Crystal structure prediction with the supramolecular synthon approach: Experimental structures of 2-amino-4-ethylphenol and 3-amino-2-naphthol and comparison with prediction†

Archan Dey, Narendra Nath Pati and Gautam R. Desiraju*

Received 27th June 2006, Accepted 31st August 2006
First published as an Advance Article on the web 12th September 2006
DOI: 10.1039/b609101f

Two earlier crystal structure predictions (CSP) of aminophenol compounds are checked against experimental structure determinations. One of the predictions classified originally as ''good'' is verified and is now seen to be somewhat correct. The other prediction, which was classified earlier as ''unclear'', is incorrect. The two experimental crystal structures are characterized by small and large synthons both of which contain O–H···N and N–H···O hydrogen bonds. The evolving nature of the CSP exercise is noted.

Introduction

Crystal structure prediction (CSP)[1–4] is the computational prediction, from the molecular structure, of the space group and the positional parameters of the atoms in the crystal structure. CSP is a major scientific problem today and is of great difficulty. A number of crystal structures are obtained computationally using a selected force field and the experimental structure is hidden generally amongst the 100 lowest energy structures. When the experimental structure is also the thermodynamic structure, accurate force fields may reveal this structure as the global minimum. When the experimental structure is a higher energy kinetic structure, a purely computational technique is often inadequate. In these cases, we have suggested a knowledge-based alternative, the supramolecular synthon approach to CSP.

At the core of this methodology is the *supramolecular synthon*,[5] which is a structural unit smaller than the complete crystal but which encapsulates a sufficient amount of critical structural information so that it serves as a realistic model for the entire crystal. The synthon is a kinetic entity; in this methodology the computational results are biased manually with synthon information from a database of known crystal structures to incorporate the kinetic factors. Synthons in this database are loosely classified as ''small'' and ''large'' based on their complexity. The absence of a small synthon in a predicted structure is a negative factor and is justification for its down-ranking or elimination. The presence of a large synthon in a predicted structure is a positive factor and is grounds for its up-ranking. The highest ranked structures in this re-ranked list are taken as the predictions.

School of Chemistry, University of Hyderabad, Hyderabad, 500 046, India. E-mail: gautam_desiraju@yahoo.com

† Electronic supplementary information (ESI) available: List of training set and test set compounds in the CSP (Scheme SI1); polymorph prediction results for the ten lowest energy crystal structures of **1** and **2** (Table SI1); and pairs of compounds with both methyl and ethyl derivatives, and phenyl and naphthyl derivatives, which have crystal structures in the CSD (Tables SI2 and SI3). See DOI: 10.1039/b609101f

We have shown earlier that such synthon based CSP (with the COM force field) works well for rigid aminophenols and related compounds.[6] In our earlier study, CSP was performed for nine amino-hydroxy compounds with unknown crystal structures using a training database of the 10 isomeric methylaminophenols and the three simple unsubstituted aminophenols. Upon prompting from a referee, we experimentally verified the prediction in two of the cases (8-amino-2-naphthol, 4-aminocyclohexanol). In the present paper, we report the experimental crystal structures of two more of these nine compounds, namely 2-amino-4-ethylphenol (**1**) and 3-amino-2-naphthol (**2**), and compare the experimental and predicted structures. The molecular structures of all the nine compounds and of the database compounds are given in the ESI.†

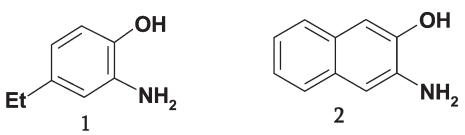
Experimental

Synthesis

Compound **1** was synthesized by nitration of 4-ethylphenol (Aldrich) followed by reduction. Compound **2** was purchased from Alfa Aesar. Single crystals of **1** and **2** were obtained from acetone–hexane (2 : 1) and EtOAc–MeCN (1 : 1) respectively, but the crystals of **1** were not the best for X-ray work (see below).

