SUPPLEMENTARY SECTION

Rational design of RAR selective ligands revealed by RAR crystal structure

Pierre Germain[*], Sabrina Kammerer[*], Efrén Pérez[1], Carole Peluso-Iltis, David Tortolani[2a], F. Christopher Zusi[2b], John Starrett[2b], Philippe Lapointe[2c], Jean-Paul Daris[2c], Anne Marinier[2c], Angel R. de Lera[1], Natacha Rochel and Hinrich Gronemeyer[‡]

Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)/CNRS/INSERM/ULP, BP 10142, 67404 Illkirch Cedex, C. U. de Strasbourg, France;

and

[1] Universidade de Vigo, Facultade de Química, 36200 Vigo, Spain;

and

[2a] Bristol-Myers Squibb, Pharmaceutical Research Institute, PO Box 4000, Princeton, NJ 08543-4000; [2b] Bristol-Myers Squibb, 5 Research Parkway, Wallingford, Ct 06492, USA; and [2c] Bristol-Myers Squibb, 100 Boul de L'Industrie, Candiac, PQ, Canada

1. SOFTWARE USED

CHARMM (Brooks *et al*, 1983; MacKerell *et al*, 1998)

QUANTA package (Molecular Simulations Inc., 1997)

DENZO and SCALEPACK from the HKL2000 package (Otwinowski and Minor, 1997)

CNS solve package (Brunger *et al*, 1998)

O (Jones *et al*, 1991)

GOLD (Jones *et al*, 1997)

VOIDOO (Kleywegt and Jones, 1994)

PYMOL (http://www.pymol.org)

SETOR (Zhang and Koshland, 1996)

MOLSCRIPT (Kraulis, 1991)

RASTER3D (Merrit and Murphy, 1994)

MSMS (http://www.scripps.edu/~sanner)

DINO (http://www.bioz.unibas.ch/~xray/dino)

2. LIGANDS AND PLASMIDS

TTNPB	Biomol
BMS453, BMS641, BMS701, BMS987, BMS009	US patent 6,624,154; World patent # 98-046,228; World patent # 00-017,147; European patent # 661,259
UVI2007 3-Chloro-4-[2-(5,6,7,8-tetrahydro-5,5,8,8-tetramethylnaphthalen-2-yl)propenyl]benzoic acid	synthesized as described below
pET15b-RARγ LBD	gift of Thierry Lerouge (IGBMC)
(RARE)3x-tk-luc	gift of Patrick Balaguer (U. Montpellier)

3. PROTEIN PURIFICATION, CRYSTALLIZATION AND DATA COLLECTION

The His-tagged hRARγ LBD (173-409) was purified to 98% homogeneity as described (Bourguet *et al*, 1995). TTNPB was added at 2 fold molar excess, incubated overnight at 4°C and concentrated to 7.5mg/ml. Crystallization was at 17°C (sitting-drop vapor-diffusion method). The 0.15ml reservoir contained 5% PEG8000, 0.05M MES (pH 5.6), 0.2M KCl, 0.01M $MgCl_2$; drops were formed by mixing 1.5µl of protein solution with 1.5µl of reservoir solution. One orthorhombic crystal was frozen in liquid ethane after cryoprotection using a mixture of 25% glycerol in reservoir solution. X-ray diffraction data were collected (Synchrotron Light Source, Zürich, Switzerland) and processed with DENZO and SCALEPACK from the HKL2000 package.

4. STRUCTURE REFINEMENT.

Structure determination and refinement was done with CNS solve package (Brunger *et al*, 1998) using the hRARγ LBD complex (PDB ID, 3LBD) as search model. The top solution had a correlation coefficient of 58 (next highest solution, 18). Data between 20 Å and 2.1 Å were included in a rigid body refinement with CNS. A first simulated annealing followed by

an individual temperature factor refinement and a least square minimization led to a R_{free}/R_{crys} of 32.7%/27.8%. The calculated electron density maps (α_A-weighted $2F_{obs}-F_{calc}$ and $F_{obs}-F_{calc}$ maps) allowed placement of the ligand that was subsequently included in the refinement. The final steps comprised essentially manual rebuilding, conformational changes of side chains using the program O (Jones *et al*, 1991) and individual isotropic temperature factor refinement. The electron density maps were finally calculated and allowed location of the solvent molecules. The final model contains 232 residues (1831 non-hydrogen atoms), one ligand (26 non-hydrogen atoms), and 128 water molecules and has been refined to R_{crys}=21.3% and R_{free}=25.6% (10% of the data excluded from the refinement). According to PROCHECK (Laskowski *et al*, 1993) the refined model shows no Ramachandran plot outliers, 97.7% and 6.3% of the residues in the most favored and additionally allowed regions, respectively. No clear density was obtained for residues 362-364 and 175-177.

