FA1-MS10-P07

A Graphical Front End for the ARP/wARP Software in Macromolecular. Gerrit G. Langer[a], Victor S. Lamzin[a]. [a]*European Molecular Biology Laboratory (EMBL) Hamburg, c/o DESY, Building 25a, Notkestrasse 85, 22603 Hamburg, Germany.*
E-mail: gerrit@embl-hamburg.de

Traditionally crystallographic protein modelling and particularly model completion involves graphical display software for manual user interference. During the last decade, a number of software packages - such as ARP/wARP [1] - emerged that do a large part of the crystallographic map interpretation and model building automatically without much manual work. This has tremendously shortened the time in which a macromolecular model could be built, from months and sometimes years to days or even a few hours. Having done a computer intensive map interpretation step there is generally no elaborate graphical communication or reporting framework, which would indicate how a model has progressed from a visually hardly interpretable density to an almost complete set of polypeptide fragments. As a result, the modelling needs to be viewed, judged and completed where needed elsewhere. At the same time, the recent past has shown that programs that allow semi-automatic modelling coupled with instantaneous graphical display - such as Coot [2] - are becoming very popular.
To bridge the gap between the automation that the ARP/wARP package offers and intuitive communication via graphical components, we have developed a graphical front end to ARP/wARP. It allows to run less time consuming tasks from the ARP/wARP suite, e.g. fitting ligands [3] or secondary structure [4]. The user can choose the input from the graphical items and get results displayed immediately. The approach behind such a graphical tool is to not only provide in total more information than logfiles do on their own, but also at the same time to instantaneously allow the user to choose from minimal intervention to very informed guidance of the software.

[1] Perrakis, A., Morris, R., Lamzin, V.S., *Nat. Struct. Biol.*, **1999**, 6, 458, [2] Emsley, P., Cowtan, K., *Acta Cryst D*, **2004**, 60, 2126, [3] Zwart, P.H., Langer, G.G., Lamzin, V.S., *Acta Cryst D*, **2004**, 60, 2230, [4] Langer, G.G., Cohen, S.X., Perrakis, A., Lamzin, V.S., Nature Protocols, **2008**, 3, 1179.

Keywords: ARP/wARP; molecular graphics; ligand fitting

FA1-MS10-P08

Homology Modeling of VEGF-D as a Basis for Structural and Functional Analysis. Lenita Viitanen[a], Pyry I. Toivanen[b,c], Tiina Nieminen[b,c], Annamari Alitalo[b], Miia Roschier[b], Suvi Jauhiainen[b], Johanna E. Markkanen[b], Olli H. Laitinen[b], Tomi T. Airenne[a], Tiina A. Salminen[a], Mark S. Johnson[a], Kari J. Airenne[b], Seppo Ylä-Herttuala[b]. [a]*Department of Biochemistry and Pharmacy, Åbo Akademi University, Turku, Finland.* [b]*Department of Biotechnology and Molecular Medicine, A.I. Virtanen Institute for Molecular Sciences, Kuopio, Finland.* [c]*Ark Therapeutics Oyj, Kuopio, Finland.*
E-mail: lenita.viitanen@abo.fi

Vascular endothelial growth factors (VEGFs) are key growth factors that induce angiogenesis and lymphangiogenesis during embryogenesis and maintain vasculature during adulthood. They are potential therapeutics for induction of blood vessel formation when normal blood flow is compromised. Most members of the VEGF family exist as covalently bound antiparallel homodimers. The mature form of human VEGF-D (VEGF-D$^{\Delta N\Delta C}$) is, however, predominantly a non-covalent dimer, although the cysteines, forming the intersubunit disulfide bonds in other VEGFs, are conserved in VEGF-D. However, VEGF-D has an extra cysteine at the subunit interface, compared to the other VEGFs. In order to investigate the importance of this cysteine, as well as other residues at the subunit interface, we constructed homology models of VEGF-D, based on the known crystal structures of VEGF-A [1] and the VEGF-F VR-1 [2]. The homology models were used as a basis for the design of site-specific mutations at the subunit interface of VEGF-D. Recombinant VEGF-D$^{\Delta N\Delta C}$ and mutant VEGF-D variants were expressed in cells and their structural and functional properties were studied and compared. Our results show that mutation of specific residues at the VEGF-D dimer interface, affects properties such as dimerization, biological activity, receptor binding affinity and protein stability [3].