X-Ray crystallography

X-Ray data for the compounds **1** and **2** were collected on a Bruker SMART diffractometer using Mo Kα radiation. Compound **2** posed no special problem and the data could be processed adequately with the SHELXTL program.[7] For **1**, the crystal quality was poor and the R_{int} value of 18% is not satisfactory. However, the structure could be solved and refined to a level where the hydroxy and amino H-atoms could



Scheme 1

Table 1 Crystallographic data and synthon information

	1	2
Chemical formula	$C_8H_{11}NO$	$C_{10}H_9NO$
Crystal system	Monoclinic	Monoclinic
Space group	$P2_1/c$	$P2_1/c$
a/Å	14.806(13)	15.1054(13)
b/Å	7.155(6)	6.0153(5)
c/Å	7.820(7)	8.4629(7)
β/°	94.861(14)	99.5820(10)
Z	4	4
V/Å^3	825.5(12)	758.24(11)
D_{calc}/mg m^{-3}	1.104	1.394
R_1 [$I > 2\sigma(I)$]	0.0806	0.0415
wR_2	0.2201	0.1121
GOF	0.927	1.048
Synthon	**I**, **II**, **III** and **IV**	**I**, **II**, **III** and **IV** (distorted)
Packing coefficient (%)	64.9	75.5

be located in difference Fourier maps and refined isotropically. The other H-atoms were fixed in geometrically sensible positions. The *R*-factor is just about acceptable (0.0806) and the packing of molecules is obtained to an accuracy that is relevant to assessing the structure prediction. Crystallographic details for both compounds are given in Table 1. CCDC reference numbers 612505 and 612506. For crystallographic data in CIF or other electronic format see DOI: 10.1039/b609101f

Computational

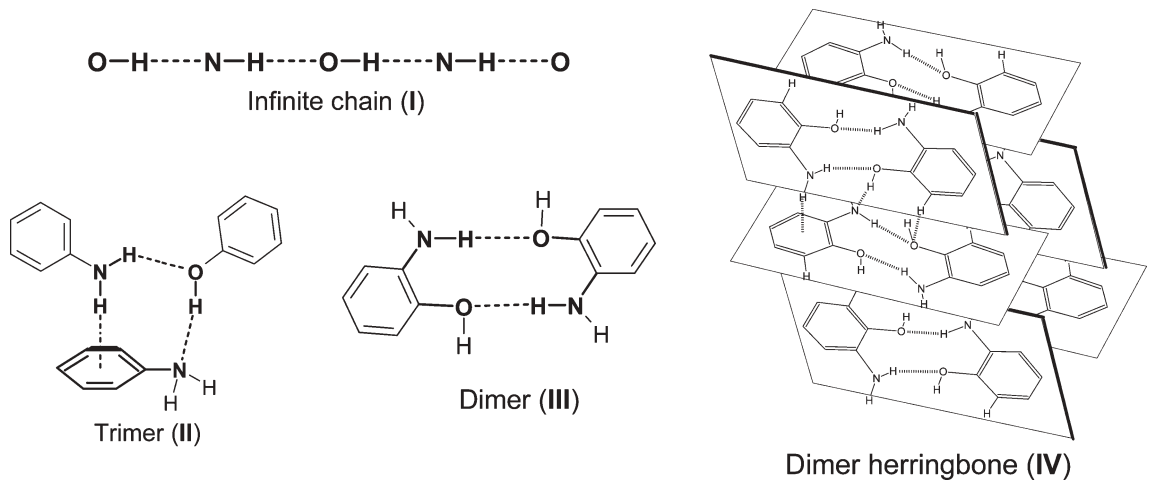
The Polymorph Predictor (PP) results were taken from our previous paper,[6] and are given in Table SI1.† Lattice energy minimization of experimental structures was carried out with version 4.8 of the *Cerius*[2] molecular modelling[8] environment running on Silicon Graphics workstations. The COSET program was used for all the PP analysis (RMSD calculations for the experimental and predicted structures).

