ARTICLES

High-resolution structure prediction and the crystallographic phase problem

Bin Qian[1]*, Srivatsan Raman[1]*, Rhiju Das[1]*, Philip Bradley[1], Airlie J. McCoy[2], Randy J. Read[2] & David Baker[1]

The energy-based refinement of low-resolution protein structure models to atomic-level accuracy is a major challenge for computational structural biology. Here we describe a new approach to refining protein structure models that focuses sampling in regions most likely to contain errors while allowing the whole structure to relax in a physically realistic all-atom force field. In applications to models produced using nuclear magnetic resonance data and to comparative models based on distant structural homologues, the method can significantly improve the accuracy of the structures in terms of both the backbone conformations and the placement of core side chains. Furthermore, the resulting models satisfy a particularly stringent test: they provide significantly better solutions to the X-ray crystallographic phase problem in molecular replacement trials. Finally, we show that all-atom refinement can produce *de novo* protein structure predictions that reach the high accuracy required for molecular replacement without any experimental phase information and in the absence of templates suitable for molecular replacement from the Protein Data Bank. These results suggest that the combination of high-resolution structure prediction with state-of-the-art phasing tools may be unexpectedly powerful in phasing crystallographic data for which molecular replacement is hindered by the absence of sufficiently accurate previous models.

High-resolution prediction of protein structures from their amino acid sequences and the refinement of low-resolution protein structure models to produce more accurate structures are long-standing challenges in computational structural biology[1]. The refinement problem has become particularly important in recent years, as the continued increase in the number of experimentally determined protein structures, together with the explosion of genome sequence information, has made it possible to produce comparative models of a large number of protein structures with wide utility[2]. Ideally, these models would consistently approach the resolution offered by X-ray crystallography, enabling precise drug design and a deeper understanding of catalysis and binding. Accurate high-resolution models can, in principle, be achieved by searching for the lowest energy structure given the sequence of the protein. However, despite progress[3], the large number of degrees of freedom in a protein chain and the ruggedness of the energy landscape produced by strong atomic repulsion at short distances greatly complicate this search for sequences lacking close homologues of known structure.

An important application for predicted structures is to help solve the X-ray crystallographic phase problem[4,5]. Converting X-ray diffraction data into electron density maps of proteins requires the inference of phases associated with each diffraction peak. Although phase estimates can be obtained through the preparation of heavy atom derivatives, the problem can be solved without additional experimental information by the technique of molecular replacement[4,5] given a structure model that has high structural similarity (better than 1.5 Å root-mean-squared (r.m.s.) deviation) to the crystallized protein over a large fraction of the molecule. As an example of the stringency of this condition, models of protein structures derived from nuclear magnetic resonance (NMR) data typically do not give good molecular replacement models for crystallographic data on the same proteins[6]. Perhaps the most successful approach to molecular replacement is the use of previous crystal structures of highly sequence-similar (>40%) templates as search models. In cases of lower sequence similarity, structure prediction tools can frequently help build comparative models that give better molecular replacement solutions; however, the success rate drops rapidly as the template sequence identity falls below 30%[4,5]. In cases where structurally similar experimental models are not available, *ab initio* phasing techniques have had some success for targets with simple folds of high symmetry[7,8] or with new structures that have been rationally designed from first principles[9], but *ab initio* phasing of diffraction data for natural globular proteins remains an unsolved problem.

In this study, we present a new energy-based rebuilding-and-refinement method that consistently improves models derived from NMR, from sequence-distant templates, and from *de novo* folding methods. The final models include high-resolution features not present in the starting models, including the packing of core side chains. Bringing together these results from all-atom structure prediction with state-of-the-art algorithms for molecular replacement and automated rebuilding[10–12], we show that distant-template-based and *de novo* models can reach the accuracy required to solve the X-ray crystallographic phase problem.

Targeted rebuilding-and-refinement protocol

We have developed a new approach for refining protein models that combines the targeting of aggressive sampling to regions most likely to be in error with powerful global optimization techniques. The new protocol is outlined in Fig. 1a. The first step of this protocol is the energy-based optimization of an input ensemble of models using the previously described Rosetta all-atom refinement method. This method combines Monte Carlo minimization with side-chain remodelling to relieve inter-atomic clashes and to optimize side-chain packing and hydrogen bonding, as encoded by an all-atom force field[13,14]. Briefly, in each Monte Carlo move, a random perturbation to the protein backbone torsion angles is followed by discrete

[1]University of Washington, Department of Biochemistry and Howard Hughes Medical Institute, Box 357350, Seattle 98195, USA. [2]Department of Haematology, University of Cambridge, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Hills Road, Cambridge CB2 0XY, UK.
*These authors contributed equally to this work.

optimization of the side-chain conformations[14,15], which allows efficient crossing of side-chain torsional barriers. Then, quasi-Newton optimization of the side-chain and backbone torsion angles is carried out before the decision on whether to accept the move. Because of the final minimization, each point on the landscape is mapped to the closest local minimum, flattening energy barriers[16]. Although making it possible to recognize near-native predictions based on their low energies[1,13], this all-atom refinement alone does not consistently produce significant improvements in model quality (Supplementary Fig. 1).

The second step in the new protocol is the identification of regions of variation in the ensemble of refined models. We have found a marked correlation between the extent of variation in the coordinates of a residue in the refined structures and the deviation of the coordinates of the residue in the refined models from the native structure. An example is shown in Fig. 1b, c: positions exhibiting small variance across the models are usually quite close to the correct structure, whereas positions for which the variance is large often deviate considerably from the native structure. This correlation arises from the relatively short range of the force field and the energy gap between the native structure and the models: because the energy of the entire system is roughly equal to the sum of its parts, for most portions of the protein, the correct conformation will be lower in energy than non-native conformations. Regions of the protein that can access the native conformation are likely to converge on this conformation and thus exhibit less variation, whereas locally incorrect conformations are likely to be spread throughout the landscape and exhibit more variation. We observe this correlation for many different proteins in both the cartesian coordinates and the internal torsion angles; a related principle has recently been used in the Pcons method for assessing protein models[17].

The third step in the new protocol targets aggressive sampling to the regions most likely to be in error. A fragment-based segment rebuilding method (see Supplementary Material) is used to rebuild completely regions of models with relatively high variation in the model population. Because the precise regions that are incorrect cannot be identified unambiguously, we carry out many independent calculations in which different segments in the higher variation regions are randomly selected for complete rebuilding. The partially rebuilt models are then subjected to the Rosetta all-atom refinement protocol described above[13,14]. In the segment rebuilding process, side chains are initially represented as soft interaction centres and the connectivity of the chain is temporarily broken, thus permitting the traversal of much larger barriers than those crossed by all-atom refinement alone.

As indicated in Fig. 1a, if the lowest energy refined structures have not converged, the rebuilding-and-refinement protocol is applied iteratively using a selection process inspired by natural evolution to guide convergence on the global minimum. At each iteration, a subset of models that are low in energy yet structurally diverse is chosen to seed the next round; the regions to be rebuilt are determined on the basis of the backbone variation in the selected population. Bringing together ideas from tabu search[18] and conformational space annealing[19], the selection process alternates between the propagation of a structurally diverse population into the next round (diversification) and focusing in on the lowest energy regions of the energy landscape explored thus far (intensification). The lowest energy models after ten iterations are selected as the final predictions. As illustrated in Fig. 1d, models with progressively lower energies and more native-like structures can be obtained with increasing number of iterations; results on a number of refinement problems are summarized in Supplementary Fig. 2.