Data collection	
X-ray source	SLS, beamline X06SA
wavelength, Å	0.8856
resolution, Å	2.1
completeness, %	98.7
multiplicity	3.9
Refinement	
Rfree, % (9.7% of reflections)	25.35
Rcryst, %	21.62
non hydrogen atoms	
protein	1831
ligand	26
water molecules	128
Stereochemistry	
Rmsd bond length, Å	0.006
Rmsd bond angles, °	1.136
Rmsd improper angles, °	0.753
Rmsd dihedral angles, °	18.99
Ramachandran plot, regions	
Most favored, %	93.7
Additionally allowed, %	6.3

Table 1. Statistics of crystallographic analysis and refinement

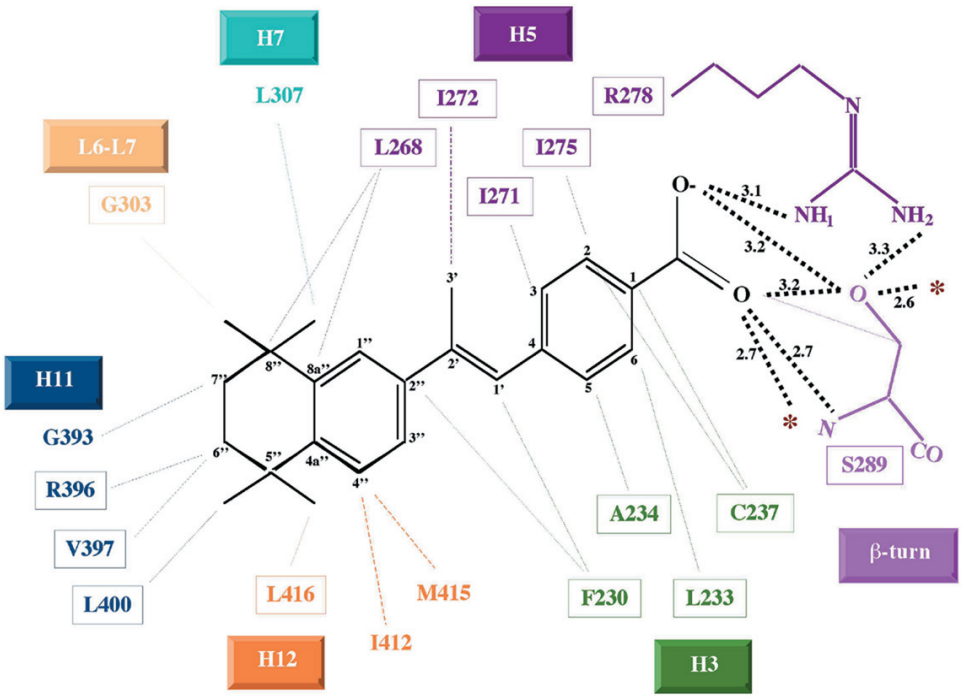


Figure 1 : Schematic diagram of the Van der Waals contacts (4,5 Å cut-off) and hydrogen bonds between protein and ligand. For each ligand carbon numbered only the shortest contact is shown. A solid box denotes residues within a 4.0 Å distance. The lengths of hydrogen bonds are indicated. Residues belonging to the same helix are given in the same colour code.

