

# Coulombic and dispersive factors in the molecular recognition of peptides: *PIXEL* calculations on two NNQQ (Asn-Asn-Gln-Gln) crystal polymorphs

**Angelo Gavezzotti**Dipartimento di Chimica Strutturale e  
Stereochimica Inorganica, Università di Milano,  
via Venezian 21, 20133 Milano, ItalyCorrespondence e-mail:  
angelo.gavezzotti@unimi.itReceived 18 February 2008  
Accepted 19 June 2008

The crystal-packing and cohesive energies in the structures of two polymorphs of the title tetrapeptide have been analyzed using molecule–molecule energies calculated using the *PIXEL* method. Coulombic energies are non-empirical and are much more accurate than those calculated using point-charge methods. The results explain and rationalize the cohesion and mutual recognition of these peptide molecules, with a clear distinction between polar and dispersive contributions, shedding light on subtle differences between polymorphic arrangements. For systems of the present size, the necessary calculations can be carried out on a personal computer and require quite acceptable computing times. Although an extension to larger peptides is problematic for obvious reasons, it is suggested that this type of analysis could be a valuable and practical tool in the understanding of the principles of peptide aggregation.

## 1. Introduction

There has recently been interest in the formation of amyloid-like fibrils from small peptides, with the determination of their structures at atomic resolution (Nelson *et al.*, 2005) and an analysis (Sawaya *et al.*, 2007) of the overall structural types for these ‘zippers’, which are of potential clinical interest for their connection with degenerative diseases. Clearly, a quantitative determination of the implied energies with a distinction between purely Coulombic or first-order polarization terms on one side and dispersion on the other would be desirable. The rigorous calculation of electrostatic and total energies between macromolecules in crystals is becoming possible using the ‘aspherical atoms databank’ approach (Li *et al.*, 2006; Dominiak *et al.*, 2007) in the framework of symmetry-adapted perturbation theory (SAPT). Density functional theory calculations for some peptides have been presented (Tsemekhman *et al.*, 2007). A semiempirical method has been developed for the calculation of intermolecular energies based on numerical integration over *ab initio* determined charge densities (henceforth referred to as the *PIXEL* method; Gavezzotti, 2003) and has been successfully applied to several topics in the crystal packing of organic molecules (see, for example, Dunitz & Gavezzotti, 2005). In this communication, the application of *PIXEL* analysis to the separation and characterization of intermolecular interactions in two polymorphs of the NNQQ tetrapeptide is described. The results show a larger contribution of dispersion in the formation of stacks, while the interdigitation that leads to interlayer cohesion and eventually to the formation of insoluble fibrils is dominated by Coulombic contributions. This type of analysis can yield valuable information on the nature of cohesive forces in peptide aggregates.

## 2. Methods

The space-group information, unit-cell parameters and atomic coordinates of the non-H atoms of the two peptides (Sawaya *et al.*, 2007; see Table 1) were retrieved from the Protein Data Bank entries. Methylene H atoms were generated according to standard geometrical rules, while N–H H atoms were placed in reasonable positions (all N–H distances 1.0 Å; NH<sub>2</sub> groups, planar geometry, all

**Table 1**

Crystal parameters (Sawaya *et al.*, 2007) and *PIXEL* energies of the two crystal forms of NNQQ.

A reduced cell was used for form 2.

	Space group	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	$\beta$ (°)	<i>d<sub>x</sub></i> (g cm <sup>-3</sup> )
Form 1	<i>P2</i> <sub>1</sub> , <i>Z</i> = 2	4.854	16.014	15.546	96.91	1.391
Form 2	<i>P2</i> <sub>1</sub> <i>2</i> <sub>1</sub> <i>2</i> <sub>1</sub> , <i>Z</i> = 4	4.915	15.479	30.552	—	1.436

Coulombic, polarization, dispersion and repulsion contributions (kJ mol<sup>-1</sup>).

