in order to evaluate the predictions of their methods according to the EVA criteria. These can be submitted in contact prediction format or as PDB files together with the corresponding experimental structure.

Another possibility of the system is to perform evaluations with a variable distance cutoff for the definition of contact which can be chosen by the user (apart from the default—8 Å between C-beta atoms). This option can be accessed also from the main page of EVAcon, through the link 'Choose a different contact distance and see results for the evaluated targets'. This option is also available for the evaluation of predictions submited by the user [point (iii) above].

It is still too soon to establish a clear ranking of the performance of the servers, due to the small number of weeks that the evaluation has been active. Despite this, the assessment yields very similar results for the contact specialist methods and the 3D structure prediction servers. For some protein targets, contact specialists are performing much better than the 3D prediction methods evaluated by EVAcon. For instance, protein 1pxe chain A (zinc binding domain of the neural zinc finger transcription factor 1), where GPCpred (5)

Table 1. Servers under EVAcon evaluation

Name	Method	Server location and URL	Reference
Contact prediction servers			
PROFcon	Back propagation neural network that combines information from alignments, from one-dimensional predictions, from the region between two contacting residues, and from the average properties of the entire protein chain	CUBIC, Columbia University (http://www.predictprotein.org/submit_profcon.html)	–
GPCpred	Evolutionary algorithm (genetic programming) that selects residues and residue pairs likely to make contacts based solely on local sequence patterns extracted with the help of self-organizing maps	Stockholm Bioinformatics Center (http://sbcweb.pdc.kth.se/cgi-bin/maccallr/gpcpred/submit.pl)	(5)
CMAPpro (5 different versions)	Recurrent neural network implementations of a class of Bayesian networks called generalized input-ouput Hidden Markov Models (GIOHMMs)	Institute for Genomics and Bioinformatics, University of California-Irvine (http://www.ics.uci.edu/~baldig/)	(3)
PDGcon	Contact predictions based on correlated mutations	Protein Design Group, CNB-CSIC (http://www.pdg.cnb.uam.es:8081/pdg_contact_pred.html)	(1)
3D structure prediction servers			
FUGUE	Program for recognizing distant homologues by sequence-structure comparison. It utilizes environment-specific substitution tables and structure-dependent gap penalties, where scores for amino acid matching and insertions/deletions are evaluated depending on the local environment of each amino acid residue in a known structure.	Department of Biochemistry, University of Cambridge (http://www-cryst.bioc.cam.ac.uk/~fugue/prfsearch.html)	(11)
WURST	Threading server with a structural scoring function, sequence profiles and optimized substitution matrices	Center for Bioinformatics, University of Hamburg (http://www.zbh.uni-hamburg.de/wurst/index.php)	(20)
SAM-T99 and SAM-T02	Iterative Hidden Markov Model-based method for finding proteins similar to a target sequence	Biomolecular Engineering, University of California-Santa Cruz (http://www.cse.ucsc.edu/research/compbio/HMM-apps/T99-query.html) (http://www.cse.ucsc.edu/research/compbio/HMM-apps/T02-query.html)	(21,22)
LIBELLULA	Neural network approach that improves the selection of correct folds from fold recognition results given by SAM-T99 and 3D-PSSM (23)	Protein Design Group, CNB-CSIC (http://www.pdg.cnb.uam.es/servers/libellula.html)	(24)

has an accuracy of 50% with a coverage of contacts predicted of 27.3%, twice the accuracy and coverage of FUGUE (11), the best performing threading server for this protein. Another interesting example is the case of the target 1xjh chain A (redox switch domain of the *Escherichia coli* Hsp33), where the best result is produced by CMAPpro_band (3), with a prediction accuracy of 66.7% and a coverage of 20%, and where the best threading prediction, in this case done by WURST (20), is only able to achieve an accuracy of 58.3% and a coverage of 17.5%. These results seem to confirm our initial observations (8), and if confirmed in the following weeks with additional proteins and methods, it will demonstrate the utility of adding contact prediction methods to the set of tools used in fold recognition approaches.

