Journal of Computer-Aided Molecular Design **18:** 101–118, 2004.

Gaussian mapping of chemical fragments in ligand binding sites

Kun Wang[1,3], Marta Murcia[1,2], Pere Constans[1,4], Carlos Pérez[1,5] & Angel R. Ortiz[1,2,*]
[1]*Department of Physiology & Biophysics, Mount Sinai School of Medicine, One Gustave Levy Pl., Box 1218, New York, NY 10029, USA;* [2]*Bioinformatics Unit, Centro de Biología Molecular 'Severo Ochoa', Universidad Autónoma de Madrid, Cantoblanco, E-28949 Madrid, Spain;* [3]*Current address: Laboratory for Molecular Modeling, Division of Medicinal Chemistry and Natural Products, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7360, USA;* [4]*Current address: Department of Chemistry, Rice University, Houston, TX 77005-1892, USA;* [5]*Current address: PharmaMar S.A., Calera 3, Tres Cantos, E-28760 Madrid, Spain*

Received 2 September 2003; accepted in revised form 5 March 2004

Key words: pharmacophore determination, protein–ligand docking, structure-based drug design, virtual screening

Summary

We present a new approach to automatically define a quasi-optimal minimal set of pharmacophoric points mapping the interaction properties of a user-defined ligand binding site. The method is based on a fitting algorithm where a grid of sampled interaction energies of the target protein with small chemical fragments in the binding site is approximated by a linear expansion of Gaussian functions. A heuristic approximation selects from this expansion the smallest possible set of Gaussians required to describe the interaction properties of the binding site within a prespecified accuracy. We have evaluated the performance of the approach by comparing the computed Gaussians with the positions of aromatic sites found in experimental protein–ligand complexes. For a set of 53 complexes, good correspondence is found in general. At a 95% significance level, ∼65% of the predicted interaction points have an aromatic binding site within 1.5 Å. We then studied the utility of these points in docking using the program DOCK. Short docking times, with an average of ∼0.18 s per conformer, are obtained, while retaining, both for rigid and flexible docking, the ability to sample native-like binding modes for the ligand. An average 4–5-fold speed-up in docking times and a similar success rate is estimated with respect to the standard DOCK protocol.

Abbreviations: RMSD – root mean square deviation; ASA – Atomic Shell Approximation; LSF – Least-Squares Fitting; 3D – three-dimensional; VDW – Van der Waals.

Introduction

For large-scale docking in virtual screening applications, the ligand binding site is often preprocessed to a set of interaction points, representing in a simplified way the spatial neighborhoods likely to be occupied by the atoms of high-affinity binders. These transformations are convenient, as they allow the use of fast algorithms, such as those based on graph-matching techniques [1]. Preprocessing is inspired by the lock and key concept of ligand–receptor interactions, and follows typically one of these two approaches: either the space adjacent to the binding pocket is chemically mapped with appropriate probes and interaction points are defined based on energy or statistical criteria, or a negative image of the protein binding region is generated and approximated with a set of spheres mimicking shape complementarity in a geometrical sense. For the first technique, various potential energy functions and knowledge-based approaches are available. Molecular mechanics programs such as GRID [2–6], MCSS [7–9], or the Autogrid program within Autodock [10], among others, use atom probes or functional groups to estimate the interaction energy of the probe with the site, allowing the identification of energetically favorable positions for various lig-

[*] To whom correspondence should be addressed. E-mail: aro@cbm.uam.es; Fax: 34-91-497-4799

and functionalities. X-SITE [11], SuperStar [12] or DrugScore [13], on the other hand, use knowledge-based potentials derived from the analysis of crystals to find favorable positions for chemical groups. When shape complementarity is sought, a popular technique is to compute the Connolly surface [14, 15] of the macromolecular cavity and then proceed to find a set of spheres able to fill the volume of the cavity. The sphere centers in this 'negative image' of the site represent putative ligand atom positions. This is the technique used within the SPHGEN program in the DOCK package [16–18].

When using interaction points, the time required for a docking calculation increases with the number of points to match. Optimally designed sets of interaction points can therefore help to increase the performance of virtual screening computations. By *optimal design* we refer to the smaller set of points carrying maximal information about the interaction properties of the pocket. For example, by the method used in their construction, negative image spheres carry limited energetic and chemical information beyond their primary geometric information. Furthermore, for a given enclosed volume, it is possible to find redundant sets of spheres; sets filling the cavity within similar occupancies but having different numbers of elements and different configurations. Conclusive demonstration of the importance of optimally placing interaction points has been provided by Zavosdszsky et al. [19], who have recently shown that improving the representation of hydrogen-bonding and hydrophobic interaction points by a knowledge-based approach improves the quality of docking and the docking scores of known ligands. Similarly, Joseph-McCarthy and Alvarez [20] have shown with DOCK that biasing the search using points located in local energy minima allows for more effective sampling of the target site.

The problem is how to optimally select these points. The rugged energy landscape of the protein–ligand interaction makes it difficult to automatically select them on the basis of the local properties of the energy distribution. Compression algorithms able to summarize the global energy density over the binding site are required. Research on such algorithms is receiving increasing attention. Nissink et al. [21] have presented anisotropic Gaussian-type descriptors to approximate IsoStar [22] propensity distributions. Rantanen et al. [23], on the other hand, have modeled propensity data with Gaussian mixtures within a Bayesian framework. So far, more emphasis has been placed on the fitting properties than on the number of fitting variables. Here, we address the problem of transforming a grid energy map to a set of interaction points, specifically seeking a minimum number of variables able to fit the scatterplot within a prespecified error. We first map the binding of molecular fragments to ligand binding sites [24] by computing a grid-based potential similar to the ones produced by GRID or MCSS, and then proceed to transform the map to a reduced set of Gaussian functions which contain the interaction centers and associated radii we seek to deduce. In what follows we provide a description of the methodology and describe the main computational experiments we have used to validate it.

Methods

Computation of the molecular electron densities of small organic molecules

Our initial tests of fitting performance were carried out using molecular electronic densities, since these afford an easy and intuitive evaluation of the performance of the methodology. For the computation of the molecular electron density in isolated, small molecules, we used a promolecular representation according to the Atomic Shell Approximation (ASA) [25, 26]. Within this approximation, densities are presented with the simple general form of Equation 1, where shell occupation n_i is constrained to positive values.

$$f(\mathbf{x}) = \sum_{a} \sum_{i \in a} n_i c_i e^{-\frac{(\mathbf{R}_a - \mathbf{x})^2}{2\sigma^2}} \tag{1}$$

Grid description of the binding space and energy computation

We use a fragment positioning method to determine energetically favorable positions for various chemical fragments. First, appropriate chemical fragments are docked in the binding site. The protein–ligand intermolecular energy is pre-computed using an underlying 3D grid of 1 Å [27]. The potential we have used consists of non-bonded interaction energies (in kcal/mol) computed with the AMBER force field using an all-atom model [28]:

$$E_{MM} = \sum_{i}^{Nprot} \sum_{j}^{Nlig} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + 332 \frac{q_i q_j}{\varepsilon r_{ij}} \right] \tag{2}$$

A_{ij} and B_{ij} represent the van der Waals (VDW) parameters of the atom types to which atoms i and j belong, q_i and q_j are the partial charges (in electron units) of atoms i and j, respectively, and r_{ij} is the distance (in Å) between them. A dielectric constant of $\varepsilon = 4$ or $\varepsilon = r_{ij}$ was used to scale down the electrostatic term. For each fragment, a complete enumeration of all possible orientations of the rigid molecular fragments in the active site of the rigid protein is carried out. The molecule is translated within the box using a grid spacing of 1.0 Å, and at each grid point, a complete sampling of the rotational space is achieved by computing all non-degenerate sets of Euler angles obtained with a resolution of 27° [29]. At each rotational and translational point, the fragment is subjected to a rigid body off-lattice energy minimization using the SIMPLEX algorithm from Nelder and Mead [30]. Then, the minimum energy found for the fragment is stored for that grid point, providing the discrete function $f(\mathbf{x})$ to be subjected to Gaussian fitting in order to define the interaction points for that particular fragment.

The computed grid contains those regions with favorable interactions for a given probe with the binding site. From here, we want to obtain a simpler description in the form of a set of Gaussian centers with associated radii accounting for the same information. We want the set of interaction points deduced to be optimal or close to optimal, i.e., the minimal set of Gaussians able to account for the energy surface stored in the grid within a prespecified error. In the particular case of the tests reported in this paper (Tables 2 and 3), the grid has been defined from the X-ray structure of the complexes, defining a box enclosing the bound ligand and imposing a distance of at least 5 Å between any atom of the ligand and the box edges.

