Electronic Supplementary Information

Racemic Progesterone: Predicted *in silico* and produced in the solid state

Robert W. Lancaster, Panagiotis G. Karamertzanis, Ashley T. Hulme, Derek A. Tocher, Douglas F. Covey, Sarah L. Price

Contents

1. Computational methodology and additional detailed results

Five conformations of progesterone were generated by constraining torsion angle φ1 (figure 1) to the values identified in the relaxed scan of figure S1 and optimizing the rest of the molecular geometry at the HF/6-31G (d,p) level of theory.

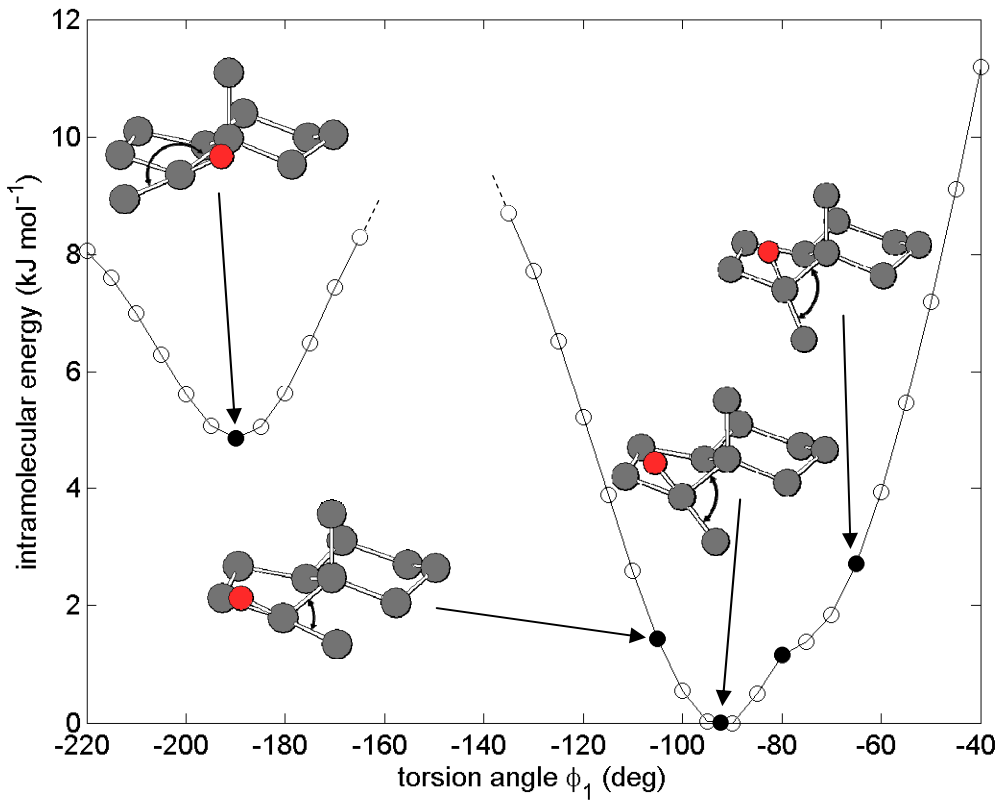


Figure S1 MP2/6-31G(d,p) intramolecular energy as a function of the rotation of the keto group for a molecular fragment comprising the C ring (saturated with hydrogen atoms), D ring and side chain of progesterone. Solid circles correspond to side chain conformations used in searching for stable packing arrangements; the molecular diagrams shown are drawn with the C20-C17 bond perpendicular to the plane of the figure. The discontinuity at -80 ° is due to the 21-methyl group rotation switching between local minima.

