LETTERS

Super-resolution biomolecular crystallography with low-resolution data

Gunnar F. Schröder[1,2], Michael Levitt[2] & Axel T. Brunger[2,3,4,5,6]

X-ray diffraction plays a pivotal role in the understanding of biological systems by revealing atomic structures of proteins, nucleic acids and their complexes, with much recent interest in very large assemblies like the ribosome. As crystals of such large assemblies often diffract weakly (resolution worse than 4 Å), we need methods that work at such low resolution. In macromolecular assemblies, some of the components may be known at high resolution, whereas others are unknown: current refinement methods fail as they require a high-resolution starting structure for the entire complex[1]. Determining the structure of such complexes, which are often of key biological importance, should be possible in principle as the number of independent diffraction intensities at a resolution better than 5 Å generally exceeds the number of degrees of freedom. Here we introduce a method that adds specific information from known homologous structures but allows global and local deformations of these homology models. Our approach uses the observation that local protein structure tends to be conserved as sequence and function evolve. Cross-validation with R_{free} (the free *R*-factor) determines the optimum deformation and influence of the homology model. For test cases at 3.5–5 Å resolution with known structures at high resolution, our method gives significant improvements over conventional refinement in the model as monitored by coordinate accuracy, the definition of secondary structure and the quality of electron density maps. For re-refinements of a representative set of 19 low-resolution crystal structures from the Protein Data Bank, we find similar improvements. Thus, a structure derived from low-resolution diffraction data can have quality similar to a high-resolution structure. Our method is applicable to the study of weakly diffracting crystals using X-ray micro-diffraction[2] as well as data from new X-ray light sources[3]. Use of homology information is not restricted to X-ray crystallography and cryo-electron microscopy: as optical imaging advances to subnanometre resolution[4,5], it can use similar tools.

A grand challenge in structural biology is to determine atomic structures of large macromolecular complexes. Unfortunately, growth of well-ordered crystals needed for high-resolution X-ray crystallography is often precluded by inherent flexibility and disordered solvent, lipids and other essential components; diffraction often is weak, anisotropic and has an effective resolution of worse than ~4 Å. Atomic interpretation of resulting electron density maps is limited to fitting rigid models. There is a need for accurate atomic structures from low-resolution diffraction data to reach mechanistic conclusions that critically depend on individually resolved residues.

X-ray crystal structures can achieve 'super-resolution', where the estimated coordinate accuracy is better than the resolution limit of the diffraction data (typically by 10 times), by imposing constraints when interpreting observed diffraction data and electron density maps. Super-resolution arises from the excluded volumes of atoms: the

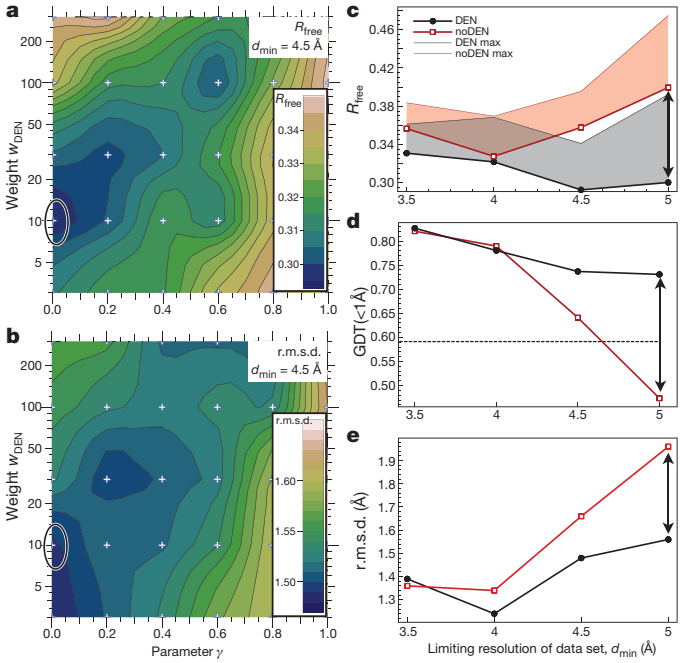

Figure 1 | Results for the penicillopepsin test calculations using the MLHL target function (experimental phase information). In all panels, black lines refer to DEN refinements, whereas red lines refer to noDEN refinements. **a**, How the (γ, w_{DEN}) grid-search determines the values that give the best R_{free} value for the synthetic diffraction data set at d_{min} = 4.5 Å. The R_{free} value is contoured using values calculated on a 6 × 5 grid (marked by small '+' signs) where the parameter γ was [0.0, 0.2, 0.4, 0.6, 0.8, 1.0] and w_{DEN} was [3, 10, 30, 100, 300]. For each parameter pair, we performed an extensive refinement protocol (Methods). The contour plot shows clear minima and maxima with the value of R_{free} varying from 0.295 to 0.35. **b**, The contour map of the all-atom r.m.s.d. between the target structure PDB 3APP and the DEN-refined structure (repeat with the lowest R_{free} value) at each grid point in **a**. Again there are clear minima and maxima with the r.m.s.d. varying from 1.47 to 1.60 Å. **c**, The R_{free} value as a function of d_{min} of the four synthetic diffraction data sets. Thick lines mark the lowest R_{free} values obtained from the ten repeats using the optimum parameters; the corresponding thin lines mark the highest R_{free} values. For the synthetic data sets at $d_{min} \geq 4$ Å, DEN refinement performs much better than noDEN, reaching lower R_{free} values. **d**, The variation of Zemla's GDT(<1 Å) score[16], which measures structural similarity to the target structure PDB 3APP, as a function of d_{min}; the dashed line indicates the GDT score of the initial model. At all resolutions, DEN outperforms noDEN and gives GDT values that are more favourable (higher) than those of the initial structure. **e**, How the r.m.s.d. of all atoms to the PDB 3APP target structure varies versus d_{min} of the four synthetic diffraction data sets. Once again DEN gives lower r.m.s.d. values, especially at low resolution. The DEN-refined models used in **d** and **e** correspond to the best models among ten repeats as assessed by R_{free} (filled black circles in **c**). Black ellipses in **a** and **b** indicate values corresponding to the structure with lowest R_{free} value obtained for d_{min} = 4.5Å.

[1]Institut für Strukturbiologie und Biophysik (ISB-3), Forschungszentrum Jülich, 52425 Jülich, Germany. [2]Department of Structural Biology, Stanford School of Medicine, D100 Fairchild Building, 299 West Campus Drive, Stanford, California 94305, USA. [3]Howard Hughes Medical Institute, [4]Department of Molecular and Cellular Physiology, [5]Department of Neurology and Neurological Sciences, [6]Department of Photon Science, Stanford University, James H. Clark Center E300, 318 Campus Drive, Stanford, California 94305, USA.

scattering objects are always further apart than half the wavelength of X-ray radiation typically used (1–2 Å). This atomicity leads to a solution of the phase problem for small molecule crystals[6], and it allows estimation of coordinate errors[7]. Assuming that polymers have standard chemical bond lengths and bond angles extends this concept to the resolution characteristic of macromolecular crystallography[8,9].

