

# Wavelet compression of three-dimensional maps gsk GlaxoSmithKline



## **Richard Martin<sup>1</sup>**, Eleanor Gardiner<sup>1</sup>, Val Gillet<sup>1</sup> and Stefan Senger<sup>2</sup>

Example

(DeLano, 2006).

A map reconstructed using varying amounts of

data (orange), overlaid on the original (blue). The

ligand is *1uom* from the ESR1 inhibitor set, whose map

has been calculated using the GRID program

(Goodford, 1985) and the dry probe. Surface calculated

at an energy threshold of -5 using PyMOL software

0.01% of the coefficients -

obvious artefacts (e.g. left)

severe distortion and

#### Introduction

Three-dimensional maps can convey a great deal of information about a molecule, highlighting regions of favourable interaction, hydrophobicity etc, which cannot be observed in less intricate twodimensional representations. However, if these maps are large and finely sampled they can be costly to store and analyse. Therefore, by compressing the data significantly these maps could be stored more efficiently, and allow for a more rapid comparison.

The wavelet transform (Sundling et al., 2006) is a recent adaptation of the Fourier method which seeks to bypass the limitations imposed by the sinusoidal Fourier basis. Wavelets have been shown to be capable of achieving high compression ratios whilst maintaining accuracy of the restored data; they have seen application across a range of scientific disciplines, and are in common usage in the JPEG2000 image standard (Skodras et al., 2001).

Here we present the effects of wavelet compression and reconstruction upon similarity rankings and image quality within three different data sets of pre-aligned ligands.

#### Method

Compression was achieved by repeatedly applying the *Daubechies* 4-tap wavelet transform upon the data alternately in each dimension, and setting the lowest n% of the resultant (non-zero) coefficients to zero. We applied this method to three data sets (Chen et al., 2006): CDK2 (57), HIV (18) and ESR1 (13), using maps generated by four GRID probes: hydrophobic (dry), water, sp3 NH amine cation (N1+), and sp2 phenolate oxygen (O-), zeropadded to resolution 64<sup>3</sup>. The maps were wavelet transformed, compressed at a range of ratios *n* and reconstructed.

The distance between two maps was defined as the sum of the absolute difference between every voxel (volume element) in the query map and its counterpart in the target. Using this definition similarity rankings for each compression ratio and data set were obtained using each ligand in turn as query and compared with the original data using Spearman's Rank Correlation Coefficient (SRCC).



1% of the coefficients general shape is preserved, but with loss of quality



Figure 1, The mean correlation between similarity rankings of compressed and original data.

### Results

Reconstructing the maps using 10% of the data or more results in maps near identical to the originals, demonstrating that a compression ratio of up to 90% allows for highly accurate data reconstruction.

The mean correlation between original rankings and those obtained using 1% of the data is consistently greater than 0.9, and greater than 0.8 using only 0.1% (see Figure 1). Hence even 99% compression (perhaps even 99.9%) has little impact on similarity, however beyond this ratio the visual quality of the map begins to be compromised considerably (see Examples).

Surprisingly, the HIV set produced correlations over 0.7 at 99.99% compression with the exception of the dry probe, which was typically found to produce the weakest correlations (see Figure 1).

#### Conclusions

Wavelet transforms provide a rapid and effective method for the efficient electronic storage of compressed volumetric data, and the similarity rankings between reconstructed maps using the Daubechies 4-tap wavelet have been shown to correlate strongly with those of the original data even at high compression ratios.

#### Some future work

•Investigate efficacy of different wavelet bases, e.g. Haar, Daubechies 12-/20-tap, Coiflets, Symmlets. •Similarity comparison of wavelet compressed maps without performing reconstruction.

Rotation of compressed maps for alignment.

#### References

- Chen, Q., Higgs, R.E. & Vieth, M. (2006). Journal of Chemical Information and Modelling, 46 (5), 1996-2002
- DeLano, W.L. (2006) "The PyMOL Molecular Graphics System". DeLano Scientific LLC, San Carlos, CA, USA, http://www.pymol.org
- Goodford, P.J. (1985). Journal of Medicinal Chemistry, 28 (7), 849-857. Skodras, A., Christopoulos, C. & Ebrahimi, T. (2001), IEEE Signal Processing Magazine, 18, 36-58.
- Sundling, C.M., Sukumar, N., Zhang, H., Embrechts, M.J. & Breneman C.M. (2006). Reviews in Computational Chemistry, 22.

#### Affiliations and acknowledgements

- Dept. of Information Studies, University of Sheffield, Western Bank, Sheffield S10 2TN, UK
- GlavoSmithKline Medicines Research Centre Computational and Structural Chemistry, Stevenage, Herts SG1 2NY, UK.
- software licensed Discovery GRID is from Molecular (http://www.moldiscovery.com/)
- This project is funded by BBSRC and GlaxoSmithKline