Improving NMR models

As a first test of the new rebuilding-and-refinement method, we sought to improve the accuracy of protein structure models derived from moderate-resolution NMR experiments. NMR is an important method for determining structures of proteins at atomic resolution that has the advantage of not requiring crystals. In some cases, however, NMR models can contain errors due to either insufficient data or ambiguities in interpretation of the input NMR spectra[20]. We applied the method outlined in Fig. 1a to ten ensembles of NMR models deposited in the Protein Data Bank (PDB) for which independently determined high-resolution X-ray crystal structures

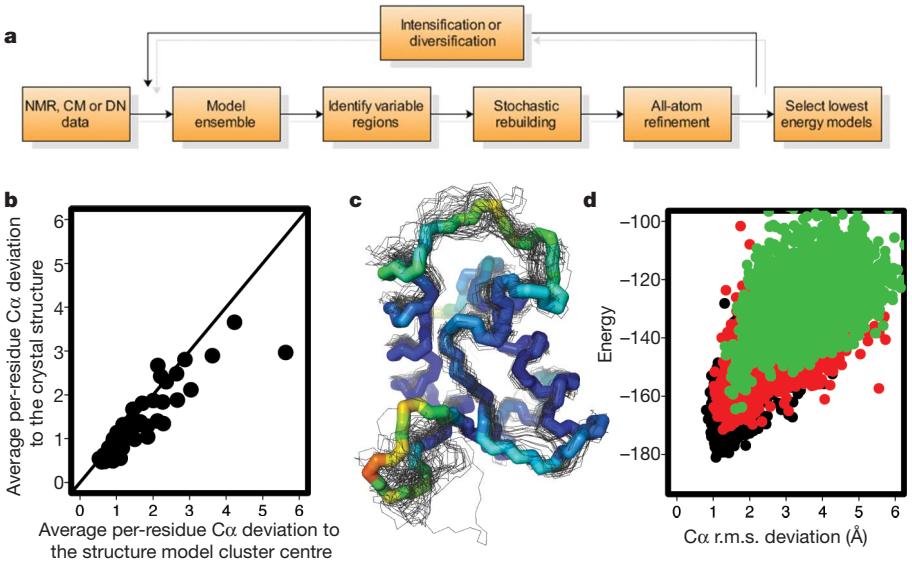

Figure 1 | Overview of the rebuilding-and-refinement method. a, Schematic diagram of the rebuilding-and-refinement method applied to structures from NMR, from comparative modelling (CM) and from *de novo* (DN) modelling approaches. **b**, Strong correlation between the per-residue backbone conformation variation in the model ensemble and the deviation from the native structure for target T0199 from the sixth critical assessment of structure prediction (CASP6). **c**, Superposition of the native structure of CASP6 target T0199 with 50 low-energy all-atom refined models. The native structure backbone is shown as a thick line, and the models are shown as thinner lines. Residues in the native structure are coloured by the average per-residue Cα r.m.s. deviation to the native from 4.5 Å (red) to 0.5 Å (blue). **d**, Iterative rebuilding and refinement yields low-energy native-like models. The energy and the Cα r.m.s. deviation of models generated during three iterations of the loop-relax protocol are displayed for iteration 1 (green), iteration 4 (red) and iteration 7 (black). The Rosetta all-atom energy includes the enthalpy plus the solvation contribution to the entropy but not the configurational entropy.

provide tests of model accuracy[21,22]. Regions with high variation in initial all-atom refined ensembles were stochastically rebuilt as well as regions assessed as poorly packed (see Methods) to allow for possible over-convergence of the initial NMR ensemble in regions with incorrect constraints.

In eight of the ten cases, the lowest energy refined model was closer to the crystal structure than any member of the starting NMR ensemble (typically 20 members) in terms of backbone agreement, as assessed by GDT-HA (geometric distance test (high accuracy)[23]). Comparison of the best of five lowest energy refined models to the NMR ensemble indicates improvement in backbone accuracy and core packing in all cases (see Table 1 and Supplementary Figs 3 and 4). In addition, the quality of the lowest energy models was consistently better than the starting NMR models in terms of clash score, number of rotamer outliers and number of backbone (Ramachandran) outliers, as assessed by the MolProbity server (Supplementary Table 2)[24]. Four examples of this energy-based structural improvement are shown in Fig. 2a–d. It should be noted that no NMR data were included in these rebuilding-and-refinement tests; judicious use of experimental NMR information to focus all-atom refinement (for example, using inferential structure determination[22]) could yield still better results.

As noted above, NMR structures often do not give good molecular replacement models for crystallographic data[6], and we hypothesized that the all-atom refined models would yield better solutions. Indeed, we found such improvement in molecular replacement scores for all eight cases in which diffraction data were publicly available (Table 1), using the sensitive and widely used Phaser software[10]. Furthermore, using phases from the molecular replacement trial with the highest translation function *Z*-score, electron density maps were generated and in seven of the eight cases the widely used ARP/wARP[11] or RESOLVE[12] automatic map tracing programs could build the majority of the residues with no human intervention (Table 1). An example of the improvement in density is shown in Fig. 3a, b. These results suggest that all-atom rebuilding and refinement may be a powerful supplement to existing strategies of trial-and-error trimming of NMR ensembles to improve molecular replacement solutions for crystallographic data[6].

Improved blind predictions based on templates

As a further challenging test, we used the new energy-based rebuilding-and-refinement method to make blind structure predictions for 26 proteins with lengths less than 200 residues that had distant homologues (sequence identity lower than 30%) with known structure during the seventh Critical Assessment of Techniques for Protein Structure Prediction (CASP7). Ensembles of starting models based on different alignments to one or more of these distant homologues were generated as described in the Supplementary Information, and the rebuilding-and-refinement protocol was carried out with several rounds of iteration to explore more broadly conformational space (Fig. 1a). Five representative low-energy structures from the final population were submitted to the CASP organizers. For 18 of the 26 cases, at least one of these 5 models was closer to the correct structure than the closest homologous structure in the PDB, as assessed by the GDT-HA score[25]. Marked improvement was observed in seven cases, with a 10–30% increase in this measure of model quality (see Table 1).

Table 1 | Improvement of model accuracy and molecular replacement by a rebuilding and refinement protocol

	X-ray structure	Starting model*	Length (*n*)†	Sequence identity to best template (%)‡	GDT-HA§		TFZ‖ in molecular replacement		Auto-traced residues (backbone, side chain)¶	
					Best template	Refined model	Best template	Refined model	Best template	Refined model
NMR	1hb6	2abd	86	N/A	0.58	0.79	4.1	11.3	12, 0	80, 80
	1who	1bmw	94	N/A	0.59	0.68	5.7	8.3	25, 12	47, 44
	1gnu	1kot	119	N/A	0.64	0.73	6.6	10.6	62, 53	82, 78
	1a19	1ab7	89 (2)	N/A	0.63	0.78	3.7	8.8	31, 20	48, 37
							4.5	12.5	14, 0	44, 35
	1fvk	1a24	189 (2)	N/A	0.49	0.69	3.4	6.9	66, 50	97, 91
							4.3	12.4	55, 43	85, 68
	1mzl	1afh	93	N/A	0.60	0.66	4.6	5.1	36, 29	58, 44
	1tvg	1xpw	143	N/A	0.63	0.74	4.3	6.7	15, 6	103, 86
	2snm	2sob	97	N/A	0.45	0.48	3.8	4.8	17, 16	43, 37
	1agr	1ezy	129	N/A	0.49	0.76		N/A#		N/A#
	1abq	1awo	56	N/A	0.58	0.83		N/A☆		N/A☆
CM	2hhz (T0331)	1ty9A	149	14.5	0.49	0.58	5.4	8.8	28, 24	68, 63
	2hr2 (T0368)	2c21C	158 (6)	14.8	0.57	0.67	6.0	5.4	37, 37	20, 14
	2hq7 (T0380)	2fhqA	145 (2)	25.4	0.58	0.69	4.4	6.6	47, 23	92, 83
							4.6	14.2	30, 17	60, 59
	2ib0 (T0385)	1jgcB	170 (2)	7.8	0.62	0.69	5.1	7.9	63, 37	56, 56
							5.8	15.5	50, 2	52, 52
	2hi0 (T0329_D2)	1rqlA	92 (2)	8.8	0.52	0.67		N/A#		N/A#
	2hcf (T0330_D2)	1lvhB	75	14.1	0.51	0.65		N/A#		N/A#
	2hi6 (T0357)**	1aco	132	8.4	0.45	0.52		N/A**		N/A**
DN	2hh6 (T0283)	2b2j	112	3.6	0.22	0.64	5.4	9.0	26, 12	112, 112