Low-resolution X-ray diffraction data at 5 Å contains, in principle, sufficient information to determine the true structure (the 'target structure'), because the number of observable diffracted intensities exceeds the number of torsion-angle degrees of freedom of a macromolecule (W. A. Hendrickson, personal communication). Although an exhaustive conformational search in torsion-angle space against the diffraction data should lead to an accurate structure at 5 Å resolution, such a search is computationally intractable. Our approach aids the search by adding known information to the observed data at low resolution. Instead of adding just generic information about macromolecular stereochemistry (idealized chemical bond lengths, bond angles and atom sizes that heralded the era of reciprocal-space restrained refinement[8,9]), we also add specific information for the particular macromolecule(s) or complex, deriving this information from known structures of homologous proteins or domains (the 'reference model').

The target structure often differs from the reference model by large-scale deformations, related to the approximate conservation of local polypeptide geometry as sequence and function evolve. How can such deformations be mathematically described? An early approach[10] used low-frequency normal modes, shown to reproduce large-scale collective changes in structures with very few degrees of freedom[11]; it has been used to refine protein structures with low-resolution X-ray or cryo-electron microscopy data[12,13]. Here we take a very different approach. Instead of choosing special collective degrees of freedom, we use an extension of our deformable elastic network (DEN) approach[14]. DEN fits models into cryo-electron density maps, allowing large deformations such as hinge bending. DEN defines springs between selected atom pairs using the reference model as the template. The equilibrium distance of each spring (the distance at which its potential energy is minimum) is initially set to the distance between these atoms in the starting structure for refinement. As torsion-angle molecular dynamics against a combined target function (comprising diffraction data, DEN, and energy, equation (1) in Methods Summary) proceeds, the equilibrium lengths of the DEN network are adjusted to incorporate the distance information from the reference model. The degree of this adjustment is controlled by a parameter, γ (Methods). Here we extend DEN to homology models, or more generally, any reference model, such as a predicted structure.

We first tested our method on a model system, the protein penicillopepsin whose structure had been determined to $d_{min} = 1.8$ Å resolution (PDB ID 3APP)[15]. Synthetic low-resolution data sets were generated at 3.5, 4.0, 4.5 and 5.0 Å resolution (Methods). Optimum values for the γ and w_{DEN} parameters used for DEN refinement were obtained by a grid search against R_{free} (see Fig. 1a for refinement at 4.5 Å resolution). With this standard protocol, referred to here as DEN, the R_{free} optimum is found at (γ, w_{DEN}) = (0, 10) (marked by the black ellipse in Fig. 1a). As a control, we performed a refinement using exactly the same protocol but with the DEN potential set to zero; this corresponds to a second standard protocol, referred to here as 'noDEN'. We assess the quality of the resulting models by comparing the structures resulting from the DEN and noDEN refinements to the target structure (the 1.8 Å-resolution crystal structure of penicillopepsin, PDB 3APP). Figure 1b shows a contour plot of the all-atom root-mean-square difference (r.m.s.d.) between PDB 3APP and the corresponding DEN refined structures from Fig. 1a. The r.m.s.d. shows good agreement with the R_{free} values. Thus, the lowest R_{free} value should be a good predictor for the (γ, w_{DEN}) pair that gives the optimum structure in cases when a high-resolution target structure is not known. The resulting electron density maps (Supplementary Fig. 1) are greatly improved, showing better connectivity and sidechain definition compared with noDEN refinement.

DEN refinement dramatically improves the structure compared to noDEN over a wide range of low-resolution data sets (Fig. 1c–e, Table 1), and with and without experimental phase information (compare Fig. 1 and Supplementary Fig. 2): the DEN R_{free} values (Fig. 1c) are nearly independent of the limiting resolution of the synthetic data sets (black), whereas they steadily increase for noDEN (red). For the data set at 5 Å resolution, DEN improves R_{free} by 0.1 (Fig. 1c, black double-arrow). The global distance test (GDT) score[16] measures the fraction of atoms that fit the target structure well and thus focuses on the more accurate part of the structure (Fig. 1d). For data sets at $d_{min} > 4$ Å, the GDT scores dramatically worsen for the structures refined without DEN: the resulting GDT score is worse than that of the initial model (Fig. 1d, dashed

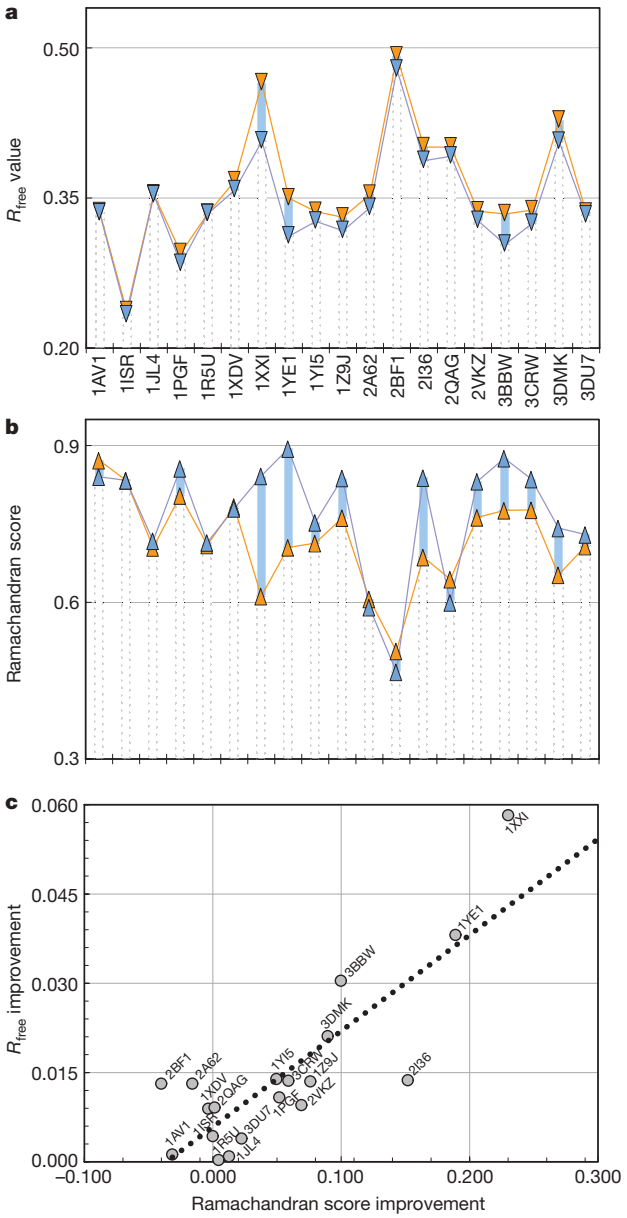

Figure 2 | Re-refinement of 19 low-resolution PDB structures. **a**, R_{free} values of PDB structures refined with DEN (blue) and without DEN (noDEN, orange). In every case, the DEN-refined structure has the lower R_{free} value. For each protein, (γ, w_{DEN}) parameter optimization was performed (Methods, Supplementary Fig. 4), and the structure with the lowest R_{free} value used for analysis. **b**, Fraction of residues in the favoured region of the Ramachandran plot as determined by Molprobity[29], termed here Ramachandran score. **c**, Significant correlation (correlation coefficient, 0.83) is seen between R_{free} improvement and Ramachandran score improvement for DEN versus noDEN.

