5. M. M. C. Lai, K. V. Holmes, in *Fields' Virology*, D. M. Knipe, P. M. Howley, Eds. (Lippincott. Williams. and Wilkins, Philadelphia, PA, 2001).
6. Y. Xu *et al.*, *J. Biol. Chem.* **279**, 49414 (2004).
7. Y. Xu *et al.*, *J. Biol. Chem.* **279**, 30514 (2004).
8. V. M. Supekar *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17958 (2004).
9. W. Weissenhorn *et al.*, *Mol. Membr. Biol.* **16**, 3 (1999).
10. M. J. Moore *et al.*, *J. Virol.* **78**, 10628 (2004).
11. X. Xiao, S. Chakraborti, A. S. Dimitrov, K. Gramatikoff, D. S. Dimitrov, *Biochem. Biophys. Res. Commun.* **312**, 1159 (2003).
12. T. M. Gallagher, M. J. Buchmeier, *Virology* **279**, 371 (2001).
13. W. Li *et al.*, *Nature* **426**, 450 (2003).
14. H. Hofmann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7988 (2005).
15. P. Towler *et al.*, *J. Biol. Chem.* **279**, 17996 (2004).
16. S. K. Wong, W. Li, M. J. Moore, H. Choe, M. Farzan, *J. Biol. Chem.* **279**, 3197 (2004).
17. G. J. Babcock, D. J. Esshaki, W. D. Thomas Jr., D. M. Ambrosino, *J. Virol.* **78**, 4552 (2004).
18. W. Li *et al.*, *EMBO J.* **24**, 1634 (2005).
19. Y. Guan *et al.*, *Science* **302**, 276 (2003).
20. H. D. Song *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2430 (2005).
21. J. Sui *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2536 (2004).
22. E. N. van den Brink *et al.*, *J. Virol.* **79**, 1635 (2005).
23. Y. He, H. Lu, P. Siddiqui, Y. Zhou, S. Jiang, *J. Immunol.* **174**, 4908 (2005).
24. Materials and methods are available as supporting materials on *Science* Online.
25. R. K. Williams, G. S. Jiang, K. V. Holmes, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 5533 (1991).
26. G. S. Dveksler *et al.*, *J. Virol.* **65**, 6881 (1991).
27. H. Kubo, Y. K. Yamada, F. Taguchi, *J. Virol.* **68**, 5403 (1994).
28. A. Bonavia, B. D. Zelus, D. E. Wentworth, P. J. Talbot, K. V. Holmes, *J. Virol.* **77**, 2530 (2003).
29. Chinese SARS Molecular Epidemiology Consortium, *Science* **303**, 1666 (2004).
30. W. Li *et al.*, *J. Virol.* **78**, 11429 (2004).
31. J. Sui *et al.*, *J. Virol.* **79**, 5900 (2005).
32. Y. He *et al.*, *Biochem. Biophys. Res. Commun.* **324**, 773 (2004).
33. F. Li, W. Li, M. Farzan, S. C. Harrison, data not shown.
34. B. Sainz Jr., J. M. Rausch, W. R. Gallaher, R. F. Garry, W. C. Wimley, *J. Virol.* **79**, 7195 (2005).
35. Y. Xu *et al.*, *Biochemistry* **43**, 14064 (2004).

36. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

37. We thank staff at the Advanced Light Source beamlines 8.2.1 and 8.2.2 for assistance and M. Berardi and E. Settembre for discussions. This work was supported by NIH grants CA13202 (to S.C.H.) and AI061601 (to M.R.F.). S.C.H. is an Investigator in the Howard Hughes Medical Institute. Coordinates and structure factors have been submitted to the Protein Data Bank with accession number 2AJF.

Supporting Online Material
www.sciencemag.org/cgi/content/full/309/5742/1864/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References

22 June 2005; accepted 11 August 2005
10.1126/science.1116480

Toward High-Resolution de Novo Structure Prediction for Small Proteins

Philip Bradley, Kira M. S. Misura, David Baker*

The prediction of protein structure from amino acid sequence is a grand challenge of computational molecular biology. By using a combination of improved low- and high-resolution conformational sampling methods, improved atomically detailed potential functions that capture the jigsaw puzzle–like packing of protein cores, and high-performance computing, high-resolution structure prediction (<1.5 angstroms) can be achieved for small protein domains (<85 residues). The primary bottleneck to consistent high-resolution prediction appears to be conformational sampling.

