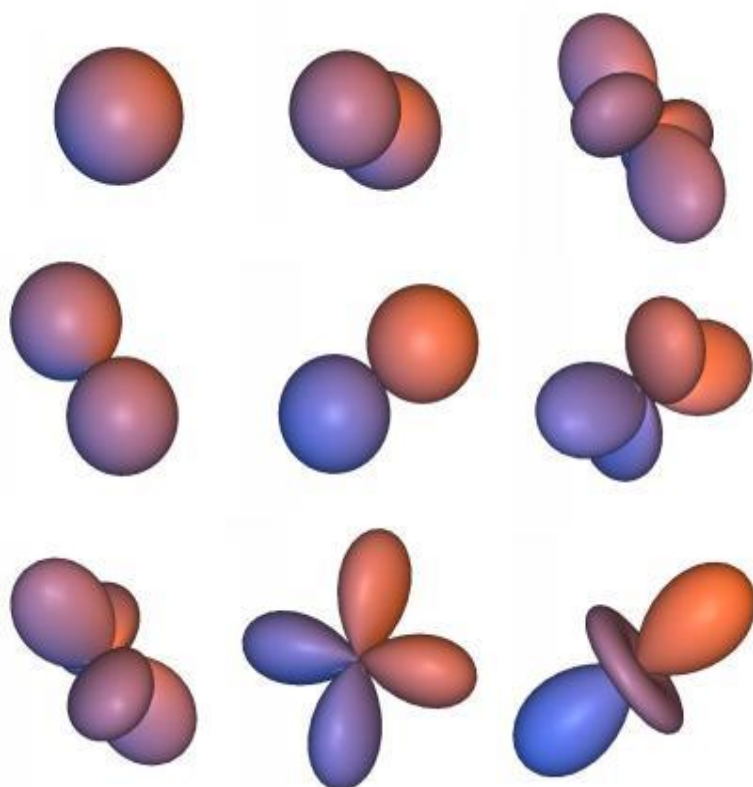


## Hex Gallery

This page (written by Dave Ritchie) contains some example images from *Hex*. Mostly the pictures were created directly as screen-shot JPEG files, although several are composite screen-shots put together with [The Gimp](#).

### Spherical Harmonics



Spherical harmonics are special functions of the spherical coordinates, theta and phi. These functions can be thought of as "standing waves on a sphere". They are characterised by two quantum numbers, L and M, which together determine the number and spatial arrangement of nodes in each function. The image on the left shows some of these functions.  $Y(L=0,M=0)$  is just a sphere (top left), and  $Y(L=2,M=0)$  is at bottom right. The figures above the diagonal have  $M>0$ , and those below have  $M<0$ . In general, there are  $2L+1$  allowed values of M for a given L. Adding another row and column to

the image would give 7 functions with  $L=3$ , etc. To use spherical harmonics to represent molecular surface shapes, we need to keep going up to about  $L=25$ , or higher...

### Spherical Harmonic Molecular Surfaces

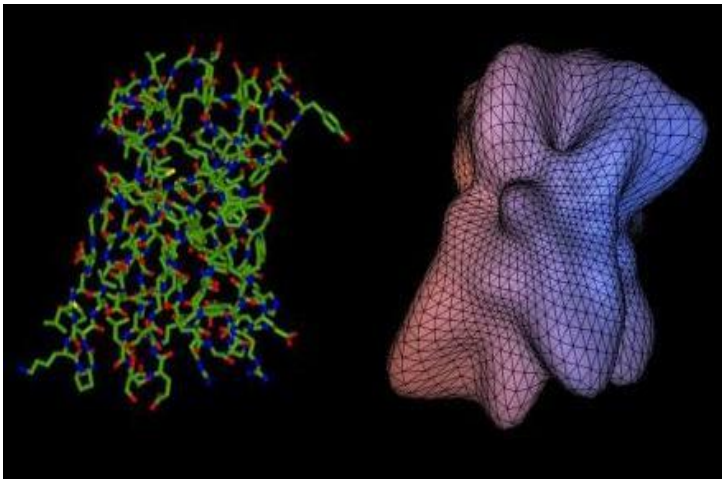
A surface in 3D space can be approximated by encoding the radial distance of surface points from the origin as a sum of spherical harmonic functions up to some limiting value of L:

$$r(\theta, \phi) = \sum_{l=0}^L \sum_{m=-l}^l a_{lm} y_{lm}(\theta, \phi) \quad (1)$$

where the choice of upper limit  $L$  determines the accuracy of the representation. Each coefficient,  $a_{lm}$ , can be determined by an integral:

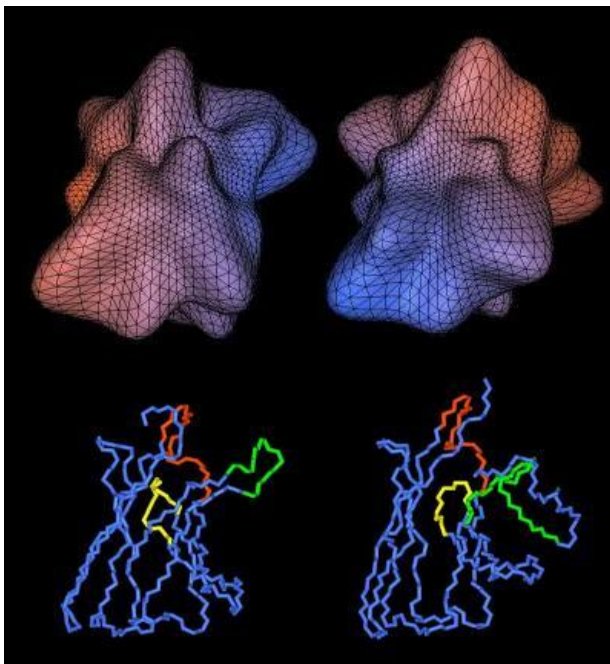
$$a_{lm} = \int_0^{2\pi} \int_0^\pi r(\theta, \phi) y_{lm}(\theta, \phi) \sin \theta d\theta d\phi \quad (2)$$

We have developed a fast way to estimate such integrals using a sampling scheme based on an icosahedral tessellation of the unit sphere.



This image represents the resulting spherical harmonic surface approximation of the molecular surface of an antibody VH domain, using harmonics up to  $L=14$ . Except for the region near the phenylalanine side chain at top right, the surface approximation is a good estimate of the true molecular surface. The error here is a result of the surface being doubly-valued with respect to radial projections from the chosen origin.

## Molecular Shape Comparison using Spherical Harmonics

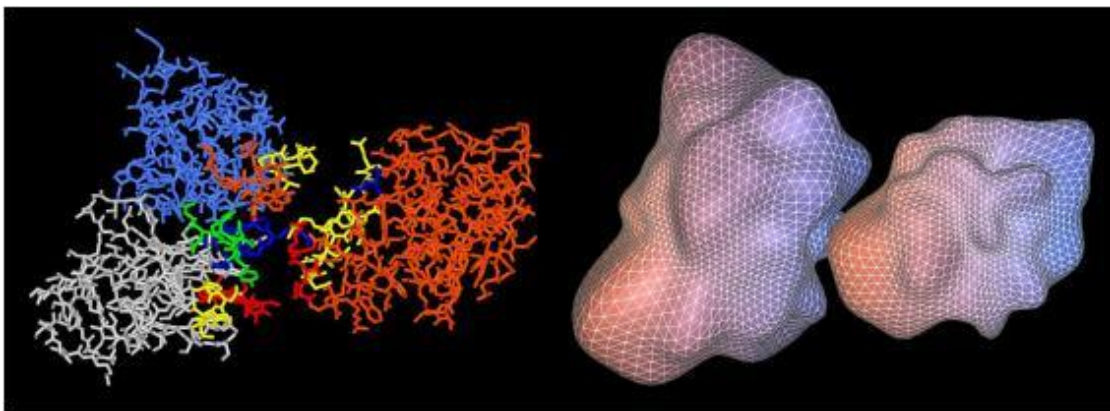


Pairs of spherical harmonic surfaces can be compared by minimising the "distance", between their respective harmonic expansion coefficients [1]. This is very much like a conventional 3D Fourier correlation, except that here, the correlation is a function of the 3 Euler rotation angles ( $\alpha, \beta, \gamma$ ), instead of the more usual Cartesian translations. The upper images show the minimum surface distance orientation of two antibody VH domains (HyHel-5 and KOL), separated by 36 Angstrom in the x-axis. The lower images show the backbone traces of the two molecules in the same orientation. We have effectively matched the secondary structures using only geometric surface information.