[1] Wiesmann C., Fuh G., Christinger H.W., Eigenbrot C., Wells J.A., de Vos A.M., **1997,** *Cell.* 91, 695-704 [2] Suto K., Yamazaki Y., Morita T., Mizuno H., **2004**, *J.Biol.Chem.* 280, 2126-31 [3] Toivanen P.I., Nieminen T., Viitanen L., Alitalo A., Roschier M., Jauhiainen S., Markkanen J.E., Laitinen O.H., Airenne T.T., Salminen T.A., Johnson M.S., Airenne K.J., Ylä-Herttuala S. **2009,** Novel Vascular Endothelial Growth Factor D (VEGF-D) variants with increased biological activity. *Submitted.*

Keywords: homology modelling of proteins; growth factors; mutagenesis

FA1-MS10-P09

Sticky Triangles for Experimental Phasing. Tobias Beck[a], Tim Gruene[a], George M. Sheldrick[a]. [a]*Department of Structural Chemistry, Georg-August-Universität Göttingen, Germany.*
E-mail: tbeck@shelx.uni-ac.gwdg.de

Obtaining phase information for the solution of macromolecular structures is still one of the major challenges in X-ray crystallography. 5-Amino-2,4,6-triiodoisophthalic acid (I3C) and 5-amino-2,4,6-tribromoisophthalic acid (B3C) represent a novel class of compounds that may be used for heavy-atom derivatization of biological macromolecules. I3C and B3C contain an easily recognizable arrangement of three anomalous scatterers (iodine or bromine, respectively) and three functional groups for hydrogen bonding to the protein.
I3C was synthesized and incorporated into test protein crystals using cocrystallization and soaking techniques [1,2]. Single-wavelength anomalous dispersion (SAD) phasing was successfully carried out with in-house data.
The bromine derivative B3C, suitable for multi-wavelength anomalous dispersion (MAD) experiments, was synthesized and employed for experimental phase determination using synchrotron data.
The three halogen atoms present in I3C and B3C form an equilateral triangle that can be easily identified in the heavy atom substructure. A new version of SHELXD that takes this information into account has been developed and is currently being tested.
New functional groups are currently tested to improve the

binding capabilities of I3C and B3C. It was noted earlier that small molecules similar to B3C and I3C can promote crystal growth. A crystallization screen with a set of phasing tools having different functional groups should shed some more light on this issue.

[1] Beck, T. & Sheldrick, G.M. *Acta Crystallogr. Section E* **2008**, 64, o1286. [2] Beck, T., Krasauskas, A., Gruene, T. & Sheldrick, G.M. *Acta Crystallogr. Section D* **2008**, 64, 1179-1182.

Keywords: heavy-atom derivatives; experimental phasing; anomalous dispersion

FA1-MS10-P10

Molecular Replacement Using ab *initio* Polyalanine Models Generated with ROSETTA. Martyn Winn[a], Daniel Rigden[b], Ronan Keegan[a]. [a]*Computational Science and Engineering Department, STFC Daresbury Laboratory, UK.* [b]*School of Biological Sciences, University of Liverpool, UK.*
E-mail: martyn.winn@stfc.ac.uk

The success of the molecular-replacement method for solving protein structures from experimental diffraction data depends on the availability of a suitable search model. Typically, this is derived from a previously solved structure, sometimes by homology modelling. Recently, Baker, Read and coworkers have demonstrated a successful molecular-replacement case based on an ab initio model generated by ROSETTA [1]. We have looked at a number of additional test cases in which ab initio models generated using modest computational resources give correct molecular-replacement solutions [2]. Polyalanine models were generated using ROSETTA v.2.1.2, while side chains were added using SCWRL. Molecular replacement trials were performed with PHASER. Success of MR is judged by comparison with the deposited structure, as well as automatic model re-building and refinement with ARP/wARP. Unsuccessful cases are also reported for comparison and the factors influencing the success of this route to structure solution are discussed.