Table 1 | DEN refinement improves structures refined against four synthetic data sets

Target function	Resolution (Å)	R_{free}			$R_{free} - R_{work}$		Ramachandran score		
		DEN	noDEN	Improvement	DEN	noDEN	DEN	noDEN	Improvement
MLHL	3.50	0.331	0.357	0.0256	0.05	**0.09**	0.783	**0.783**	0.0000
MLHL	4.00	0.322	**0.328**	*0.0058*	0.07	**0.09**	0.754	0.772	*−0.0184*
MLHL	4.50	**0.293**	0.358	0.0651	**0.02**	0.11	0.702	0.632	0.0699
MLHL	5.00	0.300	0.400	**0.0991**	**0.02**	0.14	**0.790**	0.599	**0.1912**
MLF	3.50	*0.378*	0.390	0.0123	0.10	0.11	0.757	0.699	0.0588
MLF	4.00	0.347	0.391	0.0445	0.09	0.15	0.732	0.658	0.0735
MLF	4.50	0.348	0.413	0.0655	0.08	0.12	0.702	*0.544*	0.1581
MLF	5.00	0.341	*0.425*	0.0841	*0.13*	*0.18*	*0.599*	0.551	0.0478
Average	4.25	0.332	0.383	0.0503	0.07	0.12	0.727	0.655	0.0726
Minimum	3.50	**0.293**	**0.328**	*0.0058*	**0.02**	**0.09**	*0.599*	*0.544*	*−0.0184*
Maximum	5.00	*0.378*	*0.425*	**0.0991**	*0.13*	*0.18*	**0.790**	**0.783**	**0.1912**

Refinement starts from a homology model of penicillopepsin (PDB 3APP) that was built using the endothiapepsin structure (PDB 4APE) as a template with an initial r.m.s.d. of 1.7 Å. DEN refinements were performed (Methods). DEN-refined structures are dramatically improved over noDEN structures, especially at low resolution (>4 Å), with an average improvement of 0.078 in R_{free} for resolutions of 4.50 and 5.00 Å, with (MLHL) or without (MLF) phases. At these same resolutions, the secondary structure definition also improved for DEN structures, as shown by a higher Ramachandran score (as determined by Molprobity[29]). At the higher resolutions of 3.50 and 4.00 Å, the Ramachandran score only improves without phase information, which shows that DEN provides little new information at higher resolution when experimental phase information is available. As expected, R_{free} values are lower when using phase information for both DEN and noDEN refinements, with an average improvement of 0.042 for DEN and 0.045 for noDEN. In each column, bold font marks the most-favourable maximum or minimum value (high Ramachandran score or a low *R*-value); italic font marks the least-favourable value.

line). In contrast, the GDT score of the DEN-refined models is consistently high. The r.m.s.d. to the target structure (PDB 3APP) (Fig. 1e) is also significantly smaller with DEN. These improvements persist even when refinement cycles are added to the protocol without DEN (that is, with w_{DEN} set to zero) (Supplementary Fig. 3).

In a broader test, we applied our method to 19 existing structures for which only low-resolution X-ray diffraction data are available (worse than 4 Å). To focus on DEN's core strengths, we chose to re-refine the existing low-resolution structures with the help of a reference model that contains higher-resolution information. To minimize bias, we automated the re-refinement which is expected to limit structure improvement; as discussed below, much better results could be obtained by an investigator familiar with the structure and differences from the reference model.

For each selected PDB structure, a reference model was built by homology modelling on templates manually selected by simultaneously satisfying the three criteria of high sequence identity, high resolution and large number of matched residues (Supplementary Tables 1 and 2). On average, 86% of the residues could be modelled. In some extreme cases (PDB 1AV1, 2VKZ and 2BF1), the main chain r.m.s.d. of the template to the corresponding low-resolution PDB structure was around 10 Å, in which case structural similarity is likely to be limited and significant improvement is not expected. We included these cases to see whether DEN can lead to improvements (PDB 2VKZ and 2BF1, see below), and show that even in the worst case (PDB 1AV1), DEN does not lead to a deterioration of the structure.

The R_{free} values of the DEN refined structures (Fig. 2a, Table 2 and Supplementary Fig. 4) all improved relative to the noDEN structures. Eleven structures show an improvement of over 0.01, four an improvement of over 0.02, and the best an improvement of 0.058 (PDB 1XXI), a 12% improvement. The difference between *R* and R_{free} is on average 0.018 smaller for DEN compared with noDEN (Table 2); this indicates that overfitting is significantly reduced by DEN. Both the minimum and the maximum R_{free} values are generally lower for DEN than for noDEN (Supplementary Table 3), indicating that relevant, low-R_{free} regions of conformational space are better sampled.

The Ramachandran score shows that DEN refinement generally improves the secondary structure compared with noDEN (Fig. 2b and Table 2), with an average increase of 0.05. The largest improvement (0.23, or 37%) is again seen for PDB 1XXI. There is high correlation between

