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