It has been known for more than 40 years that the three-dimensional structures of proteins are completely determined by their amino acid sequences (*1*), and the prediction of protein structure from amino acid sequence—the "de novo" structure prediction problem—is a long-standing challenge in computational biology and chemistry. Although there are notable exceptions, the majority of protein structures are likely to be at global free-energy minima for their amino acid sequences. The de novo protein structure prediction problem hence is to find the lowest free-energy structure for a specified amino acid sequence. The problem is challenging because the size of the conformational space to be searched is vast (*2*) and because the accurate calculation of the free energies of protein conformations in solvent is difficult.

Although there has been considerable progress in low-resolution de novo protein structure prediction (*3*), both the accuracy and the reliability of the structural models produced by these methods is fairly low: C_α-RMSDs (root mean square deviation of alpha-carbon coordinates after optimal superposition) of ~4 Å with incorrect packing of the amino acid side chains. Achieving higher resolution requires both more physically realistic energy functions and better conformational searching; the problem is difficult because the more realistic the energy function, the more rugged the landscape, and thus the more difficult it is to search. Here, we show that high-resolution de novo structure prediction can be achieved by generating structurally diverse populations of low-resolution models and refining these structures in the context of a physically realistic all-atom energy function.

Critical to high-resolution structure prediction is a force field for which native structures are low in free energy compared with non-native structures and a refinement protocol that can efficiently navigate the corresponding free-energy landscape. We have developed an all-atom force field (*4*) that focuses on short-range interactions—primarily van der Waals packing, hydrogen bonding, and desolvation—while neglecting long-range electrostatics. The high-resolution refinement protocol (*5*, *6*) is designed to search in the local neighborhood of a starting model for low-energy structures. The protocol consists of multiple rounds of Metropolis Monte Carlo with minimization (*7*); each trial consists of a random perturbation of one or several backbone torsion angles, fast side-chain optimization using a rotamer representation (*8*, *9*), and a gradient-based minimization of the energy function with respect to backbone and side-chain torsion angles. In this way, the continuous space of backbone conformations and the discrete set of side-chain packing arrangements are searched simultaneously. Details on the energy function and methods are provided in (*10*).

Figure 1 and fig. S1 illustrate the challenge of high-resolution de novo structure prediction. All-atom refinement trajectories begun at the native state produce models (refined natives) that sample a deep near-native free-energy basin. Although these structures typically have lower all-atom energies than do non-native structures, Rosetta de novo models—built from an extended-chain starting conformation—do not sample close enough to the native structure to fall into this narrow energy well during all-atom refinement. The narrow widths of the native basins reflect the fact that nativelike side-chain packing can be disrupted by even relatively small backbone perturbations. Thus, the critical step in high-resolution structure prediction is generating low-resolution models that are within the "radius of convergence" of the native free-energy minimum using the all-atom refinement protocol. This is challenging, because the low-resolution search integrates out the side-chain degrees of freedom to smooth the energy landscape and hence lacks the detail necessary to reliably discriminate nativelike models, leading to false minima. We attempt to overcome this problem by generating low-resolution models for a large number of se-

University of Washington, Department of Biochemistry, and Howard Hughes Medical Institute, Box 357350, Seattle, WA 98195, USA.

*To whom correspondence should be addressed. E-mail: dabaker@u.washington.edu

Table 1. Benchmark proteins and results. Protein Data Bank (PDB) (*18*) or Structural Classification of Proteins (SCOP) (*19*) ID is given in column 1 (*10*). Protein length, fraction alpha helix, and fraction beta strand are given in columns 2 to 4. C_α-RMSD values for the model with the lowest all-atom energy in rounds 1 and 2 are given in columns 5 and 6, respectively (*20*). RMSD values calculated over all heavy atoms in the protein core (*21*) are given in parentheses. Column 7 reports the best C_α-RMSD of the centers of the largest five clusters when the low-energy models from round 1 are clustered.

