

***Ab Initio* potential grid based docking: From High Performance Computing to *In Silico* Screening**

Marc R. de Jonge¹, H. Maarten Vinkers¹, Joop H. van Lenthe², Frits Daeyaert¹,

Ian J. Bush³, Huub J. J. van Dam⁴, Paul Sherwood⁴, Martyn F. Guest⁵

¹ MolMo Services BVBA, Campus Blairon 424, B-2300 Turnhout, Belgium
corresponding author: E-mail marc@molmo.be

² Theoretical Chemistry Group, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands

³ Science and Technology Facilities Council, Daresbury Laboratory, CSE, ARCG, Warrington, WA4 4AD, United Kingdom

⁴ Science and Technology Facilities Council, Daresbury Laboratory, CSE, CCG, Warrington, WA4 4AD, United Kingdom

⁵ Advanced Research Computing Division (ARCCA), Cardiff University, 39-41 Park Place, Cardiff CF10 3BB, Wales United Kingdom

Abstract

We present a new and completely parallel method for protein ligand docking. The potential of the docking target structure is obtained directly from the electron density derived through an *ab initio* computation. A large subregion of the crystal structure of Isocitrate Lyase, was selected as docking target. To allow the full *ab initio* treatment of this region special care was taken to assign optimal basis functions. The electrostatic potential is tested by docking a small charged molecule (succinate) into the binding site. The *ab initio* grid yields a superior result by producing the best binding orientation and position, and by recognizing it as the best. In contrast the same docking procedure, but using a classical point-charge based potential, produces a number of additional incorrect binding poses, and does not recognize the correct pose as the best solution.

1 Introduction

1.1 Docking

Molecular docking is one of the best established computational methods to predict the affinity of small molecules towards a target enzyme or protein structure. The basic tenet of the method is that for a biologically relevant effect the molecule has to find an energetically favourable binding geometry, orientation, and position on or inside the target. This goes both for classical 'lock-and-key' interactions[1], where the geometry for the enzyme is independent of the substrate, and for the related 'induced-fit' type interactions[2], where the lock geometry is forced by the substrate or an allosteric effector. The docking method has become a widely used method for rational drug design because, theoretically at least, it allows the evaluation of completely novel compounds against any target system for which the 3D structure is known in sufficient detail.

As drugs are very often enzyme inhibitors that need to bind to their target in a specific way, pharmaceutical research has produced many computational methods for determining this optimal geometry of the complex formed by inhibitor and target enzyme in its search of new drugs. In essence all docking methods generate a score

for the interaction between the systems by a function that provides a simplified representation of the target or of some properties of the binding cavity inside the target, where this target is based on an X-ray, or occasionally NMR, derived structure of the enzyme. The scores use either a molecular mechanics (also known as force-field) based classical model for the interaction energy (see for instance [3,4,5,6]) or an empirical model that assigns a numerical score to the different recognized fragments of the ligand (for instance [7,8]) or the interacting atom types [9] and deriving a quality of fit measure from known protein-substrate and protein-inhibitor complexes.

Many improvements have been made to these docking methods during the past years, especially with respect to the global optimization strategies by which positions, orientations, and internal geometries of the docked ligands are generated, and with respect to the level of detail with which molecules or fragments are parameterized. However, so far we find one aspect of optimization curiously lacking: Even the most elaborate physical scoring functions determine electrostatic potentials from a point-charge model for the enzyme interior, also for the very closest contacts between enzyme and inhibitor. Obviously in the mid-1990's deriving the potential grid from such a model was the maximum that was practically achievable for rational drug design purposes. But, with current hardware we think this particular simplification of the problem may no longer be warranted. Therefore we will develop and apply a new scoring function for docking that uses a full *ab initio* computed electrostatic potential for a more interaction energy estimate as docking score. By this we hope to avoid some of the artefacts introduced by the classical approach, especially the large number of computed docking poses that needs to be taken into account to avoid missing the correct one.