5. DOCKING EXPERIMENTS : VALIDATION OF THE GOLD ALGORYTHM

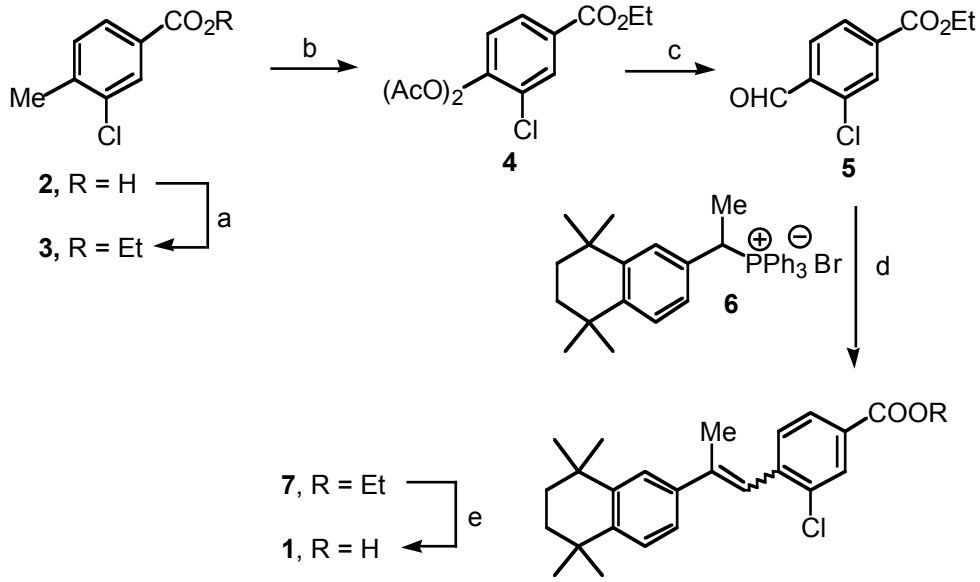
To validate the energy function of GOLD for retinoid docking the fitness scores were determined for docking back TTNPB into the RAR pocket determined by crystal structure analysis, and 9-cis RA and ATRA in the LBP of RAR(PDB 3lbd and 2lbd respectively). For these positive controls GOLD calculated fitness scores between 69 and 72. Inspection of the individual solutions predicted by GOLD revealed in all cases an accurate positioning of the ligand with the carboxylate properly anchored at RAR$_2$R$_{269}$ and S$_{280}$ and correct contacts within the LBP. The rms deviation values between the experimentally determined model and the corresponding solutions by GOLD ranged between 0.2 Å and 0.6 Å. As negative control we asked GOLD to dock a panRAR antagonist (BMS009) into the holo pocket of RAR The corresponding scores were below zero reflecting obvious steric clashes. These results support that the GOLD energy function is useful to predict proper docking of retinoids into RAR LBPs.

6. SYNTHESIS OF UVI2007

3-Chloro-4-[2-(5,6,7,8-tetrahydro-5,5,8,8-tetramethylnaphthalen-2-yl)propenyl]benzoic Acid (UVI2007, 1) was synthesized following the procedure previously described for the parent retinoid TTNPB (Loeliger *et al*, 1980), namely a Wittig condensation of phosphonium salt **6** and ethyl 3-chloro-4-formylbenzoate **5** (Schultz *et al*, 1976) (Scheme 1). The preparation of aldehyde **5** required the oxidation of the methyl group of the corresponding halogenated benzoate **3** (obtained by quantitative esterification of the commercially available carboxylic acid **2**). Generation of the ylide derived from **6** by treatment with *n*-BuLi at 0 ºC, followed by addition of aldehyde **5** and stirring to 25 ºC for 2 h provided the arotenoid skeleton of **7**. In contrast to the non-halogenated aldehyde, which afford a single geometric isomer, the Wittig reaction involving **5** is non-stereoselective, and provided a 1:1 mixture of

E/*Z* stilbene isomers which could not be separated and was used in the next step. Saponification of the mixture obtained above by heating with 2M KOH in MeOH provided a similar mixture of carboxylic acid isomers **1**, which were separated by HPLC (Waters Spherisorb S5NH_2, 2 mL/min, 10x250 mm, 2.5:98:0.5 MeOH/CH_2Cl_2/AcOH, t_R= 10 min (*Z*); t_R= 15 min (*E*)).