DISCUSSION

The prediction of 3D contacts between amino acids is an active field of research, in which new methods are continually appearing.

We introduce here EVAcon, an automatic system for the continuous evaluation of contact predictions. With the EVAcon evaluation, we want to help developers to test their new methods, and compare their results with others. The dynamic part of the system is particularly interesting for them, since it allows the production of common subsets of proteins adequate for the equilibrated evaluation of the methods. But above all, we think EVAcon can be very useful for biologists as a way to check which servers produce better predictions (or predictions with certain characteristics of coverage, accuracy, ...), based on an independent statistically reliable comparison. One of the reasons for that is that the comparison between methods based on the original publications is difficult since they are based on different datasets.

EVAcon can also be useful to test the potential use of contact prediction as additional restraints incorporated in other methods, for example, as assistance in the selection of correctly folded models. We also think that the system can be of interest for protein modellers searching for the best method from which to obtain additional structural information.

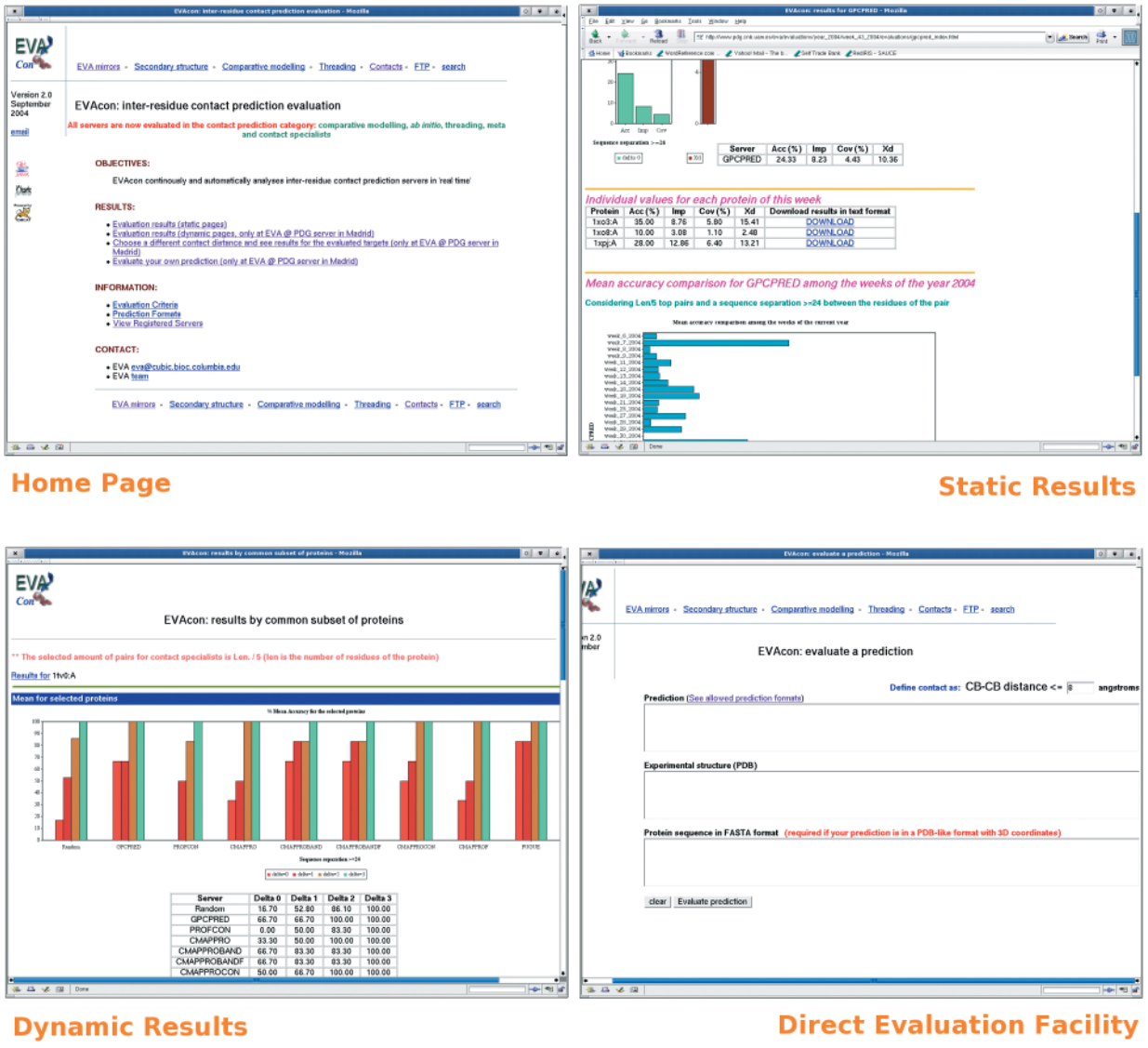

Figure 1. Interfaces of the main EVAcon modules. The system is composed of three different modules: One of them provides a fast view of the results through static pages that are regenerated every week. The second one is composed of a query-based system that allows a flexible display of the results. The third module is a service for the direct evaluation of contact predictions that can be submitted by users and method developers.