Table 2 | DEN refinement improves low-resolution structures in the PDB

PDB ID	Resolution (Å)	No. of residues	R_{free}			$R_{free} - R_{work}$		Ramachandran score			Comments on differences
			DEN	noDEN	Improvement	DEN	noDEN	DEN	noDEN	Improvement	
1AV1	4.00	804	0.335	0.336	0.0012	0.07	0.07	0.840	**0.872**	**−0.0314**	
1ISR	4.00	448	**0.233**	**0.237**	0.0043	0.07	0.07	0.833	0.833	0.0000	
1JL4	4.30	557	0.353	0.354	0.0009	*0.12*	0.11	0.718	0.705	0.0127	
1PGF	4.50	1,102	0.284	0.295	0.0108	0.08	0.11	0.856	0.804	0.0519	Small throughout the chains
1R5U	4.50	3,517	0.334	0.335	*0.0003*	0.05	0.05	0.714	0.710	0.0046	
1XDV	4.10	1,517	0.358	0.367	0.0089	*0.12*	0.11	0.780	0.783	−0.0034	
1XXI	4.10	3,532	0.407	0.465	**0.0582**	0.05	0.12	0.842	0.612	0.2301	Large (~4 Å domain motions)
1YE1	4.50	574	0.312	0.350	0.0381	0.08	*0.15*	**0.894**	0.705	0.1890	Small throughout
1YI5	4.20	1,356	0.323	0.336	0.0139	0.07	0.09	0.758	0.709	0.0497	Local in several chains
1Z9J	4.50	821	0.317	0.331	0.0135	0.07	0.09	0.838	0.762	0.0761	Large in chain A (domain motion)
2A62	4.50	319	0.340	0.353	0.0131	0.07	0.09	0.590	0.606	−0.0159	
2BF1	4.00	304	*0.479*	*0.492*	0.0131	*0.12*	0.12	0.467	*0.507*	*−0.0400*	
2I36	4.10	962	0.387	0.401	0.0137	0.02	0.03	0.839	0.687	0.1520	Local in chain B
2QAG	4.00	702	0.392	0.401	0.0091	0.02	**0.02**	0.616	0.614	0.0016	
2VKZ	4.00	10,941	0.327	0.337	0.0095	0.05	0.07	0.832	0.762	0.0692	Large in subdomain placements
3BBW	4.00	543	0.304	0.334	0.0304	**0.01**	0.04	0.876	0.776	0.0998	Significant local
3CRW	4.00	485	0.324	0.338	0.0136	0.09	0.11	0.836	0.777	0.0589	Large in one domain (hinge motion)
3DMK	4.19	2,127	0.407	0.428	0.0211	0.08	0.11	0.742	0.653	0.0896	Throughout, ref. model only 50%
3DU7	4.10	1,839	0.332	0.336	0.0039	0.09	0.09	0.730	0.707	0.0225	
Average	4.19	1,708	0.345	0.359	0.0146	0.07	0.09	0.768	0.715	0.0535	
Minimum	4.00	304	**0.233**	**0.237**	*0.0003*	**0.01**	**0.02**	*0.467*	*0.507*	*−0.0400*	
Maximum	4.50	10,941	*0.479*	*0.492*	**0.0582**	*0.12*	*0.15*	**0.894**	**0.872**	**0.2301**	

PDB structures (19) were re-refined with and without DEN (Methods). The tested proteins show a wide range of sizes, extending from 304 residues for 2BF1 to 10,941 residues for 1VKZ. The final R_{free} and $R_{free} - R_{work}$ values, as well as Ramachandran scores, are shown. In all cases, DEN refinement shows improvement of R_{free} as compared with noDEN; 11 out of 19 cases show an R_{free} improvement that is larger than 0.01. In 15 of the 19 cases DEN refinement also improves the Ramachandran score (four exceptions are 2BF1, 1AV1, 2A62 and 1XDV). As would be expected R_{free} is larger than R_{work} (the *R*-factor that was optimized), with average differences of 0.07 and 0.09 for DEN and noDEN refinement, respectively. In each column, bold font marks the most favourable maximum or minimum value (high Ramachandran score or low *R*-value); italic font marks the least favourable value. The comments refer to the differences between the reference (ref.) models and the corresponding DEN-refined crystal structures for the cases with $\gamma < 1$ (compare Supplementary Table 4). Two particular examples of these differences are shown in Supplementary Fig. 5.

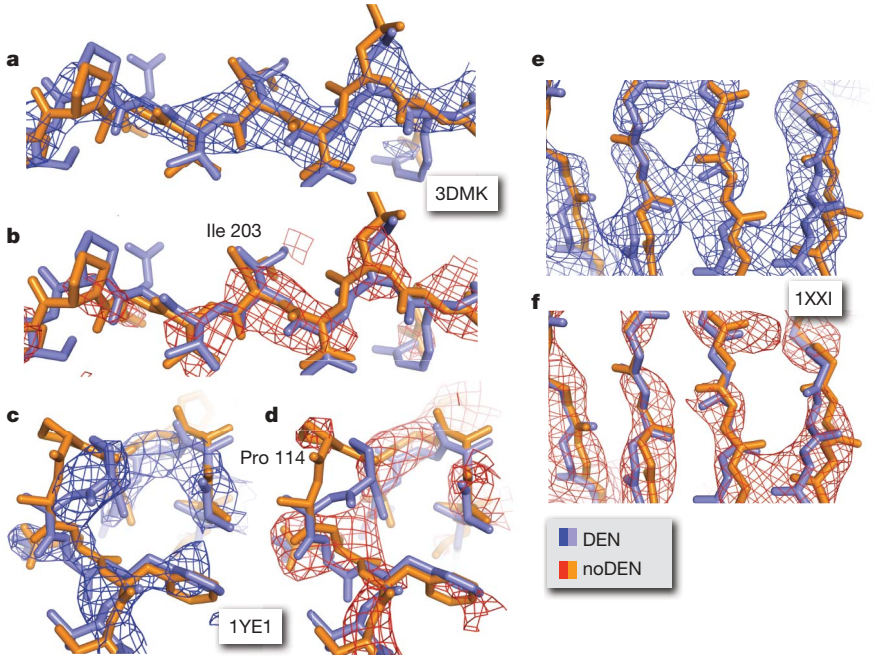

Figure 3 | Electron density map improvement upon DEN refinement for three structures, PDB 3DMK, 1YE1 and 1XXI. The 1YE1 (**c**, **d**) and 1XXI (**e**, **f**) structures are among the cases that benefit most from DEN refinement, whereas the 3DMK (**a**, **b**) structure showed only moderate improvement of the R_{free} value (Table 2). Nevertheless, in all three cases DEN refinement dramatically improves the electron density maps. The structures refined with DEN (DEN, in blue) and without DEN (noDEN, in orange) are superimposed, and the corresponding phase-combined σ_A-weighted $2F_o - F_c$ electron density maps are shown in blue and red, respectively. The density maps for PDB 3DMK and 1XXI were *B*-factor sharpened ($B_{sharp} = -50$ Å^2) and the contour level was set to 1.5σ.

R_{free} and the Ramachandran score improvements (Fig. 2c). The four cases where the Ramachandran score has slightly worsened (PDB 1AV1, 1XDV, 2A62 and 2BF1) are all cases with an optimal value of $\gamma = 1.0$ (Supplementary Table 4). In these (and five additional cases with $\gamma = 1.0$), the reference model is ignored, as it does not provide useful distances (Methods). As expected, the average R_{free} improvement in these nine cases is small (0.0061, Supplementary Table 4). In contrast, for the ten cases with $\gamma < 1$, the average R_{free} improvement is significant (0.022, Supplementary Table 4). These ten successful cases cover a variety of differences between the reference model and the crystal structure, including large (sub)domain motions, hinge motions, local structural differences, or differences throughout (Table 2 and Supplementary Fig. 5).

We calculated electron density maps from experimental intensities combined with model phases from the DEN- and noDEN-refined structures. In the three cases shown (Fig. 3), the noDEN backbone density is broken in several places (red), making it difficult to correctly trace the backbone. In contrast, the DEN maps show a continuous backbone density (blue). The DEN-refined coordinates also show clear improvements; for example, with DEN, Pro 114 in the PDB 1YE1 structure (Fig. 3c, d) is shifted by 3.2 Å into well-defined electron density (blue), whereas very little density is visible for noDEN (red). Such improved interpretability of electron density maps indicates that the phases calculated from DEN-refined structures are superior to those from noDEN-refined structures.