1.2 Potential grid docking

In practical docking problems the interaction energies between a few thousand protein atoms of the target and up to a hundred atoms of the inhibitor need to be evaluated. This has to be repeated for each geometry-orientation point needed in the search algorithm. In a typical pharmaceutical *in silico* screening study this has to be subsequently repeated for many thousands of molecules. The number of mathematical operations for a single energy is $O(10^6)$ ($N_{\text{target}} \cdot N_{\text{ligand}}$), including a square root for the distance. The total number of single-point energy computations will therefore easily reach $O(10^{11})$. Some approximations are therefore necessary because even the simple expression of an interatomic interaction using only Coulomb and Van der Waals interactions (1) is expensive to compute.

$$E_{i,j} = C \frac{q_i q_j}{r_{i,j}} - \frac{B_{i,j}}{r_{i,j}^6} + \frac{A_{i,j}}{r_{i,j}^{12}} \quad (1)$$

In (1) the $A_{i,j}$ and $B_{i,j}$ describe the Van der Waals parameters for the interaction between target atom i and ligand atom j (See for instance [5]). They can be approximated by a geometric mean of coefficients depending on one atom type only:

$$A_{i,j} \simeq \sqrt{a_i} \sqrt{a_j} \quad (2)$$

The grid based algorithms can then reduce the number of operations through precomputing the protein part of the potentials. As the normal situation in docking is that the target enzyme has $O(10^4)$ atoms whereas the ligands have $O(10^2)$ atoms, the reduction of computation is in practice better than the reduction from N^2 to N .

dependence that the simple i-j interaction would suggest. The three precomputed Coulomb and Van der Waals dispersion and repulsion grids consist of the sum over all target atoms i of (1):

$$\begin{aligned} V_{cou\ k,l,m} &= C \sum_{i=1}^{N_{target}} \frac{q_i}{r_{i,klm}} \\ V_{disp\ k,l,m} &\simeq \sum_{i=1}^{N_{target}} \frac{\sqrt{b_i}}{r_{i,klm}^6} \\ V_{rep\ k,l,m} &\simeq \sum_{i=1}^{N_{target}} \frac{\sqrt{a_i}}{r_{i,klm}^{12}} \end{aligned} \quad (3)$$

(Another approach for grid generation, with a separate grid for each target atom type 'i', is used in [3], this improves accuracy but requires much more memory).

The interaction energy of a ligand atom j located at grid cell k,l,m with all target atoms is then as simple linear combination of three grid values:

$$E_j \simeq V_{cou\ k,l,m} q_j - V_{disp\ k,l,m} \sqrt{b_j} + V_{rep\ k,l,m} \sqrt{a_j} \quad (4)$$

In this study we will replace the V_{cou} grid, as above derived from the partial point charges of the target atoms, with an *ab initio* computed electrostatic potential. We will compare the resulting energy scores and docking poses of a known bound inhibitor in the *ab initio* and the point charge based potentials.

1.3 Test target systems

For testing the approach we will use a simple and a more complicated system: a single water molecule and the strongly polar, Mg^{2+} containing, binding site of *mycobacterium tuberculosis* Isocitrate Lyase (ICL), as described in Protein Databank [10] entries 1F61, 1F8I and 1F8M [11]. Hydrogens were added to the crystal structure, and the geometry was optimized by means of a molecular mechanics relaxation of the internal strain.

2 Methods

2.1 Existing flexible ligand - rigid enzyme docking

We will use a genetic algorithm for flexible ligand docking as implemented in GACS[12] for the docking experiments. The GACS program takes precomputed grids from any source as input for its search for optimal binding orientation and geometry. The program has been used with both 'soft' (4-8) or 'hard' (6-12) Lennard-Jones potentials (Equation 3). For this study we will only use the 6-12 potential; as we are docking a known bound ligand back into its binding site there is no need for the 'fuzziness' of the soft potential that is often useful in poorly constrained or flexible targets. The docking of ligands is performed by means of an evolutionary optimization[13] where the fitness of the orientation is the grid-based energy of the compound, incremented with an internal strain component (resulting from the torsions and out of plane forces of the flexible ligand). The real-valued chromosomes that describe a docking orientation in the evolution thus consist of 6+n real numbers: three Euler angles, a 3D displacement vector, and n torsion/out of plane angles. This

algorithm has been employed, with classical electrostatics, in the past in many docking studies and is robust with respect to protein and ligand electronic properties.

2.2 GAMESS-UK grid generation

The most important reason *ab initio* grids are normally not used for docking is the sheer size of the computational problem. In order to produce a good electronic structure a very large number of atoms has to be taken into account: The size of the enzyme test system requires a careful setup of the system; the complete enzyme of ICL has about 18000 atoms, but from classical electrostatics models we know that a sub-region with a radius of 16Å or of about 2000 atoms is generally sufficient for docking studies.