## Protein Docking using Spherical Harmonics?

So can spherical harmonics be used to calculate how a pair of proteins might fit together, or "dock"?

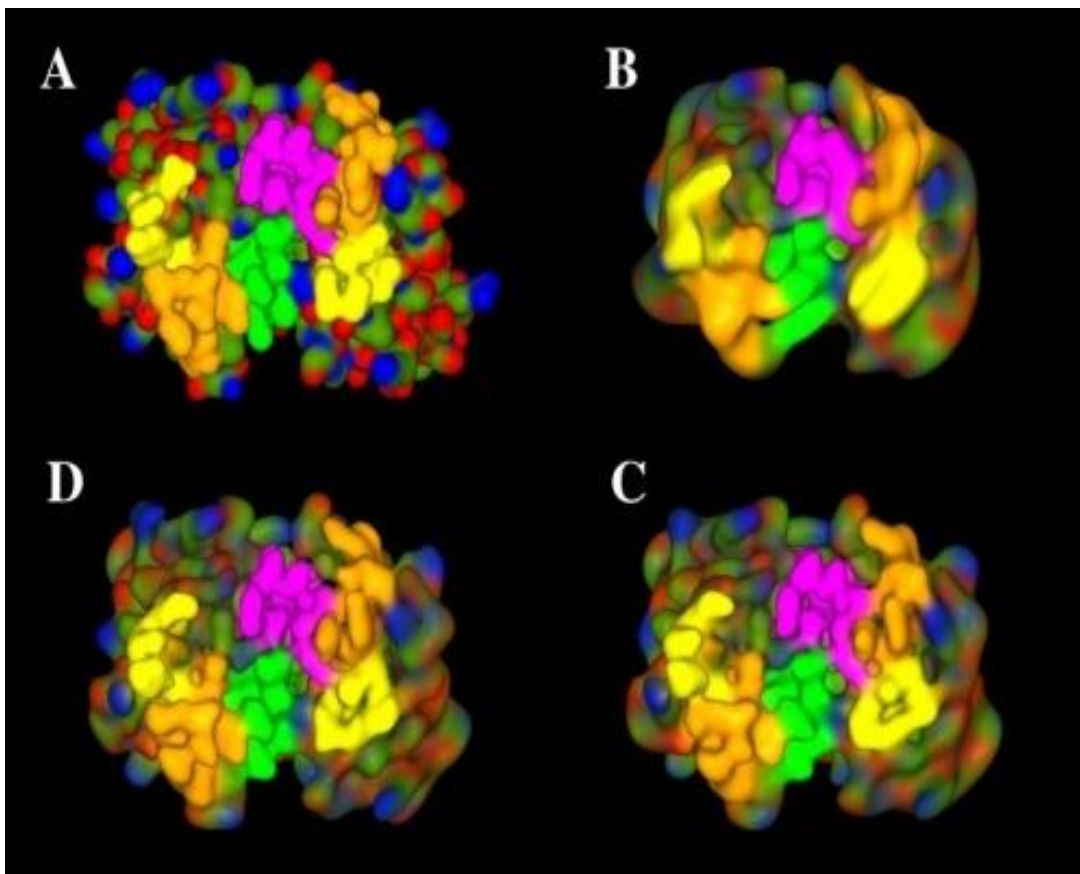
The image below shows the antibody HyHel-5 (left) and its lysozyme antigen (right) in the docked orientation, but separated by 5 Angstroms to give a better view of the interface. The  $L=12$  harmonic surfaces to the right clearly show a high degree of shape complementarity at the interface. Arguably, if two proteins are to fit together at the detailed atomic level, they would be expected to show a high degree of surface complementarity even using these low resolution representations. In other words, we should be able to use low resolution surface shape representations to predict feasible binding orientations for proteins.



However, compared to superposing similar shapes it seems much harder, mathematically, to calculate how to contrapose complementary surface shapes. The main reasons are that we now need to use two distinct coordinate systems, one for each protein, and we would need to devise a way to calculate the "distance" between just those parts of the surfaces that come into contact.