The optimization was followed by a single point energy evaluation at the MP2/6-31G(d,p) level and the resulting energy was used to calculate the intramolecular energy penalty $\Delta E^{\text{intra}}(\varphi_1) = E^{\text{intra}}(\varphi_1) - \min E^{\text{intra}}$, where the most stable conformation of energy $\min E^{\text{intra}}$ corresponded to φ_1=-93.3°. The intramolecular energy penalty for the five conformations used in the searches for stable packing

arrangements are shown in table S1. All *ab initio* calculations were performed with the suite of programs GAUSSIAN.[1] These conformers were held rigid throughout the generation and lattice energy minimization of putative crystal structures.

Table S1 Intramolecular energy penalty for the five conformations of progesterone used in rigid-body searches for low lattice energy crystal structures

Conformation (torsion φ_1, °)	HF/6-31G(d,p)	MP2/6-31G(d,p)[a]
-65.0	2.84	2.91
-80	1.53	1.45
-93.3 (*in vacuo*)	0.00	0.00
-110	1.87	2.29
-190	1.82	3.90
[a] Single point MP2/6-31G(d,p) energy at the HF/6-31G(d,p) constrained optimized minimum		

Table S1 shows that electron correlation is important in determining the relative stability of the two conformational minima for progesterone and this is probably because of the intramolecular dispersion interactions between the steroid rings and the side chain at position 17.

Initial densely-packed crystal structures were systematically generated in 40 common co-ordination environments with one crystallographically independent molecule spanning 20 of the most frequently occurring space groups using the program MOLPAK.[2] The lattice energy of up to 125 initial guesses per coordination environment was subsequently minimized with respect to the lattice lengths, angles, molecular positions and orientations using DMAREL.[3] The subset of enantiomorphic space groups included in the search ($P1$, $P2_1$, $C2$, $P2_12_12$, $P2_12_12_1$) cover 93.1% of the enantiomorphic structures in the Cambridge Structural Database. (A search of version 5.27, Nov 05 + Jan 06 update, for structures in the 65 enantiomorphic space groups returned 43535 hits (filters: only organics, no errors, no polymeric), of which 40532 were in the space groups included in the search). The electrostatic interactions were modelled *via* a set of atomic multipoles calculated from a distributed multipole analysis of the MP2/6-31G(d,p) charge density[4] using all terms up to hexadecapole. The repulsion-dispersion contribution to the lattice energy was calculated using an empirical exp-6 atom-atom potential with the parameters for the atomic types C, N, O and H (hydrogen connected to carbon) taken from Williams.[5,6] The multipole-multipole contributions were calculated with the Ewald summation technique for all slowly convergent terms, i.e. charge-charge, charge-dipole and dipole-dipole interactions. All higher order electrostatic interactions were summed to a 15 Å cut-off distance between the centres of mass of the molecules. This distributed multipole representation ensures an accurate description of the electrostatic interactions between the unperturbed charge densities of the molecules and is particularly important for the reliable modelling of the weak interactions of the carbonyl groups. The accuracy of this electrostatic model has shown to be vital in accurately modelling directional interactions, such as hydrogen bonds.[7]

The energy minimization was performed within the space group in which the initial guess was generated and so the obtained stationary points may correspond to saddle points once the symmetry constraints are removed. Saddle points were detected *via* the examination of the second derivative matrix for negative eigenvalues, or the presence of unstable elastic constants or negative phonon frequencies[8] and were rejected as the prediction of crystal structures with more than one molecule in the asymmetric unit remained beyond the scope of the present study.

For each search structure, duplicates were discarded following comparison of the Niggli reduced cell parameters,[9] lattice energies and densities. Essentially the same crystal structure was often found in more than one search, although the differences in the molecular conformation led to discrepancies in the unit cell parameters and markedly different simulated powder X-ray diffraction patterns. Despite the overall differences in the unit cells these structures may well lead to the same minimum once the intramolecular degrees of freedom and thermal effects are considered and thus were systematically identified by the similarity of their 15-molecule coordination sphere environments[10] with increased tolerance (80%). For each cluster, the structure with the lowest crystal energy was retained and the rest discarded. Following this clustering procedure, we identified 149 unique crystal structures of progesterone, within 15 kJ mol^{-1} of the global minimum, whose crystal energies (lattice energy + intramolecular energy penalty) and unit cell volume per molecule, are plotted in figure 2. The thorough consideration of the side chain flexibility and the superior intermolecular potential constitute a significant improvement over the previous crystal structure prediction study of progesterone[11] which showed limited predictive ability.