ID	L	%α	%β	Round 1	Round 2	Cluster	Protein name
1b72A	49	69	0	0.8 (0.8)	1.1 (0.9)	1.0	Hox-B1 homeobox protein
1shfA	59	5	40	11.1 (9.0)	10.8 (8.5)	10.9	Fyn tyrosine kinase
1tif_	59	22	37	5.3 (2.3)	4.1 (2.8)	3.8	IF3-N
2reb_2	60	61	20	1.2 (0.9)	2.1 (1.6)	1.3	RecA
1r69_	61	63	0	2.1 (2.4)	1.2 (1.5)	1.7	434 repressor
1csp_	67	4	53	5.1 (4.5)	4.7 (4.2)	5.1	Cold-shock protein
1di2A_	69	46	33	2.6 (2.3)	2.6 (2.2)	1.9	RNA binding protein A
1n0uA4	69	43	24	9.9 (8.3)	10.2 (8.1)	2.7	Elongation factor 2
1mla_2	70	34	37	8.4 (7.3)	8.7 (8.1)	7.2	Malonyl-CoA ACP transacylase
1af7__	72	72	0	10.1 (7.9)	10.4 (8.1)	1.7	Cher domain 1
1ogwA_	72	26	33	2.7 (2.3)	1.0 (1.0)	2.6	Ubiquitin
1dcjA_	73	31	27	3.2 (2.2)	2.5 (2.4)	2.0	Yhhp
1dtjA_	74	39	27	1.0 (0.8)	1.2 (0.9)	1.8	KH domain of Nova-2
1o2fB_	77	38	27	10.1 (8.7)	N/A	10.3	Glucose-permease IIBC
1mkyA3	81	32	24	3.2 (3.6)	6.3 (6.1)	3.7	Enga
1tig_	88	35	35	4.1 (4.2)	3.5 (3.4)	2.4	IF3-C

quence homologs in addition to the target sequence. Each homolog has a slightly different landscape in the low-resolution potential and produces a characteristic set of models due to variable hydrophobic patterning, loop lengths, and local structural biases (fig. S2). Models for each of the homologs are then mapped back to the target sequence, producing a large and structurally diverse starting population for all-atom refinement.

This approach was first tested on prediction target T0281 from the Sixth Critical Assessment of Techniques for Protein Structure Prediction (*3*) (CASP) experiment. Target 281 is a 70-residue alpha-beta protein with predicted secondary structure consisting of an N-terminal alpha helix, two or three beta strands, and two additional alpha helices. Rosetta de novo simulations with the target sequence generated a family of topologies characterized by a two- or three-stranded antiparallel sheet with alpha helices packed on both sides. When sequence homologs of the target were folded, several new topologies were found in which the helices packed together on one side of a three-stranded beta sheet (*11*). We picked clusters of models from both the target and the homologs for all-atom refinement; models for homologs were mapped back to the target sequence using the Rosetta loop modeling protocol (*12*). The low-energy models after all-atom refinement were clustered, and the lowest energy member of the largest cluster (which originated in simulations of one of the sequence homologs) was submitted as our first prediction (Fig. 2). When the experimental structure was released, this model was found to have a C_α-RMSD of 1.6 Å, making it perhaps the most accurate blind de novo structure prediction in the history of the CASP experiment.

To test this approach further, we constructed a benchmark of 16 small proteins with relatively deep multiple sequence alignments (Table 1)

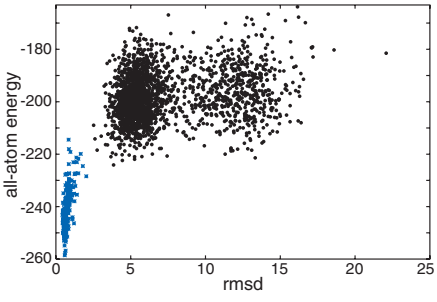

Fig. 1. Free-energy landscape for the small protein barstar (PDB code 1a19). Rosetta all-atom energy (*y* axis) is plotted against C_α-RMSD (*x* axis) for models generated by simulations starting from the native structure (refined natives, blue points) or from an extended chain (de novo models, black points). The free-energy function includes the entropic contribution to the solvation free energy but not the configurational entropy.

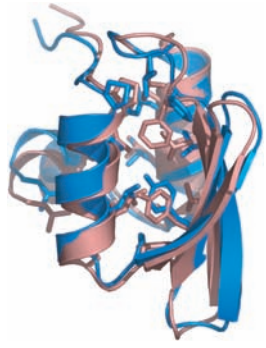

Fig. 2. 1.6 Å C_α-RMSD blind structure prediction for CASP6 target T0281, hypothetical protein from *Thermus thermophilus* Hb8. Superposition of our first submitted model for this target in CASP6 (blue) with the crystal structure (red; PDB code 1whz) showing core side chains. This figure was generated in PyMOL (*22*).