Results and discussion

Review of previous CSP results and their justification

Scheme 2 shows the structures of the synthons pertinent to the discussion in this paper. Synthon **I** is classified as ''small'' and **IV** may be considered to be ''large''. Synthons **II** and **III** are intermediate in complexity. The following section summarises the predictions for **1** and **2** in our earlier paper,[6] and also provides some justification for them. In that paper, the predictions were classified as ''good'', ''bad'' or ''unclear''. The ''good'' and ''bad'' categorization was based on the presence or absence of favourable indicators such as low energy, high density, clean demarcation from other structures and the presence of large synthons. The ''unclear'' predictions were so designated when: (i) the indicators were mixed or; (ii) if the database of known compounds was either too limited or of insufficient similarity to the test compound. In this context, the difficulty in judging the degree of similarity/dissimilarity between the training and test compounds is a real problem because one is speaking of supramolecular rather than molecular similarity, and one cannot rule out unexpected levels of complexity.[9] In any event, global minima for both **1** and **2** were selected as predictions because the best synthons were also seen in these lowest energy structures (Table SI1†). The reader will note that the synthon approach has not changed the rank in these two cases (#1 → #1). However, we have argued in our previous paper[6] that double confirmation of the prediction is welcome: indeed, *any* CSP result may be called into question at the present time and the synthon approach strengthens a prediction from a completely independent standpoint.

Among the 10 lowest energy structures for **1**, the three best structures (1st, 2nd, 3rd) are widely separated in energy terms from the rest. The 2nd, 3rd, 4th, 8th, 9th and 10th were additionally down-ranked because of the presence of unlikely synthons containing the disfavoured O–H···O and N–H···N interactions. Of the best three structures, the 1st has a higher density than the 2nd and 3rd, and it also contains the large synthon **IV**. So, the 1st structure (in space group *C*2/*c*) was selected as the most probable. However, the prediction was still classified as ''unclear'' because we were concerned about issues of similarity/dissimilarity between compound **1** and the training database compounds. The latter largely contains methyl-substituted aminophenols. A CSD survey showed that there are only seven pairs of compounds wherein a Me group is replaced by an Et group, and none of these is an isostructural



Scheme 2

 This journal is © The Royal Society of Chemistry 2006

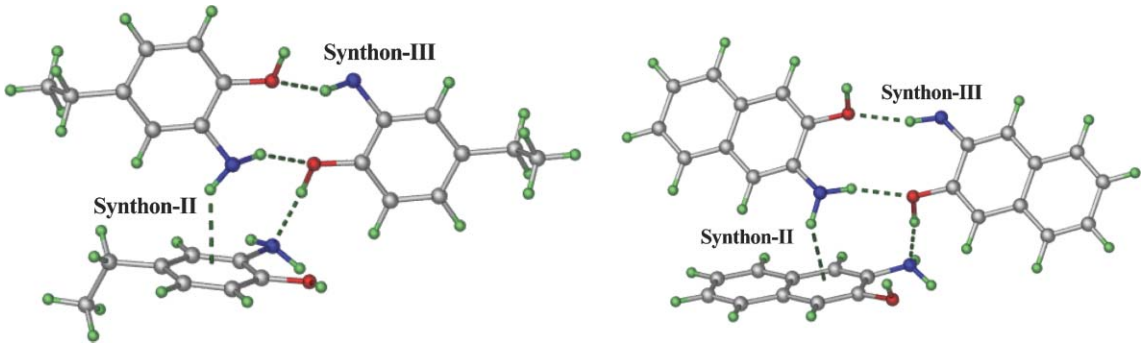


Fig. 1 Synthons **II** and **III** in the title compounds.

pair (Appendix, Table SI2†). The question therefore is whether a methyl derivative is a good supramolecular model for the corresponding ethyl derivative. Not being able to answer this question properly or completely, we were cautious about our prediction.