* PDB accession numbers for the closest previously known template (comparative modelling (CM) and *de novo* modelling (DN)) or for the NMR structure.
† Length of sequence in crystal structure (number of monomers in asymmetric unit, *n*).
‡ Number of sequence-identical residues across regions structurally aligned within 4 Å[34] divided by the length of the shorter sequence.
§ Fraction of residues in model superimposable on crystal structure with high accuracy (see Supplementary Information and ref. 23). This value is the average of four numbers: the numbers of residues aligned between model and experimental structure within 0.5 Å, within 1 Å, within 2 Å and within 4 Å. For the CM cases, GDT-HA was determined for the residues structurally aligned between the native structure and the closest template. For the NMR cases, the GDT-HA comparison presented for the best template is between the first member of the deposited NMR structural ensemble and the crystal structure.
‖ *Z*-score of Phaser log-likelihood translation function for molecular replacement solution. For CM and DN cases, molecular replacement for the best template was carried out using a mixed-model based on the best possible structural alignment between the native structure and template structure[4]; no such alignment was carried out for the refined model, however. The TFZ scores for the next best model submitted by all other CASP7 predictors were 5.4 (T0331), 6.0 (T0368), 4.4 (T0380), 5.1 (T0385) and 6.9 (T0283). For NMR cases, the presented results are from molecular replacement with the full deposited NMR ensemble and from each of the lowest energy 25 refined models (see also Supplementary Table 1). In the NMR cases, the best-TFZ structure from the deposited ensemble (see Supplementary Table 1) typically gave slightly worse results in subsequent automatic tracing than using the full ensemble, as expected[6]. In cases with multiple monomers present in the asymmetric unit, *Z*-scores for each monomer are presented, except for T0368, for which decreasing TFZ scores for molecular replacement of additional monomers after the first one indicated the solutions to be ambiguous.
¶ Number of automatically traced residues starting with molecular replacement phases given by Phaser that match the deposited crystal structure within 2 Å. In all cases, tracing and refinement was carried out with the ARP/wARP[11] and RESOLVE[12] programs, with the better results from the two programs presented.
# Predicted model is for the smaller of two domains present in the crystal structure and is thus not sufficient for molecular replacement.
☆ Structure factors not deposited in the PDB.
** Solved by NMR spectroscopy.

This is a particularly notable result because improving on the best template structure has been a long standing challenge for comparative modelling—owing to the high dimensionality of conformational space, there are many more ways to degrade a reasonably accurate model than to improve it. Superpositions of the closest homologous structure, the submitted refined models and the native structure for cases with the greatest improvement are shown in Fig. 2e–h. The improvement in the refined structures is evident even in core secondary structural elements.

Out of the seven high-resolution predictions, there were four targets for which diffraction data were available and the modelled sequence constituted the entire crystallized construct, enabling tests of molecular replacement. In each of these cases, we found that the best previous templates in the PDB failed to produce clear-cut molecular replacement solutions (Phaser *Z*-scores greater than 7), even after using knowledge of structurally alignable regions and a side-chain truncation approach to trim back the search models to their most accurate atoms[4]. Other template-based models submitted to CASP7, based on methods that typically did not use aggressive all-atom refinement, gave similarly low molecular replacement scores (Table 1). For three of the four cases, however, the refined models that we submitted for CASP7 gave significantly better molecular replacement solutions than the best template (Table 1). For these targets, the maps produced by combining phases from the blindly predicted model with the experimental diffraction amplitudes were of sufficient quality to permit the automatic chain-tracing program RESOLVE[12] to build a large fraction of each structure with high accuracy (Table 1). An example of the marked improvement in electron density on using the refined models is shown in Fig. 3c, d.

Ab initio phasing by *ab initio* modelling

To the best of our knowledge, a *de novo* structure prediction for a natural protein with an asymmetric, globular fold has never been used successfully for molecular replacement. However, the accuracy of *de novo* prediction methods has been improving rapidly. In particular, the use of all-atom refinement to follow low-resolution modelling by the Rosetta *de novo* modelling method[13] led to several blind predictions in CASP7 for proteins of all-α, all-β and α+β secondary structure classes that placed most of the backbone elements and core side chains with high accuracy (see Fig. 4a–c)[25]. This progress in *de novo* modelling, along with the successes above with refined NMR and template-based models, encouraged us to attempt molecular replacement with an exceptional prediction for the 112-residue α-helical CASP7 target T0283.

The best of five models for T0283 blindly predicted without the use of templates matched the subsequently released crystal structure (2hh6[26]) with a Cα r.m.s. deviation of 1.4 Å over 90 residues (Fig. 4c). The closest previously known fold in the PDB, identified from structure superpositions by CASP7 assessors (2b2j[27]), was significantly different from the T0283 crystal structure, aligning 70 residues with a Cα r.m.s. deviation of 3.1 Å (note also the poor GDT-HA score in Table 1).

After truncating the Rosetta prediction to a consensus core (residues 10 to 88, for which four of the five submitted models coincided to within 2.5 Å Cα r.m.s. deviation), molecular replacement by Phaser showed clear features for the omitted amino- and carboxy-terminal helices (see Supplementary Fig. 5 and caption). Starting from this molecular replacement solution, the ARP/wARP software was able to complete the structure automatically, tracing all 112 residues correctly. The final result (Fig. 4d) is in excellent agreement with the structure deposited in the PDB, which used phases experimentally derived by selenium single-wavelength anomalous dispersion, with an r.m.s. deviation of 0.13 Å for all 112 Cα atoms. In contrast, attempts to solve the structure by molecular replacement with the closest existing 'template' 2b2j failed to produce a clear-cut phasing solution (Table 1), even when knowledge of the optimal superposition was used to trim this search model back to the 70 residues that aligned best to the actual structure. It will be of great interest to investigate whether this result can be generalized to rapidly phase diffraction data for proteins of new folds.