The calculations also show that progesterone is unlikely to adopt the conformation of the second minimum in the solid state, as the energy of the most stable putative structure that contains it is 6.5 kJ mol^{-1} less stable than the global minimum. This is in accord with statistical evidence on steroid conformations[12] and can be explained by the shielding of the 20-keto group by adjacent methyl and methylene groups of the bulk steroid in this conformer preventing the formation of stabilizing carbonyl···carbonyl interactions.

The five most stable racemic and enantiomorphic predicted structures are shown in table S2. To facilitate the analysis of relative stabilities the crystal energy is partitioned into its repulsion-dispersion, electrostatic and intramolecular energy components. It is well established that the inversion operator usually leads to more stable packing arrangements in accordance with statistical trends first recognized by Wallach more than a century ago.[13,14] However, in the case of progesterone, the experimental variation in density is small, and the variation of the repulsion-dispersion contribution to the crystal energy among the racemic and chiral crystal structures is limited, which implies that the stabilization of the centrosymmetric (racemic) crystals is not due to more efficient packing following Kitaigorodsky's Principle of Close Packing.[15] Contrary to chemical intuition, Table S2 indicates that the electrostatic interactions play a decisive role in determining the relative stability of the putative crystal structures. The two structures with the lowest electrostatic energy (first and third racemic structures) both exhibit planar, antiparallel carbonyl···carbonyl arrangements with particularly energetically favourable geometry (C···O distance of 3.2 Å), which is generated by inversion centres. This is discussed further in the next section of the ESI. The comparison of the closest crystal structures generated within the search and the experimental structures in Table S2 also demonstrates that the torsion angles defining the methyl-keto conformation differ significantly in the three known forms and from the closest conformational model used in the search. Further refinement of the molecular structure under the influence of the packing forces, by optimizing the crystal energy with respect to the two torsions defining the methyl-keto group, are in progress using a new methodology.[16] Preliminary results show that the refinement of the packing density allowed by varying these two angles further compresses the range of densities and repulsion-dispersion energies of the low

energy structures shown in Table S2 and increases the role of the electrostatic energy in favouring the new racemic structure. Finally, modelling the molecular distortions under the packing forces reveals that both side chain and 21-methyl group rotations in the enantiomorphic structures deviate from their *in vacuo* conformations significantly more compared to the racemic structure.

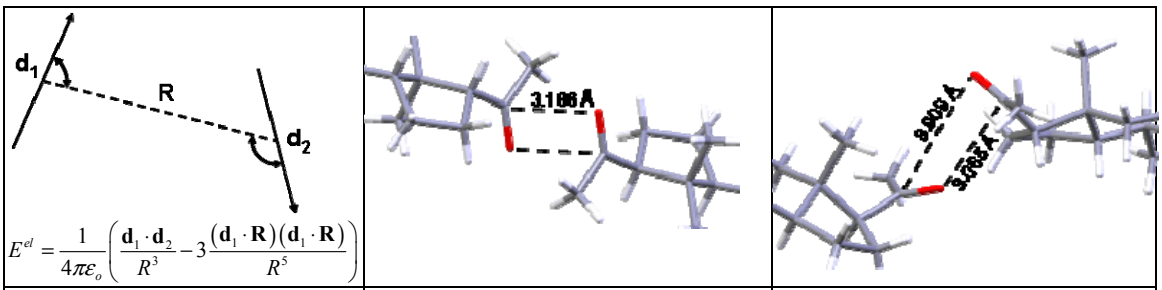
Table S2 Most stable predicted racemic and enantiomorphic progesterone structures. The three experimentally determined structures are also given for comparison with the closest structures found in the search (in bold).