[1] Qian et al., **2007**, *Nature (London)*, 450, 259–264. [2] D.J Rigden, R.M Keegan and M.D Winn, **2008**, *Acta Cryst.* D64 1288-1291.

Keywords: protein crystallography; molecular replacement; *ab-initio* structure determination

FA1-MS10-P11

Multiple Open Reading Frames GARP Content and a 32 Letter Genetic Code. William L. Duax[a], Robert Huether[a], David Dziak[a]. [a]*Hauptman-Woodward Medical Research Institute, Buffalo, NY 14203.*
Email: duax@hwi.buffalo.edu

We have unequivocal evidence that a 32 letter genetic code was a precursor to the "standard" genetic code, and that ribosomal proteins, tRNA sythetases and many species of GC-rich bacteria arose when only codons ending in G and C had sense definitions. The genome of *A. dehalogenans* (Adhal) has 4,346 coding sequences composed of 1,518,523 codons, a GC content of 75% and GC occupancy of the third base position of **97%.** Over 31% of the proteins in Adhal are annotated as hypothetical genes. Over 94% of them have multiple open reading frames (MORFs) and 25% have Val or Leu residues as start codes. The presence of MORFs, ambiguity of start and stop code identification and the questionable assumptions that (1) the longest possible open reading frame (ORF) is the probable protein coding frame and (2) the "standard" genetic code is appropriate to decoding this genome has resulted in misidentification of hundreds of nonsense sequences as proteins and errors in the true lengths of hundreds of functionally annotated proteins. We demonstrate that the combined frequency of appearance of four amino acids (Gly, Ala, Arg, and Pro-**GARP**) whose codons are composed of only guanine (G) and cytosine (C) nucleotides can be used to distinguish real proteins from nonsense sequences. The GARP content of 40,000 proteins in the Protein Data Bank ranges from 10% to 45% with an average of **21%.** The average GARP content of the putative proteins in Adhal is **44%.**Over 2000 putative proteins, including 830 ORFans claimed to be unique to Adhal, have GARP content of greater than 45% and most of these are probably nonsense. The rare appearance or complete absence of twelve codons ending in A or T from the genes of 55 ribosomal proteins and 20 tRNA synthetases indicate that these nucleotide triples are nonsense codes in Adhal. Most of these codons do not have fully cognate tRNAs. Tracking the location of the rarely used codons in functionally annotated proteins we find that over 50% of them are within 10% of the total protein length from the C- or N-terminus. Errors in genome assembly, coding frame selection and start and stop code identifications, GARP content analysis and the identification of probable nonsense codons allows us to identify which hypothetical proteins are nonsense and which functionally annotated proteins are partially nonsense. Supported by a gift from Roy Carver, Jr.

Keywords: bioinformatics; genome analysis; protein identification

FA1-MS10-P12

DIBER: Protein, DNA, or Both? Grzegorz Chojnowski[a,b], Matthias Bochtler[a,c]. [a]*International Institute of Molecular and Cell Biology, Warsaw, Poland.* [b]*Institute of Experimental Physics, University of Warsaw, Warsaw, Poland.* [c]*Schools of Chemistry and Biosciences, Cardiff University, United Kingdom.*
E-mail: Mbochtler@iimcb.gov.pl, BochtlerM@Cardiff.ac.uk

In co-crystallization experiments with protein and DNA, it is not always clear whether the crystal contains the complex or one component alone. The CCP4 compatible program DIBER was written to make a prediction before phase information becomes available. Our method relies on the detection of characteristic fiber diffraction peaks in spite of their sampling by the reciprocal lattice of the 3D crystal. In order to detect the presence of DNA, we search for characteristic groups of strong, neighbouring reflections that are attributable to in-phase scattering of the B-DNA bases in a thin shell around 3.4 Å resolution. Although our method uses information only from a small subset of reciprocal space, its predictions are more reliable than molecular replacement score based predictions.

Keywords: molecular replacement; DNA and protein crystallography; DNA-protein complexes