Transformation of the grid energies to reduced Gaussian expansions

Once we have an adequate representation of the negative image of the receptor binding site in the form of a discrete function $f(\mathbf{x})$, evaluated in a suitable region of the space $\mathbf{x} \in \Re^d$, we focus our attention to obtaining a fitting algorithm capable of minimizing the required size of the Gaussian expansions, while keeping the fitting accuracy within a given error. The problem can be stated formally as follows: given a discrete representation of a function $f(\mathbf{x})$, $f(\mathbf{x}) := \{\mathbf{x}_i, f(\mathbf{x}_i)\}_{i=1}^N$ with $\mathbf{x} \in \Re^d$, and an error bound $\varepsilon > 0$, find a *minimal* set of normalized Gaussian functions:

$$\left\{ g_j(\mathbf{x}) = a_j e^{-\frac{(\mathbf{x}_j - \mathbf{x})^2}{2\alpha_j^2}} \right\}_{j=1}^{M} \tag{3}$$

such that their linear combination fits the function $f(\mathbf{x})$ within a given error ε:

$$\left\| f(\mathbf{x}) - \sum_j^M c_j g_j(\mathbf{x}) \right\| < \varepsilon \tag{4}$$

It has been shown that deciding whether a given (ε,M)-approximation to $f(\mathbf{x})$ exists is a NP-complete problem [31]. Furthermore, finding the M-optimal (ε,M)-approximation (i.e., the *minimal* set) has been shown to be NP-hard [32]. Fortunately, we do not need to solve the problem exactly; we are satisfied with a good approximation to it. We have devised an approximation method able to provide suboptimal solutions in reasonable computation times. Our approach is based on splitting the problem in two steps: first, we provide a method to find the optimal set of coefficients, centers and bandwidths $\{c_j, \mathbf{x}_j, \alpha_j\}_{j=1}^M$, for a given M; then, we describe an approximation to the minimal M satisfying the above inequality for a given ε. We proceed to discuss each of the two steps.

1. Finding the optimal set of coefficients, centers and bandwidths. Since five parameters need to be simultaneously established per center, the first step in the procedure, as stated, is highly non-linear and ill-conditioned. To work around this difficulty, we have made use of *discrete basis set representations*. With an *a priori* fixed basis set, we linearize the first part of the problem, which is then reduced to finding the appropriate coefficients for each member of the basis set. Assuming that a set of known Gaussian functions of size M has been predefined (our basis set), our goal is to linearly combine the Gaussians to minimize the following norm:

$$\min \left\| f(\mathbf{x}) - \sum_j^M c_j g_j(\mathbf{x}) \right\| \tag{5}$$

The sum is over the M Gaussians in our basis set. The problem is reduced to finding the M coefficients c_j. Equation 5 can be expressed in matrix notation and, after standard manipulations, the vector of coefficients can be obtained as:

$$\mathbf{c} = \mathbf{S}^{-1}\mathbf{t} \tag{6}$$

Equation 6 is a typical Least Squares Fitting (LSF) problem. Elements t_i and S_{ij} in vector **t** and matrix **S**, respectively, are computed as:

$$t_i = \sum_{k}^{Np} f(\mathbf{x}_k) g_i(\mathbf{x}_k) \qquad (7)$$

$$S_{ij} = \sum_{k}^{Np} g_i(\mathbf{x}_k) g_j(\mathbf{x}_k) \qquad (8)$$

The sum is over the number of grid points N_p.

2. Determination of the minimal M. In a first step, a sequence of subspaces is built such that the first subspace – the coarser level of resolution in the sequence – is contained within the second one, which in turn is contained in the third one, and so on. The procedure generates an overcomplete basis set, termed the dictionary, and stops when the original grid is fitted to a desired accuracy. From here we shall select the minimal M, by selecting the smallest number of functions from the subset able to fit the grid equally well. This makes use of the *orthogonal matching pursuit* [31] algorithm, originally described by Davis et al. [32]. It is a greedy algorithm devised to produce suboptimal function expansions by iteratively selecting from the dictionary the Gaussian function with the largest overlap (inner product) with the grid. The search process for the best match is repeated with the grid residue from the previous iteration until it reaches the specified tolerance.

Let $D = \{g_\gamma\}_{\gamma\in\Gamma}$ be a predefined dictionary of Γ Gaussian functions of unit norm ($\|g_\gamma\| = 1$), and$f$our potential to fit (for the sake of clarity, we drop all grid dependencies to simplify nomenclature, i.e., $f \equiv f(\mathbf{x})$). Let us define the inner product of the grid and any function in the dictionary as:

$$< f, g_{\gamma_i} >= \sum_{k}^{Np} f(\mathbf{x}_k) g_{\gamma_i}(\mathbf{x}_k) \qquad (9)$$

The matching pursuit algorithm starts by approximating the grid in the first iteration with the Gaussian producing the largest projection in our dictionary, so that:

$$\begin{cases} g_{\gamma_0} = \arg\max_{g_\gamma \in D}\{< f, g_\gamma >\} \\ f =< f, g_{\gamma_0} > g_{\gamma_0} + R_f^0 \end{cases} \qquad (10)$$

The grid is then updated to the residue, i.e., $f = R_f^0$ and the procedure is repeated. Because an inner product is used, the residue obtained at each iteration has squared norm as small as possible. At the n iteration, the grid is approximated as:

$$f = \sum_{k=0}^{n-1} < R_f^k, g_{\gamma_k} > g_{\gamma_k} + R_f^n \qquad (11)$$

It is guaranteed that in the limit, as n approaches infinite, R_f^n tends to zero. If the functions in the basis set are not orthogonal (such us the Gaussians in our case), convergence is slow, since every new added function introduces additional components in the previous fittings. Faster convergence rates can be obtained by orthogonalizing first the directions of projection [31, 32]. We implemented this orthogonalization step using Gram-Schmidt. Let us denote each orthogonal function already used in the fitting as u. At step k, each added Gaussian is first orthogonalized with respect to the $k-1$ functions already in use, as follows:

$$u_k = g_{\gamma_k} - \sum_{p=0}^{k-1} \frac{< g_{\gamma_k}, u_p >}{\|u_p\|^2} u_p \qquad (12)$$

The approximation to the function converges exactly, at most in N steps, so that the function can be approximated, before the addition of the last Gaussian, as:

$$f = \sum_{n=0}^{N-1} \frac{< R_f^n, g_{\gamma_n} >}{\|u_n\|^2} u_n + R_f^N \qquad (13)$$

With error:

$$\left\| R_f^N \right\|^2 = \|f\|^2 - \sum_{n=0}^{N-1} \frac{\left| < R_f^n, g_{\gamma_n} > \right|^2}{\|u_n\|^2} \qquad (14)$$

The GAGA algorithm

The above procedure has been implemented in a program, which we now proceed to describe. The grid interaction energies are computed as specified above. Van der Waals energies above 5 kcal/mol are set to zero, in order to avoid fitting artifacts. The fitting program then uses as input the grid file to be subjected to Gaussian mapping, together with a fitting error to meet (usually a R_{factor} of 0.5, see Equation 15) or a number of Gaussians to obtain (15 by default). Then, the algorithm carries out the following set of sequential steps:

1. *Face-centered cubic lattice building:* The algorithm starts by setting up a face-centered cubic lattice spanning the grid to fit. Note that this lattice is different

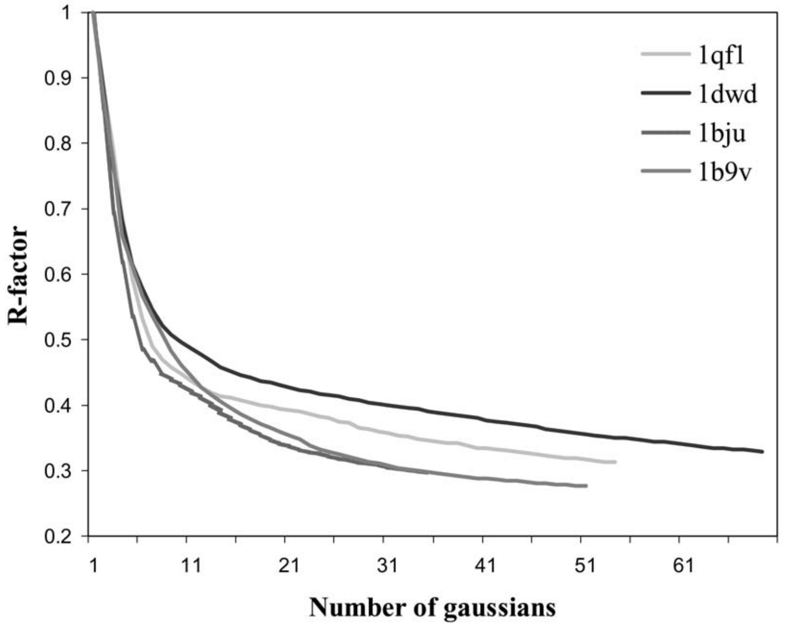


Figure 1. R_{factor} as a function of the number of Gaussian centers introduced by the matching pursuit algorithm in the fitting process of ASA-based electronic densities. Results for four different molecules (indicated in the figure and corresponding to those in Figure 2) are shown.