	<i>E</i> (Coul.)	<i>E</i> (Pol.)	<i>E</i> (Disp.)	<i>E</i> (Rep.)	<i>E</i> (Total)
Form 1	-934	-324	-261	451	-1067
Form 2	-867	-246	-262	327	-1051

**Table 2**

Details of the molecule–molecule energies in the crystals of the two forms of NNQQ.

Distance: distance between centres of mass of the molecular pair (Å). See text and Fig. 1 for explanations of symbols. Total, Coulombic and dispersion energies (kJ mol<sup>-1</sup>) are also shown against distance in Figs. 2 and 3.

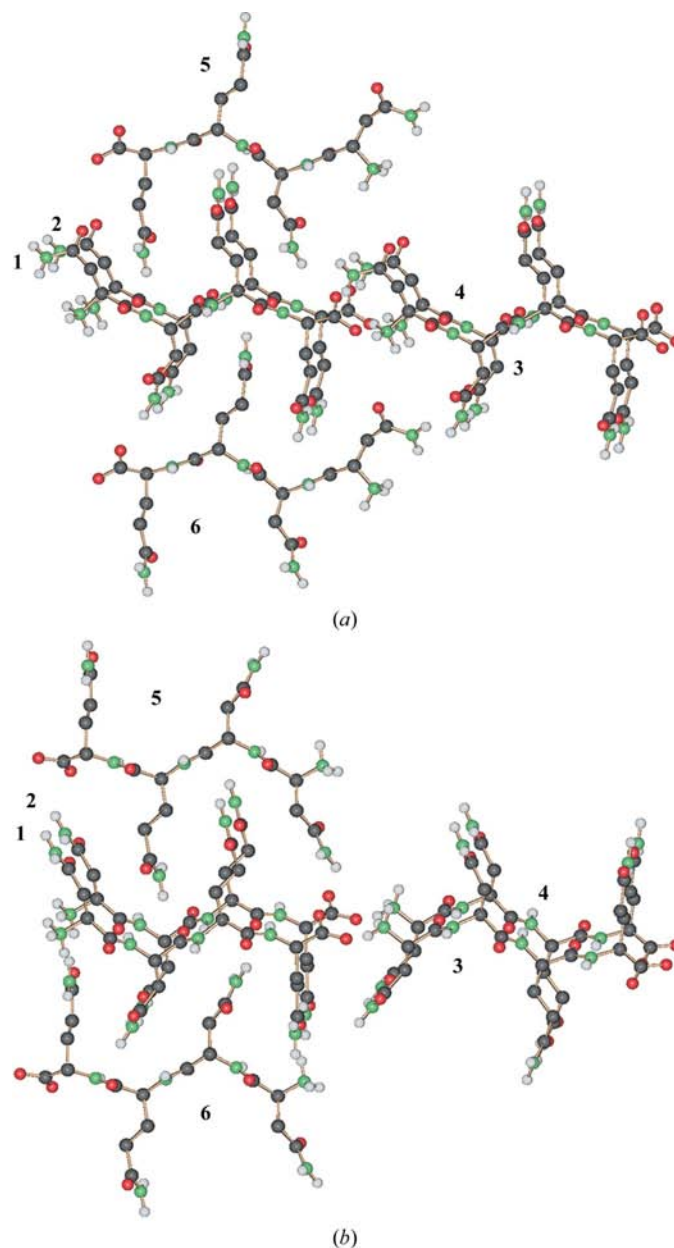
Distance	Coul.	Pol.	Disp.	Rep.	Total	Symbol
<b>Form 1</b>						
4.854	-23.4	-139.4	-98.8	187.2	-74.4	<i>T</i> <sub>T</sub> <sup>1</sup>
8.722	-142.0	-43.8	-60.3	42.1	-204.0	<i>S</i> <sup>1</sup>
8.767	-181.1	-44.3	-47.0	31.3	-241.1	<i>S</i> <sup>0</sup>
9.708	78.2	-3.8	-0.9	0.0	73.4	<i>T</i> <sub>T</sub> <sup>2</sup>
11.063	-78.3	-3.7	-1.4	0.0	-83.4	<i>S</i> <sup>2</sup>
11.170	-47.1	-4.1	-1.3	0.0	-52.4	<i>S</i> <sup>1</sup>
15.546	-252.6	-59.5	-18.7	39.2	-291.5	<i>T</i> <sub>L</sub> <sup>0</sup>
15.719	-317.8	-103.1	-21.0	147.4	-294.6	<i>T</i> <sub>L</sub> <sup>1</sup>
16.834	-71.8	-3.8	-0.5	0.0	-76.1	<i>T</i> <sub>L</sub> <sup>1</sup>
<b>Form 2</b>						
4.915	-68.3	-122.8	-101.9	161.8	-131.1	<i>T</i> <sub>T</sub> <sup>1</sup>
8.004	-201.7	-81.5	-63.4	69.5	-277.1	<i>S</i> <sub>A</sub> <sup>0</sup>
8.299	-200.0	-50.9	-53.2	44.8	-259.3	<i>S</i> <sub>B</sub> <sup>0</sup>
9.830	72.2	-3.2	-0.9	0.0	68.1	<i>T</i> <sub>T</sub> <sup>2</sup>
10.601	-72.6	-4.6	-1.6	0.0	-78.7	<i>S</i> <sub>A</sub> <sup>1</sup>
10.825	-84.8	-4.2	-1.4	0.0	-90.3	<i>S</i> <sub>B</sub> <sup>1</sup>
15.446	-208.0	-52.0	-17.2	28.6	-248.6	<i>T</i> <sub>L</sub> <sup>1</sup>
15.502	-248.0	-46.6	-16.0	17.6	-293.0	<i>T</i> <sub>L</sub> <sup>0</sup>