How does DEN increase the accuracy of the refined structure? For the penicillopepsin test case at 4.5 Å resolution, we analysed the distances between atom pairs not well defined by the diffraction data, specifically those with large root-mean-square fluctuations (r.m.s.f.) between the ten models of the noDEN refinement repeats (Fig. 4 inset). These distances are much closer to the distances in the target structure (PDB 3APP) for DEN compared with noDEN, showing that DEN provides information for distances that are not well defined by the diffraction data.

Performance can be much improved by manually selecting cut-off criteria and structural elements used for DEN. For the unligated SIV gp120 structure[17] (PDB 2BF1), we restricted the DEN network to the main chain and Cβ-atoms of the reference model (HIV gp120–antibody complex at 2.0 Å resolution[18], PDB 2NXZ) and to regions of the structure considered reliable predictors of SIV gp120 structure (at least 35.8% local sequence identity, Supplementary Table 2). Refinement with optimum DEN parameters resulted in a 4%-lower R_{free} value and 8%-higher Ramachandran score. With such judicious manual choice of the network, DEN used the reference model distances ($\gamma = 0.4$, rather than $\gamma = 1$ for automated DEN), and produced a more accurate structure as assessed by R_{free}.

Cross-validation with R_{free} allows determination of the optimum parameter values (particularly γ) yielding more accurate models at low resolution even when no high-resolution model is available. DEN can be applied to predicted structures, which have shown promise in molecular replacement[19], and to RNA/DNA. DEN can be easily modified in future developments: for example, individual weights

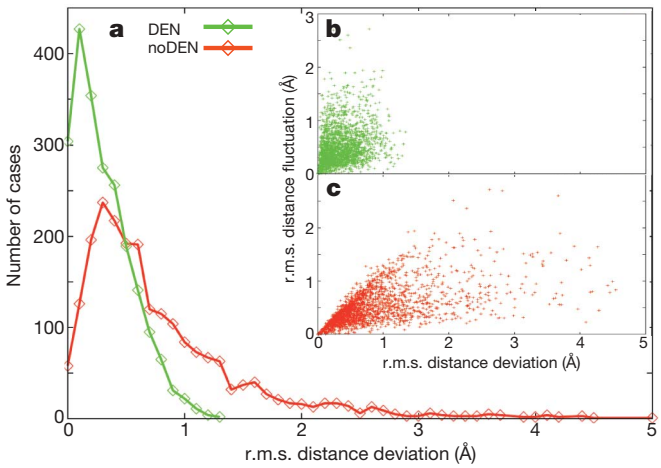

Figure 4 | DEN provides information for degrees of freedom that are weakly defined by the experimental diffraction data. **a**, DEN (green) and noDEN (red) histograms of r.m.s. distance deviation. This quantity is the r.m.s. deviation of DEN restraint distances in the target structure (PDB 3APP) from those in the ten refinement repeats (starting from the PDB 4APE initial model with $d_{min} = 4.5$ Å, the MLHL target function[23] and DEN optimum parameters (γ, w_{DEN}) = (0,10); see Fig. 1a). The largest r.m.s. distance deviation is much smaller for DEN compared with noDEN. Inset, the r.m.s. fluctuations of each distance over the ten repeats of noDEN refinement are plotted against r.m.s. distance deviation for DEN (**b**, green) and noDEN (**c**, red). Large r.m.s. fluctuation values (>1.5 Å) represent the DEN distances that are not well defined by the diffraction data. For DEN, these distances have small r.m.s. distance deviation values (<1.0 Å), whereas for noDEN they have large values. Restraint distances are much closer to the distances in the target structure for DEN, which effectively provides information missing from low-resolution experimental data.

for DEN distances could account for model error, variations in a family of homologous structures, or predicted loop conformations. Criteria for selection of distances can also be modified, as done manually for PDB 2BF1.

METHODS SUMMARY

The total energy function. This consists of a weighted sum of three terms

$$E_{\text{total}} = E_{\text{geometric}} + w_{\text{a}} E_{\text{ML}} + w_{\text{DEN}} E_{\text{DEN}}(\gamma) \quad (1)$$

where $E_{\text{geometric}}$ is a 'geometric' or stereochemical energy function commonly used for macromolecular crystal structure refinement[20], E_{ML} is a maximum likelihood target function that incorporates experimental X-ray amplitude (and optionally phase information)[21–23], $E_{\text{DEN}}(\gamma)$ is the DEN potential (Methods), and w_{a} and w_{DEN} are relative weights. Geometric energy functions have been used for refinement of macromolecules since their first introduction for energy refinement[24] and application to X-ray refinement[9]. The refinement protocol uses torsion-angle dynamics[25] against E_{total} and *B*-factor refinement (Methods), and was repeated multiple times.

For DEN, the target sequence must be sufficiently close to a homologous sequence (sequence identity at least 30%), which means that the target and homologue will be structurally similar. It also requires that the homologue structure was determined at sufficiently high resolution (at least 3.5 Å resolution), so that it will contain useful specific high-resolution information about the target. Homology models for the target sequence were constructed using standard well-accepted methods such as SegMod[26] or MODELLER[27]. Often, multiple homology models were combined to cover the entire target structure even when it consists of multiple domains and polypeptide chains.

Our approach is a major advance over conventional modelling of low-resolution X-ray diffraction data by fitting rigid bodies[28], as it accounts for deformations of the models while at the same time using a minimal set of variables (the single-bond torsion angles); for five cases, our re-refinement achieved a substantial improvement in R_{free} over rigid-body refined structures (Supplementary Table 1). Optionally, we turn off the DEN potential during the last refinement repeats to assess the robustness of the improvement achieved by DEN. The radius of convergence of DEN refinement is very large: in tests, automatic correction of polypeptide chain register in α-helices was observed, a notoriously difficult problem for macromolecular refinement.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 October 2009; accepted 10 February 2010.
Published online 7 April 2010.