Table 1. GAGA performance in electron density fitting. The table shows various measurements of performance for the Gaussian fitting algorithm in reproducing the molecular shapes displayed in Figure 1. The first column shows the PDB ID of the ligand used in the fitting experiment; the second gives the number of atoms in the ligand (*nat*), coincident with the number of functions used by the fitting algorithm (*nf*); the third shows the fitting error between densities (before and after the Gaussian approximation); and the last column reports the RMSD (expressed in Å) of the ligand coordinates after maximizing the electron density similarity, as measured by the Carbó index [33].

PDB	nat/nf	R_{factor}	RMSD
1qf1	53	0.313	0.126
1dwd	69	0.327	0.341
1bju	34	0.297	0.241
1b9v	50	0.276	0.104

from the initial grid containing the potential, which uses a cubic lattice. A face-centered cubic lattice is used here because it can provide an optimal packing of spheres in 3D cartesian space, leading to more compact solutions.

2. *Dictionary generation:*

(i) Distribute the initial set of Gaussians over the face-centered cubic lattice. Initially, 4 normalized Gaussians per axis are placed, with bandwidth $\alpha = \max(L_x,L_y,L_z)/\sqrt{2}$ and L_x,L_y,L_z being the grid dimensions in x, y and z, respectively. Functions overlapping with receptor atoms are removed (we use AMBER radii for the receptor plus a tolerance factor of 1.1). From this starting configuration of Gaussians a first LSF solution is computed (with Equation 6), providing the initial set of Gaussian coefficients and an initial error.

(ii) Recursively divide the Gaussian distance by half, multiply by two the number of Gaussians per axis, and set the bandwidth to $\alpha = \max(L_x,L_y,L_z)/2^{Rs}\sqrt{2}$, where R_s is the recursion step. At each step, compute the LSF solution to the input density. Stop when the input error level or maximum number of Gaussians have been reached. By default, the program uses a R_{factor} of 0.5 (Equation 15) or a maximum of 15 Gaussians. Collect all Gaussians in a single dictionary of functions.

3. *Orthogonal Matching Pursuit:*

(i) Compute the overlap with the target function of all functions in the dictionary (Equation 9). Then find the Gaussian function providing maximum overlap with the input grid (Equation 10). Update the function to fit with the computed residual.

(ii) Find the next best Gaussian and orthogonalize with respect to all previously selected functions using Gram–Schmidt (Equation 12). Update the

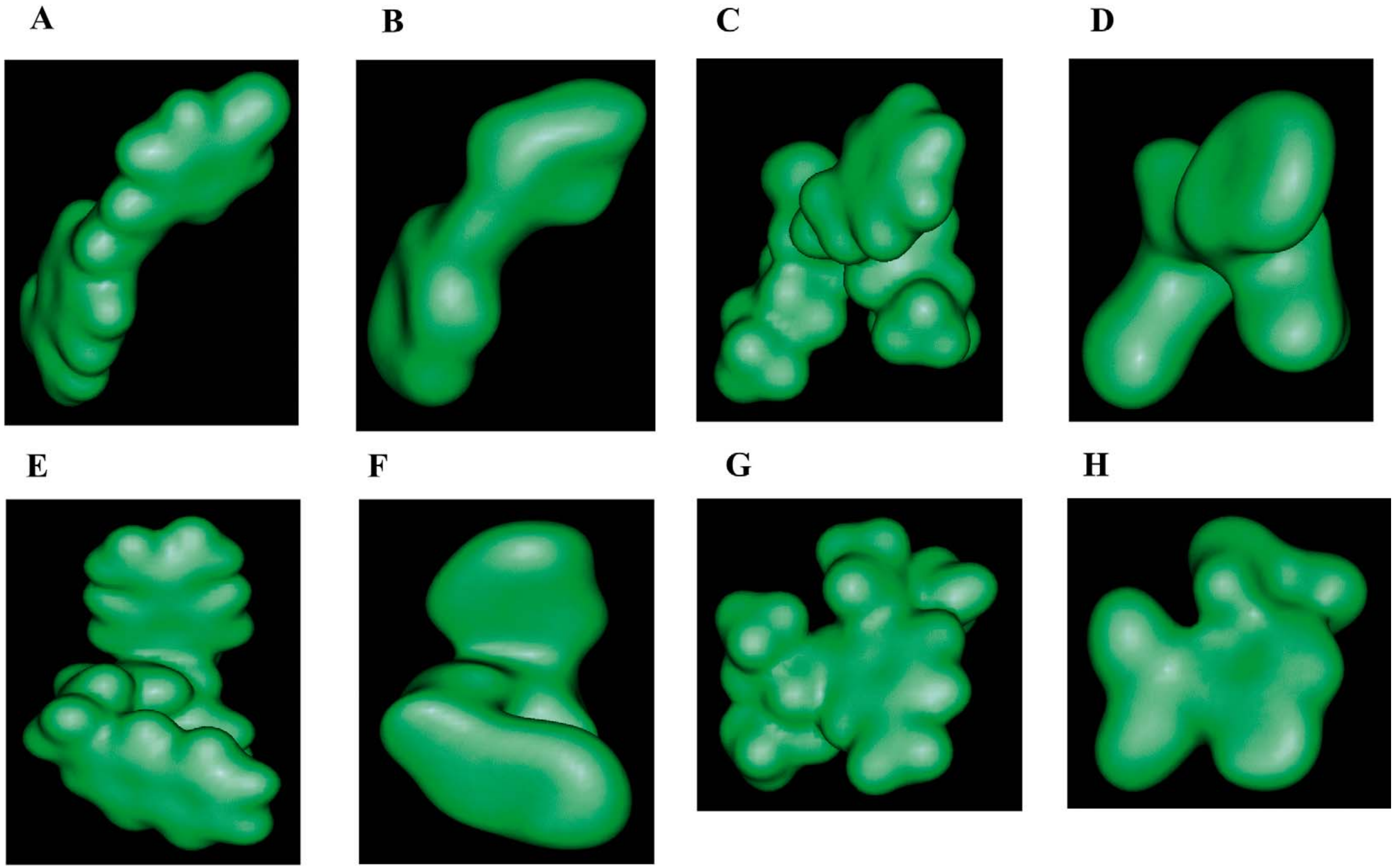


Figure 2. Gaussian fitting for a set of small molecules. ASA-based electron densities have been contoured at a density level of 0.80 (A,C,E,G), as well as for rebuilt densities after Gaussian fitting (B, D, F, H). Molecules in the figure correspond to the ligands in the following pdb files: 1bju (A,B); 1qf1 (C,D); 1dwd (E,F); 1b9v (G,H). Chart 1 shows their molecular structure.

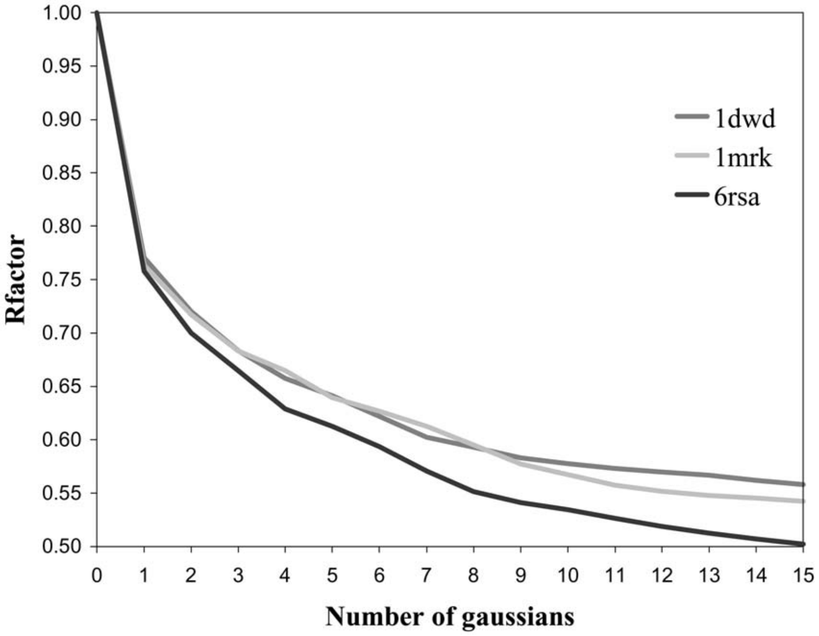


Figure 3. R_{factor} as a function of the number of Gaussian centers introduced by the matching pursuit algorithm. Grid energies were previously obtained using a benzene probe, as indicated in the Methods section. Results for three different proteins are shown, with PDB [37] identification indicated in the figure. The resulting Gaussians were then used in the docking experiments summarized in Tables 2 and 3.

residual. If neither the target error level (Equation 14) nor the number of target Gaussians are satisfied, repeat the procedure. The grid is finally approximated with Equation 13.