angles 120°; NH<sub>3</sub><sup>+</sup> groups, staggered, all angles tetrahedral). The uncertainties in the location of these atoms may change the calculated energies by a few kJ mol<sup>-1</sup>, *i.e.* not in such a way as to alter the general conclusions. The molecular structures are shown in supplementary Figure S1<sup>1</sup>. The *PIXEL* method (for the most recent version of the package, see Gavezzotti, 2008) starts from a point-by-point calculation of the charge density at the HF/6-31G level using the *Gaussian* package (Frisch *et al.*, 2003). The density is then contracted to 36 100 pixels of charge density for each molecule. The Coulombic and first-order polarization energies are calculated using ordinary electrostatics formulae over all pixels and nuclei. The Coulombic term is non-empirical since no parameters enter its evaluation; it falls off as the inverse first power of distance. The polarization term is supposed to take care of the fact that the calculation for approaching molecules in fact uses the charge densities of the isolated molecules, an approximation that also neglects dynamic polarization; this term falls off as the electric field does; that is, as the inverse third power of distance. The dispersion energy is evaluated using a London-type expression among all pixels using locally distributed polarizabilities and falls off as the inverse sixth power of distance. The repulsion term

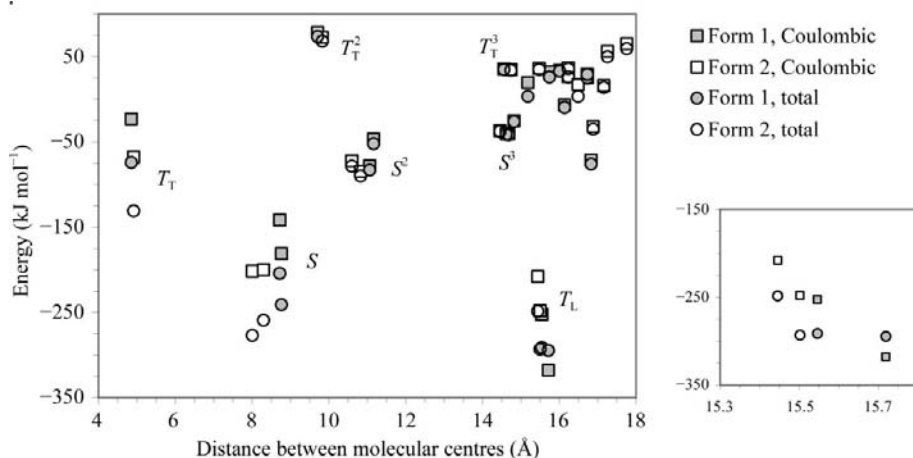
<sup>1</sup> Supplementary material has been deposited in the IUCr electronic archive (Reference: BE5104). Services for accessing this material are described at the back of the journal.

is calculated as proportional to the overlap between charge densities, with overlap integrals being calculated by numerical integration.