In this study we have computed the potential function with GAMESS-UK, a package of *ab initio* programs [14]. The *ab initio* grid generation entails first calculating the wave function and subsequently analyzing that through the generation of potential and electron density grids. The most time consuming part is the calculation of the wave function, This involves an direct implementation [15] of either a DFT (B3LYP [16]) or a Hartree-Fock algorithm [17]. The 2-electron integrals are calculated as needed, multiplied by density matrix elements (P in (5)) and added into the Fock matrix elements. The most common approach to parallelize this problem is through replicated data parallelism: The same program runs on every node of the system, each maintaining an identical copy of the data structures. The computational load is divided across the processors, and inter-processor communication phases are introduced to share the results so that all processors have the same data at the start of the next phase. It involves the contraction of blocks of 4-index two-electron integrals with the density matrix P (Eq. 5).

$$F_{\mu\nu} = H_{\mu\nu}^0 + P_{\mu\nu} \sum_{\lambda\sigma} \langle \mu\nu | \sigma\lambda \rangle - \frac{1}{2} \langle \mu\lambda | \sigma\nu \rangle \quad (5)$$

For each block of integrals, with a range of indices $\mu\nu\sigma\lambda$, a number of elements of P is accessed and the result is summed into the appropriate elements of F. All nodes can work independently until the construction is complete, at which point a global summation of contributions to F is required. This approach towards parallelization is readily implemented and has been in use since 1986 [18].

However, as merely parallelizing the Fock-matrix builder being the most time consuming part of the calculation, does not allow scaling beyond 16 to 32 processors, we have progressively parallelized more of the code by introducing the Global Array [19] and Parallel Eigenvalue solver PeIGS [20] tools developed at Pacific Northwest National Laboratories. These allow similarity transforms and diagonalization to be executed in parallel. Scalability limitations now arise from the structure of the underlying serial program, specifically its heavy use of scratch file I/O and the costs of maintaining the replicated data structures. In some phases of the calculation, matrices which are temporarily distributed across the nodes (for example, in the parallel matrix multiply) must be re-replicated for the next phase of the calculation. In the current calculations the SGI ALTIX machine at SARA [21] was employed. The scaling behaviour of the code is shown to be perfect up to 30 processors (Fig. 1). Beyond that, the scaling deteriorates; this is probably due to the architecture of the machine (which employs 2 processors per node). The memory requirement in

replicated data parallel programs is the same per processor as that of the serial code. On the platform used this results in a limit of about 10000 basis functions, since a density matrix and a Fock matrix have to be stored and manipulated. A shared memory implementation of the parallel code is currently being developed at the STFC Daresbury Laboratory [22] to extend this upper limit [23]. This facility has a shared memory architecture machine that needs to store only one density and Fock matrix per node with 32 processors. An illustration of the performance of the new code in this environment, and for some larger cutouts of the enzyme, is given in figure 2. The poor result of the 20 Å duplicated memory run was due to the fact that only 12 cpu's of the node are effectively used because of memory limitations. After determining the HF solution the final parallel computation of the grid values is fast and efficient.

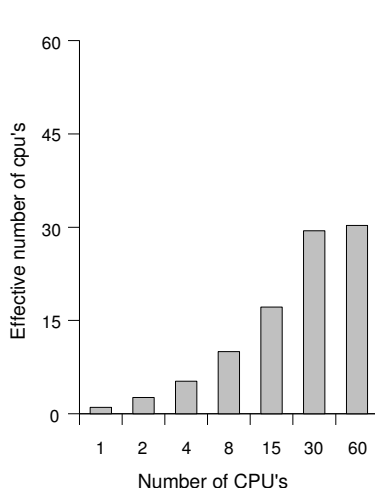


Figure 1: Scaling of a Hartree-Fock calculation of a 12 Å radius region of ICL: 737 atoms and 4500 orbitals.

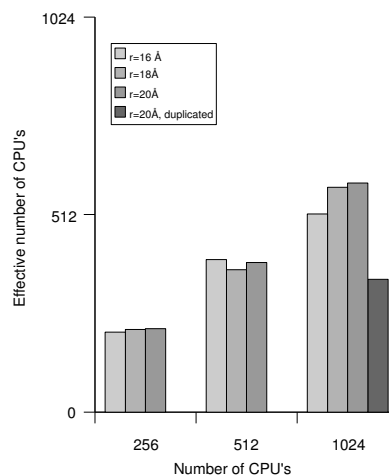


Figure 2: Scaling of Hartree-Fock calculation of 16, 18, and 20 Å radius subsets of the ICL enzyme with shared memory code.