Testing the Gaussian mapping approximation

We have carried out several types of tests of the Gaussian mapping approach. Three questions have been addressed. The first was to know how well the GAGA approach could fit an arbitrary grid file. The second was to study how well the fitted pharmacophoric points could reproduce the actual location of corresponding chemical groups in known protein–ligand complexes (hot spot location) in some well-documented cases. The third was to study the effect that the optimally deduced points could have in speeding up and increasing the accuracy of ligand docking.

1. Grid fitting: Initial studies about the fitting quality were carried out with simple 1D and 2D functions. For these cases we knew, by numerical analysis, the minimal number of Gaussians required to fit the function under a given error. The algorithm showed a good performance for these simple cases (data not shown). Subsequently, we turned to the problem of analyzing the ability of the method to reproduce molecular shapes, as specified by computed electron densities. Specifically, for a set of molecules, we compared the molecular shape given by their electron density with the electron density obtained upon fitting with GAGA the grid containing this density to a number of Gaussians equal to the number of atoms found in the original molecule. Upon fitting, we quantified the similarity between both densities in terms of the R_{factor}:

$$R_{factor} = \frac{\sqrt{\sum_{i=1}^{N_p}\left[f(\mathbf{x}_i)^2 - \left(\sum_{j=1}^{M} c_j g_j(\mathbf{x}_i)\right)^2\right]}}{\sqrt{\sum_{i=1}^{N_p} f(\mathbf{x}_i)^2}} \qquad (15)$$

where N_p is the number of grid points and M is the number of Gaussian functions selected. Finally, we measured the RMSD of the molecular coordinates of the molecules obtained after a previous optimal superimposition of both grids. Optimal superimposition was obtained by maximizing the similarity of the densities according to the Carbó index while performing a complete sampling of rotational space with a step size of 27°. The Carbó index [33 and references therein] is defined as:

$$C_{AB} = z_{AB}(z_{AA}z_{BB})^{-1/2} \qquad (16)$$

The similarity function z_{AB} of densities A and B, depending upon the orientation tensor Ω, is defined as the generalized projection:

$$z_{AB}(\Omega) = \int f_A(\mathbf{x}) f_B(\mathbf{x}; \Omega) d\mathbf{x} \qquad (17)$$

2. Evaluating hot spots. We have studied the ability of the Gaussian mapping approach to determine the position of aromatic interaction sites. For this we studied, in a set of 53 ligand–receptor complexes (Tables 2 and 3), the distance between the predicted interaction sites and the corresponding aromatic moieties in the ligands of the complexes. As a metric of the agreement, we studied the minimum distance between the set of predicted interaction points and the set of aromatic centers found in the ligand. We also studied the statistical significance of the computed distances. For that, we randomly placed in the grid box as many interaction centers as Gaussians were originally placed in the binding site by our algorithm. We then computed the distances between interaction sites and the aromatic centers in the ligand. We repeated the procedure 10,000 times, and we counted in how many cases we obtained a distance equal to or better than that observed in the real case. The P-value is the fraction of this number with respect to the total number of trials.

3. Ligand docking. We have evaluated the Gaussian mapping approach using only hydrophobic/aromatic fragments. First, they are difficult cases for compression algorithms, as these interactions are usually rather spread out over the binding site, lacking the strong concentration of energy typical of, e.g., hydrogen bonds. In fact, matching algorithms typically run into problems if hydrophobic/aromatic fragments are placed in the active site [34]. Second, in contrast to e.g. electrostatic interactions, the underlying molecular mechanics force field is able to do a reasonable job describing them, allowing to focus on the properties of the Gaussian description and not on the details of how the interaction is being computed. Third, one of the goals of this paper is to explore the minimum set of points required for successful docking; the aromatic interactions, due to their ability to summarize the shape of the ligand, are ideal candidates. For a set of 53 different protein–ligand complexes (Table 2), we have computed Gaussian centers for aromatic spots using a benzene probe, with standard AMBER types and charges of −0.155 for C and 0.155 for H. Centers and their associated widths for the Gaussians were generated by first running CGRID and CDOCK [27] to generate the grid maps, which were processed with GAGA. Then, the generated Gaussian centers were used as surrogates of the spheres required by the DOCK program, which was used both in rigid and flexible docking modes.

For flexible docking, we first created a flexible library of conformers of the ligand. Only torsion angles in rotatable bonds were considered, with a rotatable bond defined as a single or exocyclic double bond having at least one non-hydrogen neighbor on either side of the bond. Rotations affecting oxygen atoms in terminal groups, such as carboxylates, phosphates, sulfonyl, etc... were skipped. Only rotameric states were considered, with the following dihedral angles: 60°, 180°and −60° for sp^3-sp^3 bonds; 0°, 60°, 120°, 180°, −60°, and −120° for sp^3-sp^2 bonds (when symmetry is present on sp^2 only 60°, 180° and −60° are considered to avoid redundancies); and 0°, 180° for sp^2-sp^2 bonds (again, symmetry existence is checked). For sp^2-sp^2 rotatable bonds attached to aromatic systems 0°, 90°, 180°, −90° angles were used (when symmetry is present on sp^2 only 0° and 90° are scanned to avoid redundancies); on the other hand, oxygen-sp^3 bonds in esters, and ethers were treated as regular sp^3-sp^3 bonds except in cases where the sp^3 center is located in a ring (e.g. ribose, glucose derivatives), where eclipsed conformations are a common trend. For these we consider 0°, 60°, 120°, 180°, −60° and −120°.

All possible dihedral angle combinations were generated and the corresponding intramolecular energy of each of the resulting conformations is computed based on non-bonding 12-6 Lennard-Jones terms, without considering 1-2 and 1-3 interactions, and with 1-4 interactions scaled down by a factor of 2, as is customary within the AMBER force field [28]. No energy minimization was performed in any of the conformations, thus a VDW energy cutoff of 5 kcal/mol is used to cap each pairwise interaction. Only conformations with computed VDW energies within 30 kcal/mol of the global minimum were saved for docking. To reduce the combinatorial explosion, for sp^3-sp^2 bonds the program first evaluates the local energy, up to 1-4 interactions, associated with each of its 6 dihedral angles. The three lowest energy values are then used in the combinatorial search. This procedure proved very successful in keeping the combinatorial size under control, while having minor effects on the rate of bioactive conformation generation and docking accuracy (data not shown).

Prior to docking, ligands were parametrized according to the AMBER force field. Atom types were

Table 2. Results from the rigid-body docking experiments with DOCK 3.0 using GAGA generated Gaussians for 10 different complexes. PDB ID of the complex (PDB); number of Gaussian points selected in the binding site (Np); R$_{factor}$ obtained with those points; the best scored conformation found in the docking search and its corresponding RMSD (Best CONT columns); the minimum RMSD conformation found in the search and the corresponding CONTACT score (Best RMSD columns); the number of orientations saved by DOCK with low energies (Saved); and the percentage of those with less than 2 Å RMSD with respect to the X-ray structure (% Succ.).

PDB	Np	R$_{factor}$	Best CONT		Best RMSD		Saved	%Succ.
			RMSD	CONT	RMSD	CONT		
1fjs	13	0.51	1.50	257	1.50	257	350	0.3
1ajv	12	0.61	0.85	241	0.85	241	1	100.0
2tsc	11	0.54	9.68	104	7.29	70	43	0.0
3ert	11	0.64	0.68	151	0.63	138	17	100.0
1hsb	12	0.53	0.73	141	0.46	124	280	9.6
1rds	14	0.51	3.51	94	3.17	64	501	0.0
1b9v	11	0.49	0.51	169	0.51	169	148	2.7
1ppc	11	0.51	3.27	129	2.16	77	903	0.0
1kel	10	0.52	0.97	186	0.63	178	41	19.5
2fox	11	0.57	7.85	121	5.79	58	87	0.0
1fax	12	0.54	1.20	205	0.97	200	85	11.8
1xka	12	0.56	0.95	204	0.95	204	10	40.0
1dwd	13	0.56	1.29	150	1.29	150	229	0.4
1rt2	11	0.72	0.97	235	0.50	216	12	100.0
1mts	11	0.56	1.12	175	0.87	156	193	4.2
3dfr	11	0.70	1.49	222	0.44	198	6	100.0
3tpi	10	0.60	0.79	157	0.47	154	17	82.4
3cla	12	0.64	4.75	89	1.80	40	1282	0.3
1dwc	13	0.53	1.33	182	0.63	136	251	2.0
1rt1	11	0.66	0.48	166	0.29	153	9	100.0
4dfr	12	0.56	0.88	123	0.79	121	142	2.1
2dbl	12	0.54	9.75	118	3.29	83	426	0.0
1fpu	11	0.74	0.63	212	0.63	212	2	100.0
1dbm	13	0.52	1.15	216	0.87	187	606	0.8
1snc	13	0.56	6.18	139	1.00	125	252	2.0
1tni	10	0.47	4.65	106	1.80	80	226	2.7
1pph	12	0.49	1.98	155	1.98	155	1062	0.1
1srj	14	0.56	6.88	186	2.56	96	1544	0.0
1f0r	12	0.57	1.11	163	1.11	163	316	1.9
1b9t	13	0.48	0.70	131	0.61	122	914	1.3
1bjv	11	0.67	5.37	115	2.86	95	471	0.0
1rnt	12	0.52	4.28	121	1.71	82	2275	0.1
1rob	12	0.49	0.76	132	0.76	132	2082	1.4
2ak3	10	0.54	2.10	116	0.85	110	243	7.8
1bju	13	0.59	1.98	129	1.00	117	297	2.0
1ejn	11	0.67	0.66	187	0.66	187	22	13.6
1c5c	11	0.69	4.03	169	0.78	153	11	18.2
1f3d	12	0.58	6.42	115	0.63	97	198	18.2
1mld	12	0.49	1.28	109	0.69	104	18	27.8
1mrk	13	0.54	0.95	157	0.50	152	940	6.9
1wap	8	0.56	0.37	126	0.37	126	3	66.7
2cmd	11	0.48	1.23	101	1.07	101	43	18.6
6rsa	14	0.51	1.46	113	0.77	105	680	6.5

Table 2 (continued).