The 10 lowest energy structures of compound **2** were clustered into three categories. These are: (i) structures with unlikely synthons (4th, 5th, 6th); (ii) structures without large synthons (7th, 8th, 9th, 10th); and (iii) structures with large synthons (1st, 2nd, 3rd). From the last category, the 1st (in space group $C2/c$) was selected as the most probable, and indeed the 2nd or 3rd choices are practically the same in terms of packing. All four favourable indicators (low energy, high density, clean demarcation and presence of large synthons) enabled a classification of this prediction as "good". Questions of structural similarity between compound **2** and the training database compounds are also not a problem (Appendix, Tables SI1, SI3†). There is a precedent in which a pair of Me-groups *ortho* to each other on a phenyl ring are supramolecularly equivalent to a second benzene ring annulated to the first.[10] Accordingly, the best mimic of compound **2** should be 2-amino-4,5-dimethylphenol. Considering that both 2-amino-4-methylphenol and 2-amino-5-methylphenol are in the training database, this was taken as an encouraging sign. There is also evidence for benzene → naphthalene homology in the literature (Table SI3†). Accordingly, 2-aminophenol should be a good model for compound **2**; we note that it, too, is in the training database.

Experimental crystal structures of compounds 1 and 2

Compound **1** crystallizes in the space group $P2_1/c$ with $Z' = 1$. The structure contains synthons **II** and **III** (Fig. 1). The N–H···π bridge in synthon **II** to the centroid of an adjacent phenyl ring is comparable (*D, d,* θ, 3.23 Å, 2.48 Å, 140°) to that in 2-aminophenol (3.26 Å, 2.46 Å, 146°). The larger synthon **IV**, seen in 2-aminophenol and 2-amino-4-methylphenol, is also present in **1** and contains a C–H···O interaction. However, the overall packing is quite different from these compounds and confirms that the H → Me homology does not extend to the ethyl derivative (Fig. 2).

Compound **2** takes space group $P2_1/c$ with $Z' = 1$. As in **1**, synthons **II** and **III** are present (Fig. 1). The larger synthon **IV** is distorted with an elongated N–H···π interaction (3.34 Å, 2.63 Å, 137°) and without any C–H···O interaction. This distortion is measured in terms of the tilt-angle between the average planes of the dimer units in the synthon. The values for 2-aminophenol, 2-amino-4-methylphenol, **1** and **2** are 87, 86, 84 and 45°, respectively (Fig. 3).

Verification of predictions and assessment of our methodology

The prediction is verified by overlaying a 10-molecule cluster in the predicted (Pred) and experimental (Expt) crystal structures. Also compared are the lattice energies, the (reduced) cell dimensions and the crystal packing. Our initial observation was that there is no overlap between the Pred and Expt for compound **1**. The prediction is incorrect. For compound **2**, the situation is better (Fig. 4) but only just (RMSD 0.2827). At this stage, we suspected the accuracy of the force field used (COM). The reader will note that with a #1 → #1 re-ranking for both **1** and **2**, our successes and failures are as much a comment on the energy minimization (force fields) as the synthon based re-ranking. The COM force field performs better than others for aminophenols, but this still does not mean that it is good

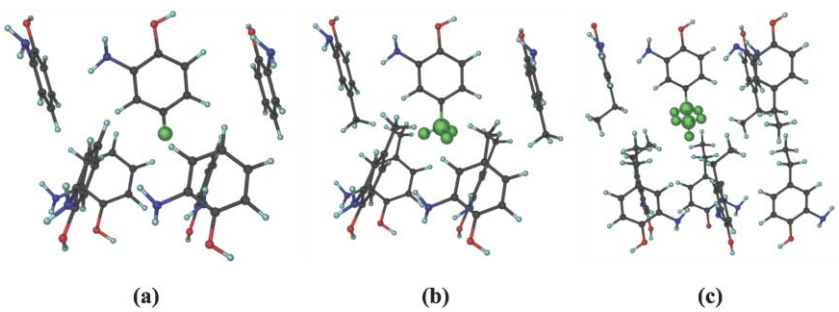


Fig. 2 Close packing of molecules in: (a) 2-aminophenol, (b) 2-amino-4-methylphenol and (c) 2-amino-4-ethylphenol, **1**. Note that the first two compounds are isostructural but that **1** is quite distinct.