Scheme 1



a) EtOH, H_2SO_4, 90 ˚C, 4 h, 75%. b) AcOH, Ac_2O, H_2SO_4, CrO_3, 0 ˚C, 1 h. c) H_2SO_4, H_2O, EtOH, 100 ˚C, 45 min, 39% (two steps). d) *n*-BuLi, 0 to 25 ˚C, 1.5 h. e) 2M KOH, MeOH, reflux, 30 min, 70% combined yield.

Ethyl 3-Chloro-4-methylbenzoate (3). To a solution of acid **2** (1.5 g, 8.79 mmol) in EtOH (45 mL) was added conc. H_2SO_4 (1 mL), and the resulting solution was heated to reflux for 2 h. After cooling down to 25 ºC, it was diluted with EtOAc, and the organic layer was extracted with EtOAc (3x). The combined organic layers were washed with an aqueous saturated $NaHCO_3$ solution (3x) and with brine, dried over Na_2SO_4, filtered and evaporated to

dryness. The residue was purified by column chromatography (silicagel, 90:10 hexane/EtOAc) to afford ester **3** (1.32 g, 75% yield). **^{1}H NMR** ($CDCl_3$, 400.13 MHz) δ 1.39 (t, J = 7.1 Hz, 3H, $CH_2C\underline{H}_3$), 2.42 (s, 3H, CH_3), 4.35 (q, J = 7.1 Hz, 2H, $C\underline{H}_2CH_3$), 7.28 (d, J = 7.9 Hz, 1H, H_5), 7.81 (dd, J = 7.9, 1.6 Hz, 1H, H_6), 8.00 (d, J = 1.6 Hz, 1H, H_2) ppm. **^{13}C NMR** ($CDCl_3$, 100.62 MHz) δ 14.3 (q), 20.3 (q), 61.1 (t), 127.7 (d), 129.7 (s), 130.1 (d), 130.8 (d), 134.5 (s), 141.3 (s), 165.5 (s) ppm. **IR** (NaCl): ν 2982 (w, C-H), 1722 (s, C=O), 1284 (s), 1253 (s), 1111 (m) cm^{-1}. **MS (EI$^+$)**: *m/z* (%) 200 ($[M]^+$, 2), 198 ($[M]^+$, 7), 170 (16), 155 (21), 153 (100), 125 (28), 89 (15). **HRMS**: calcd. for $C_{10}H_{11}{}^{35}ClO_2$ $(M)^+$: 198.0448; found: 198.0440.

Ethyl 3-Cloro-4-formylbenzoate (5). Conc. H_2SO_4 (1.16 mL) was slowly added to a cooled (-5 ºC) solution of ester **3** (1.0 g, 5.05 mmol), acetic acid (7.75 mL) and acetic anhydride (7.75 mL). Then CrO_3 (1.36 g, 13.6 mmol) was added portionwise and stirring was maintained for 20 min at 0 ºC. The reaction mixture was diluted with EtOAc and H_2O, and the organic layer was extracted with EtOAc (3x). The combined organic extracts were washed with an aqueous saturated Na_2CO_3 solution (2x) and brine, dried over Na_2SO_4, filtered and evaporated to dryness. The residue was dissolved in a mixture of H_2O (2.5 mL) and EtOH (2.5 mL), and conc. H_2SO_4 (0.25 mL) was added. The solution was refluxed for 45 min, cooled down to 25 ºC, diluted with H_2O, and extracted with EtOAc (3x). The combined organic extracts were washed with an aqueous saturated $NaHCO_3$ solution (3x) and brine, dried over Na_2SO_4, filtered and evaporated to dryness. The residue was purified by column chromatography (silicagel, 85:15 hexane/EtOAc), to afford aldehyde **5** (0.415 g, 39% yield). **^{1}H NMR** ($CDCl_3$, 400.13 MHz) δ 1.42 (t, J = 7.1 Hz, 3H, $CH_2C\underline{H}_3$), 4.41 (q, J = 7.1 Hz, 2H, $C\underline{H}_2CH_3$), 7.97 (d, J = 8.1 Hz, 1H, H_5), 8.06 (dd, J = 8.1, 1.3 Hz, 1H, H_6), 8.12 (d, J = 1.3 Hz, 1H, H_2), 10.52 (s, 1H, CHO) ppm. **^{13}C NMR** ($CDCl_3$, 100.62 MHz) δ 14.2 (q), 62.0 (t),