PDB	Np	R_{factor}	Best CONT		Best RMSD		Saved	%Succ.
			RMSD	CONT	RMSD	CONT		
7tim	9	0.71	4.76	107	4.76	107	5	0.0
1cbs	11	0.60	2.20	110	1.79	59	13	15.4
1fen	10	0.63	2.03	127	0.32	96	30	73.3
2cbs	10	0.59	9.87	109	0.38	99	17	64.7
1dbb	13	0.49	6.92	115	1.59	70	1205	0.3
1die	11	0.51	2.49	116	2.17	100	65	0.0
1tnh	10	0.48	3.91	84	1.09	75	156	30.1
1tnl	10	0.52	4.24	96	1.84	75	282	0.4
1d3h	12	0.70	7.34	156	0.77	121	185	13.0
1flr	12	0.57	1.38	204	1.28	161	64	28.1

assigned automatically [35], and monopolar charges were fitted to reproduce the molecular electrostatic potential computed with the MOPAC program [36] using the AM1 Hamiltonian [37]. Docking with DOCK 3.0 was performed using standard conditions with the CONTACT scoring function. We measured the RMSD of the lowest energy solution, the score of the lowest energy solution, and the computer time required to achieve it. We also measured the lowest RMSD found during the docking simulation and the fraction of poses below the 2 Å cutoff of RMSD with respect to the crystal structure.

Results

Quality of the Gaussian-based fitting of grid data

The evolution of the R_{factor} with the number of selected Gaussians for the case of fittings to ASA-based electron densities (see Methods) can be observed in Figure 1. Molecules in Figure 1 are found in Chart 1. We show the original ASA-based electron density and the GAGA-fitted one using identical contour levels (0.80, arbitrary units). Fitting calculations were carried out restricting the maximum possible number of Gaussian functions stored in the dictionary to the number of atoms present in the original molecule, so as to make results directly comparable. As can be observed, molecular shapes are correctly reproduced with the fitting algorithm, although the fitted contours are smoother, with the individual atomic features partially lost. The quality of the final fittings as a function of the number of Gaussians can be observed in Figure 1. Quantitative data for the four molecules represented in Figure 1 are summarized in Table 1. The RMSD of the molecular coordinates after optimal superimposition of the electron densities oscillates between 0.1–0.3 Å. This suggests that if shape features of a protein binding site can be captured in a similar way, the resulting Gaussians may become useful as pharmacophoric descriptors in screening and docking ligands in protein binding sites. We devote the next sections to explore this question.

Detection of hot spots in binding sites

The relationship of the position in the binding site of the predicted hot spots with the actual location of these functional groups in ligands bound to the target has been studied. Since positioning of hydrophobic and aromatic interaction points is challenging [34], we have focused our attention to the agreement between predicted and observed aromatic interaction sites. Figures 4 and 5 summarize the results for the set of 53 complexes extracted from the Protein Data Bank (PDB) [38] and studied in this paper. We have computed the minimum distance between the predicted sites and the geometry center of each aromatic ring in the ligand, and we have evaluated the significance of the closest distance by computing the likelihood of obtaining in a random way a distance equal to or smaller than the one observed. Details of the calculations can be found in the Methods section. Figure 5 shows that a confidence level of ~95% corresponds to a distance between the Gaussians and the aromatic centers of ~1.5 Å, with ~65% of the interaction points having an aromatic binding site within 1.5 Å (Figure 4). We conclude that the interaction points found by Gaussian mapping for aromatic sites are reasonably accurate,

Table 3. Results from the flexible docking experiments with DOCK 3.0 using GAGA generated Gaussians for 10 different complexes. This table shows, together with the PDB ID of the corresponding complex, the minimum RMSD of the isolated ligand with respect to the X-ray ligand conformation (Min RMSD); the best scored conformation found in the docking search and its corresponding RMSD (Best CONT columns); the minimum RMSD conformation found in the search and the corresponding CONTACT score (Best RMSD columns); number of conformers used in the docking calculation (Scanned); the number of them saved by DOCK with low energies (Saved); and the docking time per rotamer in seconds (measured with the timex command in a SGI R12000 processor).

PDB	Np	Min	Best CONT		Best RMSD		Scanned	Saved	Time/rot
		RMSD	RMSD	CONT	RMSD	CONT			(s)
1fjs	13	1.36	2.86	243	2.00	209	1538	100	0.22
1ajv	12	1.04	10.08	178	9.04	131	1383	31	0.18
2tsc	11	0.67	3.36	183	2.51	143	1251	100	0.08
3ert	11	0.44	1.53	188	0.97	164	1177	100	0.20
1hsb	12	0.34	5.61	144	0.86	138	663	100	0.10
1rds	14	0.72	5.18	149	2.21	124	648	100	0.14
1b9v	11	0.46	4.19	177	0.89	145	609	100	0.34
1ppc	11	1.60	8.04	184	2.88	154	577	100	0.21
1kel	10	0.74	4.05	190	1.24	166	445	100	0.09
2fox	11	0.68	1.19	258	1.19	258	349	100	0.12
1fax	12	0.78	8.49	217	1.84	188	348	100	0.07
1xka	12	0.66	3.09	209	1.69	193	338	100	0.07
1dwd	13	1.45	3.15	194	2.41	163	298	100	0.28
1rt2	11	0.69	1.34	228	1.09	195	279	5	0.15
1mts	11	1.05	2.22	201	2.19	198	192	100	0.06
3dfr	11	0.93	2.09	244	1.81	243	177	75	0.04
3tpi	10	0.39	2.91	179	0.82	160	177	100	0.06
3cla	12	0.33	5.48	115	3.43	96	171	100	0.15
1dwc	13	2.36	5.90	180	3.64	123	165	100	0.18
1rt1	11	0.43	1.53	182	0.72	176	162	100	0.13
4dfr	12	0.73	6.81	151	1.59	126	148	100	0.08
2dbl	12	0.57	9.50	162	1.86	121	111	100	0.20
1fpu	11	0.30	1.65	217	1.41	217	106	32	0.06
1dbm	13	0.56	2.40	244	1.13	138	68	68	0.32
1snc	13	0.42	4.97	202	1.87	139	66	66	0.19
1tni	10	0.56	5.07	116	1.77	101	65	63	0.04
1pph	12	0.95	4.09	162	2.89	142	60	60	0.22
1srj	14	0.12	2.60	194	2.01	152	58	58	0.70
1f0r	12	1.13	2.52	185	2.52	185	51	51	0.12
1b9t	13	0.35	5.65	144	1.11	130	47	47	0.58
1bjv	11	0.51	10.52	150	2.14	122	34	34	0.12
1rnt	12	0.80	5.55	135	2.38	122	27	27	0.37
1rob	12	0.25	0.86	141	0.79	129	26	26	0.31
2ak3	10	0.24	2.25	149	0.96	115	26	26	0.15
1bju	13	0.33	8.18	153	1.00	152	25	25	0.09
1ejn	11	0.54	3.15	177	2.09	121	18	18	0.07
1c5c	11	0.20	5.86	167	0.72	164	9	9	0.14
1f3d	12	0.30	6.57	124	1.07	112	9	9	0.14
1mld	12	0.09	1.72	113	1.29	111	9	9	0.06
1mrk	13	0.85	2.81	180	1.40	148	9	9	0.35
1wap	8	0.41	2.59	124	0.81	114	9	7	0.07
2cmd	11	0.10	1.04	106	1.04	106	9	9	0.07
6rsa	14	0.38	1.69	130	1.58	117	9	9	0.29

Table 3 (continued).