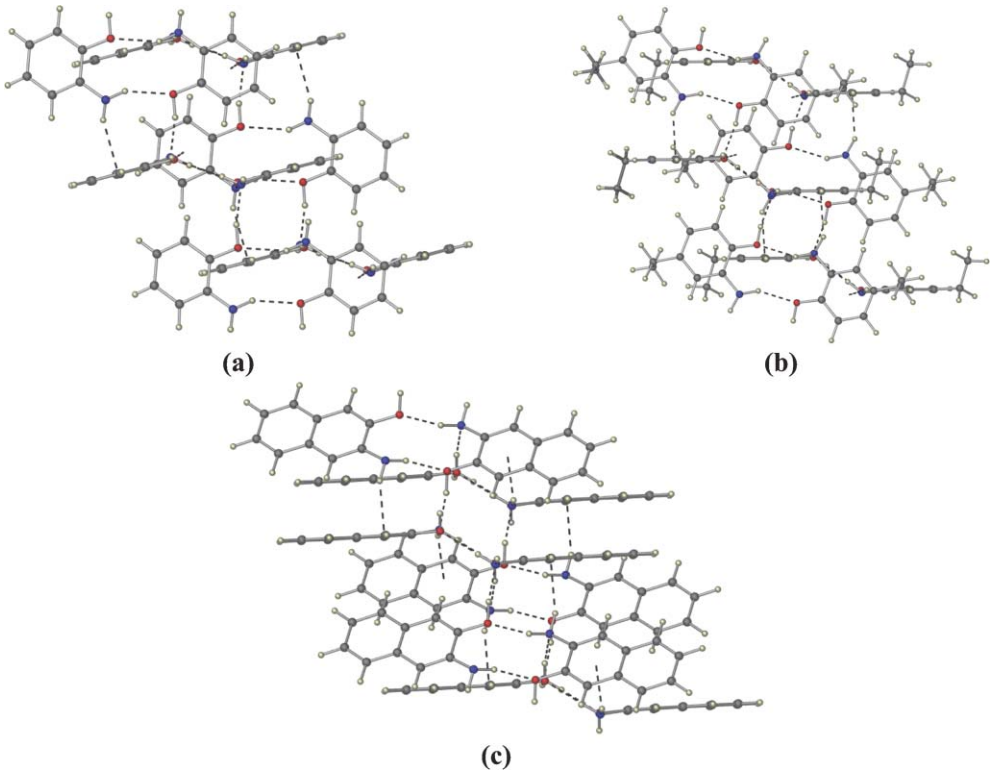


Fig. 3 Synthon **IV** in (a) 2-aminophenol, (b) ethylphenol **1** and (c) naphthol **2**.

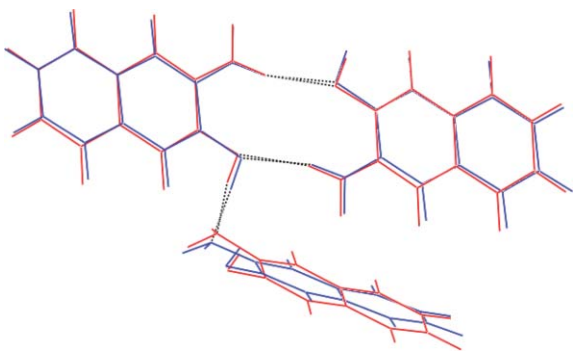


Fig. 4 Overlay diagram of predicted (blue) and experimental structure (red) of aminonaphthol **2**.

enough. Clearly, the synthon-based method will give vague and unclear results if the force field is not at some threshold level of accuracy. Noting this factor, we compared the predicted structure (Pred) with the minimized experimental structure (Expt Min) rather than with the experimental structure (Expt). With this modification (Fig. 5), the situation improves considerably for compound **2** (RMSD 0.0458) but compound **1** still behaves unacceptably (0.2895). This is a clear indication that the force field itself is inadequate, but that compound **1** has additional problems. A comparison of the Expt Min and Pred structures is given in Table 2. Both compounds (**1** and **2**) are predicted in a space group ($C2/c$) that is different from the experimental space group ($P2_1/c$). However, there is a similarity in the reduced cells, reduced cell volumes and lattice energies.