128.1 (d), 129.3 (d), 131.7 (d), 135.0 (s), 136.3 (s), 137.7 (s), 164.4 (s), 189.2 (d) ppm. **IR** (NaCl): □ 2983 (w, C-H), 1726 (s, C=O), 1698 (s, C=O), 1277 (s), 1108 (m) cm^{-1}. **MS (EI^+)**: *m/z* (%) 214 ($[M]^+$, 5), 212 ($[M]^+$, 38), 186 (28), 185 (26), 184 (74), 183 (67), 169 (63), 168 (11), 167 (100), 141 (22), 139 (68), 111 (51), 75 (88). **HRMS**: calcd. for $C_{10}H_9{}^{35}ClO_3$ $(M)^+$: 212.0240; found: 212.0248.

Ethyl 3-Chloro-4-[2-(5,6,7,8-tetrahydro-5,5,8,8-tetramethylnaphthalen-2-yl)-propenyl]-benzoate (7). To a solution of phosphonium salt **6** (0.32 g, 0.61 mmol) in THF (1 mL), was added dropwise, at 0 ºC, *n*-BuLi (0.39 mL, 1.45 M in hexane, 0.57 mmol). After stirring for 1 h, aldehyde **5** (0.1 g, 0.47 mmol) in THF (0.4 mL) was added, and the mixture was allowed to warm up to 25 ºC. After 2 h stirring, the mixture was poured over MeOH/H_2O and extracted with hexane (3x). The combined organic extracts were washed with brien, dried over Na_2SO_4, and evaporated to dryness. Crystallization of the residue with hexane/EtOAc provided the desired product as an inseparable mixture of geometric isomers.

3-Chloro-4-[2-(5,6,7,8-tetrahydro-5,5,8,8-tetramethylnaphthalen-2-yl)propenyl]benzoic Acid (1). A solution of 2M KOH in MeOH (14 mL) was added to ester **7** obtained above (173 mg, 0.42 mmol), and the mixture was heated to reflux for 1 h. After cooling down to 25 ºC, 10% HCl was added, and the mixture was extracted with CH_2Cl_2 (3x). The combined organic layers were washed with H_2O and brine, dried over Na_2SO_4 filtered and evaporated to dryness. The residue was purified by chromatography (silicagel, 90:10 CH_2Cl_2/MeOH), to afford acid **1** (112 mg, 70% yield), as a mixture of geometric isomers which were separated by HPLC (Waters Spherisorb $S5NH_2$, 2 mL/min, 10x250 mm, 2.5:98:0.5 MeOH/CH_2Cl_2/AcOH, t_R= 10 min (*Z*); t_R= 15 min (*E*) and crystallized from hexane/EtOAc, to afford *E*-**1** as a white solid (mp: 175-177 ºC), and *Z*-**1** as a white solid (mp: 158-160 ºC).

Data for *E*-**1**. **^{1}H NMR** ($CDCl_3$, 400.13 MHz) δ 1.25 (s, 6H, 2xCH_3), 1.28 (s, 6H, 2xCH_3), 1.65 (s, 4H, 2xCH_2), 2.14 (s, 3H, CH_3), 7.19 (s, 1H), 7.28 (m, 2H), 7.42 (m, 2H), 7.93 (d, *J* = 8.0 Hz, 1H, H_5) , 8.09 (d, *J* = 1.1 Hz, 1H, H_2) ppm. **^{13}C NMR** ($CDCl_3$, 100.62 MHz) δ 17.5 (q), 31.8 (q, 2x), 31.9 (q, 2x), 34.2 (s), 34.4 (t), 35.0 (s), 35.1 (t), 123.1 (d), 123.4 (d), 124.2 (d), 126.6 (d), 127.9 (d), 128.6 (s), 130.8 (d), 131.1 (d), 134.5 (s), 139.7 (s), 140.9 (s), 142.4 (s), 144.8 (s), 144.9 (s), 170.9 (s) ppm. **IR** (NaCl): ν 2959 (m, C-H),), 2861 (m, C-H), 1693 (s, C=O), 1598 (m), 1415 (m), 1296 (s) 1261 (s) cm^{-1}. **MS (EI$^+$)**: *m*/*z* (%) 384 ([M]$^+$, 18), 382 ([M]$^+$, 50), 369 (35), 368 (24), 367 (100), 272 (11), 178 (11). **HRMS**: calcd. for $C_{24}H_{27}{}^{35}ClO_2$ $(M)^+$: 382.1700; found: 382.1700.