PDB	Np	Min	Best CONT		Best RMSD		Scanned	Saved	Time/rot
		RMSD	RMSD	CONT	RMSD	CONT			(s)
7tim	9	0.45	1.96	114	1.96	114	9	4	0.06
1cbs	11	0.32	1.54	118	1.54	118	4	4	0.20
1fen	10	0.47	9.38	141	1.24	111	4	4	0.20
2cbs	10	0.10	4.58	119	0.79	112	4	4	0.18
1dbb	13	0.47	6.75	147	2.12	128	3	3	0.56
1die	11	0.33	2.52	115	2.52	115	3	3	0.13
1tnh	10	0.06	3.92	88	1.41	80	3	3	0.15
1tnl	10	0.10	4.25	106	2.25	96	3	3	0.14
1d3h	12	0.04	2.40	155	2.24	151	2	2	0.39
1flr	12	0.41	1.28	183	1.28	183	2	2	0.44

Table 4. Comparing GAGA and SPHGEN performances to generate interaction points for docking with DOCK 3.0. Two different complexes, 3dfr and 6rsa, have been used for this comparison. Interaction points with GAGA were generated as described in Methods. Interaction points with SPHGEN were obtained from the demos available in the DOCK package distribution (61 interaction points for 3dfr and 47 for 6rsa). See caption of Table 3 for description of the columns.

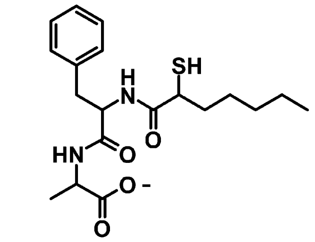
PDB	Best CONT		Best RMSD		Scanned	Saved	Percent	Time/rot
	RMSD	CONT	RMSD	CONT			success	(s)
GAGA 3dfr	2.09	244	1.81	243	177	75	2	0.06
6rsa	1.69	130	1.58	117	9	9	56	0.39
SPHGEN								
3dfr	2.52	262	1.15	255	177	100	4	7.34
6rsa	1.49	144	1.44	136	9	9	22	1.16

and they will likely be helpful in docking algorithms. The next section explores this issue.

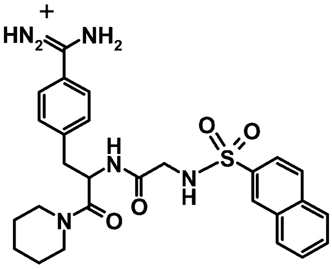
Application to molecular docking with the program DOCK

We investigated whether the small number of interaction points selected by Gaussian mapping of hydrophobic/aromatic interactions, between 11 and 12, is enough to efficiently sample native-like complexes using standard docking algorithms. Because our aromatic Gaussian centers are positioned so that they account for the maximal amount of energetic information for the probe within the binding site, we would expect them to be particularly well suited for the task. We have used a test set of 53 complexes extracted from the PDB [38] for this experiment (Tables 2 and 3), and selected a standard negative sphere-based docking algorithm such as DOCK. The Gaussian centers placed by GAGA were then used to replace the sphere centers generated by the program SPHGEN [17] within the DOCK package, as described in Methods. The rest of the computations in DOCK 3.0 were performed with standard parameters.

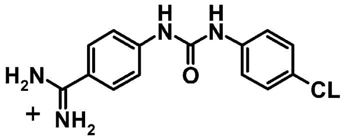
Results are summarized in Table 2 (rigid docking) and Table 3 (flexible docking). The number of selected interaction points is in all cases very similar, with a mean of 11.5, in spite of differences in shape and interaction properties of the binding sites. The evolution of the R_{factor} as a function of the number of selected Gaussians is recorded in Figure 3 for three different cases. As in the case of electron densities, decay in R_{factor} follows an exponential law, as predicted from theoretical arguments. At the automatically selected number of interaction points, the average R_{factor} is 0.57. These observations suggest that with 11–12 functions only the coherent part of the signal in the grid energies is captured by the Gaussian description. However, as observed in Figure 4, this coarse recovery is enough to approximately describe the shape of the



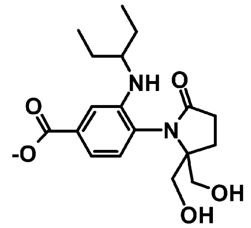
1qf1



1dwd



1bju



1b9v

Chart 1. Structure of the molecules shown in Figure 2, with their PDB [38] ID.

binding site and likely active ligands binding to it. The small set of interaction aromatic/hydrophobic points is enough to successfully sample native-like geometries. Tables 2 and 3 show that for ∼70% of the cases in flexible docking, and ∼85% in rigid docking, DOCK is able to sample geometries below 2 Å among the 100 upper ranking decoys using Gaussian mapping. The success rate in flexible docking considering only the highest scoring pose is, however, considerably lower, only ∼25% in flexible docking and ∼58% in rigid docking. Most likely this is related to the force field used to score the solutions (the CONTACT score in DOCK). Recently, in a comparative study of several docking algorithms, the Brooks group [39] obtained a success docking rate for DOCK of ∼29% in flexible docking, comparable to ours. This result suggests that our success rate for the first ranking pose is not significantly worse than the one obtained with the standard DOCK program. As for the 30% of sampling failures, ∼30% of them correspond to cases where the best sampled conformer has an RMSD above 1.0 Å from the crystal structure in the complex, suggesting that for these cases the sampling of the ligand conformational space is insufficient.

Our most important result is that, due to the small number of points selected and the coarse rotameric description, computing times with Gaussian mapping are highly reduced, while preserving sampling of native-like poses. In Table 4 we compare the computing times and resolutions using the standard spheres generated with SPHGEN and the interaction points generated with GAGA for two representative cases. These two cases were selected because they are provided as examples within the different DOCK distributions. Therefore they can be considered to provide unbiased comparisons between the results with both center generation methods (SPHGEN and GAGA). It is remarkable that, using only aromatic points, docking performances within DOCK remain similar, whether the interaction points were generated with GAGA or with SPHGEN. In contrast, docking times are dramatically shorter with GAGA, with speed-ups per conformer of ∼100 times for the case of 3dfr and of ∼3 times for 6rsa. For 3dfr and using DOCK 4.0 with incremental construction, the reported total docking times on a R10000 processor ensuring native-like sampling are 200 s [16]. The total docking time with DOCK 3.0 using our Gaussians and coarse rotamer description in a R12000, with 2% native-like poses in the upper ranking list, is 10 s (Table 4). While conditions are not directly comparable, the numbers do suggest a substantial speed-up in docking times, of a factor of 20 in this case. For the set of 53 complexes examined in this work, the average time per conformation (on a SGI R12000 processor) is 0.18 s. The average ligand is represented in our test set with ∼200 conformers, and this translates to an average time of ∼35 s per molecule. The average docking times recently reported by the Brooks group [39] with DOCK are ∼143 s on a similar machine. Therefore, a speed-up factor of ∼4–5 is expected on average while keeping similar predictive properties. Clustering of conformers [40], or the use of simplified representations for the ligand, can potentially reduce these docking times even further.

The data presented in Table 4 also suggest that, in practical applications, additional filtering might be required. Re-ranking the upper-ranking configurations selected with the CONTACT scoring using more sophisticated energy functions, such as the GB-SA method [41], might be of interest, and will be the subject of future research. While a more comprehensive study is clearly needed before drawing definitive conclusions, the results presented here are promising in pointing to Gaussian mapping as a way to increase speed in docking algorithms.

Discussion

In this paper we present a new approach to obtain a quasi-optimal design of pharmacophoric points from receptor binding sites, where each pharmacophoric point is modeled as a Gaussian center. We first generate a molecular mechanics energy map of the interaction of representative molecular probes with the residues in the active site at regular lattice locations. This part of the method shows close similarities both with GRID [2–6] and MCSS [7–9]. Then, and given this fragment-based energy grid, the second part of the procedure attempts to find a minimal number of Gaussian centers able to account for the energy distribution stored in the grid, so that each selected point carries with it maximal information content (i.e., maximal correlation) about the energy distribution of the probe in the binding site. Thus, by construction, the procedure attempts to position Gaussian centers of appropriate widths in regions of energy minima in the binding site. These are regions where the highest likelihood to find an atom of the corresponding type in the ligand can be expected. We have also shown initial evidence that these sets of centers are well suited for docking. We have, on purpose, tested the generated points based only on hydrophobic/aromatic potentials, in order to reduce the number of interaction points to an absolute minimum. Tests with the DOCK program, where the traditional set of spheres derived from a purely geometric analysis of the binding site has been replaced by a small set of fitted Gaussians, have shown good abilities to sample native-like binding modes with a reduction by a factor of ∼4–5 in computing times. Addition of hydrogen bond interaction points should increase the predictive properties with modest increments in computing times.