## Docking using Polar Fourier Correlations

One way to avoid these difficulties is to use special orthogonal radial functions, along with the spherical harmonics, to represent each protein using a pair of 3D *density* functions instead of explicit 2D surfaces. These aren't so easy to display graphically, but basically the first density function is one inside a protein's surface and zero outside. The second density function is zero everywhere except within a small surface "skin" region between the van der Waals surface and the solvent accessible surface. Good docking orientations can then be found by maximising the degree of overlap between the surface skin of one protein with the interior density of the other, and vice-versa. Steric clashes can be penalised by subtracting an interior-interior overlap term from the scoring function. To represent proteins sufficiently accurately for a docking calculation, we find its necessary to use radial functions up to order  $N=25$  or  $N=30$ , where  $N$  is the highest polynomial power in  $r$ , along with all spherical harmonics up to order  $L=N-1$ . Essentially, this means we represent the global shape of a protein using just two vectors of Fourier expansion coefficients.

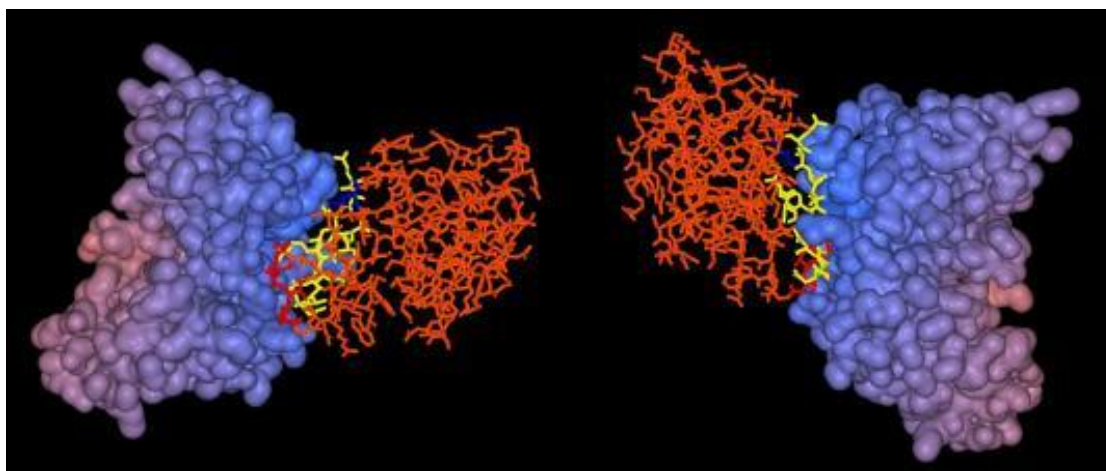


This image shows some views of an antibody Fv fragment represented as 3D polar Fourier density expansions to different orders. A: original atom-Gaussian density representation; B: polar Fourier shape density reconstructed to  $N=16$  (1,496 coeffs); C:  $N=25$  (5,525 coeffs); D:  $N=30$  (9,455 coeffs). Each density function was contoured using a density threshold which corresponds approximately to the van der Waals radii of the atoms in (A). The surfaces are drawn as tiny connected triangles and are colour-coded by atom type according to a Gaussian mixing rule (based on the distance of each triangle vertex to nearby atom positions). Hence sharp colours correspond to an accurate surface reconstruction. Note the relatively low resolution towards the edges of the molecule, even for high order expansions. This is due to the Gaussian decay factor in the radial basis functions, which dominates at large distances from the origin.

By borrowing quite a bit of mathematics from quantum mechanics, we can calculate how to rotate *and* translate these density functions by transforming only the original expansion coefficients. The overlap at each trial orientation can then be calculated simply by multiplying pairs of expansion coefficient vectors. This gives a full 6D docking correlation algorithm, which we believe is highly competitive compared to conventional grid-based Fourier correlation methods. These ideas have been implemented in the docking program *Hex*[2] which may be downloaded from the [Hex Home Page](#). In addition to the above shape correlation, *Hex* also supports a similar Fourier-based calculation of *electrostatic interaction energies*, and (as of version 3.0) a molecular mechanics force-field method of refining candidate docking orientations.