(*10*). For each protein, 15 to 50 sequence homologs were selected for folding, and low-resolution models were built for each (*10*). The sequence of the target protein was threaded onto each model, and the structure was refined with the all-atom refinement protocol described above to generate 20,000 to 30,000 all-atom models (round 1). To introduce additional diversity into the high-resolution search, we built a second set of models (round 2) by refining low-energy models from the first round with sequences from close homologs and then mapping back to the target sequence (*10*). The all-atom energy and C_α-RMSD to native are plotted for each population in figs. S4 and S5. As a stringent test of the all-atom energy function, the single lowest energy model from each round was identified and compared with the native structure (Table 1, columns 5 and 6).

For five of the proteins, the lowest energy model generated in either round 1 (three cases) or round 2 (four cases) had a C_α-RMSD to the native structure of less than 1.5 Å. The accuracy of the recapitulation of both the protein backbone and the core side chains is illustrated by structural superpositions of the lower RMSD of the round 1 low-energy model and the round 2 low-energy model onto the corresponding native structures (Fig. 3, A to E). Scatter plots of C_α-RMSD versus all-atom energy are shown in Fig. 3, G to K, and details on each of the predictions are provided in the figure legend.

For 8 of the remaining 11 proteins, the lowest energy round 1 or round 2 structure (six cases) or one of the centers of the five largest clusters of low-energy models (seven cases) were topologically correct, with C_α-RMSDs ranging from 1.5 to 5.0 Å (Table 1, columns 5 to 7), but the native side-chain packing was not captured to the extent shown in Fig. 3 (fig. S3, A to C). In one of these cases, the second-lowest energy model (fig. S3D) is quite accurate (C_α-RMSD 1.1 Å). For seven of these eight cases (*13*), and for all three of the remaining cases where topologically correct predictions were not achieved,

the failure to achieve high-quality models is due to inadequate conformational sampling. The worst of the predictions (Fig. 3F) illustrates this sampling problem: Although refined native structures have lower energies than the de novo models (Fig. 3L), there is no sampling in the native basin and a false minimum is selected.

The high accuracy of the models shown in Fig. 3, A to E, is encouraging and, along with recent success in protein design and protein-protein docking (*4*, *14*–*16*), suggests that the Rosetta all-atom potential may capture the key forces contributing to the stability of small, globular proteins. In particular, the emphasis on van der Waals interactions and hydrogen bonding and the neglect of long-range electrostatics support the view that conformational specificity is provided in large part by short-range interactions, primarily the jigsaw puzzle–like complementary packing in the protein core. The free-energy landscapes in Figs. 1 and 3, together with the ability to make predictions with C_α-RMSDs under 1 Å, suggest that conformational sampling in solution in the protein core may be restricted to a narrow ensemble centered near the crystal structure.

On the other hand, because high-accuracy models were selected for only a third of the proteins in the test set, further improvements in both the sampling methodology and the free-energy function are clearly necessary for consistent and reliable de novo structure prediction of small proteins. Conformational sampling remains the primary stumbling block, as highlighted by the lack of models with C_α-RMSDs <2.5 Å for most of the failures in our test set and the fact that the refined natives (blue points in Figs. 1 and 3 and figs. S1, S2, S4, and S5) generally have lower energies than the vast majority of de novo generated models. Improvements in sampling that reduce overconvergence in the low-resolution search should also eliminate the dependence on simulations with homologous sequences to adequately cover conformational space.