1. Davies, J. M., Brunger, A. T. & Weis, W. I. Improved structures of full-length p97, an AAA ATPase: implications for mechanisms of nucleotide-dependent conformational change. *Structure* **16,** 715–726 (2008).
2. Sanishvili, R. *et al.* A 7 μm mini-beam improves diffraction data from small or imperfect crystals of macromolecules. *Acta Crystallogr. D* **64,** 425–435 (2008).
3. Raines, K. S. *et al.* Three-dimensional structure determination from a single view. *Nature* **463,** 214–217 (2010).
4. Moerner, W. E. New directions in single-molecule imaging and analysis. *Proc. Natl Acad. Sci. USA* **104,** 12596–12602 (2007).
5. Pertsinidis, A., Zhang, Y. & Chu, S. Localization, registration and distance measurements between single-molecule fluorescent probes with sub-nanometer precision and accuracy. *Nature* (submitted).
6. Karle, J. & Hauptman, H. A theory of phase determination for the four types of non-centrosymmetric space groups 1P222, 2P22, 3P(1)2, 3P(2)2. *Acta Crystallogr.* **9,** 635–651 (1956).
7. Luzzati, V. Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr.* **5,** 802–809 (1952).
8. Hendrickson, W. A. & Konnert, J. H. A restrained-parameter thermal-factor refinement procedure. *Acta Crystallogr. A* **36,** 344–350 (1980).
9. Jack, A. & Levitt, M. Refinement of large structures by simultaneous minimization of energy and R factor. *Acta Crystallogr. A* **34,** 931–935 (1987).
10. Diamond, R. On the use of normal modes in the thermal parameter refinement: theory and application to the bovine pancreatic trypsin inhibitor. *Acta Crystallogr. A* **46,** 425–435 (1990).
11. Levitt, M., Sander, C. & Stern, P. S. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* **181,** 423–447 (1985).
12. Delarue, M. & Dumas, P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc. Natl Acad. Sci. USA* **101,** 6957–6962 (2004).
13. Tama, F., Miyashita, O. & Brooks, C. L. III. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J. Struct. Biol.* **147,** 315–326 (2004).
14. Schröder, G. F., Brunger, A. T. & Levitt, M. Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure* **15,** 1630–1641 (2007).
15. James, M. N. & Sielecki, A. R. Structure and refinement of penicillopepsin at 1.8 A resolution. *J. Mol. Biol.* **163,** 299–361 (1983).
16. Zemla, A. LGA: a method for finding 3D similarities in protein structures. *Nucleic Acids Res.* **31,** 3370–3374 (2003).
17. Chen, B. *et al.* Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature* **433,** 834–841 (2005).
18. Zhou, T. *et al.* Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445,** 732–737 (2007).
19. Qian, B. *et al.* High-resolution structure prediction and the crystallographic phase problem. *Nature* **450,** 259–264 (2007).
20. Engh, R. & Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* **47,** 392–400 (1991).
21. Bricogne, G. & Gilmore, C. J. A multisolution method of phase determination by combined maximization of entropy and likelihood. I. Theory, algorithms and strategy. *Acta Crystallogr. A* **46,** 284–297 (1990).
22. Pannu, S. N. & Read, R. J. Improved structure refinement through maximum likelihood. *Acta Crystallogr. A* **52,** 659–668 (1996).
23. Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. Incorporation of prior phase information strengthens maximum-likelihood structure refinement. *Acta Crystallogr. D* **54,** 1285–1294 (1998).
24. Levitt, M. & Lifson, S. Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Mol. Biol.* **46,** 269–279 (1969).
25. Rice, L. M. & Brunger, A. T. Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins* **19,** 277–290 (1994).
26. Levitt, M. Accurate modeling of protein conformation by automatic segment matching. *J. Mol. Biol.* **226,** 507–533 (1992).
27. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815 (1993).
28. Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S. H. A structure-factor least squares refinement procedure for macromolecular structures using constrained and restrained parameters. *Acta Crystallogr. A* **33,** 800–804 (1977).
29. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32** (Web Server issue), W615–W619 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. D. Adams, S. C. Harrison and T. D. Fenn for discussions. We also thank the National Science Foundation for computing resources (CNS-0619926), the National Institutes of Health for both Roadmap Grant PN2 (EY016525) and grant GM63718 to M.L., and the Deutsche Forschungsgemeinschaft (DFG) for support for G.F.S.

Author Contributions G.F.S. developed the computational algorithms, and G.F.S. and A.T.B. designed the computational experiments and performed all calculations and analysis. All authors wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.F.S. (gu.schroeder@fz-juelich.de) and A.T.B. (brunger@stanford.edu).

METHODS

Extension of the DEN method. We extended the DEN approach[14] so as to accommodate reference models obtained from homology modelling for multi-domain proteins. Unless specified otherwise we used automatic DEN generation to avoid the need to define the boundaries between domains. This was done by using restraint distances between N randomly selected pairs of atoms in the reference model that are separated by not more than ten residues along the polypeptide sequence and are separated by 3–15 Å in space. The value of N is chosen to be equal to the number of atoms, so the set of distance restraints is relatively sparse, with an average of one restraint per atom. If needed, the sequence separation limit of ten residues can be relaxed to include additional inter-atomic distances so as to better define β-sheets and hairpin loops, but this requires having to explicitly define domain boundaries (see the 2BF1 example discussed in the text). The specific limits on maximum distance (15 Å) and sequence separation limit along the chain (10) given above were chosen by trial and error using as test case the ribose binding protein described in ref. 14. There is a clear trade-off: when the limits are too large, the flexibility of the DEN becomes restricted, requiring large deformation of the network to fit to the diffraction data. When the limits are too small, there are few restraints so that very little information is added from the reference model. In this work, the reference models are obtained by homology modelling[19,27] (see below) and thus they are expected to have good stereochemistry and secondary structure definition.

The elastic network energy term is a sum of distance deviations to the power p over all selected atom pairs i,j

$$E_{\mathrm{DEN}}(\gamma,n) = \sum_{N \text{ pairs } i,j} \left(d_{ij} - d_{ij}^{0}(\gamma,n)\right)^{p} \qquad (2)$$

where d_{ij} is the distance between atom pair i and j in the current atomic model and $d_{ij}^{0}(\gamma,n)$ is the corresponding equilibrium distance after DEN update step n. $d_{ij}^{0}(\gamma,0)$ is the corresponding distance in the initial (starting) model used for refinement. The exponent p is set to 2 for better numerical stability or to 4 for faster convergence. The equilibrium distances $d_{ij}^{0}(\gamma,n)$ are updated every six torsion-angle molecular dynamics steps (each with a time step of 4 fs) by

$$d_{ij}^{0}(\gamma,n+1) = (1-\kappa)\,d_{ij}^{0}(\gamma,n) + \kappa\left[\gamma\, d_{ij} + (1-\gamma)\, d_{ij}^{\mathrm{ref}}\right] \qquad (3)$$

where d_{ij}^{ref} are the distances in the reference model, the parameter κ determines the speed at which the network adapts to the requirements of the total energy function (equation (1)) and the parameter γ balances the influences of the diffraction data and the reference model[14]. Note that $\gamma = 0$ corresponds to a non-deformable elastic network. Although κ is an adjustable parameter, we found that the results are insensitive to the exact choice of κ and chose to always set its value to 0.1. To allow initial relaxation of the atomic model without influence of the reference model, the κ parameter is set to zero during the first few (typically 3) refinement macrocycles (see below) and then set to one of several trial values or to the optimized value obtained by a global grid search (see below) for the remaining cycles.

The DEN potentials are weighted by a factor w_{DEN} and then added to the standard crystallographic target function E_{ML} and the geometric energy function $E_{\mathrm{geometric}}$ (equation (1)). The parameters γ and w_{DEN} are the most important adjustable parameters for DEN and their values are optimized in a global grid search against R_{free}. It is important to get the value of γ right so as to balance fitting of the diffraction data and to incorporate the most useful information from the reference model. A value of $\gamma = 0$ means that the DEN potential minima strictly move towards the distances derived from the reference model, whereas a value of $\gamma = 1$ means that no distance information from the reference model is used and the potential minimum gradually follows the coordinates of the structure as it is being fitted to the diffraction data. Clearly, we expect the most power of this method to arise when the value of γ falls somewhere between these two values.