		Crystal Energy[1] (kJ mol^{-1})					**Conventional Cell**				**Conformation**[2]	
structure[3]	space group	$U+\Delta E^{intra}$	U^{vdW}	U^{elec}	ΔE^{intra}	density (g cm^{-3})	a (Å)	b (Å)	c (Å)	β (°)	φ_1 (°)	φ_2 (°)
Racemic												
1, ak16	***P*2$_1$/*c***	**-136.08**	**-118.54**	**-17.54**	**0.00**	**1.162**	**11.221**	**7.577**	**21.171**	**93.37**	**-93.32**	**-10.22**
Racemic, 150 K	***P*2$_1$/*c***					**1.224**	**10.922**	**7.440**	**21.034**	**93.36**	**-87.79**	**-9.78**
2, ak1	*P*2$_1$/*c*	-134.64	-119.27	-15.37	0.00	1.164	15.183	7.611	21.449	133.62	-93.32	-10.22
3, de56	*C*2/*c*	-133.07	-114.35	-18.72	0.00	1.146	29.431	7.563	21.095	129.06	-93.32	-10.22
4, am2	*P*2$_1$/*c*	-131.53	-121.92	-12.52	2.91	1.162	7.584	13.923	17.076	94.39	-65.00	+23.52
5, ak113	*P*2$_1$/*c*	-130.99	-118.84	-12.15	0.00	1.163	14.038	7.478	20.340	122.72	-93.32	-10.22
Enantiomorphic												
20, aq94	***P*2$_1$2$_1$2$_1$**	**-127.74**	**-117.39**	**-10.35**	**0.00**	**1.153**	**6.278**	**12.743**	**22.645**	**-**	**-93.32**	**-10.22**
form 2[4], 150 K	***P*2$_1$2$_1$2$_1$**					**1.205**	**6.209**	**12.580**	**22.188**		**-77.98**	**-34.52**
34, aq46	***P*2$_1$2$_1$2$_1$**	**-126.26**	**-116.20**	**-12.97**	**2.91**	**1.139**	**10.176**	**12.871**	**13.999**	**-**	**-65.00**	**+23.52**
form 1, 150 K	***P*2$_1$2$_1$2$_1$**					**1.197**	**10.250**	**12.483**	**13.641**		**-65.49**	**+25.31**
54, az3	*P*2$_1$2$_1$2$_1$	-124.80	-114.30	-11.95	1.45	1.133	8.568	11.580	18.580	-	-80.00	+4.38
73, aq38	*P*2$_1$2$_1$2$_1$	-123.62	-123.62	-8.91	0.00	1.143	7.397	11.694	21.124	-	-93.32	-10.22
87, aq13	*P*2$_1$2$_1$2$_1$	-123.09	-124.54	-11.13	1.45	1.128	7.130	14.222	18.260	-	-80.00	+4.38

[1] crystal energy ($U+\Delta E^{intra}$) partitioned into its intermolecular repulsion-dispersion (U^{vdW}), intermolecular electrostatic (U^{elec}) and intramolecular (ΔE^{intra}) components
[2] methyl-keto (φ_1) and 21-methyl group (φ_2) rotation; torsion values refer to *nat*-progesterone
[3] structure name given as the rank order in rigid-body search followed by a MOLPAK identification code.
[4] form 2 is predicted to have 1.5 kJ mol^{-1} lower enthalpy than form 1, which is the thermodynamically most stable form. The energy difference is well within the errors expected due to the use of a rigid-body model, inaccuracies of the intermolecular potential and the neglect of thermal effects. Form 2 was also found in the search with the φ=-80 ° conformation, but the 21-methyl rotation in the latter conformation (+4.4 °) is considerably different from its value in the experimental structure (-34.5 °), and hence destabilized this rigid molecule approximation of form 2.