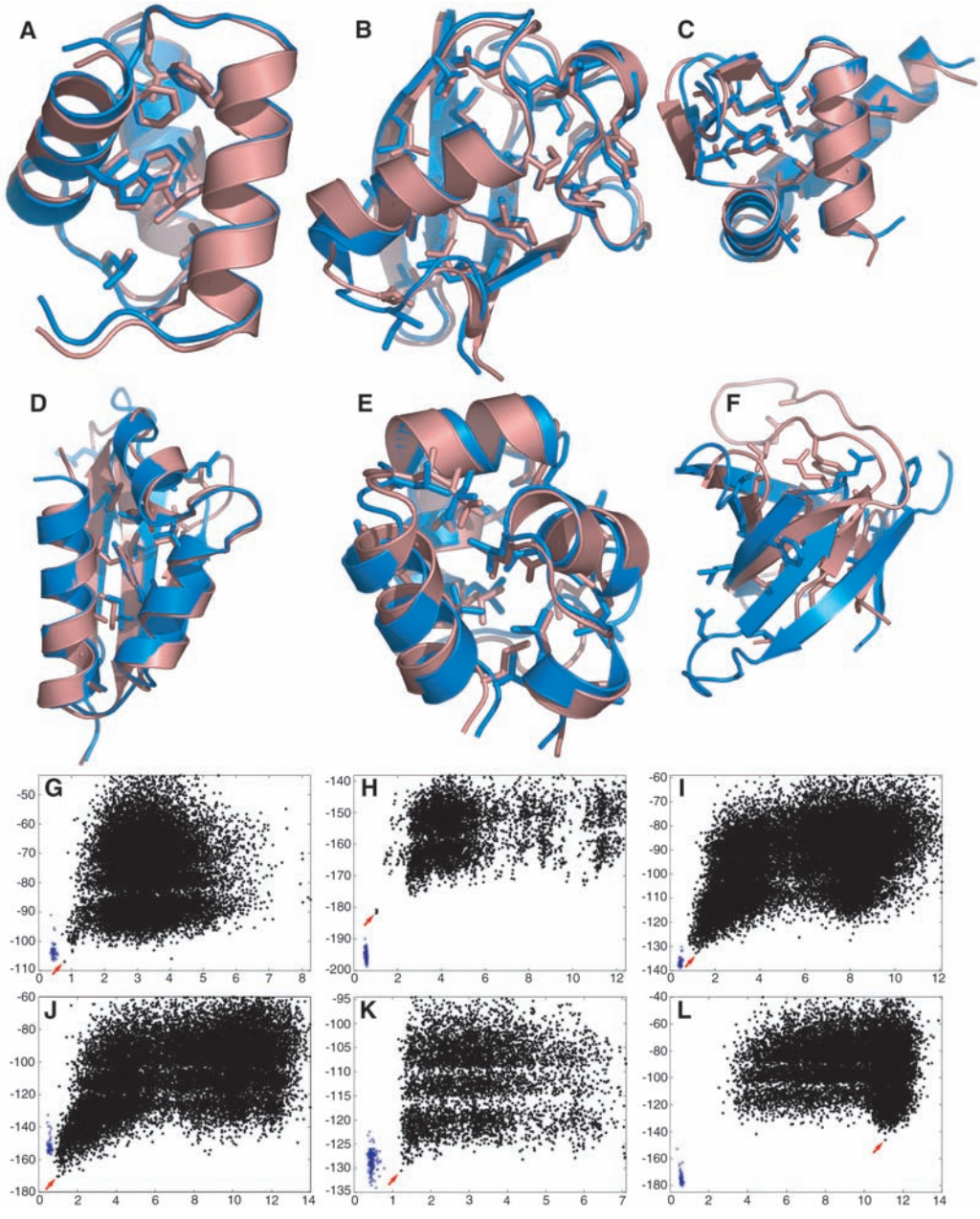

Fig. 3. High-resolution de novo structure predictions. (**A** to **F**) Superposition of low-energy models (blue) with experimental structures (red) showing core side chains. (**G** to **L**) Plots of C_α-RMSD (*x* axis) against all-atom energy (*y* axis) for refined natives (blue points) and the de novo models (black points). Red arrows indicate the lowest energy de novo models, shown in [(A) to (F)]. [(A) and (G)] Hox-B1. In the lowest energy model (A) from round 1, the aromatic side chains, particularly the central phenylalanine, overlay almost perfectly. The all-atom refinement step reduced the model C_α-RMSD to the native structure from 1.5 Å to 0.8 Å. [(B) and (H)] Ubiquitin. In the lowest energy model (B) from round 2, almost all of the core side chains overlay well, including the central partially buried lysine. [(C) and (I)] RecA. The lowest energy model from round 1 (C) has nearly all the core side chains in place; the RMSD versus energy plot (I) exhibits a broad funnel. [(D) and (J)] KH domain of Nova-2. A loop for which density is missing in three of four monomers in the crystal structure of the tetramer packs more closely on the rest of the protein in the low RMSD models than in the native monomer, in which the density was interpretable, and is responsible for the lower than native energy of models in the native basin (J). The lowest energy model after round 1 (D) has a C_α-RMSD to native of 1.0 Å with the omission of this loop, which is involved in RNA binding. [(E) and (K)] 434 repressor. The lowest energy model after round 2 (E) has a C_α-RMSD of 1.3 Å despite consistent errors in the population in one of the loops. The lowest energy models for the remaining 11 proteins in our test set were much less accurate, with side-chain packing and, in some cases, the fold incorrect. The lowest energy round 1 structure (F) for the Fyn Tyrosine kinase replaces the native diverging turn with an additional hairpin (C_α-RMSD 11.1Å); de novo models fail to sample the deep energy minimum near the native structure (L). [(A) to (F)] were created in PyMOL (*23*) and [(G) to (L)] with gnuplot (www.gnuplot.info).