For the special case of $\gamma = 1$, no distance information from the reference model will ever be used. From equation (3), it is clear that the updated network distances $d_{ij}^{0}(\gamma,n+1)$ then only depend on the equilibrium distances in the previous step $d_{ij}^{0}(\gamma,n)$ and on the distances in the current atomic model d_{ij}, but not on the corresponding distances in the reference model d_{ij}^{ref}. Such use of $\gamma = 1$ does not correspond to not using DEN restraints at all: the elastic network is still present and it will slowly follow the structural changes of the atomic model as it is being refined. In other words, the reference model defines the 'topology' of the DEN restraints, so that this method is in effect a generalized version of the use of secondary structure restraints during refinement. Indeed, even for cases with $\gamma = 1$ there is still an improvement in R_{free} compared with refinements without DEN restraints (Supplementary Table 4). This occurs because the DEN restraints still influence the conformational search. When the starting structure is already close to a good solution, DEN increases the chance to find a better minimum and DEN allows it to sample its neighbourhood more extensively, thereby increasing the chance to move towards lower energy regions.

Refinement protocol. Starting from an initial model, torsion-angle molecular dynamics simulations[25] are performed with the generalized forces derived from E_{total} (equation (1)). Use of torsion angles as a reduced set of variables also has a long history[30,31], but ref. 25 describes the first algorithm that exactly integrates the equation of motion in this reduced variable space for macromolecules. Note that with torsion-angle molecular dynamics, the term $E_{\mathrm{geometric}}$ (equation (1)) becomes a simple repulsive van der Waals term plus bond length and angle restraints for disulphide bonds as these cannot be exactly constrained in the torsion-angle dynamics method implemented in CNS[32]. The initial (starting) model can be any atomic model that is reasonably close to the crystal structure; in particular, the initial model can be set to the reference model or to a model that has been built or rebuilt manually. The weight w_{a} is chosen to yield comparable absolute values of the gradients of $(E_{\mathrm{geometric}} + w_{\mathrm{DEN}}E_{\mathrm{DEN}})$ and E_{ML} averaged over a 0.1 ps molecular dynamics simulation at 300 K (ref. 33). A typical refinement protocol used several macrocycles (usually eight) of slow-cooling (cooling rate, 50 K over 6 steps) torsion-angle molecular dynamics[25] starting at 3,000 K and ending at 0 K with a time step of 4 fs, interspersed with overall anisotropic B-factor refinement, grid-search bulk solvent model parameter optimization[34], and (optionally) grouped B-factor refinement. Optionally, the atomic van der Waals radii can be artificially reduced (typically 75%) during the first few macrocycles of the refinement protocol to improve sampling of conformational space. The entire refinement protocol was repeated a number of times (usually ten) with different random number seeds used for the velocity assignments (termed 'refinement repeats'), and the structure with the lowest R_{free} value was kept for further analysis.

The grouped B-factor refinement method (two B-values per residue, one for backbone and one for sidechain atoms) was generalized by imposing restraints between the grouped B-values similar to individual restrained B-factor refinement. To accommodate the reduction of degrees of freedom caused by imposing group constraints, the target standard deviations (σ values) for the B-factor restraints had to be increased considerably (a tenfold increase of the σ values was found by trial and error to be appropriate for most cases). We find that this 'restrained grouped' B-factor method produces lower R_{free} values for low-resolution structures than either individual restrained or conventional grouped B-factor refinement (data not shown).

Optionally, a few (typically 2) macrocycles without DEN (that is, w_{DEN} set to zero) can be added at the end of the refinement protocol. Test calculations showed that the improvements achieved by DEN (in terms of R_{free} and Ramachandran statistics) persist during these final refinement rounds (Supplementary Fig. 3).

It should be noted that positional (xyz) minimization of E_{total} can sometimes be detrimental for low-resolution refinements. Such minimization is done in Cartesian space, so that more degrees of freedom (namely the bond angles and lengths) are used relative to those used in torsion space. Such an increased number of degrees of freedom can lead to over-fitting which will be manifested by an increased R_{free} value. In addition, the presence of DEN may distort bonds and angles involving atoms that are restrained by the elastic energy term (equation (2)). Thus, we choose not to use positional minimization throughout this work.

Optimization of DEN parameters. For each protein model and set of X-ray diffraction data, the optimum values of the γ and w_{DEN} parameters were found by a global two-dimensional grid search. At each grid point, ten refinement repeats were performed with different random initial velocities, and the refinement with the (γ, w_{DEN}) pair that produced the lowest R_{free} was used for subsequent analysis. For all cases, we performed 30 combinations of six γ-values [0.0, 0.2, 0.4, 0.6, 0.8,1.0] and five w_{DEN} values [3, 10, 30, 100, 300].

Penicillopepsin test cases. Test calculations with penicillopepsin used the published coordinates (PDB ID 3APP), diffraction data, and experimental phases obtained by single isomorphous replacement[15]. We generated four synthetic diffraction data sets of increasingly lower resolution (d_{min} ranging from 3.5 to 5.0 Å) by truncating the original diffraction data and subsequently applying B-factor smoothing given by the factor $\exp(-B\sin^2\theta/\lambda^2)$ with $B = 17.5$, 26.25, 35.0 and 43.75 Å^2 for the data sets at resolutions set by d_{min} values of 3.5, 4.0, 4.5 and 5.0 Å, respectively. These specific smoothing B-factors were the smallest corrections to the diffraction data needed to obtain reasonable overall B-factors of the bulk solvent model and of the atomic model (use of much larger smoothing factors resulted in artefacts for this particular diffraction data set).

As a starting structure we chose a homology model based on PDB 4APE (endothiapepsin)[35] that has an r.m.s.d. of 1.7 Å from the penicillopepsin structure and an identity of 51.5% to its sequence. This homology model, which was generated using the automated procedure described above, serves as the reference model for DEN. The initial position and orientation of the starting model

was obtained by molecular replacement for each respective diffraction limit as previously described[36]. To facilitate comparison, the specific random selection of DEN restraint atom pairs was kept the same for all synthetic data sets.

We monitored the accuracy of our refined models as a function of the limiting resolution d_{min} of the synthetic diffraction data set used. For each of the 6×5 combinations of γ and w_{DEN} parameter values with $\gamma = [0.0, 0.2, 0.4, 0.6, 0.8, 1.0]$ and $w_{DEN} = [3, 10, 30, 100, 300]$, we performed a refinement protocol consisting of ten refinement repeats (consisting of eight macrocycles of torsion-angle refinement) and the structure with the lowest R_{free} value was used for subsequent analysis. This global search for the best values of the (γ, w_{DEN}) parameters was repeated separately for each synthetic diffraction data set. To allow proper comparison between DEN and conventional refinement, the same refinement protocol was used for a case with w_{DEN} set to zero (termed 'noDEN'). No individual or grouped *B*-factor refinement was used for all refinements with penicillopepsin.