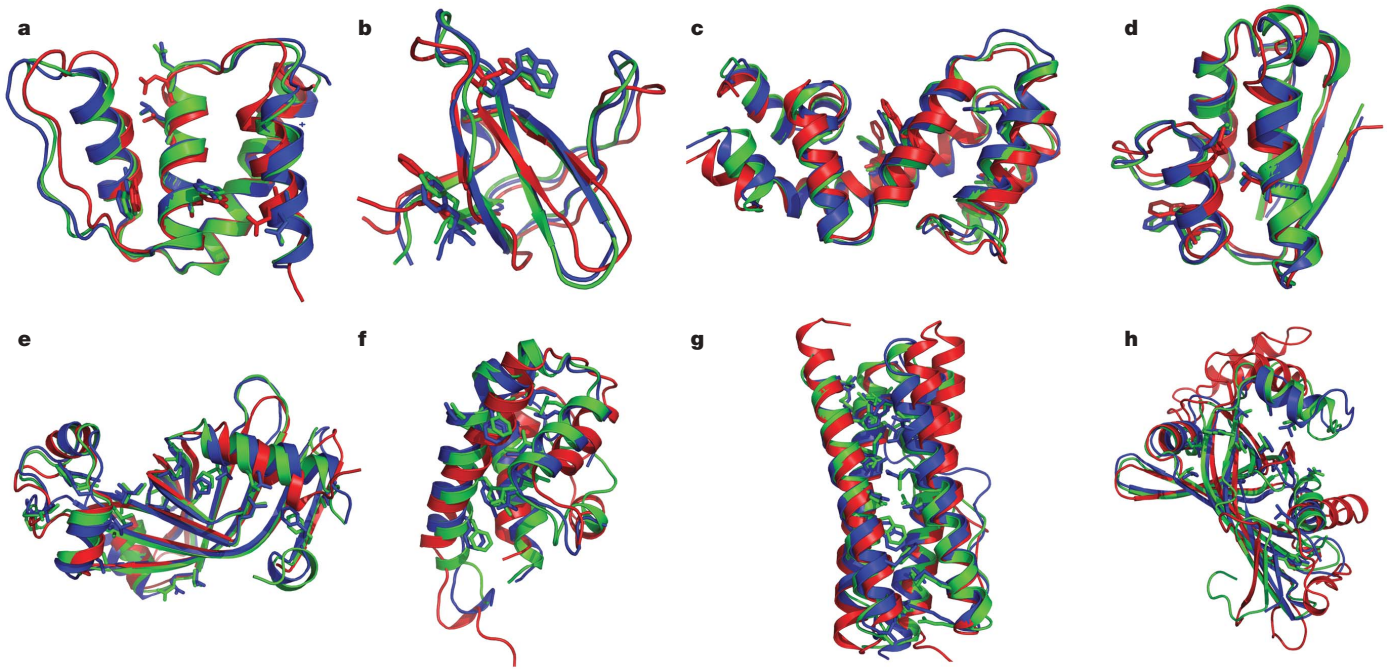

Figure 2 | Improvement in model accuracy produced by rebuilding and refinement. a–d, NMR refinement tests displaying superpositions of the crystal structure (blue), model 1 of the NMR ensemble (red) and the lowest energy all-atom refined model (green) for four NMR refinement test cases (**a**, acyl CoA binding protein, 2abd; **b**, SH3 domain of ABL tyrosine kinase, 1awo; **c**, guanine nucleotide binding protein, 1ezy; **d**, barstar, 1ab7). **e–h**, Blind predictions produced by comparative modelling, displaying superpositions of the native structure (blue), the best template in the PDB (red) and the best of our five submitted models (green) for four CASP7 targets (**e**, T0380; **f**, T0385; **g**, T0330 domain 2; **h**, T0331). A subset of the core side chains is shown in stick representation to illustrate the accuracy of core packing. Figures were prepared in PyMOL (Delano Scientific, Palo Alto, California).

Improving model accuracy and molecular replacement

The results described here show that an all-atom rebuilding-and-refinement protocol can produce protein structure models of high accuracy. The iterative protocol outlined in Fig. 1a brings together the individually quite powerful global optimization ideas underlying Monte Carlo minimization[16], tabu search[18] and conformational space annealing[19] while targeting aggressive sampling to regions most likely to be incorrect. The substantial improvements achieved in prediction quality—in several cases enabling molecular replacement phasing of X-ray diffraction data—suggest that structure prediction has matured considerably. Nevertheless, we emphasize that there is still considerable room for improvement: our high-resolution rebuilding-and-refinement protocol does not always improve starting models, and T0283 is the only CASP7 target predicted *de novo* for which the models were accurate enough for molecular replacement. We look forward to advances in both the energy function, notably the addition of configurational entropy, and in conformational sampling. The significant energy gap between the refined models and the refined crystal structure[13] for most of the cases studied here suggests that sampling is still the primary bottleneck for high-accuracy all-atom structure prediction.

At present, the Protein Structure Initiative lists hundreds of proteins with lengths less than 200 residues that have been crystallized but not yet solved. Publication of diffraction data sets that have not yielded to experimental phasing could catalyse the development of new hybrid prediction/phasing algorithms, much like the blind CASP trials have accelerated progress in the field of structure prediction. With continuing advances in high-resolution structure prediction, in molecular replacement tools, and in the interface between these two fields, we expect that *in silico* phasing will become an increasingly important component of the crystallographer's toolkit.

In the present study, aggressive all-atom refinement was carried out in the absence of any experimental information. The incorporation of experimental data into the rebuilding-and-refinement protocol could help overcome the current shortcomings in both the energy function and conformational sampling and allow more consistent high-resolution structural inference. In practical applications to molecular replacement trials, the diffraction data do not need to be set aside as a stringent *post facto* test of model accuracy, as was carried out in this study. Diffraction data without phases would be useful in screening larger numbers of trial structures for molecular replacement or in complementing the physical energy terms with diffraction-data-derived likelihood scores[28] during rebuilding and refinement. Weak phase information, for example based on anomalous scattering from intrinsic sulphur atoms[29], could also be exploited, for instance by using an initial molecular replacement model to locate the anomalous scatterer sites[10]. Although not used in the present study, NMR chemical shift, nuclear Overhauser effect, and residual dipolar coupling data can help to pinpoint regions of the models to rebuild and regions to constrain during all-atom refinement. On a larger scale, mass spectrometry techniques coupled with hydrogen/deuterium exchange[30], chemical cross-linking[31] and radical footprinting[32] show great promise for providing high-throughput, residue-level information that may rapidly constrain structure prediction and, in the absence of crystallographic data, help validate models. We anticipate that the combination of high-resolution modelling with limited experimental structural data will

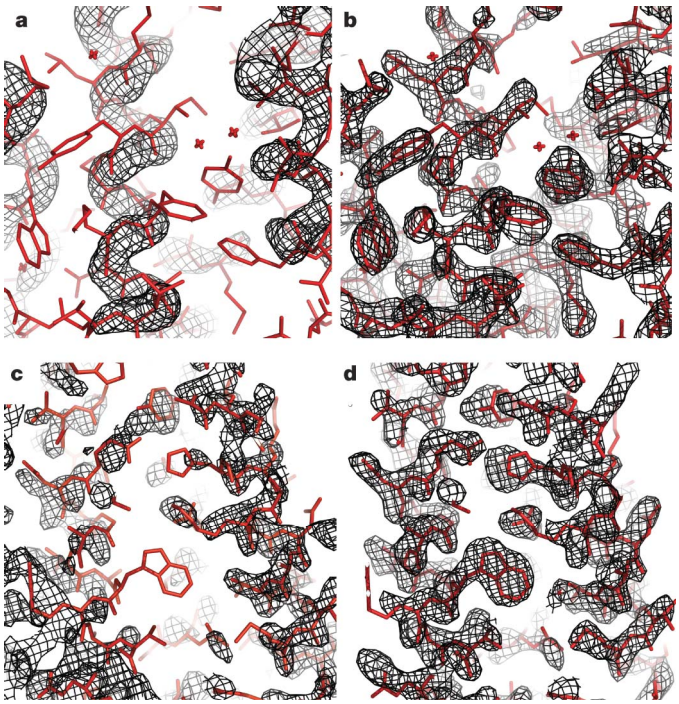

Figure 3 | Improvement in electron density using models from rebuilding and refinement in molecular replacement searches. Examples are presented for the NMR structure of acyl CoA binding protein 2abd (**a**, **b**) and CASP7 comparative modelling target T0385 (**c** and **d**). Black mesh represents electron density ($2mF_o-DF_c$; 1.5σ contour) using experimental structure factors and phases from molecular replacement with the starting model (**a** and **c**) or the refined model (**b** and **d**). The coordinates deposited in the PDB, determined using experimental phase information, are shown in stick representation. Note that the 'refinement' applied to the models refers to the all-atom energy-based protocol (see Fig. 2 and text) and not to refinement against the diffraction data. The accurate modelling of side chains by Rosetta was critical for the illustrated map improvement; molecular replacement trials gave significantly better solutions if the Rosetta-predicted side chains were retained rather than truncated.