It will allow a statistically valid comparison between contact prediction methods and general protein structure prediction methods.

EVAcon performs evaluations once a week, like the other EVA evaluation modules. This continuous assessment produces rankings that are statistically more significant than those at CASP, an experiment that takes place over just 2–3 months every 2 years, over a relatively small number of target sequences. Unlike CASP the predictions are assessed entirely by automatic methods.

We hope to be able to increase the number of servers evaluated by EVAcon and to produce an accurate evaluation of the current methods when the evaluation server has been running for a sufficient number of weeks. Furthermore, we foresee using the results of EVAcon to evaluate the relation between prediction methods and classes of proteins, under the assumption that some contact methods might lead to better predictions for some specific class of proteins or structural motifs.

ACKNOWLEDGEMENTS

We would like to thank members of the Protein Design Group (CNB-CSIC Madrid), and in particular David Juan and Michael L. Tress, for interesting discussions. Members of the EVA team, in particular Marc Marti-Renom, Mallur Madhusudhan and Andrej Sali, (UCSF) for their work in the EVA-CM system, and Ingrid Y.Y. Koh and Megan Restuccia (Columbia U.) for computer assistance. We also thank the protein structure prediction community for making their methods available as public webservers, and in particular to those that allow us to assess their results in the context of EVA. Funding to pay the Open Access publication charges for this article was provided by the contribution of the PDG,

supported in part by a grant from BIOSAPIENS (LSHC-CT-2003-505265), a grant from TEMBLOR (QLRT-2001-00015) and a grant from GENEFUN (LSHG-CT-2004-503567).

Conflict of interest statement. None declared.