What are the prospects for high-resolution protein structure prediction more generally? First, protein core prediction may be a fundamentally easier problem than prediction of the detailed structures of functionally relevant parts of proteins, such as active sites where buried charged and polar interactions are more common (*17*). Second, the computational cost of high-resolution refinement is expected to increase dramatically with chain length, and hence the refinement of models of large proteins is likely to require orders of magnitude more computing power than the ~150 CPU days required for each of the predictions in this paper. Although our results are encouraging, consistent and reliable high-resolution modeling of protein structure remains a formidable challenge.

References and Notes

1. C. B. Anfinsen, E. Haber, M. Sela, F. H. White Jr., *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1309 (1961).
2. C. Levinthal, *J. Chim. Phys.* **65**, 44 (1968).
3. J. Moult, K. Fidelis, A. Zemla, T. Hubbard, *Proteins* **53** (Suppl 6), 334 (2003).
4. B. Kuhlman *et al.*, *Science* **302**, 1364 (2003).
5. J. Tsai *et al.*, *Proteins* **53**, 76 (2003).
6. K. M. S. Misura, D. Baker, *Proteins* **59**, 15 (2005).
7. Z. Li, H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6611 (1987).
8. B. Kuhlman, D. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10383 (2000).
9. R. L. Dunbrack Jr., F. E. Cohen, *Protein Sci.* **6**, 1661 (1997).
10. Materials and methods are available as supporting material on *Science* Online.
11. Inspection of the multiple sequence alignment revealed that the target sequence had several hydrophobic residues at exposed beta-sheet positions in this topology that were replaced by polar residues in other members of the alignment, which offers a possible explanation for the absence of this topology in target-sequence models.
12. C. A. Rohl, C. E. Strauss, D. Chivian, D. Baker, *Proteins* **55**, 656 (2004).
13. The single case where non-native models had lower energies than the native is a domain of a larger protein that has a quite hydrophobic interface and may not be stable in isolation.
14. G. Dantas, B. Kuhlman, D. Callender, M. Wong, D. Baker, *J. Mol. Biol.* **332**, 449 (2003).
15. J. J. Gray *et al.*, *J. Mol. Biol.* **331**, 281 (2003).
16. C. Wang, O. Schueler-Furman, D. Baker, *Protein Sci.* **14**, 1328 (2005).
17. The physical chemistry of the packing of nonpolar atoms is considerably easier to model than the subtle trade-offs between desolvation of polar groups and the formation of buried polar interactions. Accurate modeling of functional sites may further require inclusion of explicit solvent molecules, modeling of protonation states and interactions with ligands, and polarizible electrostatics treatments. Together with the increase in the cost of refinement with chain length, this has implications for structural genomics efforts: The refinement to high resolution of comparative models of large proteins (>400 amino acids) based on ~30% sequence identity, with many buried polar and charged residues, may be a harder problem than the de novo prediction of the structures of small proteins.
18. H. M. Berman *et al.*, *Nucleic Acids Res.* **28**, 235 (2000).
19. A. G. Murzin, S. E. Brenner, T. Hubbard, C. Chothia, *J. Mol. Biol.* **247**, 536 (1995).
20. To conserve computational resources, no round 2 modeling was done for Glucose Permease IIBC because the native topology was never sampled during fragment assembly. Secondary structure predictions for all 50 homologs predicted an alpha helix in place of the N-terminal beta strand in the native structure.
21. Core residues are defined as those with <20% solvent-accessible surface area compared with an extended G-X-G peptide.
22. W. L. Delano (DeLano Scientific, San Carlos, CA, USA, 2002).