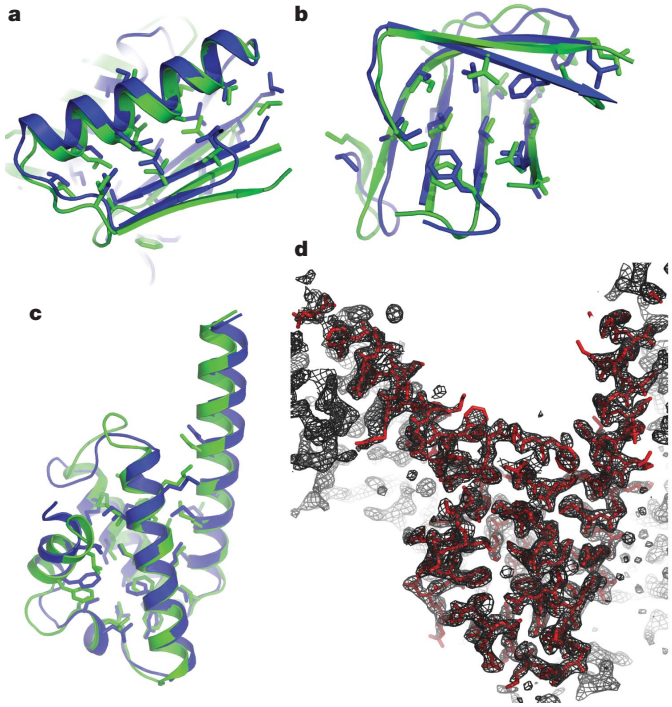

Figure 4 | *Ab initio* phasing by *ab initio* modelling. **a**–**c**, Superpositions of blind Rosetta *de novo* structure predictions (green) and the subsequently released crystal structures (blue) for CASP7 targets T0354 (**a**), domain 3 of T0316 (**b**) and T0283 (**c**). Buried side chains and backbone-aligned residues are displayed. **d**, Electron density map ($2mF_o-DF_c$; 2σ contour) produced by automatic refinement of the molecular replacement solution obtained from the T0283 structure prediction (black mesh; 1σ contour) agrees with the coordinates deposited in the PDB (red), solved with experimental phase information. The electron density map immediately after molecular replacement is shown in Supplementary Fig. 5.

become an increasingly powerful approach for characterizing the structures of biological macromolecules and complexes in the years to come.

METHODS SUMMARY

Models produced using NMR data, comparative modelling and *de novo* structure prediction were refined using the targeted rebuilding-and-refinement protocol introduced in this paper. To assess accuracy, the resulting models were compared to high-resolution crystal structures by the GDT-HA (geometric distance test (high accuracy)) score[24,33], the average percentage of Cα atoms agreeing within 0.5, 1.0, 2.0 and 4.0 Å. As a final test of accuracy and of practical utility, models were screened for suitability in phase estimation for crystallographic diffraction data using the Phaser molecular replacement software[10]. The widely used ARP/wARP[11] and RESOLVE[12] programs were then used to refine automatically the electron density maps and build density-constrained protein coordinates.

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

Received 8 May; accepted 13 September 2007.
Published online 14 October 2007.

1. Misura, K. M. & Baker, D. Progress and challenges in high-resolution refinement of protein structure models. *Proteins* **59,** 15–29 (2005).
2. Pieper, U. *et al.* MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* **34,** D291–D295 (2006).
3. Moult, J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.* **15,** 285–289 (2005).
4. Schwarzenbacher, R., Godzik, A., Grzechnik, S. K. & Jaroszewski, L. The importance of alignment accuracy for molecular replacement. *Acta Crystallogr. D* **60,** 1229–1236 (2004).
5. Giorgetti, A., Raimondo, D., Miele, A. E. & Tramontano, A. Evaluating the usefulness of protein structure models for molecular replacement. *Bioinformatics* **21** (suppl. 2), ii72–ii76 (2005).
6. Chen, Y. W., Dodson, E. J. & Kleywegt, G. J. Does NMR mean "not for molecular replacement"? Using NMR-based search models to solve protein crystal structures. *Structure* **8,** R213–R220 (2000).
7. Strop, P., Brzustowicz, M. R. & Brunger, A. T. *Ab initio* molecular-replacement phasing for symmetric helical membrane proteins. *Acta Crystallogr. D* **63,** 188–196 (2007).
8. Rossmann, M. G. *Ab initio* phase determination and phase extension using non-crystallographic symmetry. *Curr. Opin. Struct. Biol.* **5,** 650–655 (1995).
9. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302,** 1364–1368 (2003).
10. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
11. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6,** 458–463 (1999).
12. Terwilliger, T. C. Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr. D* **59,** 38–44 (2003).
13. Bradley, P., Misura, K. M. & Baker, D. Toward high-resolution *de novo* structure prediction for small proteins. *Science* **309,** 1868–1871 (2005).
14. Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **383,** 66–93 (2004).
15. Leaver-Fay, A., Kuhlman, B. & Snoeyink, J. Rotamer-pair energy calculations using a Trie data structure. In *Algorithms in Bioinformatics* (eds Casadio, R. & Myers, G.) 389 (Springer, Berlin, 2005).
16. Wales, D. J. & Scheraga, H. A. Global optimization of clusters, crystals, and biomolecules. *Science* **285,** 1368–1372 (1999).
17. Wallner, B. & Elofsson, A. Identification of correct regions in protein models using structural, alignment, and consensus information. *Protein Sci.* **15,** 900–913 (2006).
18. Glover, F. & Laguna, M. *Tabu Search* (Kluwer, Norwell, Massachusetts, 1997).
19. Lee, J., Liwo, A. & Scheraga, H. A. Energy-based *de novo* protein folding by conformational space annealing and an off-lattice united-residue force field: application to the 10–55 fragment of staphylococcal protein A and to apo calbindin D9K. *Proc. Natl Acad. Sci. USA* **96,** 2025–2030 (1999).
20. Doreleijers, J. F., Rullmann, J. A. & Kaptein, R. Quality assessment of NMR structures: a statistical survey. *J. Mol. Biol.* **281,** 149–164 (1998).
21. Grishaev, A. & Bax, A. An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation. *J. Am. Chem. Soc.* **126,** 7281–7292 (2004).
22. Rieping, W., Habeck, M. & Nilges, M. Inferential structure determination. *Science* **309,** 303–306 (2005).
23. Zemla, A. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res.* **31,** 3370–3374 (2003).
24. Lovell, S. C. *et al.* Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins* **50,** 437–450 (2003).
25. Das, R. *et al.* Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home. *Proteins* doi:10.1002/prot.21636 (25 September 2007).
26. Berman, H., Henrick, K., Nakamura, H. & Markley, J. L. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* **35,** D301–D303 (2007).
27. Andrade, S. L., Dickmanns, A., Ficner, R. & Einsle, O. Crystal structure of the archaeal ammonium transporter Amt-1 from *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA* **102,** 14994–14999 (2005).
28. Pannu, N. S. & Read, R. J. Improved structure refinement through maximum likelihood. *Acta Crystallogr. A* **52,** 659–668 (1996).
29. Dauter, Z. New approaches to high-throughput phasing. *Curr. Opin. Struct. Biol.* **12,** 674–678 (2002).
30. Englander, J. J. *et al.* Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proc. Natl Acad. Sci. USA* **100,** 7057–7062 (2003).
31. Young, M. M. *et al.* High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl Acad. Sci. USA* **97,** 5802–5806 (2000).
32. Takamoto, K. & Chance, M. R. Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 251–276 (2006).
33. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33,** 2302–2309 (2005).
34. Ortiz, A. R., Strauss, C. E. & Olmea, O. MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. *Protein Sci.* **11,** 2606–2621 (2002).

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

Acknowledgements We thank Rosetta@home participants for contributing computing power that made testing of many new ideas possible; the DOE INCITE program for access to Blue Gene/L at Argonne National Laboratory and the IBM Blue Gene Watson supercomputers; and the NCSA, SDSC and Argonne National Laboratory supercomputer centres for computer time and help with porting Rosetta to Blue Gene. We thank D. Kim and K. Laidig for developing the computational infrastructure underlying Rosetta@home; J. Abendroth for help with RESOLVE and ARP/wARP software; M. Kennedy of NESG for the NMR structure coordinates of protein 1xpw and for help with the molecular replacement calculations; and J. Abendroth, J. Bosch, J. Havranek and C. Wang for comments on the manuscript. We also thank the CASP organizers and contributing structural biologists for providing an invaluable test set for new structure refinement methods. This work was funded by the National Institute of General Medical Sciences, National Institutes of Health (to D.B.), the Wellcome Trust, UK (to R.J.R.), the Howard Hughes Medical Institute (D.B.), a Leukemia and Lymphoma Society Career Development fellowship (to B.Q.), and a Jane Coffin Childs fellowship (to R.D.).