Because of above mentioned practical constraints the number of basis functions in the quantumchemical computation needs to be limited. Obviously for the tests on the single water molecule a TZVP basis set was used without problems. But the protein structure of ICL has to be prepared with some care: After protonating the crystal structure, we first relaxed the whole system by means of 10000 steps of conjugate gradient minimization with the protein backbone atoms constrained to their original position by means of a quadratic displacement penalty of 20 kcal mol⁻¹ Å⁻². Subsequently a spherical subvolume is defined by removing (and capping whole) residues that are more than 16 Å away from the Mg²⁺ ion located in the active site.

It is this spherical volume that is modeled in GAMESS-UK. The number of atoms is still so large that not all atoms can be treated at maximum number of basis functions. We defined concentric shells with different basis functions, from TZVP in the central region, through 6-31g at 6 Å, to sto3g starting at 12 Å as illustrated in Figure 3. This shell structure reduces the size of the problem to less than 2000 atoms

with 10000 basis functions. The grid cells for the potential calculation are set to 0.2 Å in all directions, the grid size is 121x121x121 cells, centered around the Mg^{2+} ion (the potentials in this grid are computed from the whole 16Å volume).

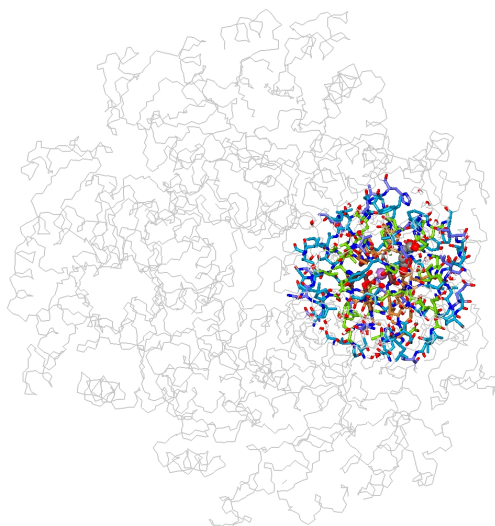


Figure 3: Whole ICL enzyme in light grey, and the 16 Å region with its different basis sets colour coded. The pink atom is the Mg^{2+} ion of the catalytic site, the grey atoms next to it are from the succinate reaction product.

3 Results and discussion

3.1 Potential field of a water molecule

To illustrate the effect of the different ways of computing an electrostatic potential we performed a small scale experiment on a single water molecule. Figure 4a shows the forcefield partial charge based electrostatic potential, and Figure 4b the same derived from NBO computed partial charges [24] from an *ab initio* HF. It is clear that the two methods, both using point charges produce very similar potentials (This is not surprising because the forcefield we used is designed to reproduce the NBO partial charges). In contrast the electrostatic potential computed directly from the QM results in Figure 4c (i.e. without projection onto point charges), is notably different. Not only inside the Van der Waals radius of the atoms, but also outside up to a distance that is very relevant for docking. Even in a 'toy' problem like this, the difference between a point-cloud description and a volume charge description is significant. This difference between the two point-charge derived potentials on one hand and the electron density based potential on the other, at 3 Å from the oxygen atom, is about 4 kCal/Mol (at the positions indicated by the black arrows in Figure 4).

The reason for this, unexpectedly large, difference lies in the fact that electrons of the atoms extend further than the Van der Waals radius, which means that, for the close contacts that are important in docking, the implicit assumption that all partial

Ab Initio potential grid based docking

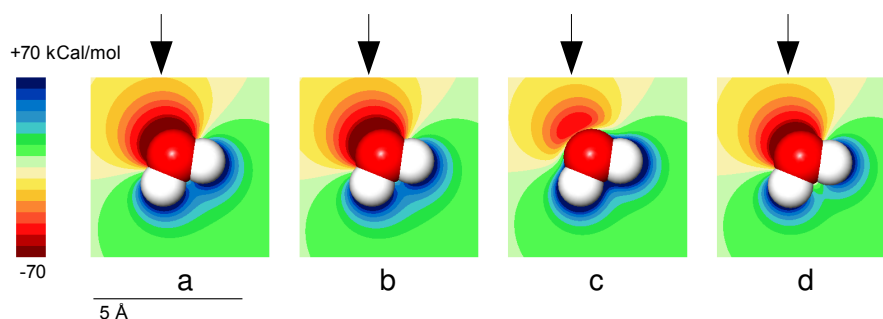


Figure 4: Electrostatic potential of a water molecule. a: Molecular Mechanics point charges, b: *ab initio* (RHF with NBO) point charges, c: *ab initio* electron density based, d: TIP4P (H_2O +dummy charge) water.