23. We thank L. Malmström for computational assistance and K. Laidig for flawless administration of the computing resources necessary for these calculations. We gratefully acknowledge support from the Howard Hughes Medical Institute (P.B. and D.B.) and the Helen Hay Whitney Foundation (K.M.S.M.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5742/1868/DC1
Materials and Methods
Figs. S1 to S5
References

20 April 2005; accepted 15 August 2005
10.1126/science.1113801

Azathioprine and UVA Light Generate Mutagenic Oxidative DNA Damage

Peter O'Donovan,[1] Conal M. Perrett,[1,3] Xiaohong Zhang,[1,2] Beatriz Montaner,[1] Yao-Zhong Xu,[2] Catherine A. Harwood,[3] Jane M. McGregor,[3] Susan L. Walker,[4] Fumio Hanaoka,[5] Peter Karran[1]*

Oxidative stress and mutagenic DNA lesions formed by reactive oxygen species (ROS) are linked to human malignancy. Clinical treatments inducing chronic oxidative stress may therefore carry a risk of therapy-related cancer. We suggest that immunosuppression by azathioprine (Aza) may be one such treatment. Aza causes the accumulation of 6-thioguanine (6-TG) in patients' DNA. Here we demonstrate that biologically relevant doses of ultraviolet A (UVA) generate ROS in cultured cells with 6-TG–substituted DNA and that 6-TG and UVA are synergistically mutagenic. A replication-blocking DNA 6-TG photoproduct, guanine sulfonate, was bypassed by error-prone, Y-family DNA polymerases in vitro. A preliminary analysis revealed that in five of five cases, Aza treatment was associated with a selective UVA photosensitivity. These findings may partly explain the prevalence of skin cancer in long-term survivors of organ transplantation.

The thiopurines azathioprine (Aza), 6-mercaptopurine (6-MP), and 6-thioguanine (6-TG) are cancer therapeutic and immunosuppressive agents. They are all prodrugs (compounds that the body converts into active drugs) requiring metabolic activation into the thioguanine nucleotides (TGNs) that are precursors for 6-TG incorporation into DNA (*1*). Experimentally, 6-TG is a surrogate for Aza because it bypasses many of the activation steps and is directly converted to TGN. The normal DNA bases do not absorb significantly at ultraviolet A (UVA) wavelengths (320 to 400 nm), whereas thiopurines do, and 6-TG has an absorbance maximum at 342 nm. 6-MP generates ROS when exposed to UVA (*2*). ROS are pernicious DNA-damaging agents (*3*), and although cells are equipped to deal with them, abrupt increases in ROS cause oxidative stress and produce mutagenic DNA lesions (*4*). The possibility that DNA 6-TG might act as an endogenous UVA chromophore and provide a source of promutagenic oxidative DNA damage prompted us to investigate the photochemical properties of 6-TG and the biological consequences of the interaction between DNA 6-TG and UVA.

HCT116 human colorectal carcinoma cells are mismatch repair–defective and tolerant of high levels of DNA 6-TG (*5*). We found that UVA generated intracellular ROS in 6-TG–treated HCT116 cells in which 6-TG replaced approximately 0.2% of DNA guanine. After uptake of CM-H2DCFDA dye and irradiation with UVA (3 kJ/m^2) [approximately equivalent to 1 to 2 min of exposure around noon in England at midsummer (*6*)], the cells emitted a green fluorescence indicating the formation of ROS. This was detected by fluorescence-activated cell sorting and by microscopy (Fig. 1A). Because ROS are highly unstable, they

[1]Cancer Research UK London Research Institute, Clare Hall Laboratories, South Mimms, Hertfordshire EN6 3LD, UK. [2]Department of Chemistry, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK. [3]Centre for Cutaneous Research, Institute of Cell and Molecular Science, Barts and The London Queen Mary's School of Medicine and Dentistry, 4, Newark Street, London, E1 2AT, UK. [4]Department of Photobiology, Guy's, King's and St Thomas' School of Medicine, St. John's Institute of Dermatology, King's College, London SE1 7EH, UK. [5]RIKEN Discovery Research Institute, Wako-shi, Saitama 351-0198 Japan.

*To whom correspondence should be addressed. E-mail: peter.karran@cancer.org.uk