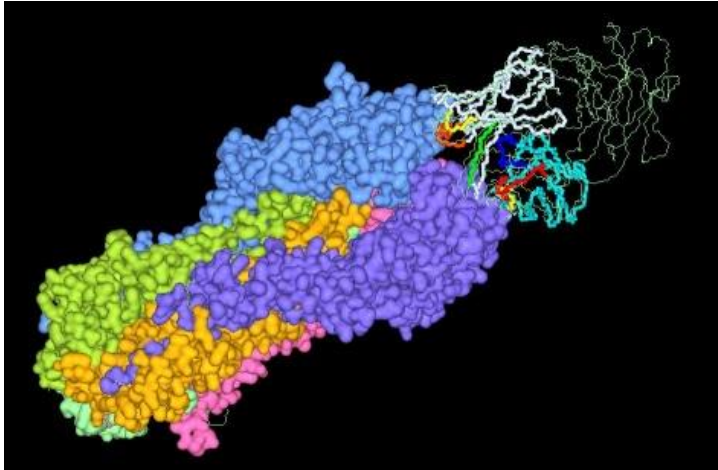
The image below shows two views of the HyHel-5/lysozyme complex, as docked by *Hex*. This is a relatively easy complex to dock, as coordinates from the bound form of the antibody had to be used, and because lysozyme changes relatively little on complexation. Results with this and similar complexes, indicate that rigid-body docking can do a remarkably good job of predicting how macromolecules might interact, provided the degree of conformational change is small and at least some information about the binding site(s) is available.



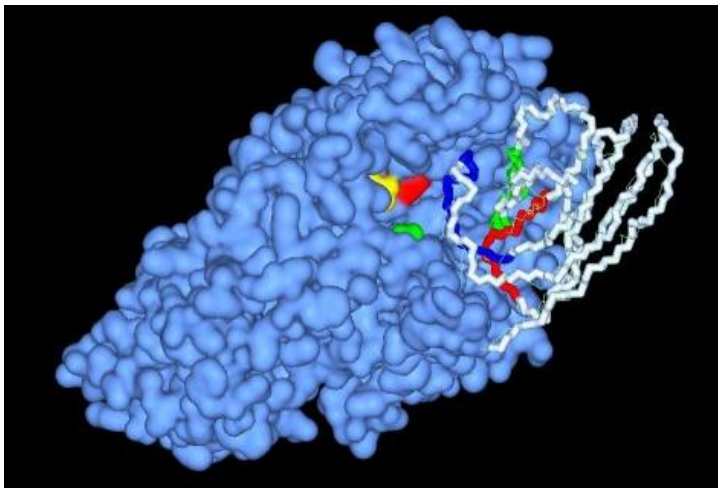
For those interested in the technicalities, the above docking was calculated starting from a randomly oriented lysozyme (1LZA) placed over the antibody hypervariable loops (3HFL). Roughly 72 million trial orientations were generated by "spinning" the molecules (5 rotational degrees of freedom), and by varying the intermolecular distance in 0.5 Angstrom steps (1 degree of freedom). These orientations were scored using shape correlations to  $N=16$ , and the best 20,000 orientations were then passed to a high resolution scoring stage which used both shape and electrostatic correlations at  $N=30$  (this stage found 5 orientations within the top 100 which were within 3 Angstroms RMS of the "right answer"). The best 1,000 of the  $N=30$  orientations were then refined using a "soft" molecular mechanics rigid body minimisation algorithm. After minimisation, the lowest energy solution (the one shown) had a main-chain deviation of 1.25 Angstroms RMS from the "right answer" (a least-squares fit of the lysozyme onto the complex). The Fourier correlation part of the calculation took about 13 minutes using a dual 800MHz Xeon Pentium III system with 512Mb RAM, and evaluated an average of 60,000 orientations per second (the peak rate at  $N=16$  was 375,000 orientations per second). Energy minimisation took about 6 minutes, refining about 2.5 orientations per second. In the "spinning" process, the lysozyme was constrained to remain near the hypervariable loops. A blind (fully unconstrained) global search would increase the execution time by roughly a factor of 7.