A related idea has been described recently by Joseph-McCarthy and Alvarez [20]. These authors have described a method to determine chemically labeled site points by automatically extracting them from a clustering of selected low-energy functional-group minima obtained with MCSS. They have also shown that the pharmacophoric site points can be directly matched to the pharmacophoric features of database molecules with DOCK to place the small molecules into the binding site. In agreement with our results, they have shown that biasing the search using points selected with energy criteria allows for more effective sampling of the target site. A difference with their work is that we attempt not only to select these energy minima, but also to ensure that the set of points selected is quasi-optimal, in the sense described in the Introduction. We show that quasi-optimality in the design allows drastic reductions in computing times without severely affecting the ability to sample native-like poses.

The idea of compressing grid-sampled energies into a reduced Gaussian set to deduce pharmacophores is not new. Previously, Klebe and co-workers [21] described a way to derive shape descriptors built from anisotropic Gaussian contributions. These were obtained from a fit to propensity distributions stored in IsoStar [22], a database of interaction geometries between a common central group and various interacting moieties extracted from small-molecule crystal structures. The different initial densities, test cases and accuracy measures used in their study and ours preclude a direct comparison. In general terms, their approach bears considerable similarities with the one presented here, although there are also some noticeable differences. Besides using statistically derived propensities instead of molecular mechanics energies, Klebe and co-workers use anisotropic Gaussian functions, instead of the isotropic Gaussians used in this work, to fit the field. Thus, in their approach, 10 parameters need to be fitted per function, while in

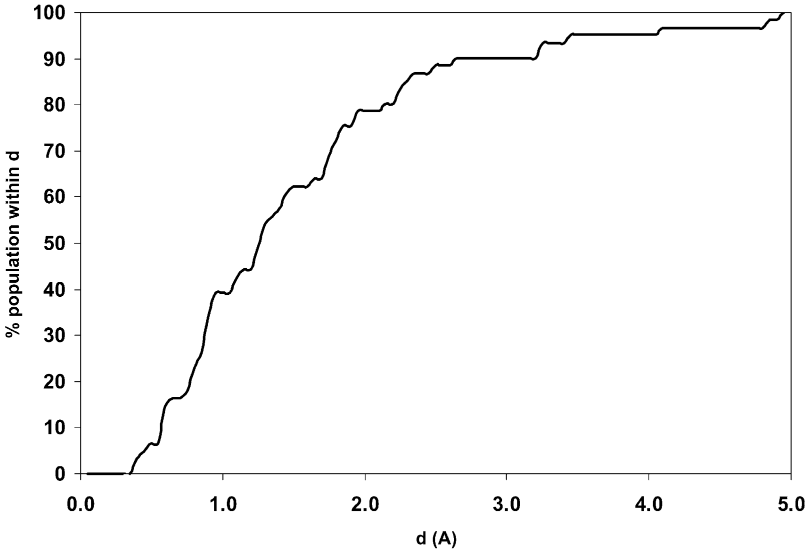


Figure 4. Cumulative distribution of the closest Gaussian-aromatic center distance for each one of the aromatic centers in the set (complexes in Tables 2 and 3), as a function of the distance.

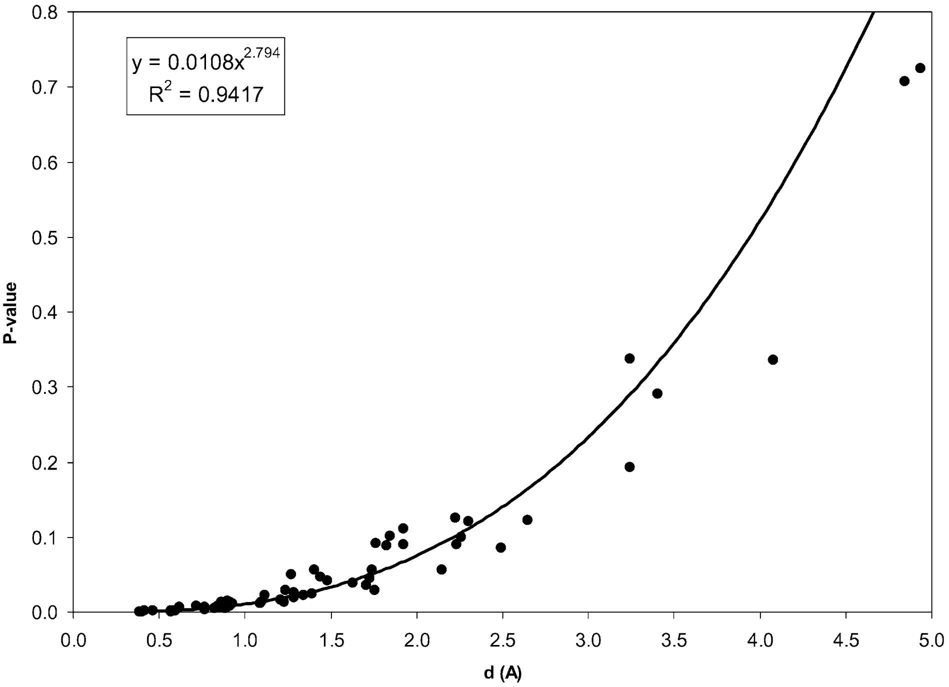


Figure 5. P-value as a function of distance for each one of the aromatic centers in the set of complexes in Tables 2 and 3. The regression fit to a power law is also shown.

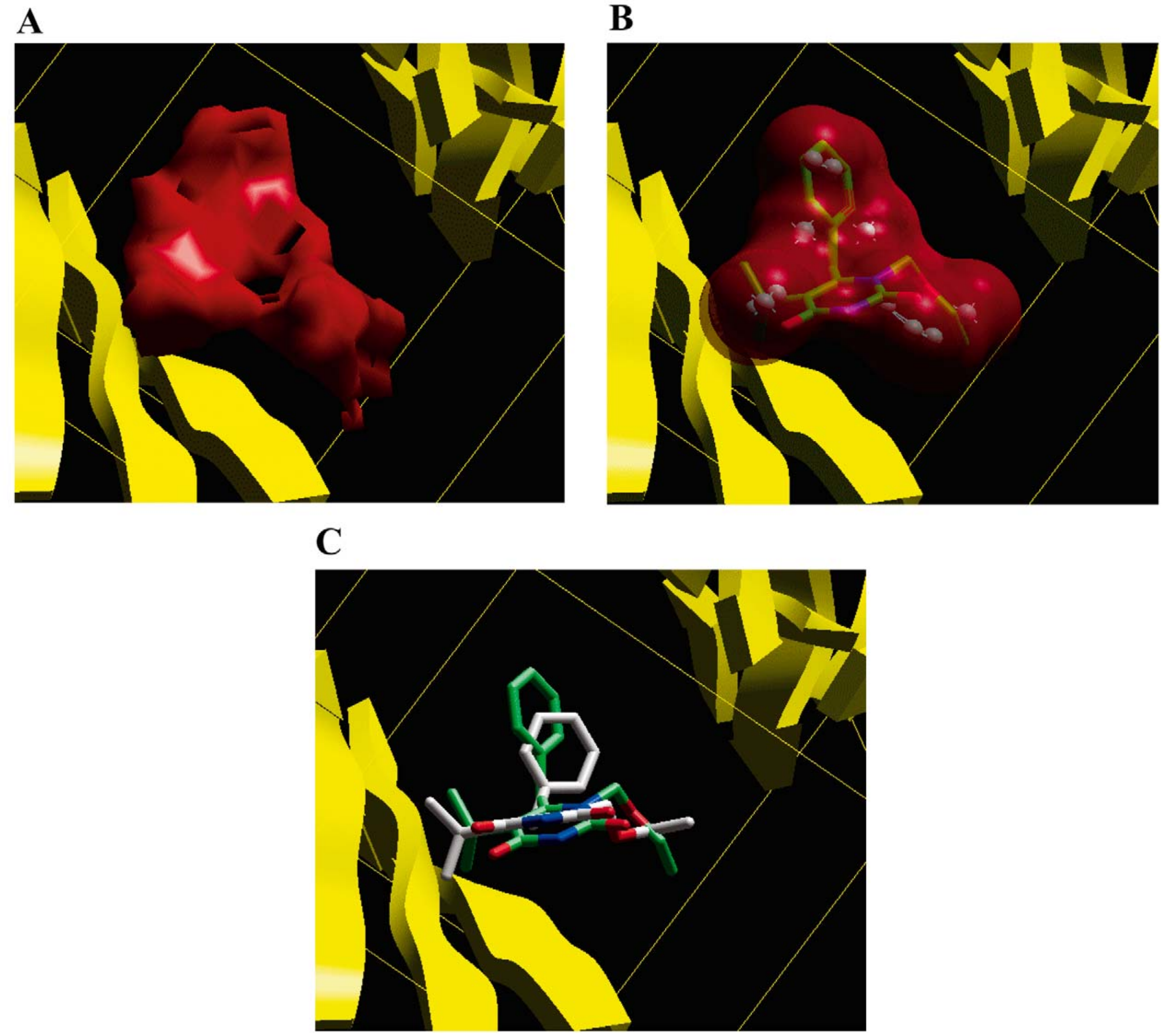


Figure 6. Gaussian mapping at work, as exemplified with 1rt1. (A) Contour (at −6 kcal/mol) of a benzene probe in the binding site, computed with CGRID and CDOCK; (B) Gaussians fitted by GAGA to the above contour. The small blue spheres depict the Gaussian centers, while the red wine envelope corresponds to a Connolly surface of the spheres using their associated bandwidth as atomic radii. Also shown in the figure is the X-ray conformation of the ligand in 1rt1, so that the closeness between ligand atoms and Gaussian centers can be appreciated; (C) superimposition of the best scored (CONTACT) docking conformation of 1rt1 upon flexible docking with DOCK using the set of interaction points in B and the X-ray structure. RMSD is 1.53 Å.

our representation 5 parameters are required. While the fitting procedure could be adapted to anisotropic Gaussians, the computational simplicity and amenability to analytical treatment of spherical Gaussians lead us to implement only the isotropic case. Another important difference lies in the position and number of Gaussians. Again, their approach did not guarantee to select a close to optimal set of centers. In any case, a detailed comparison between both approaches under similar conditions would be of interest.