Author Contributions B.Q., S.R. and R.D. contributed equally to this work. Structure predictions for NMR-based, comparative-model-based and *de novo* predictions were carried out by S.R., B.Q. and R.D. respectively, with advice and software from D.B. and P.B. Phasing trials were performed by R.J.R., B.Q., S.R. and R.D., with advice from R.J.R. and A.J.M. All authors discussed results and commented on the manuscript.

Author Information Rosetta software and source code are available to academic users free of charge at http://www.rosettacommons.org/software/. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.B. (dabaker@u.washington.edu).

METHODS

We present detailed descriptions of six methods discussed in the main text: (1) rebuilding-and-refinement protocol; (2) identification of regions to rebuild from the NMR structure ensemble; (3) preparation of blind predictions; (4) metrics for comparing models with crystal structures; (5) screening of models for suitability for molecular replacement; (6) assessing model quality with MolProbity.

Rebuilding-and-refinement protocol. We describe below the three key steps of the rebuilding-and-refinement protocol: segment rebuilding, all-atom refinement and iterative evolution.

For the first of these three steps, we used a new segment rebuilding protocol to rebuild regions with high structural variation in the model population, as these regions are often incorrect (see, for example, Fig. 1b). Because of uncertainties in the precise locations of incorrect regions, the portions of the model to be rebuilt were chosen stochastically from the regions with high variance at the beginning of each simulation. Up to 90% of all the separate regions were rebuilt in a given run—this allows for compensatory changes in interacting segments to occur.

The coordinates in the region to be rebuilt were generated using the Rosetta fragment-insertion-based *de novo* folding protocol[36]. After each fragment insertion, the decision to accept or reject was made according to the standard Metropolis criterion based on the total energy of the system. To maintain the connectivity of the protein chain, cyclic coordinate descent (CCD[37]) was used to close the chain break at a stochastically selected position of the region rebuilt. The rebuilding process was divided into ten stages. At each successive stage, an increasing chain-break score (a penalty to the deviation of the peptide bond length at the chain break from the ideal peptide bond length) was applied. In each of the first five stages, the number of fragment insertion trials was ten times the number of residues in the region being rebuilt. In a fragment insertion trial, randomly chosen nine-residue, three-residue, or one-residue fragments were inserted into randomly chosen positions in the region being rebuilt, and the Metropolis Monte Carlo criterion was used to accept or reject the newly inserted fragment based on the Rosetta low-resolution energy function[14]. In each of the five last stages, in addition to the fragment insertion trials, we also performed cyclic-coordinate-descent-based backbone torsion angle moves (CCD moves) in which the cyclic coordinate descent solution was calculated and the backbone torsion angles for five randomly picked positions in the region being rebuilt were modified according to the CCD solution.

If after the ten rebuilding stages described above any chain break remained larger than 0.2 Å, the region to be rebuilt was expanded by one residue on both sides. The above fragment insertion and chain-break closing process was repeated using a harmonic tether to the starting values of the torsion angles in the newly included regions (which may fall into regions with low variance in the starting population) and another stochastically selected chain-break position. The regions to be rebuilt were allowed to expand by up to five residues upstream and downstream of the original starting and ending positions, until chain closure was achieved. This procedure was usually sufficient to ensure the recovery of a continuous peptide chain. In very rare cases where the chain could not be closed in a rebuilt region, it was merged with an adjacent region to be rebuilt along with the fixed portion of the model between these two regions and the rebuilding process was repeated. With the added flexibility of a larger region being rebuilt, the peptide chain could essentially always be closed. Variable regions at the chain termini were rebuilt using the fragment insertion-based *de novo* protocol without steps for chain-break closure.

The segment rebuilding protocol is implemented in the 'loop_relax' subroutine in the freely available Rosetta source code.

The segment rebuilding protocol described above aggressively employs fragment insertion moves to sample a broad range of conformations. The all-atom refinement protocol—the second key step of the rebuilding-and-refinement protocol—then searches for local minima in the vicinity of the structures produced by segment rebuilding using a detailed all-atom force-field.

The Rosetta all-atom energy function is largely dominated by short-range interactions[9], primarily Lennard–Jones interactions, orientation-dependent hydrogen bonding, and the Lazaridis–Karplus implicit solvation model[38]. The torsional states of backbone and side chains are evaluated using knowledge-based potentials derived from amino-acid-specific Ramachandran maps and the rotamer probabilities and χ angle standard deviations in the backbone-dependent rotamer library developed by ref. 39.

During all-atom refinement, all the backbone and side-chain atoms in the protein are explicitly represented. The bond lengths and angles are kept fixed at ideal values[40], and the polypeptide chain is described in internal coordinates (the backbone and side-chain torsion angles). A single move in the all-atom refinement protocol consists of the following steps: (1) one of the several types of perturbations to the backbone torsion angles described below; (2) greedy optimization of the side-chain rotamer conformations ('rotamer trials'[41]) for the new backbone conformation; (3) minimization of the energy with respect to either the backbone degrees of freedom only (first half of refinement procedure) or backbone and side-chain degrees of freedom (second half of refinement procedure) using the Davidson–Fletcher–Powell (DFP) algorithm. The convergence criterion for exiting this quasi-Newton minimization was decreased from 10^{-3} to 10^{-5} during the course of refinement to enable more complete minimization in the final stages of refinement. (4) The compound move (steps 1–3) is accepted or rejected according to the Metropolis Monte Carlo criterion. These compound moves extend the Monte Carlo minimization procedure found to be quite powerful in previous studies[42] by incorporating discrete optimization of side-chain conformations; this allows energy-directed barrier hopping at the level of the side chains.

The following backbone perturbations are used at step (1) in the Monte Carlo minimization move described above and in a previous reference[14]. The 'small' and 'shear' moves are small perturbations of the backbone at five to ten randomly chosen positions. In small moves, ϕ and ψ are perturbed randomly by up to 1° in helix or strand regions or 1.5° in loop regions. In shear moves, ϕ is perturbed randomly by up to 2° in helix or strand regions or 3° in loop regions and the preceding ψ is perturbed by the same amount of degrees in the opposite direction to produce a compensatory shear motion in the peptide plane. The 'wobble' and 'crank' moves involve insertion of fragments and are more aggressively perturbing than the small and shear moves[14]. For both of these move types, the fragment set[36] is filtered to exclude those which cause a mean square deviation in the coordinates of the downstream atoms of more than 60 Å and one of the remaining fragments is chosen randomly for insertion. In wobble moves, the torsion angles belonging to the three residues immediately following the site of the one- or three-residue fragment insertion are varied to minimize the downstream perturbation still further. In crank moves, one residue is varied immediately after the insertion site, and three more residues at a site spaced by 6–20 residues from the fragment insertion site; this produces a 'crankshaft'-like movement of the intervening portion of the chain. 'Small-wobble' moves involve an initial 10–20° random change in the torsion angles of a single residue, followed by minimization of the perturbation over the three adjacent residues. The minimization of the perturbation in the wobble and crank moves is carried out using the fast gradient-based algorithm described previously[14]. After all five move types, the side chains are optimized and the energy is minimized as described in the preceding paragraph.