A model of the crystal is constructed starting from a reference molecule and generating a number of symmetry-related surrounding molecules using the space-group operators (within a certain distance threshold). The interaction energy between the reference molecule and each of the surrounding molecules is then calculated as the sum of Coulombic, polarization, dispersion and repulsion terms. The total lattice energy is the sum of all these molecule–molecule energies. For analysis of the coordination sphere in the crystal, molecule–molecule energies are plotted against the distance between molecular centres of coordinates. Such *R/E* plots immediately display the relative importance of molecular interactions between near-neighbours in crystals and have been shown to be useful in the characterization of

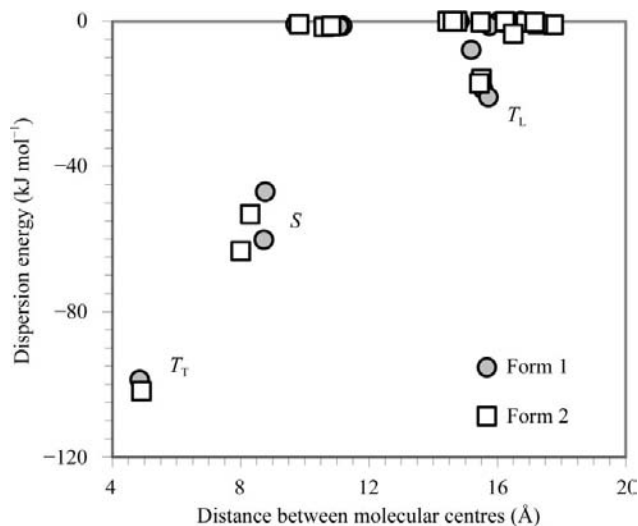


**Figure 1**  
The first coordination shell in the two crystalline forms of NNQQ. (a) Form 1; (b) form 2. 1, Reference molecule; 2, *T*<sub>T</sub>; 3, 4, *T*<sub>L</sub>; 5, 6, *S*. Nitrogen, green; oxygen, red; nonhydrogen-bonding H atoms were omitted for clarity. Structures from Nelson *et al.* (2005).



**Figure 2**

*R/E* plot: each point represents a molecular pair in the coordination sphere of the two crystal polymorphs of NNQO (symbols are defined in Fig. 1). Horizontal axis, distance between molecular centres; vertical axis, *PIXEL* Coulombic and total energies (see numerical data in Table 2). The inset on the right shows an enlargement of the lower right part of the figure.



**Figure 3**

As in Fig. 2 but for dispersion energies between the same molecular pairs (see numerical data in Table 2). The plots in this figure and in Fig. 2 give a comprehensive view of the strength and quality (Coulombic *versus* dispersive) of the crystal cohesion forces in the peptides.

crystal polymorphism (Bernstein *et al.*, 2008). Table 2 contains the numerical data that gave origin to the *R/E* maps in Figs. 2 and 3.

Computing times on an ordinary personal computer (Windows XP environment, Intel Core 2 Duo 2.40 GHz processor) are about 30 s for each molecular pair; that is, about 30 min for the complete evaluation of a lattice energy. The time required for the molecular-orbital calculation depends on the basis set; in the present case, it was about 1 h. For reproducibility of the calculations here described, the *PIXEL* computer software and input and output files are available from the author upon request. A full documentation with worked examples is available at the author's website (<http://users.unimi.it/gavezzot>).