REFERENCES

1. Gobel,U., Sander,C., Scheneider,R. and Valencia,A. (1994) Correlated mutations and residue contacts in proteins. *Proteins*, **18**, 309–317.
2. Fariselli,P., Olmea,O., Valencia,A. and Casadio,R. (2001) Prediction of contact maps with neural networks and correlated mutations. *Protein Eng.*, **14**, 835–843.
3. Pollastri,G. and Baldi,P. (2002) Prediction of contact maps by GIOHMMs and recurrent neural networks using lateral propagation from all four cardinal corners. *Bioinformatics*, **18** (Suppl. 1), 62–70.
4. Shao,Y. and Bystroff,C. (2003) Predicting interresidue contacts using templates and pathways. *Proteins*, **53** (Suppl. 6), 497–502.
5. MacCallum,R.M. (2004) Striped sheets and protein contact prediction. *Bioinformatics*, **20** (Suppl. 1), 224–231.
6. Dekker,J.P., Fodor,A., Aldrich,R.W. and Yellen,G. (2004) A perturbation-based method for calculating explicit likelihood of evolutionary co-variance in multiple sequence alignments. *Bioinformatics*, **20**, 1565–1572.
7. Aloy,P., Stark,A., Hadley,C. and Russell,R.B. (2003) Predictions without templates: new folds, secondary structure, and contacts in CASP5. *Proteins*, **53** (Suppl. 6), S436–S456.
8. Eyrich,V.A., Przybylski,D., Koh,I.Y., Graña,O., Pazos,F., Valencia,A. and Rost,B. (2003) CAFASP3 in the spotlight of EVA. *Proteins*, **53** (Suppl. 6), 548–560.
9. Rohl,C.A., Strauss,C.E., Misura,K.M. and Baker,D. (2004) Protein structure prediction using Rosetta. *Methods Enzymol.*, **83**, 66–93.
10. Jones,D.T. (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.*, **287**, 797–815.
11. Shi,J., Blundell,T.L. and Mizuguchi,K. (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.*, **310**, 243–257.
12. Lundström,J., Rychlewski,L., Bujnicki,J. and Elofsson,A. (2001) Pcons: a neural-network-based consensus predictor that improves fold recognition. *Protein Sci.*, **10**, 2354–2362.
13. Olmea,O., Rost,B. and Valencia,A. (1999) Effective use of sequence correlation and conservation in fold recognition. *J. Mol. Biol.*, **295**, 1221–1239.
14. Ortiz,A.R., Kolinski,A., Rotkiewicz,P., Ilkowski,B. and Skolnick,J. (1999) *Ab initio* folding of proteins using restraints derived from evolutionary information. *Proteins*, **37** (Suppl. 3), 177–185.
15. Ortiz,A.R., Kolinski,A. and Skolnick,J. (1998) Fold assembly of small proteins using Monte Carlo simulations driven by restraints derived from multiple sequence alignments. *J. Mol. Biol.*, **277**, 419–448.
16. Eyrich,V.A., Marti-Renom,M.A., Przybylski,D., Madhusudhan,M.S., Fiser,A., Pazos,F., Valencia,A., Sali,A. and Rost,B. (2001) EVA: continuous automatic evaluation of protein structure prediction servers. *Bioinformatics*, **17**, 1242–1243.
17. Koh,I.Y., Eyrich,V.A., Marti-Renom,M.A., Przybylski,D., Madhusudhan,M.S., Eswar,N., Graña,O., Pazos,F., Valencia,A., Sali,A. and Rost,B. (2003) EVA: evaluation of protein structure prediction servers. *Nucleic Acids Res.*, **31**, 3311–3315.
18. Holm,L. and Sander,C. (1991) Database algorithm for generating protein backbone and side-chain co-ordinates from a C alpha trace application to model building and detection of co-ordinate errors. *J. Mol. Biol.*, **218**, 183–194.
19. Pazos,F., Helmer-Citterich,M., Ausiello,G. and Valencia,A. (1997) Correlated mutations contain information about protein–protein interaction. *J. Mol. Biol.*, **271**, 511–523.
20. Torda,A.E., Procter,J.B. and Huber,T. (2004) Wurst: a protein threading server with a structural scoring function, sequence profiles and optimized substitution matrices. *Nucleic Acids Res.*, **32**, W532–W535.
21. Karplus,K., Barrett,C. and Hughey,R. (1998) Hidden Markov models for detecting remote protein homologies. *Bioinformatics*, **14**, 846–856.
22. Karplus,K., Karchin,R., Draper,J., Casper,J., Mandel-Gutfreund,Y., Diekhans,M. and Hughey,R. (2003) Combining local-structure, fold-recognition, and new fold methods for protein structure prediction. *Proteins*, **53** (Suppl. 6), 491–496.
23. Kelley,L.A., MacCallum,R.M. and Sternberg,M.J. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.*, **299**, 499–520.
24. Juan,D., Graña,O., Pazos,F., Fariselli,P., Casadio,R. and Valencia,A. (2003) A neural network approach to evaluate fold recognition results. *Proteins*, **50**, 600–608.