The all-atom refinement protocol is divided into three stages. The first is ramp-up. The ramp-up stage consists of sets of ten small and shear moves preceded by combinatorial optimization of the side-chain rotamer conformations. The weight on the repulsive part of the Lennard–Jones potential is progressively increased from 0.05 to 1.0 over eight such move sets. The gradual ramping up of the repulsive weight facilitates a smooth rearrangement of the side chains with small perturbations of the backbone and ensures a reasonably well-packed low-energy model before the more aggressive second stage. This second stage is the aggressive sampling stage: alternating wobble, small-wobble and crank compound Monte Carlo minimization moves are carried out; the total number of attempts for each move type is equal to the number of residues in the protein. A full combinatorial search over side-chain rotamer conformations is carried out after every 25 attempts of each type of move. The more aggressive nature of the moves used at this stage allows the traversal of modest energy barriers. The convergence tolerance for the DFP minimization is set to 10^{-4}. The third stage is the fine optimization stage: alternating small and shear moves are carried out, again for a total number of attempts equal to the number of residues in the protein. The more subtle backbone conformation changes brought about by these moves assist convergence on a relatively low-energy local minimum. The convergence tolerance for minimization is set to 10^{-5}. After these three stages, a final minimization with respect to all degrees of freedom is carried out with a convergence tolerance of 10^{-6}.

The refinement protocol described above is implemented in the 'fullatom_relax' subroutine in Rosetta; the CPU cost is about 20 min for a 100-residue protein on an Intel Pentium IV 1.6 GHz processor.

The challenge in refinement is to focus sampling on the lowest energy regions of the energy landscape identified up to that point while maintaining a broad enough search to avoid converging on a local energy minimum. Towards this end, we developed a protocol that balances intensification of the search in low-energy regions with diversification to maintain subpopulations exploring alternative energy minima. The approach—the third key step of the rebuilding-and-refinement protocol; that is, ensemble evolution by alternate cycles of diversification and intensification—adopts the idea of explicit control of the search intensity from tabu search[18], and is a generalization of the conformational space annealing (CSA) technique, which has achieved success in a broad range of optimization problems[19].

In both the intensification and diversification steps, an input population of 200 models was clustered using the method described in ref. 13 to identify distinct populations of structures. The clustering threshold was chosen such that the largest cluster contained 10% of the models. For each cluster, ten models were selected (if there were fewer than 10 models in a cluster, all were selected) and each model was subjected to nine independent segment rebuilding plus all-atom refinement runs initialized with different random number seeds.

In the diversification stages (iterations 1, 3, 5, 7 and 9), the models in the parent population were kept in their original cluster assignment. A newly generated model was assigned to the closest cluster if the root-mean-squared deviation over alpha carbons (C_α r.m.s. deviation) between this model and the closest cluster member was less than the current diversity threshold (see below), and the highest energy member of the cluster was thrown away. If the r.m.s. deviation between a newly generated model and its closest cluster member was higher than the current diversity threshold, then the model with the highest energy in the current parent population was thrown away, and the newly generated model formed a cluster of its own. This is analogous to speciation in natural evolution. As a model is discarded for each new model added, the population size stayed unchanged. The diversification step favours a broad exploration of the conformational space by maintaining the distinct populations of clusters: there is competition for low energy within but not between clusters. Combined with the initial clustering step, it ensures that the new population will not be dominated by overly closely related structures, which could result in premature convergence away from the global minimum.

In the intensification stages (iterations 2, 4, 6, 8 and 10), all but the lowest energy 10% of the entire population (parents plus offspring) is discarded to bring the population back to a size of 200. The remaining models from the parent population keep their original cluster assignment. A newly generated model was assigned to the closest remaining cluster if the r.m.s. deviation between this model and the closest cluster member was lower than the current diversity threshold; otherwise it formed a new cluster of its own. This stage differs from the diversification stage in that the energy-based selection is carried out across all clusters and hence higher energy clusters can be eliminated completely. This stage allows more thorough exploration of the most promising (lowest energy) regions of the energy landscape explored thus far.

The diversity threshold used to maintain distinct populations and to guide the spawning of new populations was reduced at each iteration to allow gradual convergence on the global energy minimum. The starting value was the clustering threshold in the original population, and this was reduced by 0.1 Å at each iteration. This annealing of the diversity threshold was introduced in the CSA strategy[19].

The new parent population generated by the diversification or intensification procedures was used to seed the next generation, and nine independent segment rebuilding plus all-atom refinement calculations were again carried out for each parent. After ten iterations, the low energy models were clustered and the lowest energy models in the largest five clusters were selected as the final predictions. The overall iterative procedure took approximately 2,000 CPU hours per target. For molecular replacement efforts, this computational effort would probably be significantly reduced if phasing trials with diffraction data are used to screen models.

Identification of regions to rebuild from the NMR structure ensemble. The test cases for NMR refinement were chosen to be proteins representing different fold topologies for which an NMR structure and a high-resolution crystal structure (with structure factors deposited in PDB) existed. These were chosen from the data sets used by refs 21 and 43.

For investigations of refinement of NMR structures, we rebuilt two sets of regions. The first are regions that vary within the NMR ensemble. As in the comparative modelling case, we have observed that regions that vary within the NMR ensemble are likely to be the regions that are most different from a high-resolution crystal structure. These are most likely loops that are either inherently dynamic in the NMR structure or loops that are held in place with insufficient restraints. (Applying all-atom refinement to the NMR ensembles gave essentially the same list of variable regions (data not shown).)

The second set of regions are segments that are internally consistent within the NMR ensemble but systematically under-packed. To estimate packing we used a recently developed packing metric (W. Sheffler, personal communication) based on the relative accessible surface areas of groups of atoms. For each buried atom, we compute the largest sphere tangent to that atom which can fit into empty space within the protein. A group composed of all atoms within 5 Å of the centre is defined for each sphere. For each group of atoms, accessible surface (SASA) to small and large spherical probes (radii 0.9 Å and 2 Å, respectively) is computed; given that a ball of atoms has a certain area accessible to a large sphere, less-accessible area to a small sphere indicates better packing. A summary percentile score is computed on the basis of a reference set of crystal structures, approximating the fraction of native proteins which are better packed than the scored structure.

Preparation of blind predictions. The initial set of template-based models was obtained from the 3D-Jury server[44] and subjected to all-atom refinement using the Rosetta all-atom energy function. Up to ten templates from which the very lowest energy models were derived were used as the candidate templates. Alignment ensembles between the candidate templates and the target sequence were parametrically generated using the K*Sync alignment method[45]. The alignment ensemble was turned into a model ensemble by placing the sequence of the query onto the backbone of the parent based on each alignment. Missing densities from the insertion and deletion regions of the alignment were modelled using the segment modelling protocol described in the 'rebuilding-and-refinement protocol' section. The full-chain models were then subjected to the all-atom refinement procedure as described in the same section, constrained by a set of Cα–Cα distance constraints, described next.

The Cα–Cα distance constraints were generated from the 3D-Jury[44] template-based models with the lowest Rosetta all-atom energies after all-atom refinement. A Cα–Cα pair was used to derive constraints only when the associated distance was less than 8 Å in more than 80% of the selected constraint-generating models. Upper and lower bounds for each of these pairs were determined by padding the highest and lowest of these distances by one standard deviation of the Cα–Cα distance distribution function, as described in ref. 46. For computational efficiency, we further trimmed down the number of constraint pairs by eliminating neighbouring pairs separated by one or two residues. During all-atom refinement, a penalty is applied when the Cα–Cα distances in the model exceed the upper or lower limit of the corresponding constraints. If a distance exceeds the upper or lower constraint limit by d (in Å), then the penalty E_c is d^2 when $d < 0.5$ Å, and $(d - 0.25$ Å$)$ when $d \geq 0.5$ Å. The resulting ensemble of low-energy comparative models became the inputs to further rounds of rebuilding-and-refinement (Fig. 1a).