How does one assess these results? In the critic's viewpoint, these are not correct predictions; even the experimental space group is missed. The lack of efficacy of synthon **IV** (a so-called large synthon) as a positive discriminator in synthon-based CSP of compound **1** indicates that this synthon is still too simple. How much more complex does a synthon need to be to be identified as a "fingerprint" for compound **1**? The experimental structure was not found even in the higher-energy frames, and it has a very low packing coefficient (0.65). Apart from force field problems (see below), it is probably fair to

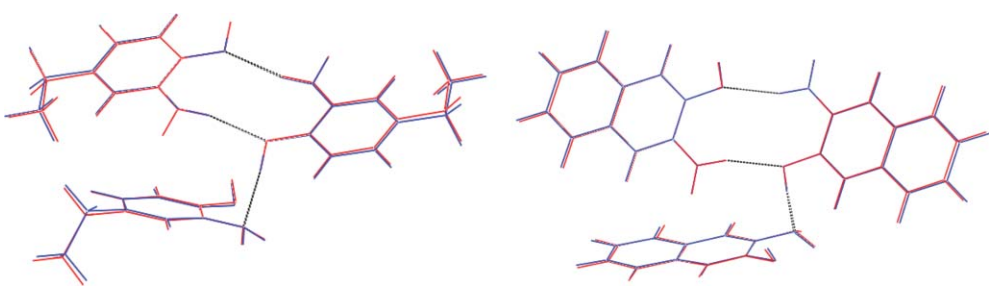


Fig. 5 Overlay diagram of predicted (blue) and minimized experimental structure (red) of **1** and **2**.

 This journal is © The Royal Society of Chemistry 2006

Table 2 Comparison of cell parameters and lattice energies for **1** and **2**

		Cell parameters (Å/°)	Vol. /Å^3	Reduced cell parameters[a] (Å/°)	Vol./Å^3)	Energy / kcal mol^{-1}	RMSD[b]
1	Expt Min ($P2_1/c$)	15.4516, 5.5834, 8.4669, 87.7660	729.91	5.583, 8.467, 15.452, 92.23	729.9	−36.484	0.2895 (no match)
	Pred ($C2/c$)	30.7885, 5.6042, 8.5279, 88.1722	1470.70	5.604, 8.528, 16.088, 103.56, 100.03, 90.00	735.3	−36.620	
2	Expt Min ($P2_1/c$)	14.7083, 5.8697, 8.8205, 101.8460	745.29	5.870, 8.820, 14.708, 101.85	745.3	68.516	0.0458 (0.2827)
	Pred ($C2/c$)	31.5755, 5.8505, 8.8893, 66.5760	1506.81	5.850, 8.889, 14.892, 97.06, 101.33, 90.00	753.4	68.744	

[a] The reduced cell calculation was carried with PLATON.[11] [b] The root mean square deviation (RMSD) calculation was carried out with the COSET program with a 10 molecule cluster. The values given are for Expt Min and Pred while the values in brackets correspond to Expt and Pred.

say that the training database is inappropriate (too small and/or not varied enough) to handle ethyl substituted aminophenols—in this respect, our earlier assessment of the prediction as ''unclear'' is completely on target.

Compound **2** presents a different problem. The Pred agrees with Expt Min rather than with Expt and this means that the force field needs fine-tuning and further investigation. Of course, if the force field is of sub-critical accuracy, then all further discussion regarding synthon-based re-ranking is practically pointless. In this particular case, we evaluated several force fields before opting for COM and in the end, we had little choice.[6] However, improvements in a force field are always possible in principle.