More generally, grid compression with simplified functionals is the subject of intensive research in different areas of computational structural biology. For example, Wriggers and co-workers have pioneered the use of vector quantization [42, 43], a mathematical technique closely related to the matching pursuits employed in this work, to model electron microscopy data. More recently, Carazo [44] and co-workers have refined the vector quantization approach with a modification that incorporates a Gaussian kernel, in order to select the points within the macromolecule that best approximate the probability density function of the original volume data. The idea is very close in spirit to the one presented here, although applied within a different mathematical framework and in the context of electron microscopy data. These areas could benefit from the generality and compactness of the matching pursuits described here.

One of the limitations of our method, in its current form, is the computational burden of the preparation phase, about 1 min on average. While this is not an obstacle when the objective is screening molecular databases, it does make the current technique uncompetitive for docking individual molecules. Nevertheless, there is considerable room to improve the computational speed, as we have not attempted to optimize the code. First, calculations within the grid generation program, CGRID [27], are slow due to the energy minimization of the fragment used as probe. Here, the use of statistical potentials [12, 45, 46], stored as look-up tables and hence faster to compute, could be of interest. Second, our docking results seem to indicate that only a coarse-level coverage of the different energy minima within the binding site is required. Therefore, it is likely that the use of non-orthogonal matching pursuits suffices for this particular application. For the coherent part of the signal, non-orthogonal matching pursuits are computationally more efficient [31, 32]. Third, since the overlap matrix used in the least-square fitting procedure is sparse, it is suitable for the application of special, fast diagonalization techniques, speeding up the calculation of the inverse.

The integration of the Gaussian descriptors with docking techniques can be extended and improved in a number of ways. One important advantage of the approach presented here over the use of the more traditional spheres or interaction points is the flexibility that they allow in the criteria used for the hot spots, accommodating different objectives in a virtual screening search.

Acknowledgements

This work has been supported by a grant awarded by Eli Lilly & Co. Research at CBMSO has been supported by grants from MCYT and FIS, and by an institutional grant from Fundación Ramón Areces. M.M. was partially supported by a postdoctoral fellowship from Fundación Ramón Areces. Insightful discussions with Dr. Michal Vieth from Eli Lilly & Co. are gratefully acknowledged. We thank Dr. Federico Gago for carefully reading the manuscript and the anonymous referees for their comments.

References

1. Brooijmans, N. and Kuntz, I.D., Annu. Rev. Biophys. Biomol. Struct., 28 (2003) 28.
2. Wade, R.C. and Goodford, P.J., J. Med. Chem., 36 (1993) 148.
3. Wade, R.C., Clark, K.J. and Goodford, P.J., J. Med. Chem., 36 (1993) 140.
4. Boobbyer, D.N., Goodford, P.J., McWhinnie, P.M. and Wade, R.C., J. Med. Chem., 32 (1989) 1083.
5. Goodford, P.J., J. Med. Chem., 28 (1985) 849.
6. Goodford, P.J., J. Med. Chem., 27 (1984) 558.
7. Miranker, A. and Karplus, M., Proteins, 23 (1995) 472.
8. Caflisch, A., Miranker, A. and Karplus, M., J. Med. Chem., 36 (1993) 2142.
9. Miranker, A. and Karplus, M., Proteins, 11 (1991) 29.
10. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J., J. Comp. Chem., 19 (1998) 1639.
11. Laskowski, R.A., Thornton, J.M., Humblet, C. and Singh, J., J. Mol. Biol., 259 (1996) 175.
12. Verdonk, M.L., Cole, J.C., Watson, P., Gillet, V. and Willett, P., J. Mol. Biol., 307 (2001) 841.
13. Gohlke, H., Hendlich, M. and Klebe, G., J. Mol. Biol., 295 (2000) 337.
14. Connolly, M.L., Science, 221 (1983) 709.
15. Connolly, M.L., J. Mol. Graph., 11 (1993) 139.
16. Ewing, T.J., Makino, S., Skillman, A.G. and Kuntz, I.D., J. Comput.-Aided Mol. Design, 15 (2001) 411.
17. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.
18. Meng, E.C., Gschwend, D.A., Blaney, J.M. and Kuntz, I.D., Proteins, 17 (1993) 266.
19. Zavodszky, M.I.S.P.C., Korde, R. and Kuhn, L.A., J. Comput.-Aided Mol. Design, 16 (2002) 883.
20. Joseph-McCarthy, D. and Alvarez, J.C., Proteins, 51 (2003) 189.
21. Nissink, J.W.M., Verdonk, M.L. and Klebe, G., J. Comput.-Aided Mol. Design, 14 (2000) 787.
22. Bruno, I.J., Cole, J.C., Lommerse, J.P., Rowland, R.S., Taylor, R. and Verdonk, M.L., J. Comput.-Aided Mol. Design, 11 (1997) 525.
23. Rantanen, V.V., Gyllenberg, M., Koski, T. and Johnson, M.S., J. Comput.-Aided Mol. Design, 17 (2003) 435.
24. Bitetti-Putzer, R., Joseph-McCarthy, D., Hogle, J.M. and Karplus, M., J. Comput.-Aided Mol. Design, 15 (2001) 935.
25. Girones, X., Amat, L. and Carbo-Dorca, R., J. Mol. Graph. Model, 16 (1998) 190.
26. Girones, X., Carbo-Dorca, R. and Mezey, P.G., J. Mol. Graph. Model, 19 (2001) 343.
27. Perez, C. and Ortiz, A.R., J. Med. Chem., 44 (2001) 3768.
28. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A., J. Am. Chem. Soc., 117 (1995) 5179.
29. Lattman, E.E., Optimal sampling of the rotation function. In: The Molecular Replacement Method, Rossmann, M.G. (ed.), Gordon and Breach, Science Publishers Inc., New York, 1972, pp. 179–185.

30. Nelder, J.A. and Mead, R., Comput. J., 7 (1965) 308.
31. Davis, G., Mallat, S. and Avellaneda, M., Constr. Approx., 13 (1997) 57.
32. Davis, G., Department of Mathematics, Courant Institute of Mathematical Sciences, New York University, New York, 1994.
33. Besalu, E., Girones, X., Amat, L. and Carbo-Dorca, R., Acc. Chem. Res., 35 (2002) 289.
34. Rarey, M., Kramer, B. and Lengauer, T., Bioinformatics, 15 (1999) 243.
35. Murcia, M. and Ortiz, A.R., J. Med. Chem., 47 (2004) 805.
36. Stewart, J.J., J. Comput.-Aided Mol. Design, 4 (1990) 1–105.
37. Dewar, M.J., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., J. Am. Chem. Soc., 107 (1985) 3902.
38. Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., Eur. J. Biochem., 80 (1977) 319.
39. Bursulaya, B.D., Totrov, M., Abagyan, R. and Brooks III, C.L., J. Comput.-Aided Mol. Design, 17 (2003) 755.
40. Joseph-McCarthy, D., Thomas, B.E.t., Belmarsh, M., Moustakas, D. and Alvarez, J.C., Proteins, 51 (2003) 172.
41. Bashford, D. and Case, D.A., Annu. Rev. Phys. Chem., 51 (2000) 129.
42. Wriggers, W., Milligan, R.A. and McCammon, J.A., J. Struct. Biol., 125 (1999) 185.
43. Wriggers, W., Milligan, R.A., Schulten, K. and McCammon, J.A., J. Mol. Biol., 284 (1998) 1247.
44. De-Alarcon, P.A., Pascual-Montano, A., Gupta, A. and Carazo, J.M., Biophys. J., 83 (2002) 619.
45. DeWitte, R.S. and Shakhnovich, E.I., J. Am. Chem. Soc., 118 (1996) 11733.
46. Gohlke, H. and Klebe, G., Curr. Opin. Struct. Biol., 11 (2001) 231.