2. Analysis of carbonyl···carbonyl interactions

Carbonyl is one of the groups with the largest bond dipole moment which is estimated at 2.5 D.[17,18] The interaction energies of two dipole moments at distance R can be readily computed as shown in figure S2.



$$E^{el}=\frac{1}{4\pi\varepsilon_o}\left(\frac{\mathbf{d}_1\cdot\mathbf{d}_2}{R^3}-3\frac{(\mathbf{d}_1\cdot\mathbf{R})(\mathbf{d}_1\cdot\mathbf{R})}{R^5}\right)$$

(a) (b) (c)

Figure S2: Electrostatic energy of two dipole moments in (a) arbitrary configuration and the carbonyl···carbonyl interactions in the (b) racemic and (c) chiral form 2.

In racemic progesterone, which corresponds to the most stable structure generated in the search, the dipoles are arranged in an almost perfectly rectangular, antiparallel manner ($\mathbf{d}_1\cdot\mathbf{d}_2=-d^2$) with 3.186 Å between the centre of the carbonyl bonds. This geometry is stabilized by approximately 5.81 kJ mol^{-1} in the dipole-dipole model, which is significant in comparison to the total lattice electrostatic stabilization energy, computed *via* the distributed multipole model, which does not exceed 20 kJ mol^{-1} in the most stable racemic and chiral putative crystal structures. Since steric effects from the functional groups bonded to the carbonyl group do not permit close packing with the more energetically favourable head-to-tail arrangement of dipoles,[19] this antiparallel arrangement is the most stable configuration for two dipole moments modelling carbonyl interactions. The dipolar stabilization due to this favourable arrangement of carbonyl groups is absent in enantiomorphic form 1 but is encountered in form 2, albeit in a distorted configuration (figure S2c) which yields an interaction energy of 3.98 kJ mol^{-1}. Notwithstanding the crudeness of the dipole model for the electrostatic energy and the problems of just considering this contribution to the carbonyl···carbonyl interaction,[20] this analysis does show that the antiparallel carbonyl···carbonyl geometry achieved in the racemic crystal structure of progesterone makes a significant contribution to its stability compared to the narrow energy range of putative structures.

3. Further experimental information

Crystallization of the racemic crystal

A small sample (ca. 2.5 mg) of enantiomeric (*ent*) progesterone was supplied. 2 mg samples (accurately weighed) of *ent*- and *nat*-progesterone were dissolved in chloroform and ether respectively. The resulting solutions were carefully added to an NMR tube with a view of initiating crystallization on the solvent interface. After several weeks a dense oil formed that was dispersed with difficulty using sonication. Redissolving the residual material in dichloromethane resulted in an emulsion being formed after which small crystals appeared on the wall of the tube. A number of cycling experiments were carried out whereby a small amount of dichloromethane was added to the apparently highly supersaturated oil. The size and quality of the resulting crystals were gradually improved to the point that a suitable specimen (small columnar block) was subsequently obtained and the structure solved in the space group $P2_1/c$ as described in the crystallographic information file for the racemic crystal structure, which is available as a separate ESI file.