### **Selected CAPRI Results**

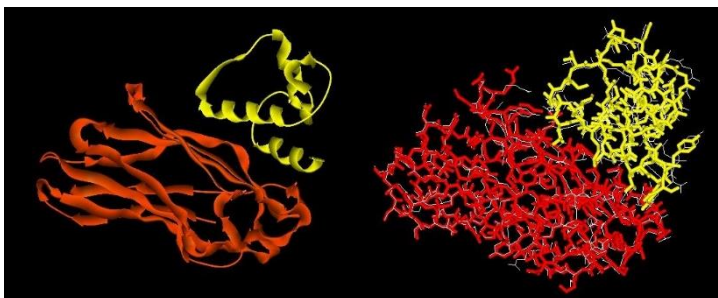
The images below show the good docking predictions obtained with *Hex* for [CAPRI](#) targets 3, 6 and 12.



CAPRI Target 3:  
Hemagglutinin/Antibody-HC63.  
This is solution no. 4 from *Hex*. It has the antibody docked onto the A and C domains of the large hemagglutinin. The deviation between the predicted and actual antibody CA atom positions (Fv atoms only) from the complex (PDB code 1KEN) was 7.4A RMS.



CAPRI Target 6: alpha-amylase/antibody-AMB9. This is solution no. 5 from *Hex*. The deviation between the predicted and actual AMD9 antibody CA atom positions from the complex (PDB code 1KXQ) was 2.2A RMS.

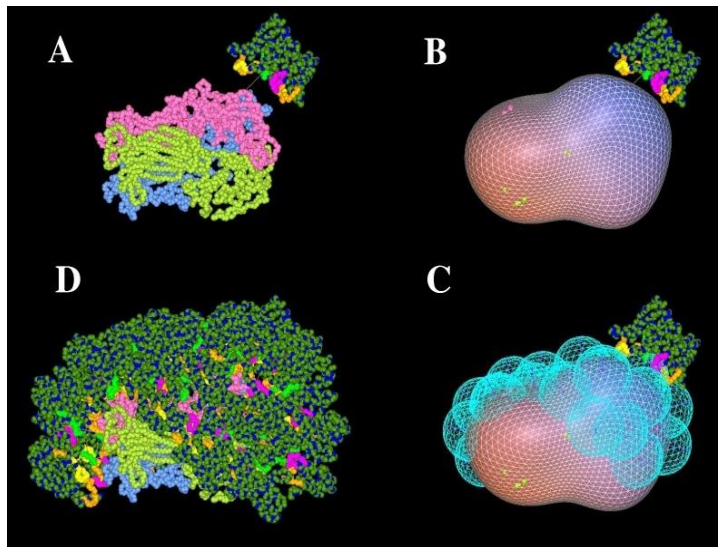


CAPRI Target 12:  
cohesin/dockerin (an unbound/unbound target). This is solution no. 6 from *Hex* shown in both ribbon and licorice drawing modes (cohesin in red, dockerin in yellow, complex in white). The deviation between the predicted and actual dockerin CA atom positions from the complex (PDB code 1OHZ) was 2.4A RMS after superposing the cohesin CA atoms of the prediction onto the complex (3.3A RMS deviation).

### Docking Very Large Molecules

As mentioned above, the radial basis functions used in *Hex* fall off rapidly beyond a certain distance (around 30 Angstroms) from the origin. Hence very

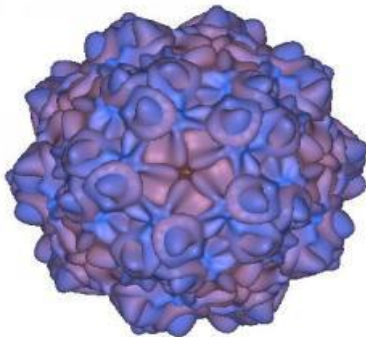
large proteins (e.g. CAPRI Target 6, above) have to be docked using multiple localised docking runs, as illustrated below.



This picture illustrates how *Hex* docks very large proteins (using the antibody-MCV/VP6 complex from CAPRI Target 2 as an example). A: The antibody hypervariable loops are initially oriented towards the large VP6 trimer; B: A low-resolution spherical harmonic surface is calculated for the VP6 trimer (2252 triangles). C: The VP6 surfaces is smothered with spheres, one per surface triangle,

and the spheres are iteratively culled until some (low) overlap threshold is reached; D: The remaining spheres are used to generate starting orientations for the antibody (here restricted to the VP6 C domain).

### Fun Stuff



This is an image of a canine viral surface coat protein (PDB code 1IJS) calculated as a low resolution spherical harmonic surface of one of the protein monomers, but redrawn 60 times to show the icosahedral symmetry of the complete capsid.