Figure 1a shows a contour plot of the resulting (best) R_{free} values for the synthetic diffraction data set with $d_{min} = 4.5$ Å for refinements with experimental phase information using the MLHL target function[23]. To investigate the influence of experimental phase information, we repeated the entire sets of refinements in the absence of experimental phase information in the crystallographic target function (MLF)[23] and found similar improvements as found for the refinements with experimental phase (Supplementary Fig. 2 and Table 1). Of course, the use of experimental phase information also improves the noDEN refinements compared to amplitude-based refinements (compare Fig. 1d, e and Supplementary Fig. 2d, e), but in any case, DEN always outperforms noDEN refinement since it adds new information to the refinement process that is not included in the diffraction data or experimental phases. For the refinement against the synthetic diffraction data set with $d_{min} = 4.5$ Å, we also tested the effect of adding two cycles with w_{DEN} set to zero at the end of the DEN refinements (Supplementary Fig. 3).

Re-refinement of low-resolution PDB structures. We randomly selected nineteen PDB structures (Supplementary Table 1) from a set of 40 that fulfilled the following four criteria: (1) the limiting diffraction of the native data set is larger than or equal to 4 Å, (2) the R_{work} value is easily reproducible within ± 0.11 (Supplementary Table 1), (3) the deposited PDB structure is an all-atom model (not just a backbone trace), and (4) a high-resolution homologue is present in the PDB. It should be noted that for PDB entry 3CRW, the deposited PDB header information indicated that d_{min} was 4 Å, although there were sparse reflections between 4 and 3 Å resolution in the deposited diffraction data file. We therefore only used diffraction data of lower resolution than 4 Å but the sparse higher-resolution data might have influenced the published PDB 3CRW structure.

We used an automated method to import the deposited coordinates into CNS[32]. Thus, only standard protein residues, ligands and modifications were recognized. The differences between published and re-calculated R_{work} values are typically much less than 5% (Supplementary Table 1) for all but three cases. For PDB 3DMK, the deposited structure was refined using 22 TLS (translation/libration/screw) groups. CNS does not support TLS refinement at present, which explains the large difference in recalculated R_{work}. Likewise, for PDB 3DU7 the difference can be explained by the use of 4 TLS groups in the deposited structure. For PDB 2QAG, the difference of 0.07 can be explained as three nucleotide ligands were excluded during the automated import into CNS. The differences for all other structures can probably be accounted for by other differences in refinement procedures.

The homology models were built as follows: the low-resolution PDB file was split into chains as defined by the chain identifier field. For each chain, a FASTA[37] sequence matching search was performed against the sequences of all known PDB structures. From the resulting list of matching sequences in the PDB, we picked templates for the homology modelling by balancing the three requirements of (1) high sequence identity, (2) high resolution, and (3) a large number of matched residues. Subsequent template-target sequence alignment was performed with the align2d procedure of MODELLER[27]. The automodel class of MODELLER was then used to build five models for each target chain. The best models were chosen based on their DOPE score as provided by the MODELLER program.

To generate the DEN reference models, the homology models for the different chains that correspond to the same target were merged into a single coordinate file. This reference model was used to define DEN restraints as described above. In some cases, the reference model has fewer residues than the original PDB structure (see Supplementary Table 2, column '% Residues in Reference Model'): no DEN restraints were defined for those residues that do not have a corresponding residue in the reference model. Note that the relative position and orientation of different chains in the reference model can be arbitrary as we did not include any DEN restraints between atoms in different chains in these calculations.

The 19 PDB structures were subjected to the same global grid-search for the optimal values of the (γ, w_{DEN}) parameter pair as for the penicillopepsin test case, and the value of R_{free} was contoured for each case (Supplementary Fig. 4). We followed exactly the same positional refinement protocol (ten repeats of eight macrocycles of torsion-angle refinement) as before, except that 50 steps of restrained group *B*-factor refinement were also used in each macrocycle. To allow proper comparison between DEN and conventional refinement, all 19 structures were also subjected to the same refinement protocol without DEN (that is, 'noDEN'). For the 19 test cases, there is no target structure determined at high resolution so we needed to evaluate our success using two criteria: the R_{free} value (Fig. 2a and Table 2), which measures the fit to the X-ray diffraction data, and the Ramachandran score (Fig. 2b and Table 2), which measures the fit of the backbone (ϕ, ψ) torsion angles of the refined structure to those observed in high-accuracy structures.

The re-refinement tests were started from the original PDB structure, rather than from the reference model as was done for penicillopepsin (see above). Accordingly, the initial equilibrium DEN distances $d_{ij}^0(\gamma,0)$ in equation (3) are set to the corresponding values in the original PDB structure. Setting the initial minimum of the elastic network potential to the coordinates of the original PDB structure ensures that the starting structure is not subjected to large forces due to distorted elastic restraints for those cases where the reference model is far away from the starting structure. Such forces on the atoms could lead to unstable molecular dynamics integration.

By default, experimental phase information is not available from the PDB structure factor file, so we chose the same refinement protocol using the MLF target function[22], ignoring any potentially existing phase information for all 19 PDB structures. Further improvements can be expected for both DEN and noDEN refinement upon inclusion of experimental phase information (compare Fig. 1 and Supplementary Fig. 2). Non-crystallographic symmetry (NCS) information was applied only for those cases where such information was provided in the deposited PDB structures (two (PDB ID 2I36 and 3BBW) of the 19 cases). In any case, comparison of these two cases with the remaining 17 cases shows that DEN achieves improvement in both the presence and absence of NCS information (Table 2).

Programs. The DEN method and refinement protocol has been implemented in a new version (v1.3) of the Crystallography and NMR System (CNS)[32,34] (http://cns-online.org/v1.3). The TMscore program[38] was used to calculate the GDT scores, the fraction of residues in the favoured region of the Ramachandran plot were determined by Molprobity[29] (termed here Ramachandran score), and molecular drawings were prepared with PyMOL[39].

30. Gibson, K. D. & Scheraga, H. A. Minimization of polypeptide energy. I. Preliminary structures of bovine pancreatic ribonuclease S-peptide. *Proc. Natl Acad. Sci. USA* **58,** 420–427 (1967).
31. Levitt, M. Protein folding by restrained energy minimization and molecular dynamics. *J. Mol. Biol.* **170,** 723–764 (1983).
32. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
33. Brunger, A. T. Crystallographic refinement by simulated annealing. Application to a 2.8 A resolution structure of aspartate aminotransferase. *J. Mol. Biol.* **203,** 803–816 (1988).
34. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2,** 2728–2733 (2007).
35. Pearl, L. & Blundell, T. The active site of aspartic proteinases. *FEBS Lett.* **174,** 96–101 (1984).
36. Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. Extending the limits of molecular replacement through combined simulated annealing and maximum-likelihood refinement. *Acta Crystallogr. D* **55,** 181–190 (1999).
37. Lipman, D. J. & Pearson, W. R. Rapid and sensitive protein similarity searches. *Science* **227,** 1435–1441 (1985).
38. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33,** 2302–2309 (2005).
39. DeLano, W. *The Pymol Molecular Graphics System* (DeLano Scientific, 2002).