### 3. Results

Fig. 1 shows the first coordination shell of the two crystal forms. The two crystal structures are fundamentally similar, as already suggested

by an inspection of the unit-cell parameters. The main constituents of the first coordination shell can be described by tracing the vectors joining the molecular centres of coordinates. They are defined as follows. (i)  $T_L$ , a 'longitudinal translation' (rectilinear displacement) parallel to the molecular-elongation axis by about 15.5 Å in the cell  $z$  direction that is obtained in form 1 by a pure cell translation and in form 2 by a screw operator which, given the overall shape of the molecule, is almost equivalent to a translation;  $T_L$  produces a confrontation between charged  $\text{COO}^-$  and  $\text{NH}_3^+$  groups. (ii)  $T_T$ , a 5 Å 'transverse translation' along the cell  $x$  direction, where 'transverse' indicates that it is perpendicular to  $T_L$ .  $T_T$  generates the stacked layers. (iii)  $S$ , a displacement of about 8.5 Å between nearest-neighbour molecules related by a crystallographic screw symmetry operation,

the translational component of which is in the cell  $y$  direction.  $S$  causes the interdigitation between adjacent layers. (Actually, in form 1 the operation consists of a screw operation along the 16 Å axis plus  $n$  translations along the short axis; in form 2 it is an  $n$ -fold application of the screw along the short axis plus a zero or a one-cell translation along the 16 Å axis; Supplementary Table S1 collects the complete detail of all symmetry operations and all calculated molecule-molecule energies.) Combinations that include these fundamental operations plus  $n$  cell translations along the short axis generate the  $T_T^n$ ,  $T_L^n$  and  $S^n$  operations.

Table 2 and Figs. 2 and 3 show the energetic features of these structure-determining molecular pairs (henceforth denoted as 'structure determinants') in the two forms. The overall similarity between the structures of the two polymorphs is borne out by the fact that the energies for similar determinants are more or less grouped together in the *R/E* map of Fig. 2. However, there are subtle differences. The  $T_L$  determinant is more stabilizing than the  $S$  determinant in form 1 and is on the same footing with  $S$  in form 2; in other words, interdigitation is more stabilizing in form 2. The  $T_L$  and  $S$  determinants have a strongly Coulombic connotation, but are also buttressed by a dispersive contribution, which is larger for the  $S$  operation (see Fig. 3), that brings the molecules together with parallel elongation axes, while the  $T_L$  operation only brings together molecular tips of opposite charge. For the second-neighbour contacts  $T_T^2$ ,  $S^2$  and  $S^3$ , the dispersive and polarization contributions fall off sharply with increasing distance and total energies coincide with the slowly declining Coulombic term.

The case of  $T_T$  is more complicated. The closest neighbour pair has a comparatively small Coulombic contribution because identical charges are brought into close contact by stacking along the short cell translation; the pair is stabilized by a large dispersion contribution, as expected when considering the close pairing with parallel elongation axes, but also by a relevant first-order polarization term. This term is always stabilizing by definition, even when charges of the same sign approach one another. The second- and third-neighbour pairs  $T_T^2$  and  $T_T^3$  are totally destabilizing because the polarization and dispersion terms have vanished. In the upper right part of the graph in Fig. 2 one sees the many destabilizing second- and third-neighbour contributions that are to be expected in the long-range interaction of a crystal formed by molecular entities with high permanent charges, just as happens in sodium chloride.