Data for *Z*-**1**. **^{1}H NMR** ($CDCl_3$, 400.13 MHz) δ 0.98 (s, 6H, 2xCH_3), 1.24 (s, 6H, 2xCH_3), 1.60 (m, 4H, 2xCH_2), 2.28 (d, *J* = 1.1 Hz, 3H, CH_3), 6.57 (s, 1H), 6.83 (d, *J* = 8.1 Hz, 1H), 6.90 (d, *J* = 1.8 Hz, 1H), 6.94 (dd, *J* = 8.1, 1.8 Hz, 1H), 7.21 (d, *J* = 8.2 Hz, 1H, H_5), 7.54 (dd, *J* = 8.2, 1.5 Hz, 1H, H_6), 8.05 (d, *J* = 1.5 Hz, 1H, H_2) ppm. **^{13}C NMR** ($CDCl_3$, 100.62 MHz) δ 25.9 (q), 31.5 (q, 2x), 31.8 (q, 2x), 33.9 (s), 34.1 (s), 34.9 (t), 35.0 (t), 122.6 (d), 124.8 (d), 126.5 (d), 127.1 (s), 127.2 (d), 127.4 (d), 130.9 (d), 131.5 (d), 133.8 (s), 137.1 (s), 142.7 (s), 143.4 (s), 144.4 (s), 144.5 (s), 170.9 (s) ppm. **IR** (NaCl): ν 2960 (s, C-H), 2861 (m, C-H),), 1691 (s, C=O), 1415 (m), 1298 (s), 1263 (s) cm^{-1}. **MS (EI$^+$)**: *m*/*z* (%) 384 ([M]$^+$, 19), 382 ([M]$^+$, 56), 396 (35), 368 (26), 367 (100). **HRMS**: calcd. for $C_{24}H_{27}{}^{35}ClO_2$ $(M)^+$: 382.1700; found: 382.1689.

6. REFERENCES CITED IN THE SUPPLEMENTARY SECTION□

Bourguet, W., Ruff, M., Bonnier, D., Granger, F., Boeglin, M., Chambon, P., Moras, D. and Gronemeyer, H. (1995) Purification, functional characterization, and crystallization of the ligand binding domain of the retinoid X receptor. *Protein Expr Purif* **6:** 604-608

Brooks, R.E., Bruccoleri, B.R., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J Comp Chem* **4:** 87-217

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5):** 905-921

Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **267:** 727-748

Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47 (Pt 2):** 110-119

Kleywegt, G.J. and Jones, T.A. (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr D* **50:** 178-185

Kraulis, P.J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Chrystallography* **24:** 946-950

Laskowski, R.A., Mac Arthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHEK : a program to check the stereochemical quality of protein structures. *J Applied Crystallogr* **26:** 283-291

Loeliger, P., Bollag, W. and Mayer, H. (1980) Arotinoids: a new class of highly active retinoids. *Eur. J. Med. Chem.* **15:** 9-15

MacKerell, A.D., Jr., Brooks, C.L., Nilsson, L., Roux, B., Won, Y., Karplus, M. and Schleyer, v.P.R., et al. (1998) CHARMM: the energy function and its parameterization with an overview of the program. *Encycl Comp Chem* **1:** 271-277

Merrit, E.A. and Murphy, M.E.P. (1994) Raster 3D version 2.0 - a program for photorealistic molecular graphics. *Acta Christallography* **D50**

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymology* **276:** 307-326

Schultz, E.M., Bicking, J.B., Deana, A.A., Gould, N.P., Strobauge, T.P., Watson, L.S. and Cragoe, E.J., Jr. (1976) (Vinylaryloxy)acetic acids. A new class of diuretic agents. 3. ((2-Nitro-1-alkenyl)aryloxy)acetic acids. *J Med Chem* **19:** 783-787

Zhang, T. and Koshland, D.E., Jr. (1996) Computational method for relative binding energies of enzyme-substrate complexes. *Protein Sci* **5:** 348-356