charge is contained within this volume is not entirely correct. Another effect is that the actual shape of the electronic structure is not really spherically symmetric (as assumed in the point charge potentials 4-a and 4-b). This latter effect also explains the relatively good potential for the four-point TIP4P [25] water model in Fig. 4d which is designed for, and commonly used in, molecular dynamics simulation. This water model has the point charge offset by 0.15 Å from the oxygen nucleus (towards the hydrogens). As a result its coulomb potential shows a better match with the *ab initio* potential at the arrow position, although it still resembles the NBO and forcefield point charge potentials closer to the oxygen atom.

3.2 Potential field of the Isocitrate Lyase enzyme binding site

The active site of ICL can be treated in the same way as the water molecule above, but as the Mg^{2+} ion is not balanced by a ligand (for docking the potential of the unliganded binding region is required) a net charge exists in this region, and consequently a net positive electrostatic potential is found everywhere near the central region. In Figure 5 the potentials computed from point-charges (a) and from the electron density (b) are shown. The red atoms in both binding sites are the oxygens of negatively charged carboxylate groups of glyoxylate and succinate in their crystal structure positions. The potential was computed without the succinate ligand, it is therefore promising for docking that the electrostatic potential shows local maxima at the position of its negatively charged carboxylate groups. There is quite a large difference in magnitude of the electrostatic potential visible, especially in close contact with the enzyme atoms. Another important feature is the generally narrower appearance of the potential extrema in the *ab initio* grid. This suggests that the energy landscape in which is scanned during the docking will have different optimal positions in the two approaches, and less freedom for (mis-)placement of compounds in the *ab initio* grid.

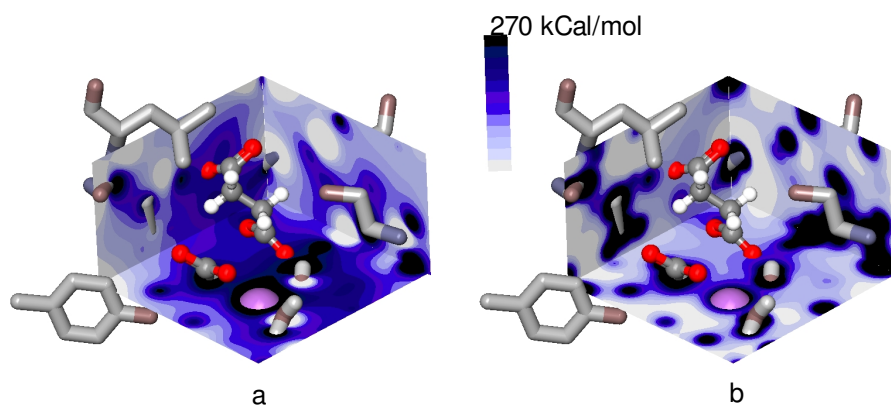


Figure 5: Three sections through the electrostatic potential of the central region of the ICL model. The point charge based model is shown in a, the ab initio computed potential in b. The Mg^{2+} ion, some relevant protein atoms, as well as the glyoxylate and succinate products from isocitrate lysis are shown.