Which form is more stable? The answer to this question is not straightforward. The calculated total lattice energies are hardly reliable for a subtle differentiation. The positions of the amino H atoms are only guessed and small displacements can change the calculated energies by up to a few (5–10) kJ mol<sup>-1</sup>. In such heavily charged species, convergence problems may be rather acute in polar space groups such as *P2*<sub>1</sub>. Intrinsic errors in the quadrature routines can cause uncertainties of a few kJ mol<sup>-1</sup>. For these reasons, all that can be said is that in all likelihood the two structures have very similar energies, presumably within a 5% difference between the two polymorphs, as is usual in organic crystal polymorphism (Gavezzotti, 2007*b*). Some further indication of the reliability of the present calculations comes from the energy breakdown (see Table 1). The lattice energies without dispersion contributions are 806 and 789 kJ mol<sup>-1</sup>, which are very close to 201 kcal mol<sup>-1</sup>, the energy calculated by density functional theory (Tsemekhman *et al.*, 2007) for the energy per monomer in an infinite fibril of a similar peptide, GNNQQNY. There is broad agreement considering that DFT is well known to underestimate or not include at all the dispersion contribution. Also, the energies per hydrogen bond quoted by Tsemekhman and coworkers, 7–9 kcal mol<sup>-1</sup>, are very similar to those from *PIXEL* (Gavezzotti, 2007*a*).

#### 4. Conclusion

We have shown that analysis of *PIXEL* molecule–molecule energies explains and rationalizes the cohesion and mutual recognition of peptide molecules, with a clear distinction between polar and dispersive contributions, shedding light on subtle differences between polymorphic arrangements. A fruitful comparison can be made with the parallel pioneering work of Tsemekhman *et al.* (2007) by DFT. *PIXEL* gives a detailed separation of Coulombic *versus* dispersive factors and requires a much smaller computational effort; the adverse sides are a lower formal numerical accuracy (as discussed above) and the semi-empirical nature of some of the *PIXEL* procedures for polarization and dispersion terms. However, these procedures are firmly anchored to the physics of the interaction, relying on charge densities rather than on more empirical quantities such as polar and

nonpolar surfaces and the like. On the other hand, *PIXEL* Coulombic energies are non-empirical and are incomparably more accurate than those calculated using point-charge methods. We emphasize that for sample systems of 100 atoms or less the necessary *PIXEL* calculations can be carried out on a personal computer and require quite acceptable computing times. We propose that *PIXEL* computer experiments on wisely chosen systems of such size can be a valuable aid in the explanation and prediction of the aggregation modes of peptides in general.

Figs. 1 and S1 were drawn using the program *SCHAKAL* (Keller, 1993). Thanks are due to Dr L. Vitagliano (CNR, Napoli) for bringing these systems to our attention.

#### References

- Bernstein, J., Dunitz, J. D. & Gavezzotti, A. (2008). *Cryst. Growth Des.* **8**, 2011–2018.
- Dominiak, P. M., Volkov, A., Li, X., Messerschmidt, M. & Coppens, P. (2007). *J. Chem. Theor. Comput.* **3**, 232–247.
- Dunitz, J. D. & Gavezzotti, A. (2005). *Angew. Chem. Int. Ed.* **44**, 1766–1787.
- Frisch, M. J. *et al.* (2003). *Gaussian 03*, revision B.04. Gaussian Inc., Pittsburgh, USA.
- Gavezzotti, A. (2003). *J. Phys. Chem. B*, **107**, 2344–2353.
- Gavezzotti, A. (2007*a*). *Molecular Aggregation*, ch. 12. Oxford University Press.
- Gavezzotti, A. (2007*b*). *Molecular Aggregation*, ch. 14. Oxford University Press.
- Gavezzotti, A. (2008). *Mol. Phys.*, doi:10.1080/00268970802060674.
- Keller, E. (1993). *SCHAKAL92. A Program for the Graphic Representation of Molecular and Crystallographic Models*. University of Freiburg.
- Li, X., Volkov, A. V., Szalewicz, K. & Coppens, P. (2006). *Acta Cryst.* **D62**, 639–647.
- Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. O., Riekel, C., Grothe, R. & Eisenberg, D. (2005). *Nature (London)*, **435**, 773–778.
- Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I., Thompson, M. J., Balbirnie, M., Wiltzius, J. J. W., McFarlane, H. T., Madsen, A. O., Rieke, C. & Eisenberg, D. (2007). *Nature (London)*, **447**, 453–457.
- Tsemekhman, K., Goldschmidt, L., Eisenberg, D. & Baker, D. (2007). *Protein Sci.* **16**, 761–764.