Conclusions

We have checked crystal structure predictions on two more aminophenols with experimental results. In the ethyl derivative **1** the prediction is not satisfactory. For the naphthyl derivative **2** the prediction is somewhat better. The differences or similarities between the predicted and experimental structures have been analysed. These results confirm earlier qualitative reasoning that a Me → Et substitutional change has a serious effect on the crystal packing whereas a phenyl → naphthyl substitutional change does not. Our assessments of our previous predictions as ''good'', ''bad'' and ''unclear'' seem to be generally correct. In particular, we will continue to reserve the descriptor ''unclear'' for a CSP whenever the training set database is not truly representative of the unknown crystal structure that is being predicted. The force field problem is also very real and illustrates that studies of complex systems (crystal structures) as a function of simple systems (molecular structures) do not lend themselves easily to routine computation. We conclude by stating that CSP is an extremely difficult problem, wherein general solutions are most likely impossible in the immediate future. Various special solutions, such as synthon-based CSP, are somewhat applicable to small homogeneous sets of compounds. Even here, these solutions are of mixed efficacy.

Appendix

Both **1** and **2** may be derived from 2-aminophenol *via* substitutional changes. Compound **1** illustrates a H → Me → Et change while **2** is obtained with a phenyl → naphthyl change. A Cambridge Structural Database (Version 5.27, including April 2006 updates) search was carried out to further understand the Me → Et and phenyl → naphthyl exchange. The results for the Me → Et exchange search are given in Table SI2.† Only 7 pairs of compounds were found. Among these, there is no isostructural pair. A corresponding search was carried out for the corresponding phenyl/naphthyl pairs (Table SI3) and 55 pairs were found. Structural analogy was seen in three of these pairs (ACANIL03/ACACTB, BOLZAD/BOLZEH and BESNUI/GAFPEJ). While the number of structures in these analyses may not be statistically significant, they hint that the phenyl → naphthyl exchange conserves the crystal structure better than the Me → Et exchange. Perhaps this indicates that the training database (of methylaminophenols and unsubstituted aminophenols) was better suited to the naphthyl derivative **2** than to the ethyl derivative **1**, explaining the different outcomes of synthon-based CSP in the two cases.

Acknowledgements

AD and NNP thank CSIR for the award of an SRF and JRF respectively. GRD thanks DST for financial support. We thank the UGC (UPE) for providing computational facilities (CMSD).

References

1 R. J. Wawak, K. D. Gibson, A. Liwo and H. A. Scheraga, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 1743.
2 G. R. Desiraju, *Nat. Mater.*, 2002, **1**, 77.
3 J. D. Dunitz, *Chem. Commun.*, 2003, 545.
4 S. L. Price, in *Encyclopaedia of Supramolecular Chemistry*, ed. J. L. Atwood and J. Steed, Marcel Dekker, New York, 2004, p. 371.
5 G. R. Desiraju, *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 2311.
6 A. Dey, M. T. Kirchner, V. R. Vangala, G. R. Desiraju, R. Mondal and J. A. K. Howard, *J. Am. Chem. Soc.*, 2005, **127**, 10545.
7 SHELXTL: Program for the Solution and Refinement of Crystal Structures, Version 6.12. Bruker AXS Inc., Madison, Wisconsin, USA, 2001.
8 *Cerius*^{2} Molecular Modelling Program Package, Accelrys Inc., San Diego, CA 92121-3752 USA.
9 G. R. Desiraju, *Curr. Sci.*, 2005, **88**, 374.
10 R. Banerjee, R. Mondal, J. A. K. Howard and G. R. Desiraju, *Cryst. Growth Des.*, 2006, **6**, 999.
11 A. L. Spek, *PLATON, A Multipurpose Crystallographic Tool*, Utrecht University, Utrecht, The Netherlands, 2002.