3.3 Docking with the two types of electrostatic potentials

Docking of one of the products of isocitrate lysis, succinate, into the binding site of the enzyme is possible with both potential grids. Figure 6 makes clear however that the two approaches are not equivalent. The compound with the dark green carbon atoms indicate the predicted best binding orientations, in Figure 6a for the point charges based potential, and in Figure 6b for the electron density based potential. An important difference is the shift of both carboxylate groups in the point charge docking result: A hydrogen bond with the coordinated water molecule next to the pink Mg^{2+} ion is lost (arrow a-1), in favour of a 'promiscuous' coulomb interaction within the not very discriminating wide positive potential region (arrow a-2). In the electron density based field the docked ligand retains the interactions that are expected from the crystal structure (a hydrogen bond with a structural water molecule at b-1 and with an arginine side chain at b-3). The point charge based docking yields three more local optima with incorrect docking poses, with a maximum docking energy difference of 13 kCal/mol. The correct orientation is also found, but scores about 2 kCal/mol worse than the lowest energy one*. The electron density grid yields only one optimal binding mode. It should be noted that, when docking small polar molecules, the number of binding orientations found is generally far more than the crystal structures indicate. The fact that, rather than four, only the one correct orientation of succinate is identified is an important advantage of docking with this new potential.

* implying 3.5% of this conformation would be present in a Boltzmann distribution at 300K.

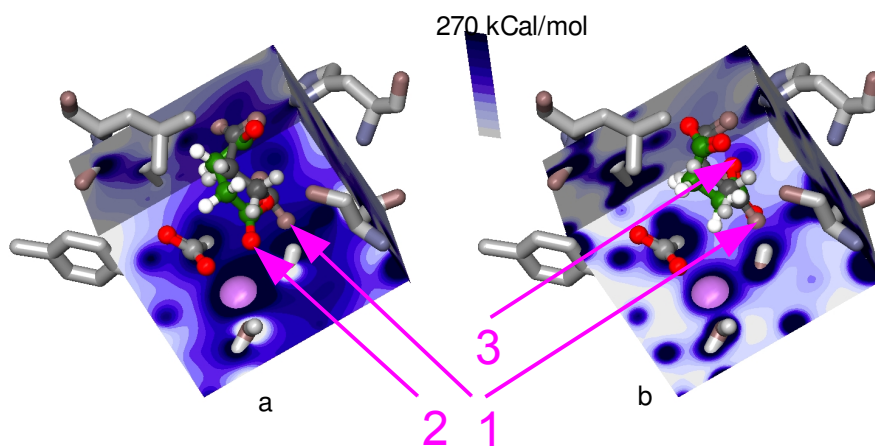


Figure 6: Docking result of succinate (green carbon atoms), a) in point charge potential field and b) in electron density potential field. For reference the dark grey crystal structure orientation of succinate is also shown in both binding sites. See text for the arrows 1,2 and 3.

4 Conclusions

We have demonstrated a parallel *ab initio* approach towards the generation of protein ligand docking grids. The whole computational procedure from grid generation to multiple ligand flexible docking can be performed in parallel. The use of electron density derived potential grids results in a substantial improvement of the quality of docking relative to a model derived from forcefield charges. In *in silico* drug design and *in silico* library screening hundreds of thousands of compounds are screened in parallel, and the amount of work spent on generating the *ab initio* grids is negligible (on the order of several thousands of cpu-hours) relative to the total computational effort involved in the subsequent docking of multiple ligand conformers of the whole library, which can easily amount a 100,000 cpu-hours.

We have also demonstrated that the essential characteristic of the method we propose is the direct generation of the grids for the docking algorithm from electron density without interposing a point-charge layer of approximation, thereby removing the artifacts of that approximation. The difference between direct grids and charge based potential fields is illustrated by using the same electronic structure but with back-projected point charges (eg. through NBO), which results in a potential field almost identical to the molecular mechanics model and not similar to the electron density grid based model. The actual docking procedure does not need to be modified in any way to make use of the *ab initio* potentials.

The close contacts between target and ligand in docking studies demand a better approximation for the electrostatic potential than the spherically symmetrical approximation commonly used.

Acknowledgments

Part of this work was carried out under the HPC-EUROPE Project (RII3-CT-2003-506079), with the support of the European Community Research Infrastructure Action under the FP6 Structure of the European Research Area Programme.

Part of this work was sponsored by the Stichting Nationale Computerfaciliteiten (NCF) for the use of supercomputer facilities, with financial support from the Netherlands Organization for Scientific Research, NWO.

We thank Luc Koymans for discussion on the protein function and protonation state.