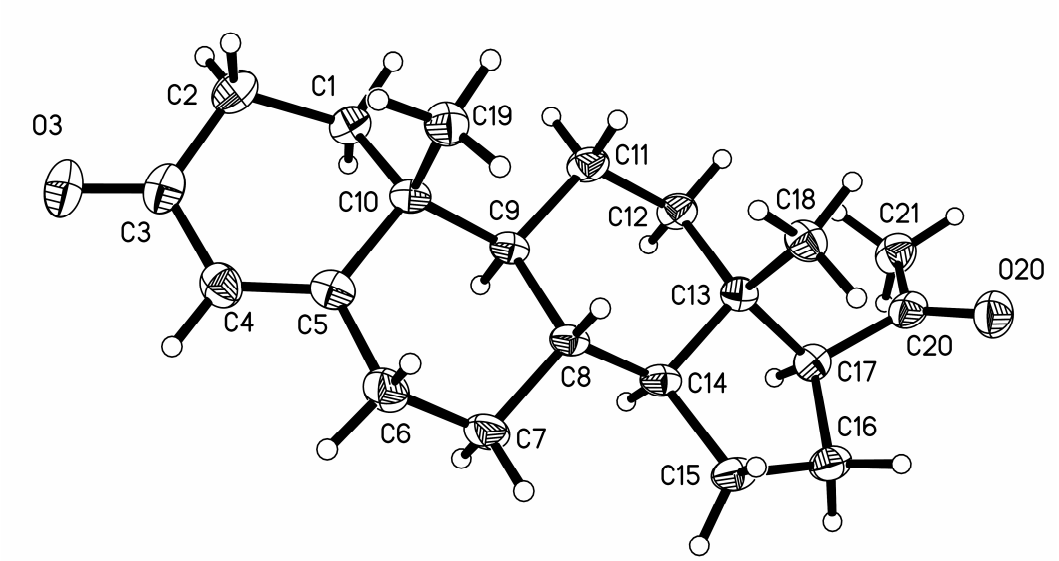


Figure S3 ORTEP plot showing thermal ellipsoids plotted at the 50% probability level for all non-hydrogen atoms. Hydrogen atoms are shown as spheres.

References to ESI

1. M. J. Frisch *et al.* GAUSSIAN 98. (A6), (1998). Pittsburgh, Gaussian Inc.
2. J. R. Holden, Z. Y. Du, H. L. Ammon, *J. Comput. Chem.*, 1993, **14**(4), 422.
3. D. J. Willock, S. L. Price, M. Leslie, C. R. A. Catlow, *J. Comput. Chem.*, 1995, **16**(5), 628.
4. A. J. Stone, M. Alderton, *Mol. Phys.*, 1985 **56**(5), 1047.
5. D. E. Williams, S. R. Cox, *Acta Crystallogr., Sect. B*, 1984, **40**, 404.
6. S. R. Cox, L. Hsu, D. E. Williams, *Acta Crystallogr. Sect. A*, 1981, **37**, 293.
7. D. S. Coombes, S. L. Price, D. J. Willock, M. Leslie, *J. Phys. Chem.*, 1996, **100**, 7352.
8. G. M. Day, S. L. Price, M. Leslie, *J. Phys. Chem. B*, 2003 **107** (39) 10919.
9. A. L. Spek. PLATON, A Multipurpose Crystallographic Tool. (2002). Utrecht, University of Utrecht.
10. J. A. Chisholm, S. Motherwell, *J. Appl. Crystallogr.*, 2005, **38**, 228.
11. R. S. Payne, R. J. Roberts, R. C. Rowe, R. Docherty, *Int. J. Pharm.*, 1999, **177**(2), 231.
12. W. L. Duax, J. F. Griffin, D. C. Rohrer, *J. Am. Chem. Soc.*, 1981, **103**(22), 6705.
13. C. P. Brock, W. B. Schweizer, J. D. Dunitz, *J. Am. Chem. Soc.*, 1991 **113**(26) 9811.
14. O. Wallach, *Liebigs Annalen der Chemie*, 1895, **286**, 90.
15. A. J. Pertsin and A. I. Kitaigorodsky, The Atom-Atom Potential Method. Applications to Organic Molecular Solids vol. 43 of Springer Series in Chemical Physics (Springer-Verlag, Berlin, 1987).
16. P. G. Karamertzanis, S. L. Price, *J. Chem. Theory Comput.*, 2006, **2**(4), 1184.
17. O. Exner, Dipole Moments in Organic Chemistry Georg Thieme Publishers, Stuttgart, 1975).
18. J. M. Stout, C. E. Dykstra, *J. Phys. Chem. A*, 1998, **102**(9), 1576.
19. A. J. Stone, *in* The Molecular Physics of Liquid Crystals, G. R. Luckhurst and G. W. Gray, Eds., (Academic Press, 1979), pp. 31-50.
20. F. H. Allen, C. A. Baalham, J. P. M. Lommerse, P. R. Raithby, *Acta Crystallogr., Sect. B*, 1998, **54**, 320.