For targets without clear templates identified by the 3D-Jury server[44], the full chain was fully modelled by fragment assembly starting from an extended chain, followed by the all-atom refinement procedure described above. The convergence of the Rosetta *de novo* prediction protocol can differ significantly for different sequence representatives of a given fold[13,47]. For T0283, one of seven tested sequence homologues gave exceptionally well converged low-energy models that, after sequence mapping, allowed structure prediction for the target sequence with the rebuilding-and-refinement protocol[13,25].

About 100,000 all-atom refined models were generated for each modelling target, requiring approximately 100,000 CPU hours. As noted above, for molecular replacement efforts, this computational effort would probably be significantly reduced if phasing trials with diffraction data are used to screen models; as the predicted models used in this manuscript were prepared as blind predictions for CASP7, such diffraction data were not available at the time of modelling.

Metrics for comparing models with crystal structures. As has been discussed previously, no metric for comparing structure models with the crystal structures is perfect[48]. In this work, we used three different structural metrics for model quality assessment. The Cα r.m.s. deviation is a widely used metric for structure comparison, but it can be distorted by large deviations in a small number of residues, especially at the termini or in long surface loops. The GDT-HA (geometric distance test (high accuracy)) score is the average percentage of Cαs in the model within 0.5, 1.0, 2.0, and 4.0 Å of the corresponding Cα coordinates in the crystal structure; we used TMalign[33] to align the structures. This metric is less sensitive than the full-chain r.m.s. deviation to deviations in poorly ordered termini and long loops, and was used in the CASP7 template-based modelling assessment.

The core residue all-atom r.m.s. deviation describes the accuracy of both the backbone and side-chain conformation prediction. We used this metric in the evaluation of NMR refinement because it can be applied to both the starting (NMR ensemble) and ending (Rosetta refined) models. In template-based modelling, this metric is not practical as the template usually does not have the same amino acid sequence as the target to be modelled.

In addition, successful molecular replacement using the predicted structure can be regarded as a stringent test for model quality assessment, as suggested in ref. 49.

Screening of models for suitability for molecular replacement. Searching for molecular replacement solutions involves applying rigid-body transformations along the six rotational and translational degrees of freedom. We carried out this search with the Phaser software, which is described in ref. 10 and references therein. For completeness, the algorithms are briefly summarized here. Phaser uses likelihood functions to judge how well molecular replacement models agree with the measured diffraction data after they have been first rotated and then also translated. Brute-force likelihood calculations over grids of orientations and

positions are computationally expensive, so fast-fourier-transform-based approximations are used to compute sets of possible solutions, which are rescored with the full likelihood targets. By using a tree-search-with-pruning strategy, almost all solutions that would be found with a full six-dimensional search are found, but with a much lower computational cost. As well, this strategy allows effective searches for multiple copies, in crystals with more than one molecule in the asymmetric unit. For each molecule to be placed, a rotation search is first carried out. A translation search is then carried out for each plausible orientation. All plausible rotation/translation solutions are checked for packing in the lattice, and solutions that pack successfully are subjected to rigid body refinement. If more than one copy is present, all plausible partial solutions are fixed in turn while carrying out rotation and translation searches for subsequent copies. In molecular replacement trials with Phaser, the clearest indication of success comes from high values of the *Z*-score (number of standard deviations above the mean), computed by comparing the log-likelihood-gain (LLG) for the peak with LLG scores for a random sample of search points.

For molecular replacement in each of the NMR modelling cases, we evaluated the combined NMR ensemble as a potential search model and compared these results to trials with the 25 lowest energy Rosetta models from rebuilding and refinement (Table 1). Furthermore, we have carried out molecular replacement trials with each of the members of the deposited NMR ensemble individually, with results given in Supplementary Table 1. Finally, for an actual search for a good molecular replacement solution, a larger set of models from rebuilding and refinement can be screened rapidly. We thus extended the search to the 1,000 lowest energy models from Rosetta rebuilding and refinement and the results, notably improved, are presented in Supplementary Table 1.

For molecular replacement in comparative modelling cases, we prepared search models from the best existing templates and from our comparative modelling predictions. For the best templates, we followed the 'mixed model' protocol described in ref. 4 for optimizing molecular replacement. Furthermore, on the basis of the 3DPAIR[50] structure alignment between the native structure and the best template structure, the template structure was trimmed to contain only the structurally alignable regions. Then the native sequence was threaded onto the backbone of the corresponding template structure, while retaining the side-chain coordinates of the identical residues between the template and native sequences. Non-identical side chains longer than serine were mutated to serine, followed by Rosetta side-chain packing protocol[51] to model the mutated serine and the shorter non-identical side chains, while keeping the identical side-chain conformation fixed. To prepare search models for these predictions, we superimposed 100 low-energy models from the final round of refinement, and defined the model that has the lowest average r.m.s. deviation to the rest of the models as the reference model. Then we calculated the average per-atom distance D_a between each of the superimposed models and the reference model. The Rosetta temperature factor is calculated as $T_e = 8\pi^2 D_a^2/3$ for each atom and inserted to the B-factor column of the refined model files. The Rosetta temperature factor is intended to represent the uncertainty in the final refined models after extensive refinement in the Rosetta all-atom force field. As suggested earlier[19], by using the B-factor effectively to smear each atom over its possible positions, the correlation of the modelled electron density with the true electron density can be maximized.

For the *de novo* modelling case, target T0283, search models for molecular replacement were trimmed according to residues for which there was consensus among submitted models. Supplementary Fig. 5 gives a more detailed description and illustration of the molecular replacement solution.

Assessing model quality with MolProbity. For the investigations of refinement of NMR models, we used the MolProbity software[24] to investigate the quality of the refined models versus that of the starting NMR ensemble. For purposes of comparison, we chose the lowest energy refined model and the first member of the deposited NMR structure. Supplementary Table 2 shows the clash score, number of rotamer outliers and number of Ramachandran outliers of the NMR and refined models. The refined models consistently have better model quality than the starting NMR structure based on these metrics.

36. Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268,** 209–225 (1997).
37. Canutescu, A. A. & Dunbrack, R. L. Jr. Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Sci.* **12,** 963–972 (2003).
38. Lazaridis, T. & Karplus, M. Effective energy function for proteins in solution. *Proteins* **35,** 133–152 (1999).
39. Dunbrack, R. L. Jr & Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6,** 1661–1681 (1997).
40. Engh, R. A. & Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* **47,** 392–400 (1991).
41. Wang, C., Schueler-Furman, O. & Baker, D. Improved side-chain modeling for protein-protein docking. *Protein Sci.* **14,** 1328–1339 (2005).
42. Li, Z. & Scheraga, H. A. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl Acad. Sci. USA* **84,** 6611–6615 (1987).
43. Garbuzynskiy, S. O., Melnik, B. S., Lobanov, M. Y., Finkelstein, A. V. & Galzitskaya, O. V. Comparison of X-ray and NMR structures: is there a systematic difference in residue contacts between X-ray- and NMR-resolved protein structures? *Proteins* **60,** 139–147 (2005).
44. Ginalski, K., Elofsson, A., Fischer, D. & Rychlewski, L. 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* **19,** 1015–1018 (2003).
45. Chivian, D. & Baker, D. Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection. *Nucleic Acids Res.* **34,** e112 (2006).
46. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815 (1993).
47. Bonneau, R., Strauss, C. E. & Baker, D. Improving the performance of Rosetta using multiple sequence alignment information and global measures of hydrophobic core formation. *Proteins* **43,** 1–11 (2001).
48. Moult, J., Fidelis, K., Rost, B., Hubbard, T. & Tramontano, A. Critical assessment of methods of protein structure prediction (CASP)–round 6. *Proteins* **61** (suppl. 7), 3–7 (2005).
49. Petsko, G. A. The grail problem. *Genome Biol.* **1,** COMMENT002 (2000).
50. Plewczynski, D., Pas, J., Von Grotthuss, M. & Rychlewski, L. Comparison of proteins based on segments structural similarity. *Acta Biochim. Pol.* **51,** 161–172 (2004).
51. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl Acad. Sci. USA* **97,** 10383–10388 (2000).