References

1. Fischer, E.: Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der Deutschen chemischen Gesellschaft* **27** (1894) 2985-2933
2. Koshland, D.E.: Application of a Theory of Enzyme Specificity to protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **44** (1958) 98-104
3. Goodford, P.J.: A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *Journal of Medicinal Chemistry* **28** (1985) 849-857
4. Von Itzstein, M., Wu, W.-Y., Kok, G.B., Pegg, M.S., Dyason, J.C., Jin, B., Van Phan, T., Smythe, M.L., White, H.F., Oliver, S.W., Colman, P.M., Varghese, J.N., Ryan, D.M., Woods, J.M., Bethell, R.C., Hotham, V.J., Cameron, J.M., and Penn, C.R.: Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363** (1993) 418-423
5. Makino, S., Kuntz, I.D.: Automated Flexible Ligand Docking Method and Its Application for Database Search. *Journal of Computational Chemistry* **18** (1997) 1812-1825
6. Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R.: Development and Validation of a Genetic Algorithm for Flexible Docking. *Journal of Molecular Biology* **267** (1997) 727-748
7. Boehm, H.-J.: The Development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex. *Journal of Computer-Aided Molecular Design* **8** (1994) 243-256
8. Wang, R., Liu, L., Lai, L., Tang, Y.: SCORE: A New Empirical Method for Estimating the Binding Affinity of a Protein-Ligand Complex. *Journal of Molecular Modeling* **4** (1998) 379-394
9. Gohlke, H., Hendlich, M., Klebe, G.: Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology* **295** (2000) 337-356
10. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E.: The Protein Data Bank. *Nucleic Acids Research* **28** (2000) 235-242
11. Sharma, V., Sharma, S., Zu Bentrup, K.H., McKinney, J.D., Russell, D.G., Jacobs Jr., W.R., Sacchettini, J.C.: Structure of Isocitrate Lyase, a Persistence Factor of Mycobacterium Tuberculosis. *Nature Structural Biology* **7** (2000) 663-668

12. Daeyaert, F., de Jonge, M.R., Heeres J., Koymans L., Lewi P., Vinkers M.H., Janssen P.A.: A pharmacophore docking algorithm and its application to the cross-docking of 18 HIV-NNRTI's in their binding pockets. *Proteins* **54** (2004) 526-533
13. Holland, J. H.: *Adaptation in natural and artificial systems*. MIT Press (1992)
14. Guest, M.F., Bush, I. J., van Dam, H.J.J., Sherwood, P., Thomas, J.M.H., van Lenthe, J.H., Havenith, R.W.A., Kendrick, J.: The GAMESS-UK electronic structure package: algorithms, developments and applications. *Molecular Physics* **103** (2005) 719-747 (see:<http://www.cfs.dl.ac.uk/gamess-uk/index.shtml>)
15. Almlöf, J., Faegri, K. Jr., Korsell, K. J.: Principles for a direct SCF approach to LCAO-MO *Ab-Initio* calculations. *Journal of Computational Chemistry* **3** (1982) 385-399
16. Stephens, P.J., Devlin, F.J., Chabalowski, C.F., Frisch, M.J.: *Ab Initio* calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *Journal of Physical Chemistry* **98** (1994) 11623-11627
17. Jensen, F.: *Introduction to Computational Chemistry*, Wiley Chichester New York (1999)
18. Guest, M. F., Harrison, R. J., van Lenthe, J. H., van Corler, L. C. H.: Computational chemistry on the FPS-X64 scientific computers. *Theoretica Chimica Acta* **71** (1987) 117-148
19. Nieplocha, J., Harrison, R.J., Littlefield, R.J.: Global Arrays: A nonuniform memory access programming model for high-performance computers. *Journal of Supercomputing* **10** (1996) 169-189
20. Dhillon, I., Fann, G., Parlett, B.: Application of a new algorithm for the symmetric eigenproblem to computational quantum chemistry. *Proceedings of the Eighth SIAM Conference on Parallel Processing for Scientific Computing* (1997)
21. NWO, Stichting Nationale Computer Faciliteiten, SARA, Amsterdam. <http://www.sara.nl/userinfo/aster/description/index.html>
22. Science & Technology Facilities Council, Daresbury Laboratory. <http://www.scitech.ac.uk/About/Struc/Locs/DL/home.aspx>
23. Bush, I.J.: *New Fortran Features: The Portable Use of Shared Memory Segments*. HPCx Technical Report, STFC Daresbury Laboratory (2007)
24. Reed, A.E., Curtiss, L.A., Weinhold, F.: Intermolecular interactions from a natural bond orbital, donor acceptor viewpoint. *Chemical Reviews* **88** (1988) 899-926
25. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., : Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79** (1983) 926-935